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

Behnaz Bojd*
University of Washington

Hema Yoganarasimhan*
University of Washington

July 18, 2019

*We are grateful to an anonymous firm for providing the data and to the UW-Foster High Performance Computing Lab for providing us with computing resources. We would also like to thank Ebrahim Barzegary, Daria Dzyabura, Anja Lambrecht, Simha Mummamaleni, and Omid Rafieian for their detailed comments which have improved the paper. Thanks are also due to seminar participants at the IESE Business School, University of California at Irvine, University of Maryland, University of Memphis, and University of Texas at Dallas, and conference participants at CIST 2018.
Please address all correspondence to: behnaz@uw.edu, hemay@uw.edu.

Abstract

We examine the effect of a user's popularity on her demand in a mobile dating platform. Knowing that a potential match is popular can increase her/his appeal. However, popular people are less likely to reciprocate, and hence users may strategically shade their revealed preferences for them 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. A key piece of information shown to users is a popularity rating, ranging from one to three stars. We quantify the causal effect of a user's star-rating on the rankings that s/he receives 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 three-star users receive worse rankings during the game, but receive more messages after. We link the heterogeneity across outcomes and user-level observables to the perceived severity of rejection concerns, and establish strategic shading as the mechanism. Further, users' rejection concerns are consistent with Step-1 bounded rationality.

Keywords: Popularity information, online ratings, strategic shading, online dating, matching markets, two-sided platforms, bounded rationality.

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, 15% of U.S. adults ($\approx$ 40 million adults) reported that they have used online dating services (Smith, 2016). 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, over the years, mobile dating apps have replaced dating websites as the dominant form of online dating because they 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 eliminates the need for users to browse and search profiles, and the second step ensures that users are not spending effort in crafting and sending emails to potential partners who have no interest in them.

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, likes) next to her/his profile. The benefits of showing users' popularity information are that – (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 or quantifies the effect of such popularity measures on users' demand in dating platforms. A large stream of literature on e-commerce and online marketplaces has shown that displaying popularity information about products/sellers can have a positive impact on their demand (Sorensen, 2007; Tucker and Zhang, 2011). But those settings did not involve inter-personal interactions. Moreover, the mechanisms at play in e-commerce markets are likely to be quite different from those in dating contexts. Hence, the extent to which these results will translate to an online dating context is not clear.

In this paper, we are interested in two key questions related to popularity information and demand in online dating.

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

In dating contexts, 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 is therefore less likely to reciprocate any interest. Thus, a user who wants to avoid the psychological costs of rejection may 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 starts with the random assignment of four men and four women to a virtual room. Then, 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 with 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 contact/engagement between previous raters and the ratee. They are, thus, pure popularity measures and do not help resolve asymmetric information about the user's quality as a date (unlike star-ratings based on purchase/experience in

[1] Rank of one denotes a preference-ranking of four, rank of two indicates a preference-ranking of three, and so on.

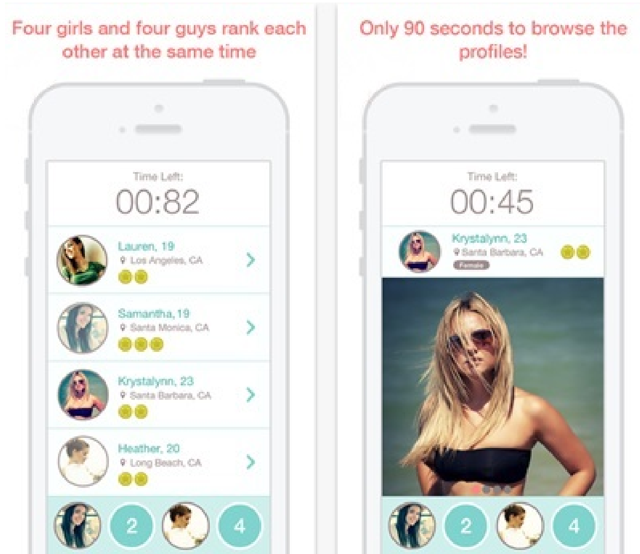

Figure 1: Screen shot of the app during a game (from the perspective of a male user). Players indicate their rank-ordered preference for the players from the opposite sex by dragging their profile pictures into the circles labeled one through four at the bottom of the app. In this example, the focal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

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 message from the matched partner after the game, (3) likelihood of receiving a reply to a message sent after the game. We face two main challenges in this task. First, a user's star-ratings and her/his unobserved attractiveness are confounded: attractive 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. Thus, we can use the *within*-person variation in star-ratings to causally infer the effect of a user's star-rating on her demand in the marketplace.

The second estimation challenge stems from the non-linearity of the three demand measures: the first measure (preference-ranking) is an ordered discrete outcome, and the other two measures (first and reply messages) are binary outcomes. We model the first measure using a fixed-effects ordered logit model, and the latter two are modeled 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. While we need fixed effects to control for the endogeneity issues discussed earlier, estimation of ordered and binary logit models with fixed-effects is tricky since there is no easy way to subtract out the unobserved user fixed-effect in a non-linear setting. To address this issue, Chamberlain proposed a general class of Conditional Maximum Likelihood (CML) estimators for non-linear models that condition on a subset of outcomes, which in turn allows them to condition-out all the fixed-effects (or nuisance parameters) and estimate only the main parameters of interest (Chamberlain, 1980). Usually, in a K outcome ordered logit model, we can derive $K-1$ consistent CML estimates. However, these $K-1$ estimates are inefficient because each of them only uses only part of the variation in the data for identification. Das and Van Soest (1999) developed an Minimum Distance (MD) estimator that combines all the CML estimators and generates both consistent and efficient estimates. We use this estimator to derive the effect of star-ratings on preference-rankings in our setting. For the two message-related binary outcome models, the CML and MD estimators are equivalent, so we simply use Chamberlain's CML for them. Note that all these estimators rely on the within-user variation in star-ratings to identify the effect of stars on outcomes, and thereby address the endogeneity issues discussed earlier.

We now discuss our main findings from the first part of the paper. 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. These results suggest that users in the platform 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 someone during a game, s/he has no information on the other person's preferences, thus the potential for being rejected (i.e., not being matched) is high. 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 can show that the negative effect of star-rating during the game is due to strategic shading.

We also provide additional evidence for strategic shading based on the heterogeneity in the effect of star ratings across user-level attributes. In particular, we show that the negative effect of star-ratings on preference-rankings is mainly driven by less-attractive users when they are ranking attractive potential partners. Since less-attractive users are more likely to have rejection concerns (especially when they are ranking attractive users), this finding corroborates our strategic shading hypothesis.

Finally, while users behave strategically given their beliefs, we find that their beliefs regarding rejection concerns during the game are not fully rational. Our results can be explained by Camerer et al. (2004)'s cognitive hierarchy model of games, which argues that users reason in steps. Specifically, our findings are consistent with Step 1 bounded-rationality. Step-1 users believe that others are Step-0 users, who will naively reveal their preferences without taking rejection concerns into account. Thus, the best response for Step-1 users is to reduce their own preference-ranking for popular users. These findings are in line with the literature in behavioral economics and bounded rationality (Nagel, 1995; Stahl and Wilson, 1995).

In sum, our paper makes three key contributions to the literature. First, we are the first to document negative returns to popularity information in online platforms. Past empirical research has mainly documented positive returns to the revelation of popularity information. 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. Third, our paper demonstrates that users exhibit bounded rationality in real-world online settings. The previous literature on bounded rationality come mainly from lab experiments; our results suggest that such behavioral effects may indeed play a significant role in platforms that involve strategic multi-player interactions.

Our results have implications for the design of online dating platforms. On the one hand, displaying popularity information can simplify users' search process and help them quickly evaluate potential partners. However, doing so can have unintended consequences on the demand for popular users. Our findings thus suggest that managers of online dating platforms should take this dampening effect of popularity information into account when designing their user-interface. More broadly, we expect our findings to be relevant to other two-sided matching markets with inter-personal rejection concerns, e.g., online labor markets.

The remainder of this paper is organized as follows. In §2, we discuss the related literature. We introduce the setting and data in §3 and §4. In §5, we present descriptive analyses on the effect of popularity information on users' demand. Next, in §6 and §7, we present our empirical specification, estimation and identification approaches, and establish the causal impact of star-ratings on preference-rankings and messages, respectively. In §8, we provide a discussion on the mechanisms driving users' behavior and examine the bounds of rationality observed in the data. Finally, in §9, we conclude with a discussion of our main findings and avenues for future research.

