Star-Cursed Lovers: Strategic Shading in Online Dating

Behnaz G Bojd*  
University of Washington

Hema Yoganarasimhan*  
University of Washington

November 10, 2018

*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. Please address all correspondence to: behnaz@uw.edu, hemay@uw.edu.
Abstract

We examine how users respond to the popularity information of potential partners in a gamified mobile dating app. On the one hand, knowing that a potential match 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 rating game, where they rank-order members of the opposite sex and are then get matched based on a Stable Matching 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 their prior received ratings. We quantify the causal impact of a user’s star-rating on the preference-ratings 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 strong support for the presence of strategic shading in our data: (1) three-star users receive lower preference-ratings during the game compared to two-star users, and (2) two-star users receive fewer first messages after the game compared to one-star users. We also show that the effect of star-ratings is heterogeneous across both outcomes and user-specific observables, and that this heterogeneity (or extent of shading) can be directly linked to the severity of rejection concerns. Our findings have implications for the design of dating platforms and measurement of mate preferences in matching markets.

Keywords: Online dating, Star ratings, Strategic shading, Gender differences, Matching markets.
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. A recent Pew study found that 27% of people in the 18-24 age group have used some form of online dating (Smith, 2016), and the most recent Singles in America Survey found that the number one meeting place for singles is now online (Safronova, 2018). Indeed, industry revenues for online dating now exceed 1.6 billion dollars in the US (Madigan, 2017; MarketWatch, 2018). 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 provide a fast and gamified way to engage with others without protracted email exchanges (Ludden, 2016).

The gamification of mobile dating apps has changed the process of finding dates in fundamental ways, the most important of which is the short decision-making time. In desktop settings, users have time to leisurely browse a potential partner’s detailed profile and take the time to compose and/or respond to long email. In contrast, in gamified dating apps, users have a very short time to decide how much they like someone, and during that time they can only process a small amount of information. To aid users’ decision-making process, dating apps often restrict themselves to showing a small subset of salient pieces of information (e.g., photo, age). In some cases, they may also display a summary measure of the popularity or attractiveness of a potential partner (e.g., star-rating, likes) next to her/his profile. The idea being that it is easier to process one cumulative popularity measure instead of parsing through detailed profile data.

In this paper, we examine how users respond to the popularity information of potential partners in dating contexts. 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 (e.g., ratings, likelihood of initiating communication) for the 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 strategically shade down her/his revealed preference for a popular user. Thus, a priori, it is not obvious whether revealing a user’s popularity information will increase or decrease her demand in matching markets. Indeed, prior research in online dating has found no evidence for strategic shading (Hitsch et al., 2010b).

Identifying strategic shading separately from true preferences is a challenging empirical problem. Consider a scenario where less-attractive users express higher preference for users similar to them.
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 preference, and is still to decide his second and fourth preferred candidates.

(i.e., less attractive users) and lower preference for attractive users. This pattern could stem either from truly homophilic (or horizontal) preferences on physical attractiveness, or because these users strategically shade their revealed preferences in order to avoid rejection by more attractive users. Thus, in order to identify strategic shading, we need information (or variables) that affect users’ beliefs about a potential partner’s popularity, but does not affect their true underlying preference for the partner.

We empirically examine the issue of strategic shading using data from a popular mobile dating application in the United States in the 2014-15 time-frame. The app is targeted towards young users and employs ideas from gamification to make the process of finding partners fun. Using the app involves playing a game where players rank members of the opposite sex. The 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. (Throughout the paper, we use the term preference-rating instead of ranking to indicate users’ ordered preference to simplify exposition.) The platform then
uses these preference-ratings as the input 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 match and chat with them, if they choose to.

A key piece of information shown to users during this process is a “star-rating” for each member of the opposite-sex. A user’s star-rating is a cumulative measure of all the preference-ratings that s/he received in the past and users with higher past preference-ratings are shown with higher stars (ranging from three to one star). 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 any actual contact/engagement between previous raters and the ratee. They are thus pure popularity measures and do not help resolve asymmetric information (unlike purchase/experience based star-ratings used in e-commerce settings like Amazon).

Our goal is to quantify the causal impact of a user’s star-rating on the preference-ratings that she receives during a game and her likelihood of receiving messages after a game. However, we have a difficult confound here: attractive users who received high preference-ratings in the past (and hence have higher stars) are also likely to receive higher ratings 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 description, fun-loving pictures). This can give rise to an upward bias in our estimates of star-rating 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 past games. Thus, we can use the within-person variation in star ratings to causally infer the effect of a user’s star-rating on the demand for her in the marketplace. That is, if the same user receives lower rankings during a game (or is less likely to receive messages after a game) when s/he is shown with three stars instead of one or two stars, then we can causally establish the presence of strategic-shading in our data.

Taking this intuition to data, we investigate the effect of a user’s popularity (or star-ratings) on three revealed preference measures with different risks of rejection: 1) preference-ratings received during a game, 2) likelihood of receiving an initial message from the matched partner after the game, 3) likelihood of receiving a response to a message sent after the game. Since the first revealed preference measure (preference-rating) is an ordered discrete outcome, we model it 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 to the star-ratings variables.

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. Chamberlain proposed a general class of Conditional Maximum Likelihood (CML) estimators for non-linear models that condition on a subset of outcome, which in turn allows them condition-out all the fixed-effects (or nuisance parameters) and estimate only 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 a subset of data for identification. Das and Van Soest (1999) developed an Minimum Distance 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 ratings in our setting. For the two message-related binary outcome models, the CML and MD estimators are equivalent. So we simply use the 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. First, we find strong evidence in favor of strategic shading in the case of preference-ratings – everything else being constant, three-star users receive lower preference-ratings compared to two-star users. Next, in the case of first messages, we find some evidence in support of strategic shading – both one-star and three-star users are more likely to receive first-messages compared to two-star users. The positive effect of one star compared to two stars suggests that there is some shading happening. Finally, in the case of reply messages, we find no evidence in support of shading, i.e., we see that star-ratings have a positive effect on the likelihood of receiving a response.

Our findings are consistent with the hypothesis that users may be shading down their preferences for popular users to avoid the psychological cost of rejection. Note that when a user is rating partners during the game, s/he has no information on the other person’s preferences, thus the potential for being rejected (i.e., not being matched) is highest. In contrast, after the game, a focal user knows that the person who s/he has been matched with must have rated her/him sufficiently high for them to be matched in the first place. This alleviates rejection concerns to some extent. In the reply case, the user has already received a message from her/his match and is considering whether to respond or not. Here rejection is not a concern at all since the other party has already expressed interest. So naturally we do not see shading in this case.

Finally, we examine the heterogeneity in the effect of star-ratings across a series of user-level observables. Interestingly, we find that the negative effect of star-ratings on preference ratings is mainly driven by younger and less-attractive users. This is again consistent with our hypothesis that shading is driven by fear of rejection since younger and unattractive users are more likely to
suffer from rejection concerns. Next, in the case of first-messages, the extent of strategic shading is moderated by the characteristics of the match; specifically, users are more likely to initiate communication with one-star matches (compared to two-star users) only when the matched partner has high baseline appeal (older and/or more physically attractive match).

In sum, our paper makes two key contributions to the literature. We are the first to 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. Second, we show that the effect of popularity ratings is heterogeneous across both outcomes and user-specific observables, and that this heterogeneity (or extent of shading) can be directly linked to the severity of rejection concerns.

Our findings are of relevance to both managers and researchers. First, from a methodological perspective, our research provides methods to separately identify mate preferences and strategic shading in dating markets. For online dating platforms, knowing who prefers who and under what circumstances can help build better partner recommendations, search functionalities, and reduce the search costs associated with finding the right partner. From researchers’ perspectives, correctly measuring mate preferences is of fundamental interest and importance in understanding the societal structure today. Second, from a substantive and policy perspective, our findings have implications for platform design, e.g., would popular users prefer a platform that prominently displays their popularity information or one that doesn’t? A large stream of prior literature has shown that prior sales and/or ratings information can have a positive impact on the demand for products on e-commerce platforms (Sorensen, 2007; Tucker and Zhang, 2011). However, we show that these findings do not always translate to two-sided dating markets where inter-personal interactions and psychological costs play a larger role.

The remainder of this paper is organized as follows. In §2 we discuss the related literature. Then we introduce the setting and data in §3 and §4 respectively. In §5 we present descriptive analyses that shows suggestive evidence for the presence of strategic shading in our data. Next, in §6 and §7 we present our empirical specifications, estimation and identification approaches, and establish the causal impact of star ratings on preference ratings and messaging behavior, respectively. Finally, in §8 we conclude with a discussion of our main findings and avenues for future research.

## 2 Related Literature

Our paper relates to a long-standing literature on the measurement and understanding of mate preferences and sorting patterns in marriage and dating markets, starting with Becker (1973). A well-established empirical pattern documented in this literature is the prevalence of strongly assortative matching patterns in equilibrium, i.e., people with similar attributes (e.g., age, physical attractiveness, education, income, race) tend to match with (or marry) each other. A key question
that researchers have grappled with here is the source of assortative matching patterns. This question is fundamentally hard because three very different mechanisms can give rise to assortative patterns (Kalmijn 1998; Browning et al. 2014). First, this could stem from search frictions. For example, college-educated people are more likely to be working in white-collar jobs and have common social-networks. Thus, a college-educated person is more likely to meet and marry another college-educated person; not because of any specific preferences, rather due to the lower search costs associated with meeting others of the same type. Second, it could be due to vertical preferences. For example, if people have vertical preferences on physical attractiveness (more attractive is preferred by everyone), then in equilibrium, people will match with others of similar attractiveness levels even in the absence of search frictions. This result is a natural outcome of any stable-matching procedure; for example, the famous Gale-Shapley matching algorithm predicts assortative matches based on vertical preferences (Gale and Shapley 1962; Becker 1973). Finally, assortative matching can also stem from purely horizontal preferences, i.e., people may simply prefer partners who are similar to them on some dimensions such as race.

