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

Behnaz Bojd*
University of California, Irvine

Hema Yoganarasimhan
University of Washington

Appendices

A Other Modeling Frameworks

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

A.1 Rank-Ordered Logit

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

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

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

where ε_{ijt} is an idiosyncratic preference shock and μ_{ijt} is a linear function of user i's observed characteristics to j and us, i's unobserved characteristics to us (η_i), and user j's characteristics, such

*Please address all correspondence to: behnazb@uci.edu, hemay@uw.edu.

that:

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

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

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

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

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

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

Similar to Equation (A.3), if we expand equation (A.4), we will have four receivers' fixed effects (η_i, η_k, η_l, and η_m). We can write the likelihood of user j giving preference-rankings to his or her potential matches in game t as:

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

where

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

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

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

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

A.2 Regression Discontinuity Design (RDD)

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

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

B Appendix for Robustness Checks

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

We consider the following linear model:

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

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

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

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

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

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

where $\overline{pref}_i = \frac{\sum_{t=1}^{T_i} \sum_j pref_{ijt}}{4 \times T_i}$, $\overline{star1}_i = \frac{\sum_{t=1}^{T_i} star1_{it}}{T_i}$, $\overline{star3}_i = \frac{\sum_{t=1}^{T_i} star3_{it}}{T_i}$, and $\overline{\epsilon}_i = \frac{\sum_{t=1}^{T_i} \sum_j \epsilon_{ijt}}{4 \times T_i}$.

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

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

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

z_i, η_i are constant across time periods, and hence their averages are the same as the variables themselves. Next, we subtract Equation (A.9) from Equation (A.8) as follows:

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

Note that all the time-invariant user-specific variables are now subtracted out and the new error term, $\epsilon_{ijt} - \bar{\epsilon}_i$, is no longer correlated with the star-ratings variables. The fixed-effects estimator is essentially a pooled OLS estimator for Equation (A.10) and it gives us consistent estimates of β_1 and β_2 under the linearity assumption. The results from this model are shown in model A3 in Table A1. Note that to keep the comparisons consistent, we only use the first 100 games of users who saw at least one star change during the observation period. Hence, model A3 in analogous to model M3 in Table 4. The results from model A3 are substantively similar to those in model M3. This suggests that our main results were not an artefact of the parametric specification of the model.

B.2 Estimation Sample

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

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

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

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

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

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

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

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

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

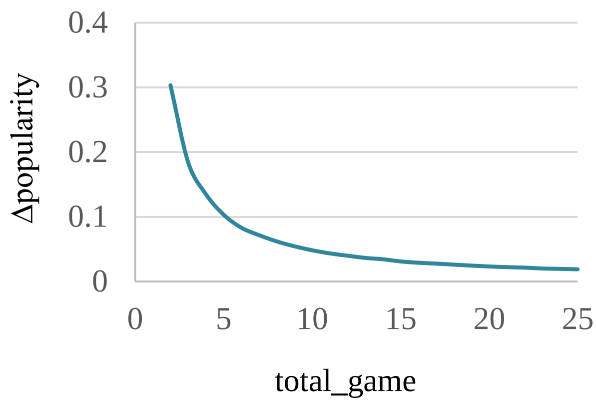

Figure A1: Absolute change in popularity score as a function of number of games played for all the users in our data for the observation period.

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

Of the 3,494 users who experience a star change in our observation period, 3,439 (98%) of them are new users who joined in the observation period ($initial\_game_i = 0$). We now compare the user-specific observables of these 3,439 new users (who went through a star-change) with those of new users who did not go through a star-change during our observation period (3,680 users). The results from this comparison are presented in Table A3.[1] Overall, there is sufficient empirical evidence to suggest that new users who experience at least one star change and those who experience no star changes are largely similar. Thus, we expect that the findings from the fixed-effects model to be applicable to the full population of users in the app.

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

Model M3 included all users who experienced a star-change and model M7 included observations where the user experienced a star-change and also initiated a message. Below, we re-estimate model

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

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

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

M3 with the sample used in model M7. We find that the results from this exercise are qualitatively similar to those presented in model M3 (see Table A4 in Appendix §B.2).