2 Related Literature

Our paper contributes to two broad streams of literature in marketing and economics.

First, it contributes to a large stream of literature that seeks to measure the effect of online popularity information on demand in e-commerce settings and online marketplaces. This research has consistently established the herding effect, i.e. shown that popularity information has a positive effect on demand/sales of products and services in a variety of contexts 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 wedding services market (Tucker and Zhang, 2011).[2] These studies have identified three underlying 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 result on the effect of popularity information, and in a previously unstudied context – dating markets. 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., 2010; Bapna et al., 2016; Lee, 2016a).

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 rejection (Cameron et al., 2013). If users shade their revealed

[2] A related stream of work examines the effect of WOM or online ratings on demand outcomes (Chevalier and Mayzlin, 2006; Sun, 2012; Yoganarasimhan, 2012, 2013). However, in these papers, the ratings are given after the interactions between the buyer and seller. Hence, they play the role of Word-of-Mouth or reputation effects, i.e., they 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.

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' beliefs on match probabilities on their revealed preference has been examined by a few papers in the literature. In an early paper, Hitsch et al. (2010) employ empirical tests and show that shading is not a concern in their setting. Nevertheless, their tests rely on aggregate data patterns and exclusion restrictions. As such, their results may not hold if we had variables that directly affect the perception of match-probability (e.g., popularity information) without affecting the attractiveness of a user at the individual-level. More recently, Yu (2018) shows that individuals become more strategically selective when they believe they have more potential matches, and less selective when they believe they have more competition. However, neither of these papers examine how revelation of popularity information affects users' demand in a dating platform, and connect it to strategic shading based on the differences in perceived rejection probabilities.

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 fun and flirtation 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. 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. A user, in fact, cannot directly access or browse other users' profiles through the app; the only way to 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 (see §4.3 for a detailed description on star-ratings). Tapping on a short profile leads to the full profile of the user. As shown in the right panel of Figure 1, full profiles typically contain a larger version of the profile picture (and possibly additional photos) and other information, such as bio, education or employment information.

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.[3]

Two points are worth noting here. 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. Thus, players do not have any visibility into their competition within each game, though they may have a sense of the general distribution of players of their own sex. 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.

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

[3]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

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 stable match, i.e., we at least have one solution 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.[4,5]

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 only one 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. 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. The intuition here is to pick the solution that, on average, gives

[4] 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.

[5] We briefly describe the Men-proposing Gale-Shapley Deferred Acceptance algorithm here: In the first iteration, each man proposes to the woman he prefers most. Then, each woman accepts the offer she prefers most. In each subsequent iteration, each unmatched man proposes to the most-preferred woman to whom he has not yet proposed regardless of whether the woman is already matched or not. Then, each woman chooses among the set of all the men who propose in this iteration as well as the one whom she is currently matched. This process is repeated until all men are matched. It can be shown that this algorithm always reaches a stable solution (Gale and Shapley, 1962).

each player her highest preference (or lowest numerical rank). Thus, the platform does not optimize for either men or women, but instead tries to pick the best globally optimal solution.

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. Each matched pair can communicate via text and/or picture and video messages, as shown in Figure 2 on the right panel. Users also have the choice to not initiate a conversation with their assigned partner and instead 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 15$^{\text{th}}$ 2014 to July 15$^{\text{th}}$ 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 start by describing the variables that characterize the time-invariant attributes associated with a user. These remain fixed for the duration of our observation period.[6]

For each user i in our data, we have information on:

1. $gender_i$: A dummy variable indicating user i's gender; is 1 for men and 0 for women.
2. age_i: User i's age.
3. bio_i: The length of user i's bio in his/her profile (i.e., number of words).
4. $education_i$: 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.

[6]In principle, some of these attributes may change over time. However, we do not observe many such changes during our period, and therefore treat them as time-invariant attributes.

5. $employment_i$: Number of positions/companies mentioned in user i's profile.

6. $initial\_game_i$: Total number of games played by user i before the data collection period.

7. $total\_game_i$: Total number of games played by user i during the data collection period.

8. 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), 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 pictures in approximately 20 minutes. In order to minimize biases due to boredom or fatigue, subjects were shown the profile pictures in a random order. 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:

9. 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.[7]

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 two employment-related information listed on her/his profile. In terms of activity, the median user had played 48 games before the data collection period and plays 18 games during the observation period. 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.

[7]The lack of pic_scores for 6,900 users does not affect our main analysis since we use a fixed-effects specification, which conditions out all user-specific variables.

Variables	Mean	Std. Dev	25^{th}	50^{th}	75^{th}	(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
$employment_i$	2.05	1.59	1	2	3	(1, 68)	15579
$initial\_game_i$	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.

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-game interactions. We describe these variables in detail below.

1. $pref_{ijt}$: An integer variable that denotes 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 in a game from one through four (as shown in Figure 1), with a rank of one indicating their highest preference and four indicating the lowest preference. 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 and more intuitive because higher values of this variable correspond to more preference (unlike rank, where lower rank indicates higher preference, which complicates exposition).

2. $match_{ijt}$: 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 set to one for only man (woman).

3. $first_{ijt}$: A dummy variable indicating whether user i receives the first message from the player he/she is matched with (denoted by j here) after game t. Note that users are not given the option to communicate with players they have not been matched with, i.e., they can only communicate with the person they have been matched with by the platform. So, by default, this variable is zero if $match_{ijt} = 0$.

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.

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

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.[8] 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 initiated message only 3380 of them receive a reply). Interestingly, 76% of the conversations are 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., 2010).

4.3 User-Game level data

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

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

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

[8]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 get matched with only one user among the four potential mates, $match_{ijt}$ becomes one once, and becomes zero three times. 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) × 32 = 3,020,652. However, some of these data points are related to users whose gender changes in the data set over the data collection period time (42 users), and we exclude them from our analysis.

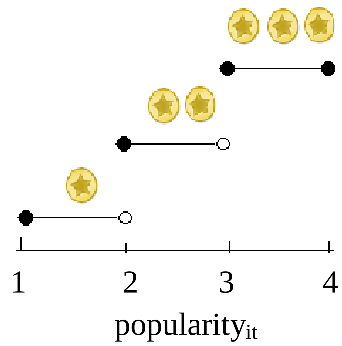

Figure 3: Pictorial representation of the star-rating rule (as a function of average preference-ranking in past games).

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

3. $star_{it}$: Indicates the star-rating that the user is shown with in game t; see Figure 1 for an example. It 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, as shown below:

$$popularity_{it} = \frac{\sum_{q=1}^{total\_game_{it}} \sum_{j=1}^{4} pref_{ijq}}{4 \times total\_game_{it}}. \tag{3}$$

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 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.

Variables	Mean	Std. Dev	25^{th}	50^{th}	75^{th}	(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.

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. We also find that the median of $total\_game_{it}$ is 59, which suggests that most users have played a good number of games before a median game in the observation period. Finally, we also see that users are shown with a two-star rating on average.

Finally, we examine the extent of variation in star-ratings within an individual. 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, 1,287 users were shown with a minimum of one star and a maximum of two stars, and 2,185 users were shown with a minimum of two stars and a 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.

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 messages, 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.[9] 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

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

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.

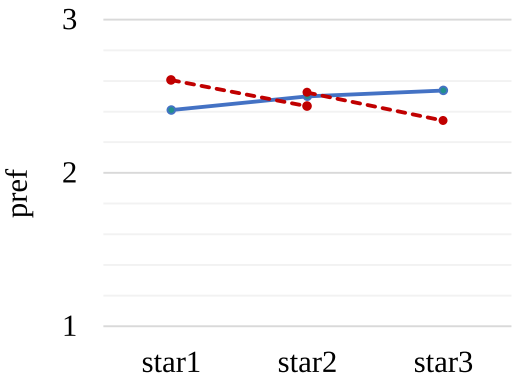

Figure 4: The relationship between star-ratings and average preference-rankings received. The solid line is for all user-game data points and the dashed lines are for within-individual data points.