Early empirical work on mating and marriage patterns mostly relied on data on equilibrium outcomes, i.e., they use data on observed marriages to estimate population-level mate preferences under the assumption of no search frictions (Wong 2003; Choo and Siow 2006). Because they lacked data on individual-level preference orderings and the search process that led to the marriage outcomes, they were unable to speak to the source of the assortative patterns. 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).

The speed-dating studies provide valuable insights into factors that affect mate preferences and how these preferences vary by gender (Kurzban and Weeden 2005; Fisman et al. 2006, 2008; Eastwick and Finkel 2008). For example, Fisman et al. (2006) show that women put greater weight on the partner’s intelligence and race, whereas men put greater weight on physical attractiveness. Similarly, using data from an online dating website, Hitsch and his co-authors show that user preferences are horizontal along some dimensions, e.g., race, education, and vertical along others, e.g., physical attractiveness, income (Hitsch et al. 2010a,b).

An important concern when measuring user preferences is the possibility of strategic behavior – users may shade down their “revealed” preference for popular or attractive users to avoid the psychological cost of rejection (Cameron et al. 2013). If users shade their revealed preferences,
and we do not explicitly account for this in the estimation, then our estimates of user preferences
will be biased. For example, if less-attractive users strategically express lower preference for
highly-attractive users, then we cannot tell if this is due to horizontal preferences on attractiveness
or due to shading. [Hitsch et al. (2010b)] employ empirical tests and show that strategic behavior 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 perception of
popularity without affecting the actual attractiveness of a user at an individual-level.

In this paper, we exploit the within-person variation in star-ratings (or popularity of the user
in the platform) and find strong evidence in support of shading, i.e., the same user receives lower
preference ratings when s/he is shown with higher stars. We are the first paper to empirically
establish the presence of strategic shading in dating markets and identify the source of shading
by examining shading in three different revealed preference measures that have different risks of
rejection. Our work thus contributes both methodologically and substantively to the literature on
mate-preferences and identification of strategic behavior.

Finally, our paper also relates to the literature on the design and impact of online dating platforms
(Bapna et al. 2016; Lee, 2016a). Our findings suggest that it may not be optimal to reveal users’
popularity information in these settings unlike e-commerce sites.

3 Setting

3.1 Mobile Dating App

Our data come from a popular online dating iOS mobile application in the United States. The app is
different from standard dating services like Match.com in two fundamental ways. First, the app (or
platform) is targeted at a younger demographic, and those using it are looking for fun and flirtation
rather long-term dating/marriage partners. Second, the app uses ideas from gamification to make
the process of finding potential partners fun. The app does not allow users to browse the profiles of
other members; instead those interested in finding a partner have to play a game. During the game,
each player rates members of the opposite sex allocated to the same game-room as him/her, and
is then allocated a match by the platform based on these ratings. Thus, the application creates a
unique dating experience by combining elements of gaming with virtual romance.

To join and use the app, users need a Facebook ID[1] 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.

[1]Users without a Facebook account cannot use this app.
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.

We now provide a detailed description of the game that users play in the app.

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 active 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 till 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 small age bandwidth. Thus, conditional on geography and age, 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, current location and their star-rating (see §4.3 for a detailed description on star-ratings). Tapping on the short profiles 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.2

Two points are worth noting here. First, players do not know the identities 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 game.

2If 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 of game-completion; over 97% of games started are completed.
same game (and their attributes). 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). Thus, while choosing their rank-orderings, they cannot use information on other player’s preferences to make their own choices.

### 3.2.3 Match Allocation

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

There are a few noteworthy points about the SMP. First, for any combination of preferences, there always exists at least one 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, have been shown that a “Men-proposing Gale-Shapley Deferred Acceptance algorithm” is optimal from the perspective of men, i.e., none of the men can do better using a different algorithm and vice-versa for women.

In our case, the platform first calculates all possible solutions for a game by considering all combinations of matches and checking for stability. While this approach is sub-optimal from a computational perspective, it has the advantage of being able to obtain all the solutions to a given SMP. Further, compute time is not an issue here since we only have four players on each side. 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 best 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 matched with and sum this number over all players. The intuition here is to pick the solution that, on average, gives each player her highest preference (or lowest numerical rank). Thus,

---

3We briefly describe the Men-proposing Gale-Shapley DA 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).
the platform does not optimize for either men or women, and 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 at the end of the game (as shown in the left 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 on 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 future, if they choose to. Finally, note that users cannot start or receive any communication from other players in the game with who they have not been matched.
3.3 Player’s Ranking Strategy

We now briefly discuss the optimal ranking strategy for a player. The standard SMP problem set-up assumes complete information and non-transferable utility. Under these conditions, depending on the matching algorithm, some parties may have an incentive to misrepresent their true preferences (Roth and Sotomayor, 1990). However, in our case, agents’ preferences are private information and they can neither collude nor can they truncate their list of preference ratings (i.e., they cannot strategically choose to only rank their top two choices and refuse to rank their bottom choices). Under such circumstances, it has been shown that the incentive to manipulate true preferences is negligible, i.e., even the worse stable matching partners give utilities that are asymptotically close to the upper bound (Demange et al., 1987; Pittel, 1989; Lee and Yariv, 2018; Lee, 2016b). Hence, we assume that users are truthful when submitting their ratings during a game.

4 Data

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

4.1 User-level Data

We 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.

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, where 1 = High-school, 2 = College, and 3 = Graduate school.
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.

In principle, many of these attributes may change over time. However, we only have a snap shot of them at the end of the data collection period; thus, we treat them as time-invariant attributes.
<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>(Min, Max)</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>age$_i$</td>
<td>21.53</td>
<td>5.41</td>
<td>19</td>
<td>21</td>
<td>22</td>
<td>(13, 109)</td>
<td>22024</td>
</tr>
<tr>
<td>bio$_i$</td>
<td>67.04</td>
<td>275.58</td>
<td>0</td>
<td>0</td>
<td>63</td>
<td>(0, 29519)</td>
<td>22948</td>
</tr>
<tr>
<td>employment$_i$</td>
<td>2.05</td>
<td>1.59</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>(1, 68)</td>
<td>15579</td>
</tr>
<tr>
<td>initial_game$_i$</td>
<td>59.50</td>
<td>64.32</td>
<td>0</td>
<td>48</td>
<td>90</td>
<td>(0, 2146)</td>
<td>24653</td>
</tr>
<tr>
<td>total_game$_i$</td>
<td>31.27</td>
<td>37.90</td>
<td>6</td>
<td>18</td>
<td>45</td>
<td>(1,1069)</td>
<td>24653</td>
</tr>
<tr>
<td>picture$_i$</td>
<td>4.26</td>
<td>1.01</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>(0, 6)</td>
<td>22669</td>
</tr>
<tr>
<td>picture.rating$_i$</td>
<td>0.00</td>
<td>0.68</td>
<td>-0.52</td>
<td>-0.09</td>
<td>0.43</td>
<td>(-2.88, 3.29)</td>
<td>17739</td>
</tr>
<tr>
<td>gender$_i$</td>
<td>(0) female: 42.45%</td>
<td>(1) male: 57.55%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>education$_i$</td>
<td>(1) high-school: 19.24%</td>
<td>(2) college: 78.12%</td>
<td>(3) graduate: 2.64%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Summary statistics of user-level data.

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

8. picture$_i$: Number of uploaded pictures in the dating profile.

In addition to the above variables, 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), in a scale of 1 to 7, with 1 being “not at all attractive” and 7 being “very attractive”. The subjects were University of Washington undergraduate students, with an equal fraction of male and female, and their ages ranged between 18-25 and 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)](https://doi.org/10.1016/S0092-8674(97)00130-1). We then take the average of all the standardized ratings that user $i$’s picture received in our study and denote it as:

9. picture.rating$_i$: The average physical attractiveness rating of user $i$’s profile picture.

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

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

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 ordered preference rating that they gave each other, their match outcome, and their post-game interactions. We describe these variables in detail below.

1. $rating_{ijt}$: An integer variable that denotes the preference rating 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.

   In practice, 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 ratings, such that rank of one denotes a preference rating of four, rank of two indicates a preference rating of three, and so on. The transformed variable $rating$ 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 makes it confusing in 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 each person $i$ in a game, this variable is one for her pairings and zero for all other pairings.

3. $first_{ijt}$: A dummy variable indicating whether user $i$ receives the first message from the player 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 $match_{ijt} = 0$.

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 $first_{ijt} = 0$. 


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

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>(Min, Max)</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>$rating_{ijt}$</td>
<td>2.5</td>
<td>1.12</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>(1, 4)</td>
<td>3008560</td>
</tr>
<tr>
<td>$match_{ijt}$</td>
<td>0.25</td>
<td>0.43</td>
<td>0</td>
<td>0</td>
<td>0.5</td>
<td>(0, 1)</td>
<td>3008560</td>
</tr>
<tr>
<td>$first_{ijt}$</td>
<td>0.05</td>
<td>0.23</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(0, 1)</td>
<td>713014</td>
</tr>
<tr>
<td>$reply_{ijt}$</td>
<td>0.08</td>
<td>0.28</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(0, 1)</td>
<td>39377</td>
</tr>
</tbody>
</table>

The summary statistics of these variables are shown in Table 2. Note that the sample size of $rating$ and $match$ reflects the fact that there are 32 data points per game. Similarly, the sample size of the $first_{ijt}$ and $reply_{ijt}$ reflect the fact there are eight players and hence eight message data points per game.

