informs
http://pubsonline.informs.org/journal/mksc

MARKETING SCIENCE
Articles in Advance, pp. 1–20
ISSN 0732-2399 (print), ISSN 1526-548X (online)

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

Behnaz Bojd,[a] Hema Yoganarasimhan[b]

[a] Paul Merage School of Business, University of California, Irvine, Irvine, California 92697; [b] Foster School of Business, University of Washington, Seattle, Washington 98195

Contact: behnazb@uci.edu, https://orcid.org/0000-0002-6115-9774 (BB); hemay@uw.edu, https://orcid.org/0000-0003-0703-5196 (HY)

Received: July 18, 2019
Revised: December 11, 2020; March 15, 2021
Accepted: March 25, 2021
Published Online in Articles in Advance: November 12, 2021

https://doi.org/10.1287/mksc.2021.1301

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

History: Olivier Toubia served as the senior editor for this article.
Supplemental Material: Data and the online appendix are available at https://doi.org/10.1287/mksc.2021.1301.

Keywords: popularity information • online ratings • strategic shading • online dating • centralized matching markets • two-sided platforms • stable matching problem

1. Introduction

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

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

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

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

However, there is no research that examines the effect of popularity information on users' demand in a two-sided dating platform. In this paper, we are interested in two related questions. First, we seek to quantify the causal effect of a user's popularity information on her or his demand measures in a centralized dating market. Second, we are interested in identifying the source of these effects (if any), that is, pin down the mechanism behind them.

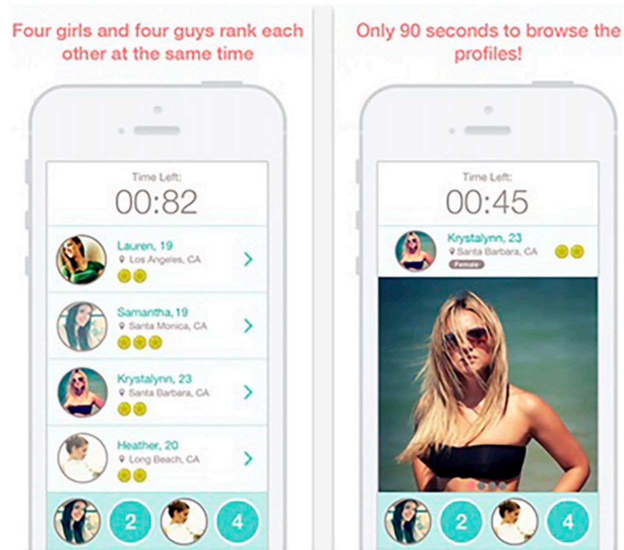
In a dating market, popularity information can have both positive and negative impact on demand. On the one hand, revealing that a potential partner is popular can increase her or his appeal, which in turn can increase a user's revealed preference for that potential partner (Hansen 1977). On the other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

We empirically examine these questions using data from a popular mobile dating app in the United States during the 2014–2015 time frame. Users in the app are matched based on games where they rank members of the opposite sex. Each game consists of four men and four women in a virtual room, where each player has 90 seconds to rank-order members of the opposite sex from one to four, with one indicating the most preferred partner and four the least (see Figure 1). (Throughout this paper, we use the term preference ranking, which is the reverse of ranking, to indicate users' ordered preferences to simplify exposition.[1])

The platform then uses these preference rankings as inputs into a stable match algorithm and matches each player in the room with a member of the opposite sex. After the game ends, users can initiate contact with their matched players and chat with them (if their matched partners reciprocate).