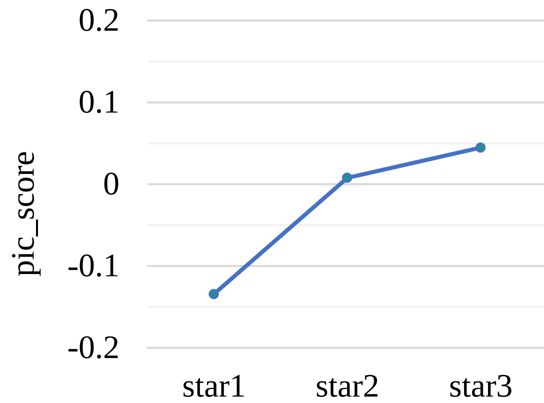

Figure 5: The relationship between star-ratings and average physical attractiveness score

One possible 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 for the same individual when s/he is shown with different star-ratings, then our comparisons are less likely to be subject to endogeneity concerns. We will expand on this theme in the next two sections, but for now, we present some graphical model-free evidence using this intuition.

First, we consider individuals who were shown with a minimum of one star and a maximum of two stars. For each of these individuals, we 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 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. The results of these comparisons are presented 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. In sum, the dashed lines in Figure 4 suggest 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.

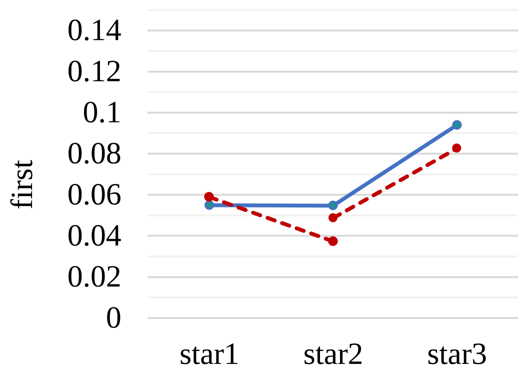

Figure 6: The relationship between star-ratings and the average likelihood of receiving the first message. Solid lines are for all user-game observations and dashed lines are for within-individual observations.

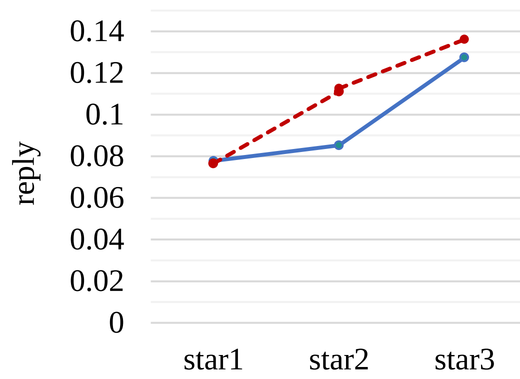

Figure 7: The relationship between star-ratings and the average likelihood of receiving a reply message. Solid lines are for all user-game observations and dashed lines are for within-individual observations.

Similarly, Figures 6 and 7 show the relationship between a user's star-rating and the likelihood of her receiving the first message and receiving a reply if she initiates a message, respectively. The solid lines show the relationship between the likelihood of receiving a message (first or reply) for all user-game observations calculated for each star-rating.[10] We see that observations where users have higher star-ratings are more likely to receive the first messages and replies.

Next, we perform a within individual analysis on users' messaging behavior. As shown by the dashed lines in Figure 6, we can see, 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. However, this effect does not carryover 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 (see dashed lines in Figure 7).

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 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 effects of star-ratings on these three revealed preference measures using econometric methods, and exploring the mechanisms driving these effects.

[10]For example, in Figure 6, the data point on the solid line for $star1$ is given by $\frac{\sum_i \sum_t (first_{ijt} \mid star_{it}=1, match_{ijt}=1)}{\sum_i \sum_t I(star_{it}=1)}$, and in Figure 7, the data point on the solid line for $star1$ is given by $\frac{\sum_i \sum_t (reply_{ijt} \mid star_{it}=1, match_{ijt}=1, first_{jit}=1)}{\sum_i \sum_t I(star_{it}=1)}$.

6 Effect of Star-ratings on Preference-Rankings

In this section, we formalize the causal impact of a user's star-rating during a game on the preference-rankings that s/he receives during the game. Since preference-rankings are ordinal, we use an ordered logit model to estimate this effect. In §6.1 and §6.2, we present the model specification and estimation. We discuss our findings in §6.3. We present some tests and robustness checks to validate the model and results 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 4 to 1, with four indicating the highest preference-ranking and one representing the lowest preference-ranking. Therefore, we use an ordered logit model relates the observed outcome variable $pref_{ijt}$ to a *latent* variable $pref^*_{ijt}$ where:

$$pref^*_{ijt} = \beta_1 star1_{it} + \beta_2 star3_{it} + \gamma z_i + \eta_i + \epsilon_{ijt}, \tag{4}$$

The latent variable $pref^*_{ijt}$ is thus modeled as a linear function of:

- $star1_{it}$, $star3_{it}$ – indicator variables for the star-rating of user i in game t, where $star2$ is considered the base.
- z_i – set of user-specific observables that can affect j's ranking of i, e.g., age of i.
- η_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.
- ϵ_{ijt} – These are factors uncorrelated to the star-rating of user i that can affect the preference-ranking s/he receives from j in game t. Three key sets of variables are subsumed here.
  - First, it includes j's attributes (both observable z_j and unobservable η_j) since there is no correlation between j and i's attributes.
  - Second, it also includes all the attributes of the other three players of i's gender who i is being compared with, in game t.

    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 attributes of users within a game.[11]

[11]In principle, because the app only considers adding new users who are within a 500 mile radius of users already

- Third, ϵ_{ijt} may include 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 assume that ϵ_{ijt}s have a logistic cumulative distribution. Although, the second point above can create correlation between ϵ_{ijt}s in one game, in §6.4.3, we show that our results are robust to such correlations.

The endogeneity concerns in this model mainly stem from the potential correlation between η_i and $star_{it}$, i.e., we expect that $E[star_{it} \cdot \eta_i] \neq 0$. We will come back to this issue when discussing estimation approaches.

We then model the relationship between $pref_{ijt}$ and $pref^*_{ijt}$ as follows:

$$pref_{ijt} = k \quad \text{if } \mu_k < pref^*_{ijt} \leqslant \mu_{k+1} \quad \forall \quad k = 1, 2, 3, 4, \tag{5}$$

where the thresholds μ_k are strictly increasing. Further, we assume that $\mu_1 = -\infty$ and $\mu_5 = \infty$. This specification is simply 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 ϵ_{ijt} as

$$F\left(\epsilon_{ijt} \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_i, \mu_k, \mu_{k+1}\right) = \frac{1}{1 + exp(-\epsilon_{ijt})} \equiv \Lambda(\epsilon_{ijt}), \tag{6}$$

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:

$$\begin{aligned} Pr\left(pref_{ijt} = k \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_i, \mu_k, \mu_{k+1}\right) &= \Lambda\left(\mu_{k+1} - \beta_1 star1_{it} - \beta_2 star3_{it} - \gamma z_i - \eta_i\right) \\ &- \Lambda\left(\mu_k - \beta_1 star1_{it} - \beta_2 star3_{it} - \gamma z_i - \eta_i\right) \end{aligned} \tag{7}$$

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

$$LL(\beta_1, \beta_2, \gamma, \eta_i, \mu_k, \mu_{k+1}) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{j=1}^{4} \sum_{k=1}^{4} \ln\left[Pr\left(pref_{ijt} = k \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_i, \mu_k, \mu_{k+1}\right)^{I(pref_{ijt}=k)}\right], \tag{8}$$

in a 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, δ 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 η_i, the remaining δ is random noise.

where N is the total number of users observed and T_i is the total number of games played by user i. Notice that the the unknown parameters in Equation (8) are $\beta_1, \beta_2, \gamma, \eta_i s, \mu_2, \mu_3, \mu_4$. We discuss their estimation in the next section.

6.2 Estimation

We are interested in estimating the effect of star-ratings (coefficients β_1 and β_2). There are two possible estimation strategies for this: (1) A pooled estimation strategy, where we ignore the user-specific unobservables η_i, and (2) a fixed-effects approach, where we allow the user-specific unobservables η_i to be arbitrarily correlated with the star-ratings. We discuss both these approaches below.

The first approach is straightforward. It simply involves pooling all the user-game data, ignoring the user-specific unobservable η_i, and then maximizing the log-likelihood in Equation (8). However, it is important to recognize that the estimates from this approach will be biased in the presence of correlated unobservables. Therefore, in the rest of this section, we focus on estimating β_1 and β_2 after controlling for η_i.