The distributions of $rating$ and $match$ are fixed by the game structure and their summary statistics are as expected. However, it is worth noting that the mean of $first_{ijt}$ is around 0.05 (from 713,014 matches, only 39,377 messages were initiated). Also, 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., 2010b).

### 4.3 User-Game level data

We now describe three user-game level variables, i.e., user-specific data that varies with each game, but is not specific to pairs of users within a game.

1. **$match\_level_{it}$**: An integer variable that denotes how much user $i$ prefers his match in game $t$; and it is calculated as follow:

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

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

Eight users participate in each game and each user receives four preference ratings from players of the opposite sex. So we have a total of $8 \times 4 = 32$ preference ratings 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 $rating_{ijt}$ and $match_{ijt}$ should be the number of games $(94,386) \times 32 = 3,020,652$. 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.
3. \( \text{star}_{it} \): The star-rating of user \( i \) in game \( t \) indicates the star-rating that the user is shown with in that particular game. See Figure[1] for an example.

Users know their own star-rating at each game. This star-rating is updated in real time after each game and is calculated as follows:

\[
\text{star}_{it} = \begin{cases} 
1, & \text{if } 1 \leq \text{popularity}_{it} < 2 \\
2, & \text{if } 2 \leq \text{popularity}_{it} < 3 \\
3, & \text{if } 3 \leq \text{popularity}_{it} \leq 4,
\end{cases} \tag{2}
\]

where popularity is defined as the average of the preference ratings that user \( i \) has received before the \( t^{th} \) game.

\[
\text{popularity}_{it} = \frac{\sum_{t=1}^{\text{total game}_{it}} \sum_{j=1}^{4} \text{rating}_{ijt}}{4 \times \text{total game}_{it}} \tag{3}
\]

Note that the platform does not reveal a user’s popularity scores to her/him or to anyone else in the platform. Instead, it only reveals users’ star-rating.

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.

The summary statistics of all the user-game level variables are shown in Table[3]. There are a few
Table 3: Summary statistics of user-game level variables.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>(Min, Max)</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>match_level&lt;sub&gt;it&lt;/sub&gt;</td>
<td>3.19</td>
<td>0.95</td>
<td>3</td>
<td>3</td>
<td>4</td>
<td>(1, 4)</td>
<td>752140</td>
</tr>
<tr>
<td>total_game&lt;sub&gt;it&lt;/sub&gt;</td>
<td>74.75</td>
<td>74.25</td>
<td>29</td>
<td>59</td>
<td>97</td>
<td>(0, 2194)</td>
<td>752140</td>
</tr>
<tr>
<td>star&lt;sub&gt;it&lt;/sub&gt;</td>
<td>2.00</td>
<td>0.10</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>(1,3)</td>
<td>745037</td>
</tr>
</tbody>
</table>

interesting points of note. First, the average match_level is 3.19. Recall that if user i is matched with her most preferred player in game t, then in that game, match_level<sub>it</sub> = 4. Instead if she is matched with her least preferred player, her match_level<sub>it</sub> = 1. If preferences were purely vertical, i.e., if all the men in a game had the same rank-ordering for women (and vice-versa), then the average of match_level would be 2.5. Instead if preferences were purely horizontal, then the mean of match_level would be 4, on average. The fact that the average of match_level lies in between (at 3.19) suggests that users’ preference ratings are a combination of vertical and horizontal attributes. We discuss these ideas in greater detail in §5.1. Next, we find that the median of total_game<sub>it</sub> 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.

5 Descriptive Analysis

We now present a series of descriptive analyses summarizing some data patterns of interest. In §5.1, we start with a broad discussion of if and how we can separate true preferences from strategic behavior in revealed preferences data. Next, in §5.2, we examine the correlation between a user’s physical attractiveness and the revealed preference for her/him. In the process, we present some suggestive evidence in support of strategic shading of preferences in our data. Finally, in §5.3, we examine the relationship between the revealed preference measures and star-ratings, which is our focal variable of interest.

5.1 What Drives Revealed Preferences: True Preferences vs. Strategic Behavior?

Understanding the factors that drive an individual’s revealed preference for a potential partner is an important first step towards gaining insights on user behavior in this market. Intuitively, the revealed preference of user j for user i is likely to be a function of: (1) the true preferences of the user j (the person giving the rating or sending the message) for user i, and (2) strategic behavior of j. Thus, revealed preferences are not the same as true preferences unless we can rule out strategic behavior. A classic example of preference shading in dating markets is when users strategically avoid very attractive or hard to get partners in order to avoid the psychological cost of rejection.

<sup>6</sup>In this case, the man and woman ranked one would be matched with each other, the man and woman ranked two would be matched, and so on. This, in turn, would lead to an average match level of \(\frac{1+1+2+2+3+3+4+4}{8} = 2.5\).
First, let us consider a simple scenario where users truthfully reveal their true preferences. In that case, j’s preferences will depend on two factors – 1) user i’s attributes and 2) interaction between user i and j’s attributes. For any attribute x, if there is a main effect of x, we can interpret x as a vertical attribute, i.e., more of the attribute is perceived as better (or worse) by all users. In contrast, if attribute x affects preferences mainly through interaction effects between x and x, then it can be interpreted as horizontal. Specifically, if users’ preferences are homophilic, then we would find that users with similar values of x prefer each other. Thus, if users truthfully reveal their preferences, we can simply regress revealed preferences on the attributes of user i and user j, and infer which attributes are vertical and which are horizontal.

However, if users strategically shade their preferences, then it is econometrically impossible to distinguish between true horizontal preferences and strategic shading without additional information and/or restrictions on the data generating process. For example, if we find that less attractive users give higher ratings or send messages to users similar to them (i.e., less attractive users), then this effect could stem from either truly homophilic preferences on physical attractiveness, or because these users strategically shaded their revealed preferences to avoid rejection by more attractive users. Thus, evidence for homophily in revealed preference is suggestive evidence for the possibility of strategic behavior. In §5.2 we take these ideas to data.

5.2 Correlation Between Revealed Preferences and Physical Attractiveness

In our data, we see revealed preference of user j for user i through three variables – (1) the rating that j gives i in game t, and (2) whether j initiates communication with i (i.e., sends first message to i) or not, conditional on being matched with him/her, or (3) whether j replies to the first message initiated by i, conditional on i and j being matched and i and j initiating communication with i.

We now examine the relationship between a user’s attractiveness ratings (from our lab study), picture_rating, and the revealed preferences for her/him using the following regression:

\[
y_{ijt} = \text{constant} + \alpha_1 \text{picture_rating}_i + \alpha_2 (\text{picture_rating}_j - \text{picture_rating}_i)^+ \\
+ \alpha_3 (\text{picture_rating}_j - \text{picture_rating}_i)^- ,
\]

where \(y_{ijt}\) represents one of the revealed preference measures: rating\(_{ijt}\), first\(_{ijt}\), or reply\(_{ijt}\).

The results from these regressions are presented in Table 4. First, we find that the main effect of picture_rating\(_i\) on rating\(_{ijt}\) is positive and significant (see Model M1). That is, users with higher physical attractiveness ratings also receive higher preference ratings. This is evidence in support of the idea that physical attractiveness is a vertical attribute. Further, the negative coefficients on the interaction measures indicate that users rate others similar to them higher. Specifically, the
### Table 4: Correlation between physical attractiveness and revealed preferences.

<table>
<thead>
<tr>
<th></th>
<th>(M1)</th>
<th>(M2)</th>
<th>(M3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ratingijt</td>
<td>firstijt</td>
<td>replyijt</td>
</tr>
<tr>
<td>picture_ratingi</td>
<td>0.07244∗∗∗</td>
<td>0.32003∗∗∗</td>
<td>0.15103∗∗∗</td>
</tr>
<tr>
<td></td>
<td>(0.00209)</td>
<td>(0.01753)</td>
<td>(0.05801)</td>
</tr>
<tr>
<td>(picture_ratingj − picture_ratingi)⁺</td>
<td>-0.00746∗∗∗</td>
<td>-0.14146∗∗∗</td>
<td>-0.40498∗∗∗</td>
</tr>
<tr>
<td></td>
<td>(0.00163)</td>
<td>(0.01920)</td>
<td>(0.05508)</td>
</tr>
<tr>
<td>(picture_ratingj − picture_ratingi)⁻</td>
<td>-0.00744∗∗∗</td>
<td>0.04739∗∗∗</td>
<td>0.40505∗∗∗</td>
</tr>
<tr>
<td></td>
<td>(0.00220)</td>
<td>(0.01638)</td>
<td>(0.05475)</td>
</tr>
<tr>
<td>Constant</td>
<td>2.50820∗∗∗</td>
<td>-2.71281∗∗∗</td>
<td>-2.49892∗∗∗</td>
</tr>
<tr>
<td></td>
<td>(0.00155)</td>
<td>(0.01474)</td>
<td>(0.04851)</td>
</tr>
<tr>
<td>Individuals</td>
<td>17739</td>
<td>17702</td>
<td>4946</td>
</tr>
<tr>
<td>Observations</td>
<td>226335</td>
<td>532606</td>
<td>33090</td>
</tr>
<tr>
<td>R-Squared</td>
<td>0.00193</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Standard errors in parentheses

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

A negative coefficient on \((picture\_rating_j − picture\_rating_i)^+\) can only be attributed to horizontal preferences since in this case j is more attractive than i and hence unlikely to worry about rejection. However, the negative coefficient on \((picture\_rating_j − picture\_rating_i)^-\) implies that if user j is less attractive than user i, during the game s/he avoids user i. This can stem from the presence of horizontal preferences, or from strategic shading (i.e., j’s avoidance of i when i is more attractive), or a combination of both. In sum, Model M1 presents evidence for all three possibilities – (a) vertical preferences, (b) horizontal preferences, and (c) strategic shading.