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

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

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

B.3 Within Game Correlation

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

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

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

B.4 Star Configuration of the Competitors in a Game

For the set of users in the estimation sample, we calculate the probability of being in a game with a specific configuration of competitors and present these probabilities in Table A6.[2] The first row considers observations where user i has a one-star rating ($star_{it} = 1$). In this case, the probability that s/he is competing with three users (i.e., the three other players of the same gender as i in game t) who all have two stars is 94.17%. Next, in observations where a user i is shown with two stars ($star_{it} = 2$), the probability of competing with three two-star players is 96.77%. And, when user i is shown with three stars ($star_{it} = 3$), this probability is 94.11%. Therefore, regardless of when a given user i decides to play a game, s/he is almost always competing with a similar configuration of players. In particular, s/he is being compared to other two-star users in over 94% of the cases. Thus, the data doesn't show any evidence that users are self-selecting entry time to avoid/obtain certain types of competitors.

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

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

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

C Conditional Log Likelihood for the Fixed-effects Logit Model

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

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

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

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

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

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

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

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

where N is the total number of users and T_i is the total number of games played by user i. Treating the η_i's as parameters and maximizing this log-likelihood via Maximum Likelihood Estimator (MLE) is inconsistent with large N and finite T due to the well-known incidental parameters problem (Neyman and Scott, 1948). As a result, the estimate of β_1, β_2 from this approach will be inconsistent. However, Chamberlain (1980) proposes a method to maximize a Conditional Log-Likelihood which gives consistent estimates. Following Chamberlain (1980), we denote s_i as the sum of all received messages (first messages or reply messages) by user i from his/her matches over time, that is:

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

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

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

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

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

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

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

D Validation Check: Truthfulness Assumption

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

In our setting, the ranking game resembles a one-to-one marriage SMP, where: (1) agents have to state their strict preference-rankings (i.e., no indifference rankings), (2) agents cannot truncate their list of preference-rankings (i.e., they cannot strategically choose to only rank their top few choices and refuse to rank their bottom choices), (3) agents cannot collude with each other, (4)

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

Table A7: Match results if users misrepresent their preferences.

agents' preferences are private (i.e,. users know their own preferences but not those of others'). Under such circumstances, it has been shown that, when a men-optimal stable matching mechanism is used, it is the dominant strategy for each man to state his true preferences, and any strategy for a woman is dominated if her stated first choice is not her true first choice; and vice-versa for women-optimal stable matching mechanism (Roth, 1989).[3] However, it has been shown that the incentive to manipulate true preferences is negligible for both sides in most real, large markets (Demange et al., 1987; Pittel, 1989; Lee and Yariv, 2018; Lee, 2016).

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

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

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

swaps her/his first and second choices. We then calculate which of her/his true preferences j will be matched with. Then, we aggregate the match outcomes over all players and all games to obtain the average probability of being matched with one's true first choice under this deviation as:

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

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

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

One possible critique of the above exercise could be that we started with the assumption that players stated-preferences are their true preferences. Therefore, we also present results from a general case, where the player's true preferences are drawn randomly (see the bottom panel of Table A7). Again, we find that deviating from truth-telling makes users strictly worse off. In sum, all our tests confirm the validity of the truth-telling assumption in our setting.

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

E Conversation History

We now examine the heterogeneous effects of star-ratings based on rank-giver's conversation history, and provide more evidence to show that the negative effect of three-star ratings during the game stems from rejection concerns. We start by defining $conversation_{jt}$ as the average number of successful conversations that user j experienced before game t. A successful conversation from j's

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

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

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

perspective is defined as one where j either received a first message from the matched partner, or received a reply to a message that s/he had initiated with the match.

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

F Physical Attractiveness

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

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

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

Next, in model A10, we re-run the analysis for unattractive receivers ($attractive_i = 0$). However, we find no significant results.[4] Thus, we find no evidence showing that users are concerned about being rejected when ranking an unattractive user.

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

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

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

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

References

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

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

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

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

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

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

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

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

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

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