A naive approach to estimation with fixed-effects is to treat the η_i's as parameters and maximize the log-likelihood in Equation (8) 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 β_1 and β_2 from this approach will be inconsistent too. Chamberlain (1980) provides an elegant solution to the incidental parameters problem for the case of binary variable 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 clarify the conditions necessary for identification, and in §6.2.3 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 $pref^k_{ijt}$ where:

$$pref^k_{ijt} = I(pref_{ijt} \geq k), \quad \text{where} \quad k = 2, 3, 4. \tag{9}$$

For example, the binary variable $pref^4_{ijt}$ indicates whether user i received a preference-ranking of 4 from user j in game t, or not. Similarly, the binary variable $pref^3_{ijt}$ 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:

$$Pr(pref_{ijt}^k = 1 \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_i, \mu_k) = 1 - \Lambda(\mu_k - \beta_1 star1_{it} - \beta_2 star3_{it} - \gamma z_i - \eta_i) \quad (10)$$

Next, 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, we denote s_i^k as the sum of all the binary transformed preference-rankings at level k received by user i over time:

$$s_i^k = \sum_{t=1}^{T_i} \sum_{j=1}^{4} pref_{ijt}^k$$

In other words, s_i^k shows the count of ones in the set of $pref_i^k$. Further, we denoted B_i^k as 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. That is:

$$B_i^k = \{d \in \{0,1\}^{4 \times T_i} \mid \sum_{t=1}^{T_i} \sum_{j=1}^{4} d_{jt} = s_i^k\} \quad (11)$$

Note that the size of $B_i^k = \binom{4 \times T_i}{s_i^k}$.[12]

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

$$Pr\left(pref_i^k \mid star1_{it}, star3_{it}, s_i^k, \beta_1, \beta_2\right) = \frac{exp\left(pref_i^k \cdot (\beta_1 star1_{it} + \beta_2 star3_{it})\right)}{\sum_{d \in B_i^k} exp\left(d \cdot (\beta_1 star1_{it} + \beta_2 star3_{it})\right)} \quad (12)$$

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

$$CLL(\beta_1^k, \beta_2^k) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \ln\left[Pr(pref_i^k \mid star1_{it}, star3_{it}, s_i^k, \beta_1^k, \beta_2^k)\right] \quad (13)$$

Since we can dichotomize $pref_{ijt}$ into three binary variables at each of the three cutoffs ($pref_{ijt}^4$,

[12]For example, consider user i who plays only 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 $\binom{4\times 2}{1} = 8$.

$pref^3_{ijt}$, and $pref^2_{ijt}$), the above CLL can be specified for each $pref^k_{ijt}$, where $k \in \{2, 3, 4\}$. Maximizing each of these CLLs gives us three separate but consistent estimates of β_1, β_2, which we denote as $\{\beta^k_1, \beta^k_2\}$, where $k \in \{2, 3, 4\}$. These are referred to as Chamberlain CML estimators.

6.2.2 Identification

Two necessary conditions need to be satisfied for the identification of $\{\beta^k_1, \beta^k_2\}$. 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 $pref^k_{ijt}$ because users with constant $pref^k_{ijt}$ do not contribute to the CLL for cut-off k (and hence identification).[13] 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 ($pref^4_{ijt}$) is constant over time even if her/his star-rating varies over time. Thus, only users for whom we have across-time variation in both the outcome variable ($pref^k_{ijt}$) and the independent variables ($star1_{it}$, $star3_{it}$) contribute to the identification of $\{\beta^k_1, \beta^k_2\}$.

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.[14]

6.2.3 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

[13] Constant $pref^k_{ijt}$ means that all elements of B^k_i are either zero or one.

[14] For individuals who have played a large number of games (large T_i) and have a large number of positive values of $pref^k_{ijt}$ (large s^k_i), calculating all combinations of outcomes can lead to numerical overflow and computational issues. For example, if user i plays 100 games ($T_i = 100$) and receives one preference-ranking of four in each game, then $s^4_i = 100$ and $\binom{4\times100}{100} = 2.24e + 96$. 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.

the application of their method for our context below.

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:

$$\hat{\beta}^{MD} = \underset{b}{\text{argmin}}(\tilde{\beta} - Mb)' var(\tilde{\beta})^{-1}(\tilde{\beta} - Mb), \tag{14}$$

where $\tilde{\beta}$ is the 6×1 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:

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

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

$$\hat{\beta}^{MD} = \{M' var(\tilde{\beta})^{-1} M\}^{-1} M' var(\tilde{\beta})^{-1} \tilde{\beta} \tag{15}$$

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).

6.3 Results

The results from our estimation exercise are presented in Table 4. We start by considering two pooled ordered logit models: (1) model M1 – a simple model that only includes star-ratings as the independent variable, and (2) model M2 – a slightly more elaborate model that includes all the user-specific observables (z_i) as controls.

In the basic pooled 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. Thus,

	(M1) (Ordered Logit)	(M2) (Ordered Logit)	(M3) (FE Ordered Logit)
$star1_{it}$	-0.14452***	-0.11431***	0.02852
	(0.02315)	(0.03021)	(0.01804)
$star3_{it}$	0.06063***	0.05578**	-0.05101***
	(0.01560)	(0.02328)	(0.01464)
Controls		✓	
μ_2	-1.09924***	-1.11220***	
	(0.00203)	(0.01533)	
μ_3	-0.00053	-0.00841	
	(0.00188)	(0.01527)	
μ_4	1.09828***	1.09372***	
	(0.00205)	(0.01529)	
Individuals	24393	11639	3494
Observations	2980148	1580848	630160

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Controls 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.

it seems like higher star-ratings lead to higher preference-rankings. This result is consistent with Figure 4 (solid line). Since a user's current star-rating is likely to be positively correlated with her/his appeal in the dating market (through physical attractiveness, age, education, etc.), in model M2 we control for all the user-specific observables. 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 shown with three stars as opposed to two stars. We do not find any significant effect of one star compared to two stars. In §6.4, we present a battery of robustness checks to confirm the validity of our empirical findings.

The main takeaway from our findings is that popularity information has a negative effect on users' demand during the game. We are the first to document negative returns to popularity in online platforms. 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 may dislike the popular users. Or, they may like popular users but avoid them due to rejection concerns: raters (rank-givers) may be concerned that popular users are harder to achieve matches with, and therefore shade their preferences for them in order to avoid rejection costs. In §8, we formalize the discussion of the mechanism behind the negative effect of popularity information, tease out these two explanations, and rule out other 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 within a given market instead of assuming positive effects based on prior work.

6.4 Robustness Checks

6.4.1 Linear Model

First, we examine whether the substantive results from §6.3 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-ratings 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 §A.1 for model details and the full table of results.

6.4.2 Estimation Sample

Next, we examine if our results are driven by the estimation sample used. Recall that the Minimum Distance estimator for the fixed-effects ordered logit model utilizes only a subset of the data for inference – data on users who went through 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. In that case, our findings would only apply to local sub-population that saw at least one star change.

We consider two validation checks to confirm that the results are not driven by the sample. First, the results from a pooled ordered logit model on the estimation sample used in the Minimum Distance estimates are similar to those obtained from the full sample. Second, we find no systematic user-level differences between users who go through at least one star change in our data compared to those who do not go through any star change. Please see Appendix §A.2 for details.

Second, recall that the effect of star-ratings on preference-ranking and replies were quite different. Our explanation of this difference was based on the differences in perceived probabilities of rejection. However, this might be due to the differences in the estimation samples used in models M3 and M6. In model M3, it includes all users who experienced a star-change, and in model M6, it includes users who experienced a star-change and those who initiated a message. As a robustness check, we therefore re-estimate model M3 with the sample used in model M6. We find that the results from this exercise are the same as those presented in M3 (see Table A4 in Appendix §A.2).

6.4.3 Within Game Correlation

Recall that ϵ_{ijt}s can include all 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 ϵ_{ijt}s in one game, if we include the observation of all competitors in one game in our model. 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 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 Table A5 in Appendix §A.3.)

6.4.4 Star Configuration in a Game

It is possible that users self-select their entry time when they expect certain types of competitors and this may affect the star configuration of the games. So we examine the star configuration of competitors in the games in our data (see Table A6 in Appendix A.4). We find that in 95.92% of the rooms (or games) all four competitors are shown with two stars. This reflects the fact that the majority of users on the platform never experience a star change. For those who do experience a star-change and are shown with three stars, all their three competitors have two stars in 4,735 games. Similarly, for those who experience a star-change and are shown with one stars, all their three competitors have two stars in 2,606 games. Other star configurations are pretty rare in our data. 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 at certain points in time etc.