Next, consider the relationship between a user’s attractiveness rating and her likelihood of receiving a first message from her/his match \((first_{ijt})\) in Model M2 and that of a reply in \((reply_{ijt})\) in Model M3. As before, \(picture\_rating\) has a positive effect of and \((picture\_rating_j − picture\_rating_i)^+\) has a negative effect, and the interpretations for the vertical and horizontal preferences here are similar to those discussed in the case of \(rating\) in the previous paragraph. Interestingly though, the coefficient of \((picture\_rating_j − picture\_rating_i)^-\) is positive in these two cases, i.e., j seems more likely to send the first message when i is more attractive than j. This suggests that rejection concerns do not play as big a role in messaging behavior. This is

---

7Though, of course, we cannot rule out shading even in this case since individuals with higher attractiveness ratings may still avoid users with lower attractiveness ratings if other attributes correlated with attractiveness also affect user preferences.
(a) Correlation between a user’s star-rating in a game and received preference ratings.

(b) Correlation between a user’s star-rating in a game and received first and reply messages.

Figure 4: Correlation between a user’s star-rating and revealed preferences.

understandable since messaging happens after matching, i.e., the other user has already expressed a sufficiently high preference for the focal user to be matched with them in the first place.

Our findings are particularly interesting for two reasons. First, they provide the first empirical evidence in support of the idea that physical attractiveness can function as both a vertical and horizontal attribute. This is especially important since previous research on mate choice finds physical attractiveness to be a purely vertical attribute (Fisman et al., 2006; Hitsch et al., 2010b). Second, we find suggestive evidence for the possibility of strategic behavior.

5.3 Correlation Between Star-ratings and Revealed Preferences

We now examine whether a user’s star-rating affect the preference ratings, first messages, and reply messages that s/he receives. Recall that user $i$’s star-rating in game $t$, $\text{star}_{it}$, is a summary measure of the preference ratings that s/he received in the past and captures $i$’s popularity in the app. The relationship between a user’s star-rating in a given game and the average preference rating that s/he receives in that game is illustrated in Figure 4a. The first bar represents the average preference rating for all user-game data points with star-rating $= 1$, i.e., $\frac{\sum_i \sum_t \sum_j (\text{rating}_{ijt} | \text{star}_{it}=1)}{4 \times \sum_i \sum_t I(\text{star}_{it}=1)}$, the second bar represents the average preference rating of user-game data points with star-rating $= 2$, and so on. Similarly, Figure 4b shows the relationship between a user’s star-rating and the probability of her receiving the first message and receiving a reply if she initiates a message.

We see that users with higher star-ratings receive higher preference ratings and are also more
Table 5: Correlation between star-ratings and attractiveness ratings.

<table>
<thead>
<tr>
<th>star</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>(Min, Max)</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.21</td>
<td>0.67</td>
<td>-0.69</td>
<td>-0.37</td>
<td>0.15</td>
<td>(-1.61, 3.08)</td>
<td>2116</td>
</tr>
<tr>
<td>2</td>
<td>-0.01</td>
<td>0.69</td>
<td>-0.53</td>
<td>-0.12</td>
<td>0.42</td>
<td>(-2.88, 3.29)</td>
<td>655258</td>
</tr>
<tr>
<td>3</td>
<td>0.14</td>
<td>0.69</td>
<td>-0.41</td>
<td>0.06</td>
<td>0.68</td>
<td>(-2.42, 3.07)</td>
<td>3871</td>
</tr>
</tbody>
</table>

(a) Comparison of average received preference ratings for individuals who were shown with a minimum of one star and a maximum of two stars.

(b) Comparison of average received preference ratings for individuals who were shown with a minimum of two stars and a maximum of three stars.

Figure 5: Comparison of average received preference within an individual.

likely to receive the first messages and replies. However, there is an obvious issue of correlated unobservables here, i.e., those 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 examine the distributions of picture_rating for all user-game data points for each star-rating (see Table 5). We find that users with higher star-ratings, on average, also have higher physical attractiveness rating. Thus, the effects shown in Figure 4 cannot be interpreted as causal.

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 revealed preferences for the same individual when s/he is shown with one star vs. two stars, then our estimate of the effect of star-rating is 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.

We now focus on users who experienced at least one change in their star-rating during our observation period. First, we consider individuals who were shown with a minimum of one star and

---

9We have 3,494 users who experiences a star change. 1,287 users were represented with a minimum star of one star and a maximum of two stars; and 2,185 users were represented 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, and they are therefore excluded from this model-free analysis.
a maximum of two stars. For each of these individuals, we calculate two averages: (1) the average of preference ratings received in games where s/he is shown with one star, and (2) the average of preference ratings received in games where s/he is shown with two stars. We then take the average of these two values across all individuals and present the results in Figure 5a. As we can see, on average, the same set of users receive higher preference ratings when they are shown with one star compared to 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 and show the results in Figure 5b. We find that, on average, the same set of users receive higher preference ratings when they are shown with two stars compared to three stars. In sum, Figure 5 suggests that higher star-ratings leads to lower preference ratings, i.e., users avoid those with higher stars! This is suggestive evidence in support of the idea that users strategically shade down their preferences for users they perceive to be more popular due to fear of rejection by such users.

Note that the direction of the effect of star-rating on preference ratings in Figure 5 are the exact opposite of what we saw in Figure 4a. This discrepancy implies that controlling for the endogeneity between star-ratings and unobserved factors that affect user attractiveness is essential to deriving the causal impact of star-ratings in our setting.

Next, we perform a similar analysis on users’ messaging behavior after the game and present the results in Figure 6. As 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. We thus see some limited evidence for strategic
shading of preferences in first messages. In contrast, in the case of reply, the same set of users are more likely to get a reply when shown with higher star ratings. In short, we find no evidence for strategic shading in replying behavior.

It is interesting that the extent of strategic shading seems to be lower when it comes to initiating communication after the game (first message) compared to ratings in the game, and is non-existent when it comes to replying to messages. These findings are consistent with our idea that users may be shading down their preferences for popular users to avoid the psychological cost of rejection. Note that when a user is rating partners during the game, s/he has no information on the other person’s preferences, thus the potential for being rejected is highest. In contrast, after the game, a focal user knows that the person who s/he has been matched with must have rated her/him sufficiently high for them to be matched in the first place. This alleviates rejection concerns to some extent. Finally, when the other person has already sent a message indicating interest, rejection is not a concern at all, and naturally we do not see shading in this case.

In sum, there is suggestive evidence in support of strategic shading in certain outcome measures in our data and some hints on the mechanism behind shading. In the rest of the paper, we focus on deriving the unbiased causal effects of star-ratings using econometric methods.

6 Effect of Stars on Preference Ratings

In this section, we formalize the causal impact of a user’s star-rating on the preference ratings that s/he receives. Since preference ratings are ordinal, we use an Ordered Logit model. In §6.1 and §6.2, we present the model specification and estimation. We discuss our findings in §6.3 and present some robustness checks in §6.4. Finally, in §6.5, we examine the heterogeneity in the effect of star-ratings across a series of user-level observables, and try to pin down the source of these effects.

6.1 Model Specification

The outcome variable, rating, is an ordinal integer value going from 4 to 1, with four indicating the highest preference rating and one representing the lowest preference rating. An Ordered Logit model relates the observed outcome variable rating\(_{ijt}\) to a latent variable rating\(_{ijt}^*\) where:

\[
\text{rating}_{ijt}^* = \beta_1 \text{star}1_{it} + \beta_2 \text{star}3_{it} + \gamma z_i + \eta_i + \epsilon_{ijt},
\]

The latent variable rating\(_{ijt}^*\) is thus modeled as a linear function of:

- \text{star}1_{it}, \text{star}3_{it} – indicator variables for the star-rating of user \(i\) in game \(t\), where \text{star}2 is considered as the base.
- \(z_i\) – set of user-specific observables that can affect \(j\)’s rating of \(i\), e.g., age of \(i\).
• \( \eta_i \) – set of unobservable (to the researcher) characteristics of user \( i \) that is visible to \( j \) and affects \( j \)’s rating of \( i \). These could include the aspects of physical attractiveness of user \( i \) not captured in our lab study (e.g., through the other photos of the user), details in her/his bio description, employment details, etc.

• \( \epsilon_{ijt} \) – These are factors uncorrelated to the star-rating of user \( i \) that can affect the rating s/he receives from \( j \) in game \( t \). Three key sets of variables are subsumed here. First, it includes \( j \)’s own attributes (both observable \( z_j \) and unobservable \( \eta_j \)) 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. So technically there is no correlation between the attributes of users within a game. Third, \( \epsilon_{ijt} \) may include idiosyncratic factors that affect \( j \)’s rating 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 \( \epsilon_{ijt} \)s are IID with a logistic cumulative distribution.