A key piece of information shown to users during and after the game is a star rating for each member of the opposite sex (ranging from one to three stars). A user's star rating is a cumulative measure of all the

Figure 1. (Color online) Screen Shot of the App During a Game (from the Perspective of a Male User)



Notes. Players indicate their rank-ordered preferences for the players from the opposite sex by dragging their profile pictures into the circles labeled one through four at the bottom of the app. In this example, the focal player has picked his first and third choices, and is yet to decide his second and fourth choices.

preference rankings that she or he has received in the past. So users who have received higher past preference rankings are shown with higher stars. Stars are thus a salient and visible indicator of a user's popularity on the platform. At the same time, they do not contain any extra information on the unobserved quality of the user because they are not based on his or her contact/engagement with previous players. Thus, they do not help resolve asymmetric information about the user's quality as a date (unlike star ratings based on purchase/experience in e-commerce settings).

Our analysis consists of two major components, which mirror our two broad research questions. To answer our first research question, we quantify the causal impact of a user's star rating on three demand measures: (1) preference rankings received during a game, (2) likelihood of receiving a first message from the matched partner after the game, and (3) likelihood of receiving a reply to a message sent after the game. The main challenge here is that users who received high preference rankings in the past (and hence have more stars now) are also likely to receive higher preference rankings now—not necessarily because of their star ratings, but because of their inherent attractiveness, which may be unobservable to the researcher (e.g., great bio descriptions, fun-loving pictures). This can give rise to an upward bias in our estimates of the effect of star ratings if we use naive estimation strategies. To overcome this challenge, we leverage the fact that a user's star rating is not static; rather, it changes over the course of our observation period as a function of her or his rankings in the previous games. Specifically, we model the first demand measure using a fixed-effects ordered logit model, and the latter two using fixed-effects binary logit models. In all these models, we allow user-specific unobservables (i.e., the fixed-effects) to be arbitrarily correlated with star ratings.

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

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

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

2. Related Literature

First, our paper relates to a large stream of literature that has established that popularity information has a positive effect on demand/sales of products and services in a variety of e-commerce settings, such as the music industry (Salganik et al. 2006, Dewan et al. 2017), books (Sorensen 2007), restaurants (Cai et al. 2009), software downloads (Duan et al. 2009), kidney transplant market (Zhang 2010), movies (Moretti 2011), digital cameras on Amazon (Chen et al. 2011), and the wedding services market (Tucker and Zhang 2011).[3] These studies have identified three mechanisms for this positive effect: (1) observational learning or quality inference based on others' actions (e.g. purchase statistics), (2) salience effect or awareness of alternative choices, and (3) network effect or increase in value of a product/service as its user base expands. In this paper, we provide the first negative effect of popularity information on demand in an online marketplace, and in a previously unstudied context: a two-sided dating market. We also present evidence for a new mechanism that can moderate the effect of popularity information: strategic shading due to rejection concerns.

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

Hitsch et al. 2010a, b; Shaw Taylor et al. 2011; Bapna et al. 2016; Lee 2016; Jung et al. 2019).

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

Finally, our work relates to the literature on two-sided matching markets. There are two types of two-sided matching markets: centralized and decentralized. In decentralized markets, there is no central matchmaker for the matching process. Instead, each agent engages in her or his own search process and makes/accepts offers over a period of time. It has been shown that these markets are prone to market failures that can lead to inefficient matching because of search costs, unraveling, and/or congestion (Roth 2008, Niederle and Yariv 2009). In particular, congestion can cause users to strategically avoid making offers to their top preferences because of rejection concerns (Roth and Xing 1997, Che and Koh 2016, Arnosti et al. 2021).

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

3. Setting

3.1. Mobile Dating App

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

The app requires users to participate in a structured matching game, which is described in detail below. Users cannot directly access or browse other users' profiles through the app; the only way to use the app is to play the ranking game described in Section 3.2.

3.2. Description of the Game

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

Therefore, conditional on geography and age, the assignment of users to games is random.

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

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

3.2.3 Match Allocation. The platform uses the rank-ordered preferences of all players in a game to derive a set of "stable matches," where the concept of stability is based on the canonical stable marriage problem (SMP): "Given n men and n women, where each person has ranked all members of the opposite sex in order 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, pp. 9–15).