7 Effect of Stars 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:

$$first_{ijt} = \begin{cases} 1, & \text{if } first^*_{ijt} > 0 \\ 0, & \text{else} \end{cases} \tag{16}$$

$$reply_{ijt} = \begin{cases} 1, & \text{if } reply^*_{ijt} > 0 \\ 0, & \text{else} \end{cases} \tag{17}$$

These latent variables are defined as:

$$first^*_{ijt} = \beta_1^f star1_{it} + \beta_2^f star3_{it} + \gamma^f z_i + \eta_i^f + \epsilon_{ijt}^f, \tag{18}$$

$$reply^*_{ijt} = \beta_1^r star1_{it} + \beta_2^r star3_{it} + \gamma^r z_i + \eta_i^r + \epsilon_{ijt}^r, \tag{19}$$

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 η_i^f and η_i^r to be arbitarily correlated to $star1_{it}$ and $star3_{it}$. Assuming that ϵ_{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 Conditional Log-Likelihoods that are independent of ηs 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 §B.

	First Message		Reply Message	
	(M4) (Logit)	(M5) (Logit FE)	(M6) (Logit)	(M7) (Logit FE)
$star1_{it}$	0.06327	0.43077***	-0.19146	-0.42609*
	(0.14248)	(0.09954)	(0.23242)	(0.25744)
$star3_{it}$	0.57401***	0.64912***	0.27474	0.28316*
	(0.08990)	(0.06280)	(0.17387)	(0.15125)
Controls	✓		✓	
Constant	-2.00281***		-1.01352***	
	(0.07918)		(0.22234)	
Individuals	11634	1972	3115	536
Observations	374727	118627	20485	8573

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Controls in Model M13 and M15 include: $gender_i$, age_i, $college_i$, $graduate_i$ pic_score_i, num_pic_i, $employment_i$, and bio_i.

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

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 controls only for the observable attributes of the (potential) receiver and M5 is a fixed-effects logit model estimated using CLL that accounts for the endogeneity between star-ratings and user-specific unobservables. 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 both three- and one-star users are more likely to receive first messages compared to two-star users. This is consistent with dashed-lines in Figure 6.[15]

In this case, the 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 users 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-star users suggests the possibility that players may value higher-star users more. Thus, these results can be explained by a combination of both higher utility for higher star users as well as lower rejection concerns, and it is hard to tease out the exact mechanism based on them.

[15] Note that we have only 1,972 users in model (M5). Although, we have 3,494 users who experienced a star-change, some of them are dropped from a fixed-effects logit model because of no variation in their outcome $first_{ijt}$.

So next, we present the results for *reply* behavior in models M6 and M7, which are analogous to M4 and M5. Interestingly, we find that compared to two-star users, one-star users are less likely and three-star users are more likely to receive a reply when they initiate contact. That is, the effect of star-ratings on preference-ranking and replies are quite different (compare models M3 and M7).[16]

The main takeaway here is that, in the case of replies, the effect of popularity information is positive and consistent with the earlier literature on herding. Intuitively, users in the reply condition are unlikely to be concerned about rejection and therefore rejection concerns may not play any role in their reply behavior. In the next section, 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 Discussion of Mechanism

We now examine the mechanism behind the effects established in §6 and §7. In §8.1, we formalize the ranking strategy of players during the game and their messaging behavior after the game. Then, in §8.2, we define strategic shading and discuss how our empirical results can be explained by strategic shading. Finally, in §8.3, we examine the rationality of strategic shading in our setting.

8.1 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).

8.1.1 Ranking Strategy During the Game

We assume that the preference-ranking that user j gives to user i is induced by j's underlying expected utilities. Let $EU(pref_{ijt})$ denote the expected utility that user j gets from giving user i preference-ranking $pref_{ijt}$, such that:

$$EU(pref_{ijt}) = U(star_{it}) \times \mathcal{P} - C \times (1 - \mathcal{P}). \tag{20}$$

Here, $U(star_{it})$ denotes the utility that user j expects to receive from a potential conversation/date with i. $U(\cdot)$ can also be a function of other observed i and j specific variables. However, we suppress them in our notation to keep the expressions simple. If j does not get matched with i, s/he incurs a psychological rejection cost of C. We assume this cost is incurred because j can infer that i did not rank him/her high enough. The cost of rejection can also be a function of i's attributes, i.e.,

[16] One possible reason for the difference in the results in models M3 and M7 could be the difference in the estimation samples used. M3 includes all users who experienced a star-change, whereas model M7 only includes users who experienced a star-change *and* those who initiated a message. 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 the same as those presented in M3 (see Table A4 in Appendix §A.2). Thus, these differences are not driven by the estimation sample used.

C can be written as $C(star_{it}, z_i)$. For instance, j may suffer higher rejection costs if i is popular (three-star) or attractive. However, this does not affect any of the arguments used to demonstrate strategic shading in §8.2.1 and therefore we simply denote it as C to keep the notation simple.[17] Finally, $\mathcal{P}$ denotes j's perceived expected probability of being matched with i conditional on giving i a preference-ranking of $pref_{ijt}$. In §8.3, we present the full expansion of $\mathcal{P}$, and show that it is a function of $pref_{ijt}$; for now it is sufficient to simply define it.

8.1.2 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.[18] 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. Since i initiated the first message, j is unlikely to have any rejection concerns when replying to i. Thus, unlike Equation (20), there is no rejection probability or cost in the expected utility that user j gets from replying to i. Thus we can write:

$$EU(reply_{ijt}) = U(star_{it}). \tag{21}$$

8.2 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 being matched with her/him ($U(\cdot)$). Instead, user j's revealed preference also takes into account the perceived probability of rejection and rejection costs. This distortion of revealed preference away from $U(\cdot)$ is referred to as *strategic shading*.

[17] In our context, C only refers to the psychological cost of knowing that other player did not rank you sufficiently high, and not opportunity costs. This is because of the following reason: in the stable matching algorithm that is used to match users, if a user does not get matched with her first choice, then it does not affect her chance of getting matched with her second choice, and so on.

[18] User j's decision to send the first message to user i is based on j's underlying expected utility, and is analogous to Equation (20). Let $EU(first_{ijt})$ denote the expected utility that user j gets from sending a first message to user i:

$$EU(first_{ijt}) = U(star_{it}) \times \mathcal{P}_f - C_f \times (1 - \mathcal{P}_f).$$

The definition of $U(star_{it})$ is the same as before. If j does not receive a reply form i, conditional on initiating a conversation with her/him, s/he incurs a rejection cost of C_f, and $\mathcal{P}_f$ denotes j's perceived expected probability of receiving a reply from i conditional on initiating a conversation with her/him. Although, users may perceive a lower probability of rejection once they have been matched with a partner, the probability of rejection is unlikely to be zero.

Strategic shading can be easily understood in our setting as follows: suppose that users value more popular users, i.e., expect higher utility (U) from dating a popular partner. However, if there is a non-zero probability of rejection and rejection costs are positive, they may reveal lower preferences for popular users. That is, users would strategically shade down their preferences for popular users in order to avoid rejection.

8.2.1 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

$$EU(reply_{ijt} \mid star_{it} = 3, first_{jit} = 1) > EU(reply_{ijt} \mid star_{it} = 2, first_{jit} = 1). \tag{22}$$

Then, based on Inequality (22) and Equation (21), we can infer that:

$$U(star_{it} = 3) > U(star_{it} = 2). \tag{23}$$

This implies that users receive higher utility from a potential conversation/date with a three-star partner compared to a two-star partner.

Next, we characterize the empirical findings from §6 (on $pref$) in Inequality (24). Recall 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. Thus, we have:

$$EU(pref_{ijt} \mid star_{it} = 3) < EU(pref_{ijt} \mid star_{it} = 2). \tag{24}$$

The above inequality is based on the assumption that users' ranking behavior during the game reflects their true preferences, i.e., preference-rankings reflect users' underlying expected utilities. We refer readers to Appendix §C for a formal statement and validation of this assumption.

Since we know from Inequality (23) that $U(star_{it} = 3) > U(star_{it} = 2)$, Inequality (24) can only be explained by rejection concerns, i.e., due to positive perceived expected probability of rejection $\mathcal{P}$ and non-zero rejection cost $\mathcal{C}$. Thus, the negative effect of star-ratings during the game can be directly attributed to rejection concerns.

8.2.2 Additional Evidence for Strategic Shading based on Heterogeneous Effects

In the previous section, we showed that the negative effect of three-star ratings on preference-rankings can be explained by rejection concerns. We now provide some additional evidence in support of this idea based on the heterogeneity in the effect of star-ratings on users' ranking behavior during the game. In particular, we examine the heterogeneity in the effect of star-ratings based on physical attractiveness.