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

We then model the relationship between \( \text{rating}_{ijt} \) and \( \text{rating}_{ijt}^* \) as follows:

\[
\text{rating}_{ijt} = k \quad \text{if} \quad \mu_k < \text{rating}_{ijt}^* \leq \mu_{k+1} \quad k = 1, 2, 3, 4,
\]

where the thresholds \( \mu_k \) 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, \( \text{rating}_{ijt} \) can take \( K = 4 \) possible values, denoted by \( k \). Because the error terms are drawn from a logistic distribution, we can write the cumulative probability function of \( \epsilon_{ijt} \) as

\[
F(\epsilon_{ijt} \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_i) = \frac{1}{1 + \exp(-\epsilon_{ijt})} \equiv \Lambda(\epsilon_{ijt}),
\]

where \( X_{it} = \{\text{star}_{1it}, \text{star}_{3it}, z_i\} \). Therefore, the probability of observing outcome \( k \) in game \( t \) for a pair of users (where \( j \) rates \( i \)) can be written as:

\[
\text{Pr}(\text{rating}_{ijt} = k \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_i) = \Lambda(\mu_{k+1} - \beta_1\text{star}_{1it} - \beta_2\text{star}_{3it} - \gamma z_i - \eta_i) - \Lambda(\mu_k - \beta_1\text{star}_{1it} - \beta_2\text{star}_{3it} - \gamma z_i - \eta_i)
\]

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

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

(8)

where $N$ is the total number of users observed and $T_i$ is the total number of games played by user $i$.

6.2 Estimation

There are two possible estimation strategies to infer the effect of star-ratings (coefficients $\beta_1$ and $\beta_2$): (1) A Pooled estimation strategy, where we ignore the user-specific unobservables $\eta_i$, and (2) a Fixed-effects approach, where we allow the user-specific unobservables $\eta_i$ to be arbitrarily correlated with the star-ratings variables. 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 $\eta_i$, 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 $\beta_1$ and $\beta_2$ after controlling for $\eta_i$.

A naive approach to estimation with fixed-effects is to treat the $\eta_i$’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 $\beta_1$ and $\beta_2$ 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. We now describe how the Chamberlain estimator can be applied to our setting.

In our case, the ordered outcome variable $rating_{ijt}$ can take $K = 4$ possible integer values, \{1,2,3,4\}. Therefore, we can transform the random variable $rating_{ijt}$ in to $K - 1 = 3$ possible binary variables $d_{rating}^{k}_{ijt}$, where:

$$d_{rating}^{k}_{ijt} = I(rating_{ijt} \geq k), \text{ where } k = 2, 3, 4.$$  (9)

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

\[
Pr(d_{\text{rating}}_{ijt}^k = 1 \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_t) = 1 - \Lambda(\mu_k - \beta_1 \text{star1}_{it} - \beta_2 \text{star3}_{it} - \gamma z_i - \eta_t)
\] (10)

Next, we denote \( s_i^k \) as the sum of all the binary transformed ratings received by user \( i \) over time:

\[
s_i^k = \sum_{t=1}^{T_i} \sum_{j=1}^{4} d_{\text{rating}}_{ijt}^k
\]

Further, let \( B_i^k \) denote 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:

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

Note that the size of \( B_i^k = \left(\frac{4 \times T_i}{s_i^k}\right)^{10} \)

Now, we can write the conditional probability of \( d_{\text{rating}}_{ijt}^k \) given \( s_i^k \) as:

\[
Pr(\{d_{\text{rating}}_{ijt}^k\} \mid \text{star1}_{it}, \text{star3}_{it}, s_i^k, \beta_1, \beta_2) = \frac{\exp(d_{\text{rating}}_{ijt}^k \cdot (\beta_1 \text{star1}_{it} + \beta_2 \text{star3}_{it}))}{\sum_{d \in B_i^k} \exp(d \cdot (\beta_1 \text{star1}_{it} + \beta_2 \text{star3}_{it}))} \] (12)

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

\[
\text{CLL}(\beta_1^k, \beta_2^k) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \ln [Pr(d_{\text{rating}}_{ijt}^k \mid \text{star1}_{it}, \text{star3}_{it}, s_i^k, \beta_1^k, \beta_2^k)]
\] (13)

Since we can dichotomize \( \text{rating}_{ijt} \) into three binary (at each of the three cutoffs \( d_{\text{rating}}_{ijt}^1, d_{\text{rating}}_{ijt}^2, \) and \( d_{\text{rating}}_{ijt}^3 \)), the above CLL can be specified for each \( d_{\text{rating}}_{ijt}^k \), where \( k \in \{2, 3, 4\} \). Maximizing each of these CLLs gives us three separate but consistent estimates of \( \beta_1, \beta_2, \)

\(^{10}\)For example, consider user \( i \) who plays only two games \( (T_i = 2) \). For \( k = 4 \), we have \( d_{\text{rating}}_{ijt}^k \in \{0, 1\} \) which denotes whether user \( i \) has received a rating of 4 from user \( j \) or not. Now, let’s consider a scenario where \( s_i^3 = 1 \), which indicates that user \( i \) has received only one rating of 4 in her games from all the players. Thus, \( B_i^4 = \{(1, 0, 0, 0, 0, 0), (0, 0, 0, 0, 0, 0), \ldots, (0, 0, 0, 0, 0, 0), (0, 0, 0, 0, 0, 0, 0, 0, 0, 1)\}, i.e., \( B_i^4 \) is the set of all possible ways that user \( i \) can get only one rating of 4 in her games. 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 ratings \((4 \times 2 = 8)\). We denote each element of set \( B_i^4 \) with vector \( d \). Also, notice that the size of \( B_i^4 \) is eight, because \( \binom{4+2}{1} = 8 \).
which we denote as \{\beta^k_1, \beta^k_2\}, where \( k \in \{2, 3, 4\} \). These are referred to as Chamberlain CML estimators.

Two necessary conditions need to be satisfied for the identification of \{\beta^k_1, \beta^k_2\}. First, we need within-user variation in \( star^1_{it} \) and \( star^3_{it} \). Intuitively, this estimator takes advantage of the variation in star-ratings and preferences “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. So if the same user \( i \) receives lower ratings 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 \( d_{rating^k_{ijt}} \) because users with constant \( d_{rating^k_{ijt}} \) do not contribute to the CLL for cut-off \( k \) (and hence identification). Thus, only users for whom we have across time in variation in both the outcome variable \( d_{rating^k_{ijt}} \) and the independent variables \( (star^1_{it}, star^3_{it}) \) contribute to the identification of \{\beta^k_1, \beta^k_2\}. We now illustrate this condition using an example. For \( k = 4 \), consider a user \( i \) who has either received a rating of 4 in all her games, or never ever received a rating of 4 in any of her games. This user does not contribute to the CLL because her outcome \( d_{rating^4_{ijt}} \) is constant over time even if her/his star rating varies over time. Intuitively, at any cut-off \( k \), only the variation above or below \( k \) is used for identification because of dichotomization; for example, the CLL for \( k = 4 \) only considers whether \( rating_{ijt} \) is greater than or equal to 4 and ignores the variation in \( rating_{ijt} \) 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.

To address the efficiency issue, Das and Van Soest (1999) proposed a Minimum Distance (MD) estimator that combines all the Chamberlain estimates. We now describe the application of their method for our context below.

Recall that we have \( K - 1 = 3 \) estimates each of \{\beta_1, \beta_2\}: \{\beta^1_1, \beta^1_2\}, \{\beta^2_1, \beta^2_2\}, \{\beta^3_1, \beta^3_2\}. Since each of these \( K - 1 \) estimates is 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. Specifically, it involves solving

\footnote{Constant \( d_{rating^k_{ijt}} \) means that all elements of \( B^k_i \) are either zero or one.}

\footnote{For individuals who have played a large number of games (large \( T_i \)) and have a large number of positive values of \( d_{rating^k_{ijt}} \), (large \( s^k_i \)), calculation of all the combinations results can result in numeric overflow and cause computational issues. For example, if user \( i \) plays 100 games \( T_i = 100 \) and in each game he receives one rating of four, then \( s^4_i = 100 \) and \( \binom{4 \times 100}{100} \) = 2.24e + 96. Therefore, we limit our empirical analysis to users’ first 100 games. Of the 3,494 users who experience a star change, only 352 (10%) users play more than 100 games. The consistency of the estimates is not affected if we choose a subset of games for players who have played a large number of games.}
the following minimization problem:

\[
\hat{\beta}^{MD} = \arg\min_b (\tilde{\beta} - Mb)'var(\tilde{\beta})^{-1}(\tilde{\beta} - Mb)
\]  

(14)

where, \( \tilde{\beta} \) is the 6 \( \times \) 1 matrix of Chamberlain estimators and \( M \) 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:

\[
\arg\min_{b_1, b_2} \begin{pmatrix}
\beta_1^2 - b_1 \\
\beta_2^2 - b_2 \\
\beta_3^1 - b_1 \\
\beta_4^1 - b_1 \\
\beta_2^3 - b_2 \\
\beta_4^3 - b_2
\end{pmatrix}' \begin{pmatrix}
var(\beta_1^2) & cov(\beta_2^2, \beta_1^2) & var(\beta_2^2) \\
cov(\beta_2^2, \beta_1^2) & \cdot & var(\beta_2^2) \\
cov(\beta_2^2, \beta_1^2) & \cdot & \cdot & var(\beta_2^3) \\
cov(\beta_2^2, \beta_1^2) & \cdot & \cdot & \cdot & var(\beta_1^1) \\
cov(\beta_2^2, \beta_1^2) & \cdot & \cdot & \cdot & \cdot & cov(\beta_2^4, \beta_1^1) & var(\beta_2^3) \\
cov(\beta_2^2, \beta_1^2) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & cov(\beta_2^4, \beta_1^1) & var(\beta_2^3) \\
cov(\beta_2^2, \beta_1^2) & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & var(\beta_2^3)
\end{pmatrix}^{-1} \begin{pmatrix}
\beta_1^2 - b_1 \\
\beta_2^2 - b_2 \\
\beta_3^1 - b_1 \\
\beta_4^1 - b_1 \\
\beta_2^3 - b_2 \\
\beta_4^3 - b_2
\end{pmatrix}
\]

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

\[
\hat{\beta}^{MD} = \left\{M'var(\tilde{\beta})^{-1}M\right\}^{-1}M'var(\tilde{\beta})^{-1}\tilde{\beta}
\]  

(15)

and its variance is given by:

\[
var(\hat{\beta}^{MD}) = \left\{M'var(\tilde{\beta})^{-1}M\right\}^{-1}
\]  

(16)

We implement this MD estimator using the Stata code developed by Hole et al. (2011).

6.3 Results and Discussion

The results from our estimation exercise are presented in Table 6. We start by considering two Pooled Ordered Logit models: (1) Model M4 – a simple model that only includes star-ratings as the independent variable, and (2) Model M5 – a slightly more elaborate model that includes all the user-specific observables \((z_i)\).

In the basic Pooled Ordered Logit model (Model M4), we see a positive and significant effect for higher star-ratings. That is, one-star users receive lower preference ratings compared to two-star users, and two-star users receive lower ratings compared to three-star users. Thus, it seems like higher star-ratings lead to higher preference ratings. This result is consistent with Figure 4a.

However, since a user’s current star-rating is likely to be positively correlated with her/his appeal in
<table>
<thead>
<tr>
<th></th>
<th>(M4)</th>
<th>(M5)</th>
<th>(M6)</th>
<th>(M7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Ordered Logit)</td>
<td>(Ordered Logit)</td>
<td>(FE Ordered Logit)</td>
<td>(Ordered Logit)</td>
</tr>
<tr>
<td>star1_{it}</td>
<td>-0.14452***</td>
<td>-0.11431***</td>
<td>0.02852</td>
<td>-0.13888***</td>
</tr>
<tr>
<td></td>
<td>(0.02315)</td>
<td>(0.03021)</td>
<td>(0.01804)</td>
<td>(0.02369)</td>
</tr>
<tr>
<td>star3_{it}</td>
<td>0.06063***</td>
<td>0.05578**</td>
<td>-0.05101***</td>
<td>0.05136***</td>
</tr>
<tr>
<td></td>
<td>(0.01560)</td>
<td>(0.02328)</td>
<td>(0.01464)</td>
<td>(0.01590)</td>
</tr>
<tr>
<td>age_{i}</td>
<td>-0.00038</td>
<td>(0.00052)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>college_{i}</td>
<td>-0.00434</td>
<td>(0.00581)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>graduate_{i}</td>
<td>-0.00550</td>
<td>(0.01305)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>picture_rating_{i}</td>
<td>0.13276***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00368)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>picture_{i}</td>
<td>0.00137</td>
<td>(0.00206)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>employment_{i}</td>
<td>-0.00018</td>
<td>(0.00136)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bio_{i}</td>
<td>-0.00001</td>
<td>(0.00000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>μ2</td>
<td>-1.09924***</td>
<td>-1.11220***</td>
<td>-1.09054***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00203)</td>
<td>(0.01533)</td>
<td>(0.00456)</td>
<td></td>
</tr>
<tr>
<td>μ3</td>
<td>-0.00053</td>
<td>-0.00841</td>
<td>-0.00036</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00188)</td>
<td>(0.01527)</td>
<td>(0.00420)</td>
<td></td>
</tr>
<tr>
<td>μ4</td>
<td>1.09828***</td>
<td>1.09372***</td>
<td>1.09067***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.00205)</td>
<td>(0.01529)</td>
<td>(0.00446)</td>
<td></td>
</tr>
<tr>
<td>Individuals</td>
<td>24393</td>
<td>11639</td>
<td>3494</td>
<td>3494</td>
</tr>
<tr>
<td>Observations</td>
<td>2980148</td>
<td>1580848</td>
<td>630160</td>
<td>630160</td>
</tr>
</tbody>
</table>

Table 6: Ordered logit estimates of the effect of star-rating on preference ratings received.

the dating market (through physical attractiveness, age, education, etc.) these results are likely to be biased upwards. So we control for all the user-specific observables in Model M5. While this changes the estimates by a small amount, the direction of the results remain unchanged. Nevertheless, both these models suffer from endogeneity concerns discussed earlier (that of $E[\text{star}_{it}.\eta_{i}] \neq 0$).

Therefore, we now focus on the results from the Fixed-effects MD estimator (see Model M6). Interestingly, here we find that the effect of star-rating is actually negative – a user gets better
preference rating when s/he is shown with two stars as opposed to three stars! We do not find any significant effect of one star compared to two stars. Overall, these results are strong evidence in support of strategic shading – users strategically give lower ratings to those whom they perceive to be very popular in order to avoid the psychological costs of rejection. Recall that if \( i \) is more popular, s/he is more likely to get matched with other popular users according to the stable matching algorithm, and hence have a lower probability of being matched with \( j \) even if \( j \) gives a high preference rating to \( i \).

Our findings are important for two key reasons. First, they provide the first empirical evidence in support of strategic shading in dating markets. While shading concerns have been discussed in the literature, none of the earlier papers have been able to causally identify them. Second, we are also the first to document negative returns to higher star-ratings in online platforms. Past empirical research has mainly documented positive gains to online rating mechanisms in a variety of contexts, e.g., book sales ([Chevalier and Mayzlin, 2006]) and freelance labor markets ([Yoganarasimhan, 2013]). In contrast, we show that high ratings can actually lead to worse outcomes for market participants in the case of online dating. Indeed, our findings suggest that we (as researchers and managers) need to understand the behavioral underpinnings of the mechanism through which star-ratings operate within a given market instead of assuming positive effects based on prior work.

Our results also have implications for empirical literature on observational learning and social effects. Note that star-ratings in our setting simply reveal information on the actions of past users, but not any information on the unobserved quality of the focal user since these star-ratings are not based on reviews after interactions between users. Thus, the only information to be gleaned from a user’s star-ratings is her/his appeal to others in the marketplace. Knowing that a user \( i \) is appealing to many other people in the market can influence user \( j \)’s preference for \( i \) in two possible ways. On the one hand, if user \( j \) wants to date someone who is wanted by others or popular in this dating market, then revealing past popularity of \( i \) should have a positive effect on \( j \)’s rating of \( i \). On the other hand, if user \( j \) is concerned that a highly popular person will be less likely to choose \( j \) as her/his top preference, and is concerned about the psychological costs of rejection, then knowing that \( i \) is popular will suppress \( j \)’s rating of \( i \). Our findings support the latter argument. Past empirical work on observational learning mostly finds evidence in support of herding, or positive returns to revealing that a product is popular ([Sorensen, 2007], [Tucker and Zhang, 2011]). However, in those settings there were no concerns about personal rejection. Our results suggest that when there are inter-personal costs associated with transactions, the direction of social effects can flip.

### 6.4 Validation and Robustness

We now present two important robustness checks to validate the model and results.
6.4.1 Linear Model

First, we examine whether the substantive results hold when we consider a simpler linear model formulation where we can directly model the outcome as a linear function of star-ratings and other relevant variables. Naturally, such a linear specification will be biased because our outcome variable is discrete and ordinal. Nevertheless, it is useful to estimate a simpler linear model and compare the results with our current specification. 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, which addressed the omitted variable by employing a within transformation. These are the linear analogs of models M4, M5, and M6. The estimates from the linear model are substantively similar to those from the Ordered Logit models. Thus, our findings are robust to the model specification. Please see Appendix §A.1 for model details and the full table of results.

6.4.2 Estimation Sample

Note that the Minimum Distance Estimator for the Fixed-effects Ordered Logit model only utilizes a subset of the data for inference; specifically it only uses data from individuals who went through at least one star change during the observation period. In principal, 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.

To check if this is indeed the case we perform two types of validation checks. 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 Model M7 in Table 6. We find that the magnitude and the direction of the estimates in Model M7 are similar to those for the full population Model M4.

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. The results from this comparison are presented in Table 7 for both female and male users. We find that the two sets of users are very similar on age and physical attractiveness ratings for both genders.

However, the main source of difference comes from users’ past experience. 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
Table 7: Comparison of attributes between groups who experienced no star change and groups who experienced at least one star change.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>25th</th>
<th>50th</th>
<th>75th</th>
<th>(Min, Max)</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$age_i$</td>
<td>No Change</td>
<td>20.88</td>
<td>4.58</td>
<td>19</td>
<td>20</td>
<td>22</td>
<td>(13, 109)</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>20.80</td>
<td>6.22</td>
<td>18</td>
<td>20</td>
<td>22</td>
<td>(13, 109)</td>
</tr>
<tr>
<td>$picture_rating_i$</td>
<td>No Change</td>
<td>-0.00</td>
<td>0.70</td>
<td>-0.53</td>
<td>-0.08</td>
<td>0.48</td>
<td>(-2.88, 2.75)</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>0.03</td>
<td>0.74</td>
<td>-0.55</td>
<td>-0.06</td>
<td>0.52</td>
<td>(-2.42, 3.07)</td>
</tr>
<tr>
<td>$initial_game_i$</td>
<td>No Change</td>
<td>85.06</td>
<td>75.32</td>
<td>20</td>
<td>74</td>
<td>126</td>
<td>(0, 764)</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>0.28</td>
<td>3.39</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(0, 100)</td>
</tr>
<tr>
<td>Male:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$age_i$</td>
<td>No Change</td>
<td>21.85</td>
<td>5.33</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>(13, 109)</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>23.49</td>
<td>8.46</td>
<td>19</td>
<td>22</td>
<td>25</td>
<td>(14, 109)</td>
</tr>
<tr>
<td>$picture_rating_i$</td>
<td>No Change</td>
<td>0.01</td>
<td>0.66</td>
<td>-0.49</td>
<td>-0.09</td>
<td>0.40</td>
<td>(-1.70, 3.29)</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>-0.07</td>
<td>0.64</td>
<td>-0.56</td>
<td>-0.19</td>
<td>0.34</td>
<td>(-1.51, 3.08)</td>
</tr>
<tr>
<td>$initial_game_i$</td>
<td>No Change</td>
<td>58.19</td>
<td>52.61</td>
<td>24</td>
<td>52</td>
<td>83</td>
<td>(0, 2146)</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>0.06</td>
<td>0.61</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(0, 14)</td>
</tr>
</tbody>
</table>

Figure 7: Change in popularity as a function of number of games played.