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 stratify the data into two groups: (i) Attractive rank-givers: data where the rank-giver's pic_score is greater than -0.09, and (ii) Unattractive rank-givers: data where the rank-giver's pic_score is less than or equal to -0.09. Then we re-run the analysis on these two strata of data separately and report the results in the first two columns of Table 6. We find that attractive users are not likely to be influenced by the star-ratings of potential partners (model M8). On the other hand, for unattractive rank-givers, the results show a negative and significant effect of $star3_{it}$ (model M9).[19] This suggests that only unattractive users avoid 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 rejection than attractive users.

To further examine the source of the effect for unattractive users, we stratify the data for unattractive rank-givers based on the physical attractiveness of the rank-receivers. We re-run our analysis on these two strata separately. The results from this exercise are shown in models M10 and M11 in Table 6. We find that there is a negative and significant effect of $star3_{it}$, only when unattractive rank-givers are ranking attractive potential partners (model M10). Again, these results suggest that our findings are driven by rejection concerns, since unattractive rank-givers are more

[19] One important thing to keep in mind is we cannot compare the magnitudes of the effects of $star1_{it}$ and $star3_{it}$ across models M8 and M9 because the variance of errors is not identified in (ϵ_{ijt}) in (ordered) logit models (Allison, 1999). For example, suppose that we have the following models for attractive and unattractive users:

$$\begin{aligned} \text{Attractive:} \quad pref^*_{ijt} &= \beta^a_1 star1_{it} + \beta^a_2 star3_{it} + \gamma^a z_i + \eta_i + \epsilon^a_{ijt} \quad \text{where} \quad \epsilon^a_{ijt} \sim GEV1(0, \sigma^2_a) \\ \text{Unattractive:} \quad pref^*_{ijt} &= \beta^u_1 star1_{it} + \beta^u_2 star3_{it} + \gamma^u z_i + \eta_i + \epsilon^u_{ijt} \quad \text{where} \quad \epsilon^u_{ijt} \sim GEV1(0, \sigma^2_u) \end{aligned}$$

Because the latent variable $pref^*$ is not observed, the variance of error terms are not identified and the above models are rescaled such that what we estimate in practice is:

$$\begin{aligned} \text{Attractive:} \quad pref^*_{ijt}/\sigma^2_a &= (\beta^a_1/\sigma^2_a) star1_{it} + (\beta^a_2/\sigma^2_a) star3_{it} + \epsilon^a_{ijt}/\sigma^2_a \\ \text{Unattractive:} \quad pref^*_{ijt}/\sigma^2_u &= (\beta^u_1/\sigma^2_u) star1_{it} + (\beta^u_2/\sigma^2_u) star3_{it} + \epsilon^u_{ijt}/\sigma^2_u \end{aligned}$$

The coefficients of $star1_{it}$ in Models M8 and M9 are not β^a_1 and β^u_1; rather they are β^a_1/σ^2_a and β^u_1/σ^2_u. We report coefficients calculated under the assumption that the variance of error is $\pi^2/3$. However, if residual variation differs between groups, comparing the magnitude of the coefficients across groups can lead to incorrect conclusions.

	(M8)	(M9)	(M10)	(M11)
			Unattractive Rank-Giver	
	Attractive Rank-Giver	Unattractive Rank-Giver	Attractive Rank-Receiver	Unattractive Rank-Receiver
$star1_{it}$	0.00064	0.02566	-0.02015	0.02220
	(0.02938)	(0.02792)	(0.05177)	(0.03998)
$star3_{it}$	-0.03107	-0.07558***	-0.07665**	0.05318
	(0.02246)	(0.02164)	(0.03190)	(0.03605)
Individuals	3444	3453	1231	1354
Observations	258527	285554	116350	131440

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Heterogeneous effect of star-rating on received preference-rankings using ordered logit fixed-effects model.

likely to expect the probability of being matched to attractive users to be lower.

In sum, we find that star-rating effects are mainly driven by users who are less-attractive than average, and especially in cases where they are considering attractive potential partners. These findings provide additional evidence in support of the our hypothesis that preference shading is driven by fear of rejection concerns.

8.2.3 Alternative Mechanisms

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

First, one possible explanation for the negative effect of three-stars during the game is the salience effect. Since most users are shown with two-stars (see Table A6 in Appendix A.4 for the distribution of stars in a game), three-star users may appear more salient and people may therefore pay more attention to them. However, this is unlikely to be the case 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 with increased salience; however we see a negative effect for three-star users. Thus, it is unlikely that these results can be explained by the salience effect.

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 user (i.e., Inequality (23)). Thus, we can rule out the explanation that users give lower preference-rankings to three-star users during the game because they dislike popular users.

Finally, a third possible mechanism for the negative effect of higher stars during the game could be the reference-point effect: when a user (rank-giver) see a potential partner with a higher popularity 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. In other words, the rank-giver may not dislike popular users, but perceive them to be less appealing conditional on their popularity rating. Similar, to the discussion above, we can rule out this explanation because such behavioral biases are not supported by the fact that three-star users receive higher number of replies after the game.

Moreover, none of these alternative explanations are consistent with the heterogeneous effects we found in §8.2.2.

8.3 Bounded Rationality: Limits of Strategic Thinking

Thus far, we have shown that players act rationally given their beliefs, i.e., conditional on their beliefs that popular players are hard to get, players respond strategically by shading their preferences for them. However, their beliefs may be mistaken. So, we now examine whether users' rejection concerns are rational.

Intuitively, if all the women (men) in a game lower their preference-ranking for a three-star man (woman), then the likelihood of being matched with him (her) should not be adversely affected. In other words, if a user can rationally infer that other players in the game may also be suffering from rejection concerns (and hence give lower preference-ranking to popular users), then they should recognize that popularity does not necessarily lead to a lower likelihood of match.

We can formalize this argument by expanding the expected probability that user j will be matched with user i, $\mathcal{P}$, from Equation (20) as:

$$\mathcal{P} = \int P\left(match_{ijt} = 1 \mid pref_{ijt}, pref_{-ijt}, pref_{-jt}(star_{it})\right) g(pref_{-jt}) dg, \tag{25}$$

where $P\left(match_{ijt} = 1 \mid pref_{ijt}, pref_{-ijt}, pref_{-jt}(star_{it})\right)$ denotes the probability that j will match with i conditional on $pref_{ijt}$ (the preference-ranking that user j gives to i), $pref_{-ijt}$ (the preference-ranking that user j gives to other players), and $pref_{-jt}$ (the preference-ranking that other users in the game give to everyone else, including i).[20] Although, j does not observe other players' preference-rankings, s/he may have beliefs about their distributions, which we denote by $g(pref_{-jt})$. Thus, we can integrate over these beliefs to obtain the expected match probability $\mathcal{P}$.

A key point to note here is that the expected probability of j being matched with i also depends on how other people in the game ($-j$) rank i (denoted by $pref_{-jt}(star_{it})$). We know from our

[20] Note that $pref_{-ijt}$ and $pref_{-jt}$ should be ideally expressed as a function of the star ratings of other players' star ratings, but we suppress them to keep the notation simple.

empirical findings in §6 that $pref_{-j't}(star_{it} = 3) < pref_{-j't}(star_{it} = 2)$ for all $j' \in -j$. Therefore, conditional on the preference-ranking that j gives to i, the rational expected probabilities of being matched with i when she is shown with three-stars should be higher than when she is shown with two-stars. However, in a fully rational world, other players will also recognize this effect, and in turn increase their preference-ranking for i. Thus, there cannot exist a Bayesian Nash equilibrium of this game where the average effect of star-rating on match probabilities is negative.

Our results thus suggest that users do not fully internalize the idea that other players in the game may also suffer from rejection concerns, and therefore express lower preference for popular users. This may be because users' beliefs about the likelihood of matches with popular users are based on their observations in the offline world and/or their bounded rationality.

Indeed, our results are consistent with experimental findings on users' behavior in guessing games, where it has been shown that most users can only reason one step ahead. The classic example here is that of Keynesian beauty contest games, where all participants are asked to simultaneously pick a number between 0 and 100, and the winner is the player(s) whose pick is closest to p times the average of all numbers submitted, where p is some fraction. The Nash equilibrium of this game is (when all players are fully rational) is zero (Keynes, 1936). However, the guesses are far from zero in nearly all experimental settings (Nagel, 1995; Bosch-Domenech et al., 2002).