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 this point using Figure 7 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 ratings 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 $z_i$, $\eta_i$, the popularity measure ($\text{popularity}_{it}$) 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).

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.

6.5 Heterogeneity in the Effect of Star-ratings

We now examine the heterogeneity in the effect of star-ratings across a series of user-level observables. This analysis is important for two reasons. First, it can give us additional insight into which types of users are more or less likely to respond to star-ratings. Thus, it can give us some insights on whether our hypothesis that avoidance of three-star users stems from fear of rejection is valid or not. Second, from the app/platform’s perspective, understanding the source of the effect of star-ratings can help managers understand the motivations of users in this market.

We consider heterogeneity along three dimensions: gender, age, physical attractiveness of the person giving the rating. Our approach to examining heterogeneous effects is to stratify the data on the variable of interest and re-estimate the Fixed-effects Ordered Logit model on the stratified data.

6.5.1 Gender

We stratify the data based on the gender of the rater (i.e., the person giving the rating) to examine whether there are any differences in how men and women respond to star-ratings. The results from this exercise are summarized in Table 8. We find that while both men and women avoid three-star users (compared to two-star users), men are also more likely to approach one-star women compared to two-star women.

One important thing to keep in mind is we cannot compare the magnitudes of the effects of $\text{star1}_{it}$ and $\text{star3}_{it}$ across models M10 and M11 because the variance of errors is not identified in $(\epsilon_{ijt})$ in (Ordered) Logit models (Allison, 1999). For example, suppose that we have the following models for women and men:

Women: $\text{rating}_{ijt}^w = \beta_1^w \text{star1}_{it} + \beta_2^w \text{star3}_{it} + \gamma^w z_i + \eta_i + \epsilon_{ijt}^w$ where $\epsilon_{ijt}^w \sim GEV(0, \sigma_w^2)$

Men: $\text{rating}_{ijt}^m = \beta_1^m \text{star1}_{it} + \beta_2^m \text{star3}_{it} + \gamma^m z_i + \eta_i + \epsilon_{ijt}^m$ where $\epsilon_{ijt}^m \sim GEV(0, \sigma_m^2)$
Table 8: Heterogeneous effect of star-rating on received preferences using Ordered Logit Fixed Effects model.

Because the latent variable $rating^*$ 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:

$$\begin{align*}
\text{Women:} & \quad \frac{rating_{ijt}^*}{\sigma_w^2} = \frac{\beta_1^w}{\sigma_w^2} star_{1it} + \frac{\beta_2^w}{\sigma_w^2} star_{3it} + \frac{\epsilon_{ijt}^w}{\sigma_w^2} \\
\text{Men:} & \quad \frac{rating_{ijt}^*}{\sigma_m^2} = \frac{\beta_1^m}{\sigma_m^2} star_{1it} + \frac{\beta_2^m}{\sigma_m^2} star_{3it} + \frac{\epsilon_{ijt}^m}{\sigma_m^2}
\end{align*}$$

The coefficients of $star_{1it}$ in Table 8 Model M10 and M11 are not $\beta_1^f$ and $\beta_1^m$, rather they are $\frac{\beta_1^f}{\sigma_f^2}$ and $\frac{\beta_1^m}{\sigma_m^2}$ respectively. In our analysis, the reported coefficients are 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. Hence, we do not directly compare the coefficients of sub-groups in any of our analyses.

6.5.2 Age

As summarized previously in Table 1, the age of the median user in our data is 21. We stratify the data into two groups based on rater’s age: (i) Older rater, data where the rater’s age is greater than 21, and (ii) Younger rater, data where the rater’s age is less than or equal to than 21. Then we re-run our analysis on these two strata separately. The results from this exercise are shown in Table 8. Interestingly, we find that older users are not influenced by the ratings of potential partners (Model M12). On the other hand, we see a negative and significant effect of $star_{3it}$ for younger raters (Model M13), i.e., only younger players strategically avoid popular users.
6.5.3 Physical Attractiveness

Finally, we examine whether a users’ response to the star ratings of potential partners varies with her/his own attractiveness. Recall that the median user has a standardized picture_rating of -0.09 (see Table 1). Based on this value, we stratify the data into two groups: (i) Attractive rater, data where the rater’s picture_rating is greater than -0.09, and (ii) Unattractive rater, data where the rater’s picture_rating 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 Table 8. We find that attractive users are not likely to be influenced by star ratings (Model M14). On the other hand in Model M15, for unattractive raters, the results show a negative and significant effect of star3_3. This suggests that only less attractive users strategically avoid popular users. This is reasonable since we expect unattractive users to be more concerned about rejection than attractive users.

In sum, we find that both men and women strategically shade down their preference ratings for three star users. Men also tend to shade up their ratings for one star women. Further, the star-rating effects are mainly driven by users who are younger and less attractive than average. These findings are consistent with our hypothesis that shading is driven by fear of rejection since younger and unattractive users are more likely to suffer from rejection concerns.

7 Effect of Stars on Messaging Behavior

So far, we have shown that users strategically shade down their preference ratings for popular users in our online dating platform. We now examine if this behavior persists even in messaging. We focus on two messaging 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 describe the model specification 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_{ij} and reply_{ij} as follows:

\[
\text{first}_{ijt} = \begin{cases} 
1, & \text{if } \text{first}_{ij} > 0 \\
0, & \text{else}
\end{cases}
\] (17)

\[
\text{reply}_{ijt} = \begin{cases} 
1, & \text{if } \text{reply}_{ij} > 0 \\
0, & \text{else}
\end{cases}
\] (18)
These latent variables are defined as:

\[
\text{first}^*_{ijt} = \beta_1^f \text{star}1_{it} + \beta_2^f \text{star}3_{it} + \gamma^f z_i + \eta^f_i + \epsilon^f_{ijt},
\]

(19)

\[
\text{reply}^*_{ijt} = \beta_1^r \text{star}1_{it} + \beta_2^r \text{star}3_{it} + \gamma^r z_i + \eta^r_i + \epsilon^r_{ijt},
\]

(20)

where the interpretations of \{\beta_1^f, \beta_2^f, \gamma^f, \eta^f_i, \epsilon^f_{ijt}\} and \{\beta_1^r, \beta_2^r, \gamma^r, \eta^r_i, \epsilon^r_{ijt}\} are similar to that in §6.1.

Further, following the same arguments, we allow for \eta^f_i and \eta^r_i to be arbitrarily correlated to \text{star}1_{it} and \text{star}3_{it}. Assuming that \epsilon_{ijt} are IID and drawn from an Extreme Value Type I distribution, the probability that user \(i\) receives a first message from user \(j\) (conditional on \(i\) and \(j\) being matched in game \(t\)) can be written as:

\[
\Pr(\text{first}ijt = 1 | \text{match}ijt = 1, X_{it}, z_i, \eta^f_i) = \frac{\exp(\beta_1^f \text{star}1_{it} + \beta_2^f \text{star}3_{it} + \gamma^f z_i + \eta^f_i)}{1 + \exp(\beta_1^f \text{star}1_{it} + \beta_2^f \text{star}3_{it} + \gamma^f z_i + \eta^f_i)}
\]

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:

\[
\Pr(\text{reply}ijt = 1 | \text{match}ijt = 1, \text{first}ijt = 1, X_{it}, z_i, \eta^r_i) = \frac{\exp(\beta_1^r \text{star}1_{it} + \beta_2^r \text{star}3_{it} + \gamma^r z_i + \eta^r_i)}{1 + \exp(\beta_1^r \text{star}1_{it} + \beta_2^r \text{star}3_{it} + \gamma^r z_i + \eta^r_i)}
\]

As in the case of the Ordered Logit model, we can use these probabilities to specify Conditional Log-Likelihoods which are independent of \(\eta\)s and then maximize the two CLLs to derive consistent estimates of \{\beta_1^f, \beta_2^f\} and \{\beta_1^r, \beta_2^r\}. Since these steps are very similar to that described in §6.2, we relegate the details to Appendix §A.2.

7.2 Results

The results for both the message outcomes are shown in Table 9. We start with a discussion of first messages (shown in models M14 and M15). Model M14 is a pooled Logit model that controls only for the observable attributes of the (potential) receiver and M15 is a Fixed-effects Logit model estimated using CLL that accounts for the endogeneity between star ratings and user-specific unobservables. In Model M14, we find that three-star users are more likely to receive first messages compared to two-star users. This result is consistent with Figure 4b. However, after controlling for the endogeneity issues, we find both three- and one-star users are more likely to receive first messages compared to two-star users (see Model M15). This is consistent with Figure 6.