To explain these results, Camerer et al. (2004) and others have argued that individuals reason in steps as follows:

- Step 0: These individuals naively state their preferences without considering others' response.
- Step 1: These individuals think one step ahead, i.e., they believe that others are Step-0 players and best respond to that.
- Step 2: And so on....

Our results are consistent with Step 1 bounded-rationality, i.e., users believe that others will naively reveal their preferences without taking rejection concerns into account, i.e., increase their ranking for popular users. Thus, the best response for Step-1 users is to reduce their own preference-ranking for popular users (given rejection concerns). Our findings are in line with Nagel (1995) and Stahl and Wilson (1995) who also find that individuals in most experimental and real-life settings exhibit Step-1 bounded rationality.

Next, we examine the extent of rationality in users' messaging behavior. We first quantify users' beliefs about the likelihood of receiving a response to a first message that she sends as a function of the star rating of the receiver. Intuitively, if a user is concerned about rejection, then s/he should be more willing to send first messages to those users who are more likely to respond. To examine if this is the case, we regress a user j's likelihood of receiving a reply from i on i's star-rating (see Table A8 in Appendix D). The results from this regression suggest that a user j is more likely to receive a reply from one and three-star matches in response to her/his first messages. These response probabilities are consistent with j's likelihood of sending a message (model 13 in Table 5). Thus, when there is no game involved (i.e., users do not need to engage in multiple steps ahead reasoning), they seem to be able to form rational beliefs. Interestingly, this can also be interpreted as Step-1 thinking, and indeed, Step-1 thinking is fully-rational when no game is involved.

9 Conclusion

In this paper, we examine how users respond to the popularity information of potential partners in a mobile dating app. On the one hand, knowing that a potential partner is popular can increase her/his appeal. On the other hand, popular people are less likely to reciprocate, and hence users may strategically shade down their revealed preference for popular users to avoid the psychological costs of rejection. In our setting, users interact with each other by playing a ranking game, where they rank-order members of the opposite sex and are then matched based on a Stable Match Algorithm. A key piece of information shown to users during this process is a star-rating for each member of the opposite sex, which is a function of the past preference-rankings received. We quantify 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. To overcome the endogeneity between a user's star-rating and her unobserved attractiveness, we employ non-linear fixed-effects models.

We find that, everything else being constant, compared to two-star users: (1) three-star users receive lower preference-rankings during the game, however, (2) three-star users are more likely to receive first and reply messages after the game. This heterogeneity across outcomes can be linked to the perceived severity of rejection concerns and can be interpreted as strategic shading. The perceived risk of rejection is highest during the game, since users have no information on the other person's preferences. In contrast, after the game, a focal user knows that the person who s/he has been matched with must have preferred her/him sufficiently high for them to be matched in the first place. This alleviates rejection concerns. We also show that the effect of star-ratings is heterogeneous across user-specific observables: users who are less-attractive than average, specifically when they are ranking attractive partners, shade their preferences for popular users. These findings are consistent with our hypothesis that shading is driven by fear of rejection since unattractive users are more likely to suffer from rejection concerns. Finally, we also find that users' beliefs regarding rejection during the game are not fully rational. Instead, our results suggest that users believe that they are playing against naive users who reveal their preferences without considering others' response.

Our paper makes a few key contributions to the literature. We are the first to document negative returns to higher star-ratings in online platforms. Second, we empirically establish the presence of strategic shading in dating marketplaces: a user's popularity can have a negative impact on the revealed preferences for her/him. Third, we establish that users exhibit bounded rationality in real-world online settings that involve strategic multi-player games.

Our findings raise many interesting questions that can serve as avenues for future research. We study a platform where users are looking for short-term fun and flirtation rather than long-term partners or marriage. Moreover, in our setting, users only have a short amount of time to process information and make their decisions. It is not clear whether our results stem from the short-term nature of the relationship or from time pressure. Thus, we cannot comment on whether (and to what extent) our results would change if users were looking for long-term relationships/marriage and/or had more time to process information. Further research on these topics can have important implications for the design of dating and matching platforms.

References

P. D. Allison. Comparing Logit and Probit Coefficients Across Groups. *Sociological Methods & Research*, 28(2):186–208, 1999.

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.

A. Bosch-Domenech, J. G. Montalvo, R. Nagel, and A. Satorra. One, Two,(Three), Infinity,...: Newspaper and Lab Beauty-Contest Experiments. *American Economic Review*, 92(5):1687–1701, 2002.

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

C. F. Camerer, T.-H. Ho, and J.-K. Chong. A Cognitive Hierarchy Model of Games. *The Quarterly Journal of Economics*, 119(3):861–898, 2004.

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. 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.

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. What Makes You Click?—Mate Preferences in Online Dating. *Quantitative Marketing and Economics*, 8(4):393–427, 2010.

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. M. Keynes. *The General Theory of Employment, Interest, and Money*. 1936.
R. Kurzban and J. Weeden. HurryDate: Mate Preferences in Action. *Evolution and Human Behavior*, 26(3): 227–244, 2005.
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.
R. Nagel. Unraveling in Guessing Games: An Experimental Study. *The American Economic Review*, 85(5): 1313–1326, 1995.
J. Neyman and E. L. Scott. Consistent Estimates Based on Partially Consistent Observations. *Econometrica: Journal of the Econometric Society*, pages 1–32, 1948.
B. Pittel. The Average Number of Stable Matchings. *SIAM Journal on Discrete Mathematics*, 2(4):530–549, 1989.
A. E. Roth. Two-Sided Matching with Incomplete Information About Others' Preferences. *Games and Economic Behavior*, 1(2):191–209, 1989.
A. E. Roth and M. A. O. Sotomayor. *Two-sided matching*. Cambridge University Press, Cambridge, 1990.
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.
A. Smith. 15% of American Adults Have Used Online Dating Sites or Mobile Dating Apps. *Pew Research Center*, pages 1–12, 2016. https://www.pewinternet.org/2016/02/11/15-percent-of-american-adults-have-used-online-dating-sites-or-mobile-dating-apps/#fnref-15504-1.
A. T. Sorensen. Bestseller Lists and Product Variety. *The Journal of Industrial Economics*, 55(4):715–738, 2007.
D. O. Stahl and P. W. Wilson. On Players Models of Other Players: Theory and Experimental Evidence. *Games and Economic Behavior*, 10(1):218–254, 1995.
M. Sun. How Does the Variance of Product Ratings Matter? *Management Science*, 58(4):696–707, 2012.
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. Yu. Search, Selectivity, and Market Thickness in Two-Sided Markets. *Working Paper*, 2018.
J. Zhang. The Sound of Silence: Observational Learning in the US Kidney Market. *Marketing Science*, 29 (2):315–335, 2010.

Appendices

A Appendix for Robustness Checks

A.1 Effect of Stars on Preference-Rankings - Linear Model

We consider the following linear model:

$$pref_{ijt} = \beta_1 star1_{it} + \beta_2 star3_{it} + \gamma z_i + \eta_i + \epsilon_{ijt} \tag{A.1}$$

The main difference of these coefficients and variables to what we discussed in §6.1 is that here they relate directly 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 two possible estimation strategies here: (1) pooled OLS, which ignores the problem of correlated unobservables, and (2) fixed-effects model, which addressed the omitted variable bias due to η_i by employing a "within" transformation to subtract out the time-invariant user-specific variables.

A pooled OLS estimation strategy consists of pooling all the data across games and users, and simply running a multiple regression on this data. We consider two pooled OLS models – (1) a simple model that only includes star-ratings variables as the independent variables, and (2) a slightly more elaborate model that includes all the user-specific observables (z_i). The results from both these models are shown in models A1 and A2 in Table A1.

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

$$\overline{pref}_i = \beta_1 \overline{star1}_i + \beta_2 \overline{star3}_i + \gamma z_i + \eta_i + \overline{\epsilon}_i, \tag{A.2}$$

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, η_i are constant across time periods, and hence their averages are the same as the variables themselves. Next, we subtract Equation (A.2) from Equation (A.1) as follows:

$$pref_{ijt} - \overline{pref}_i = \beta_1 \left(star1_{it} - \overline{star1}_i \right) + \beta_2 \left(star3_{it} - \overline{star3}_i \right) + \left(\epsilon_{ijt} - \overline{\epsilon}_i \right) \tag{A.3}$$

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.3) and it gives us consistent estimates of β_1 and β_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.