We thus find partial evidence for strategic shading in a user’s first-message behavior. Users strategically initiate contact when they are matched with one-star users; possibly with the expectation
Table 9: Effect of star-rating on messages received.

that such users may be more responsive. Interestingly, they are also more likely to reach out to three-star matches, i.e., we find some support for positive returns to star ratings. In sum, we find that users are responding to the star ratings of their matches when deciding whether to contact them or not. However, the direction of shading here is different from that in the case of preference rating. This may be due to the fact that users are more confident of their reception in this case than in the case of preference ratings. Recall that in order for \(i\) and \(j\) to be matched, \(i\) must have expressed a sufficiently high preference for \(j\) during the game. Hence, at this stage, \(j\)’s beliefs on \(i\)’s likelihood...
of rejecting her/him are likely to be lower than that before the game. Thus, \( j \) has less of an incentive to shade down her/his behavior, when it comes to contacting \( i \).

Next, we present the results for reply behavior in Models M16 and M17, which are analogous to M14 and M15 discussed above. Interestingly, we find that one-star users are less likely and three-star users are more likely to receive a reply when they initiate contact. Note that when user \( i \) has already initiated contact with \( j \), then \( j \) has no rejection concerns. Thus, in the case of replies, there is no reason for \( j \) to shade down her/his response to \( i \). Our findings here thus provide support for our earlier hypothesis that users rate higher star users lower because of rejection concerns (since this effect goes away in the case of replies).

Finally, we examine the heterogeneity in the effect of star-ratings on first messaging behavior by stratifying the data along different variables, and re-estimating the Fixed-effects Ordered Logit model on the stratified data.\footnote{We did not find any significant heterogeneity in reply behavior.} We consider three key sources of heterogeneity: (1) receiver’s age, (2) receiver’s physical attractiveness, and (3) sender’s match level. As shown in Table 10, we find that the effect of \( star_{1it} \) is only significant for older users and attractive users. Recall that in our context, the age-range of users is pretty small with median age being 21 and therefore older users are generally preferred. (Our definition of an “old” user is someone who is over 21.) Thus, we find that for users who have higher baseline appeal (are older and/or more physically attractive), their matches are more willing to initiate contact with them even when they have one-star rating. However, for less attractive/younger users, their partners are not very interested in reaching out when they have a one-star rating.

Next, in Table 11 we stratify the data into four groups based on the rating that \( j \) gave \( i \) during the game and present the results on the stratified data. For example, \( match_{level} = 4 \) implies

<table>
<thead>
<tr>
<th></th>
<th>(M18)</th>
<th>(M19)</th>
<th>(M20)</th>
<th>(M21)</th>
</tr>
</thead>
<tbody>
<tr>
<td>star_{1it}</td>
<td>0.4332**</td>
<td>0.21589</td>
<td>0.58929***</td>
<td>0.20453</td>
</tr>
<tr>
<td></td>
<td>(0.19725)</td>
<td>(0.13766)</td>
<td>(0.15871)</td>
<td>(0.15474)</td>
</tr>
<tr>
<td>star_{3it}</td>
<td>0.62000***</td>
<td>0.53255***</td>
<td>0.74309***</td>
<td>0.36395***</td>
</tr>
<tr>
<td></td>
<td>(0.12209)</td>
<td>(0.08150)</td>
<td>(0.07827)</td>
<td>(0.12347)</td>
</tr>
<tr>
<td>Individuals</td>
<td>3500</td>
<td>6972</td>
<td>4985</td>
<td>4428</td>
</tr>
<tr>
<td>Observations</td>
<td>177280</td>
<td>304757</td>
<td>240818</td>
<td>224987</td>
</tr>
</tbody>
</table>

Table 10: Effect of star-rating on likelihood of receiving a first message as a function of receiver attributes.
Table 11: Effect of star-rating on likelihood of receiving first message as a function of the match-level of the paired users.

<table>
<thead>
<tr>
<th>star1_{it}</th>
<th>star3_{it}</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M22)</td>
<td>(M23)</td>
</tr>
<tr>
<td>( match_level_{jt} = 4 )</td>
<td>( match_level_{jt} = 3 )</td>
</tr>
<tr>
<td>0.42145***</td>
<td>0.53163***</td>
</tr>
<tr>
<td>0.75568***</td>
<td>0.58915***</td>
</tr>
<tr>
<td>Individuals</td>
<td>7701</td>
</tr>
<tr>
<td>Observations</td>
<td>197680</td>
</tr>
</tbody>
</table>

Standard errors in parentheses

\* \( p < 0.1 \), \** \( p < 0.05 \), \*** \( p < 0.01 \)

that \( i \) received the highest rating from \( j \) during the game (before they were matched). Thus, if \( match\_level_{jt} = 4 \), then \( j \) got the match s/he wanted in game \( t \), and if \( match\_level_{jt} = 1 \), s/he was matched with her/his least preferred partner in the game. We find that when \( j \) liked \( i \) sufficiently enough (match levels of 3 or 4), then the results are similar to those in Table 9. However, when users do not find the partner appealing to begin with (match level of 1), then star ratings are unlikely to affect their messaging behavior.

Together, these results suggest that a user’s initial preference or interest for a matched partner has a moderating effect on the star-ratings variable.

8 Conclusion

In this paper, we examine how users respond to the popularity information of potential partners in a gamified mobile dating app. On the one hand, knowing that a potential match 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 rating 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 their prior ratings. We quantify the causal impact of a user’s star-rating on the preference-ratings 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 strong support for the presence of strategic shading in our data: (1) three-star users receive lower preference-ratings during the game compared to two-star users, and (2) two-star users receive fewer first messages after the game compared to one-star users. We also show that the effect of star-ratings
is heterogeneous across both outcomes and user-specific observables, and that this heterogeneity (or extent of shading) can be directly linked to the severity of rejection concerns. Our findings have implications for the design of dating platforms and measurement of mate preferences in matching markets.

Finally, 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. 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. Prior research seems to suggest that strategic-shading may not occur in those cases (Hitsch et al., 2010b). However, it is not clear whether our results stem from the short-term nature of the relationship or from time pressure. Further research on these topics can have important implications for the design of dating platforms and shape users’ interactions in online mating settings.
A Appendix

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

We consider the following linear model:

\[ \text{rating}_{ijt} = \beta_1 \text{star}_{1it} + \beta_2 \text{star}_{3it} + \gamma z_i + \eta_i + \epsilon_{ijt} \] (A.1)

where the interpretation of \( \beta_1, \beta_2, \gamma, \) and \( \eta_i \) are the same as that discussed in §6.1. The main difference is that here these coefficients and variables relate directly to the observed outcome instead of the latent variable \( \text{rating}^* \). 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 \( \eta_i \) 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 [A.1].

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

\[ \overline{\text{rating}}_i = \beta_1 \overline{\text{star}}_{1i} + \beta_2 \overline{\text{star}}_{3i} + \gamma z_i + \eta_i + \overline{\epsilon}_i, \] (A.2)

where \( \overline{\text{rating}}_i = \frac{\sum_{t=1}^{T_i} \sum_j \text{rating}_{ijt}}{4 \times T_i}, \overline{\text{star}}_{1i} = \frac{\sum_{t=1}^{T_i} \text{star}_{1it}}{T_i}, \overline{\text{star}}_{3i} = \frac{\sum_{t=1}^{T_i} \text{star}_{3it}}{T_i}, \) and \( \overline{\epsilon}_i = \frac{\sum_{t=1}^{T_i} \sum_j \epsilon_{ijt}}{4 \times T_i} \).

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

\[ \text{rating}_{ijt} - \overline{\text{rating}}_i = \beta_1 \left( \text{star}_{1it} - \overline{\text{star}}_{1i} \right) + \beta_2 \left( \text{star}_{3it} - \overline{\text{star}}_{3i} \right) + (\epsilon_{ijt} - \overline{\epsilon}_i) \] (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 \( \beta_1 \) and...
Table A.1: Pooled OLS and Fixed-effects estimates of the effect of user’s star-rating on preference ratings received. All standard errors are clustered at the user-level.

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 M6 in Table 6.

$\beta_2$ under the linearity assumption. The results from this model are shown in Model A3 in Table A.1.
A.2 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 \( \text{first}_{ijt} \) or \( \text{reply}_{ijt} \), and \( y_{ijt}^* \) is the corresponding latent variable as follows:

\[ y_{ijt}^* = \beta X_{it} + \gamma z_i + \eta_i + \epsilon_{ijt}, \quad (A.4) \]

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

\[ Pr(y_{ijt} = 1 \mid X_{it}, z_i, \eta_i) = \frac{exp(\beta X_{it} + \gamma z_i + \eta_i)}{1 + exp(\beta X_{it} + \gamma z_i + \eta_i)} \quad (A.5) \]

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

\[ LL(\beta, \gamma) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{k=0}^{1} \ln \left[ Pr(y_{ijt} = k \mid X_{it}, z_i, \eta_i) I(y_{ijt} = k) \right] \quad (A.6) \]

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

\[ s_i = \sum_{t=1}^{T_i} (y_{ijt} \mid \text{match}_{ijt} = 1) \quad (A.7) \]

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:

\[
B_i = \{ d \in \{0, 1\}^{T_i} \mid \sum_{t=1}^{T_i} (d_{jt} = s_i \mid \text{match}_{ijt} = 1) \}
\] (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:

\[
Pr(y_i \mid X_{it}, s_i, \beta) = \frac{\exp(y_i \cdot (\beta X_{it}))}{\sum_{d \in B_i} \exp(d \cdot (\beta X_{it}))}
\] (A.9)

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

\[
\text{CLL}(\beta) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \ln [Pr(y_i \mid X_{it}, s_i, \beta)]
\] (A.10)
References


A. Smith. 15% of American Adults Have Used Online Dating Sites or Mobile Dating Apps. Pew Research Center, pages 1–12, 2016.