A.2 Estimation Sample

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

First, we run the pooled ordered logit model on the subset of users who go through at least one star change and present the results in Table A2. We find that the magnitude and the direction of the estimates in

	(A1) (OLS)	(A2) (OLS)	(A3) (FE)
$star1_{it}$	-0.08946***	-0.07038***	0.01776
	(0.01422)	(0.01838)	(0.01126)
$star3_{it}$	0.03779***	0.03446**	-0.03100***
	(0.00971)	(0.01443)	(0.00913)
Controls		✓	
Constant	2.50031***	2.50513***	2.50006***
	(0.00113)	(0.00948)	(0.00034)
Individuals	24393	11639	3494
Observations	2980148	1580848	630160
R-Squared	0.00003	0.00260	0.00002

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
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. All standard errors are clustered at the user-level.

	(A4)
$star1_{it}$	-0.13888***
	(0.02369)
$star3_{it}$	0.05136***
	(0.01590)
μ_2	-1.09054***
	(0.00456)
μ_3	-0.00036
	(0.00420)
μ_4	1.09067***
	(0.00446)
Individuals	3494
Observations	630160

Table A2: Ordered logit estimates of the effect of star-rating on preference-rankings received (without fixed-effects), for a subset of users who experienced a star-change.

model A4 are similar to those for the full population model M1.

Second, we compare the distribution of user-specific observables for two groups of users – (1) users who saw no star change during the observation period, and (2) users who saw at least one star change in the observation period. We find that users who go through 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 experienced users who had played a large number of games in the past. It is important to note that this difference in past experience 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 than new users.

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

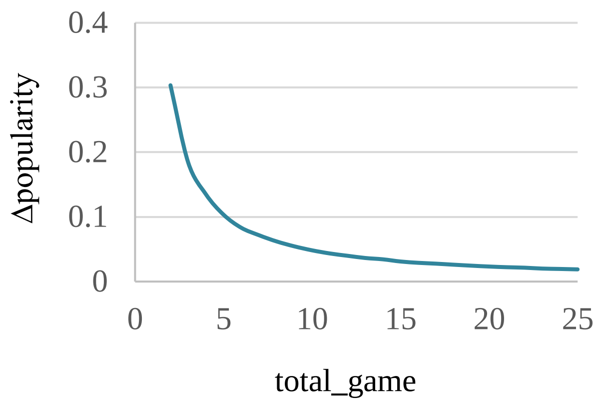

Figure A1: Change in popularity as a function of number of games played.

Next, for users who are new in the app, we find that the two sets of users who go through at least one star change and those who do not experience any star-rating change are very similar on their user-specific variable. The results from this comparison are presented in Table A3. Overall, there is sufficient empirical evidence to suggest that users who experience at least one star change and those who experience no star changes are similar on many important dimensions. Moreover, the pooled ordered logit estimates for the two subgroups are also similar. Thus, we expect the findings from the fixed-effects model can be interpreted as being largely applicable to the full population of users in the app.

Variables	Star Change	Mean	Std. Dev	Size	$Pr(\|T\| > \|t\|)$
age_i	No	21.950	7.393	2300	0.4487
	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 groups who experienced at least one star change.

	(A5)
$star1_{it}$	0.13535**
	(0.05499)
$star3_{it}$	-0.25426***
	(0.04068)
Individuals	1684
Observations	50668

Table A4: Ordered logit fixed-effects estimates of the effect of star-rating on preference-rankings received, for a subset of users who initiated a message at least once.

A.3 Within Game Correlation

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

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.

A.4 Table of Game-level Star Configurations

star2	star1	star3	Games
4	-	-	174,818
3	-	1	4,273
3	1	-	2,606
2	1	1	113
2	-	2	73
2	2	-	24
1	-	3	3
1	1	2	3
1	2	1	1

Table A6: The star configuration of four competitors in a game.

B Conditional Log Likelihood for the Fixed-effects Logit Model

To study the relationship between the users' messaging behavior with 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:

$$y^*_{ijt} = \beta_1 star1_{it} + \beta_2 star3_{it} + \gamma z_i + \eta_i + \epsilon_{ijt}, \tag{A.4}$$

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

$$Pr(y_{ijt} = 1 \mid star1_{it}, star3_{it}, z_i, \eta_i, \beta_1, \beta_2) = \frac{exp(\beta_1 star1_{it} + \beta_2 star3_{it} + \gamma z_i + \eta_i)}{1 + exp(\beta_1 star1_{it} + \beta_2 star3_{it} + \gamma z_i + \eta_i)} \tag{A.5}$$

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

$$LL(\beta_1, \beta_2, \gamma) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{k=0}^{1} \ln \left[Pr(y_{ijt} = k \mid star1_{it}, star3_{it}, z_i, \eta_i, \beta_1, \beta_2)^{I(y_{ijt}=k)} \right] \tag{A.6}$$

where N is the total number of users and T_i is the total number of games played by user i. Treating the η_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 β_1, β_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:

$$s_i = \sum_{t=1}^{T_i} (y_{ijt} \mid match_{ijt} = 1) \tag{A.7}$$

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:

$$B_i = \{d \in \{0,1\}^{T_i} \mid \sum_{t=1}^{T_i} (d_{jt} = s_i \mid match_{ijt} = 1)\} \tag{A.8}$$

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:

$$Pr\left(y_i \mid star1_{it}, star3_{it}, s_i, \beta_1, \beta_2\right) = \frac{exp\left(y_i.\left(\beta_1 star1_{it} + \beta_2 star3_{it}\right)\right)}{\sum_{d \in B_i} exp\left(d.\left(\beta_1 star1_{it} + \beta_2 star3_{it}\right)\right)} \tag{A.9}$$

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

$$CLL(\beta_1, \beta_2) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \ln\left[Pr(y_i \mid star1_{it}, star3_{it}, s_i, \beta_1, \beta_2)\right] \tag{A.10}$$

C Validation Check: Truthfulness Assumption

In §8.2, we assume that users state their preference-rankings truthfully during the game. We now formally define this assumption and establish its validity in our setting.

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, if and only if $EU(pref_{ijt}) > EU(pref_{i'jt})$.

This assumption ensures that the relationship between users' latent expected utilities for any pair of potential partners is consistent with their stated preference-ordering over them. Thus, it allows us to take the empirical patterns observed in the data (from §6) and map them to the underlying expected utilities in §8.2. In particular, it allows us to take the empirical results established in §6 to and express the relative ordering of the underlying expected utilities in Inequality (22).

Truth-telling is not always guaranteed in SMPs, and it has been shown that some parties may have an incentive to misrepresent their true preferences depending on the game settings and the matching algorithm used (Roth and Sotomayor, 1990).[21] For example, women may have incentive to mis-represent their true preferences if the platform uses a men-optimal stable matching algorithm; recall the discussion from §3.2.3.

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

[21] It has been shown that the incentive to manipulate true preferences is negligible in large markets (Demange et al., 1987; Pittel, 1989; Lee and Yariv, 2018; Lee, 2016b). However, in our game, there are only four players. So the results from the large-market literature may not generalize.

Match with	State true preferences	1^{st} and 2^{nd} preference misrepresentation	2^{nd} and 3^{rd} 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.

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 (Roth, 1989).[22] (And vice-versa for women-optimal stable matching mechanism.)

However, our platform does not use either a men-optimal or a women-optimal matching mechanism. Instead, as 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. While we cannot theoretically prove this, we now empirically establish that, on average, users cannot gain by mis-representing their preferences in our setting.

To do so, we consider 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:

$$\text{Pr(true first choice)} = \frac{\sum_{t=1}^{T} \sum_{j \in t} \text{I}(match\_level_{jt} = \text{true first choice}|pref_{jt}^{12}, pref_{-jt})}{8T}, \quad \text{(A.11)}$$

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., others' 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

[22]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).

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 $pref^{23}_{jt}$. 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 deviating 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.

D Appendix for §8.3 on Bounded Rationality

We now quantify the likelihood of that user j will receive a reply to her/his first message to a user i as a function of i's star-rating as follows:

$$reply_{jit} = \lambda_1 star1_{it} + \lambda_2 star3_{it} + \eta_i + \epsilon_{ijt} \quad \text{(A.12)}$$

The results from this regression are presented in Table A8.

	(A6)
$star1_{it}$	0.12923***
	(0.02933)
$star3_{it}$	0.09427***
	(0.01662)
Constant	0.11961***
	(0.00272)
Individuals	1972
Observations	9259

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A8: The effect of i's star-rating on j's probability of receiving a reply.