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

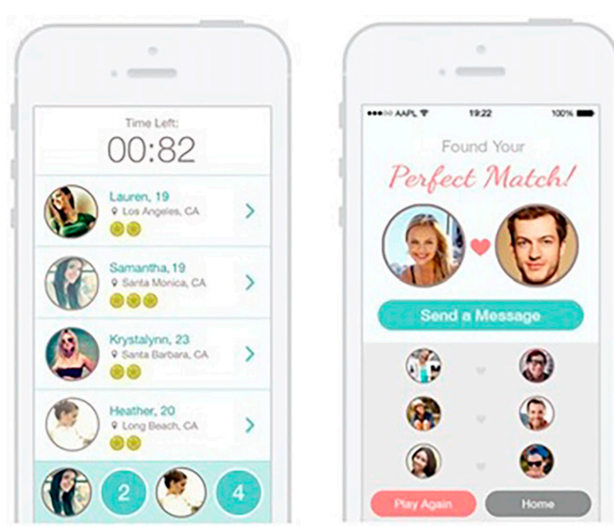
The entire matching process takes less than a second, and users can see the matches assigned to them as well as all the other matches allocated in the room (see the right panel of Figure 2).

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

4. Data

Our data comprise 94,386 games played by 24,653 unique users during the 10-month period from September 15, 2014, to July 15, 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

Figure 2. (Color online) Screen Shots of the Application Before and After a Game



variables in each of these categories and present some summary statistics on them.

4.1. User-Level Data

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

- $gender_i$: A dummy variable indicating user i's gender that is one for men and zero for women.
- age_i: User i's age.[8]
- bio_i: The length of user i's bio in his or her profile (i.e., number of words).
- $education_i$: Categorical variable that denotes the user i's highest education level (either earned or working toward), valued at one for high school, two for college, and three for graduate school.
- $employment_i$: Number of positions/companies mentioned in user i's profile.
- $initial\_game_i$: Total number of games played by user i before the data collection period.
- $total\_game_i$: Total number of games played by user i during the data collection period.
- num_pic_i: Number of uploaded pictures in the dating profile.

In addition, we also have access to the profile picture of user i. To obtain a measure of the physical attractiveness of a user's profile picture, we conducted a survey. We asked 384 heterosexual subjects in a research laboratory to rate the profile pictures of people of the opposite sex (men rated women and vice versa), on a scale of one to seven, with one being "not at all attractive" and seven being "very attractive." The subjects were undergraduate students at a large state university on the West Coast, with equal fractions of males and females, and their ages ranged between 18 and 25 (with a median age of 21). This demographic distribution closely mimics the age and gender distribution of the app users.

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

- pic_score_i: The average physical attractiveness score of user i's profile picture.

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

The summary statistics of all the user-level variables are shown in Table 1. Of the 24,653 users, 14,189 (57.55%) are male and 10,464 (42.45%) are female. The median user is 21 years old, has no bio written on her or his profile, has/is working toward a college degree, and one piece of employment-related information is listed on her or his profile. He or she has played 48 games before the data collection period and plays 18 games during it. However, there is quite a bit of variation across users in the extent of activity, with some users playing over 1,000 games during our observation period.

Finally, note that the above user-specific variables are treated as time invariant because users lack the ability to change most of their profile information after it is first imported from their Facebook profile (name, gender, age, education and employment information, and profile picture). The two pieces of information that users can change in the app are (1) the five extra pictures that they are allowed to upload (in addition to the profile picture) and (2) their short bio. However, we do not believe that this was a frequent occurrence for the reasons discussed in Section 6.3.2.

4.2. User-User-Level Data

Each game consists of eight unique users—four men and four women. For each man–woman pair in a game, we have data on the preference rankings that

Table 1. Summary Statistics of User-Level Data

Variables	Mean	SD	25th	50th	75th	(Min, max)	Size
age_i	21.53	5.41	19	21	22	(13, 109)	22,024
bio_i	67.04	275.58	0	0	63	(0, 29,519)	22,948
$employment_i$	1.29	1.60	0	1	2	(0, 68)	24,653
$initial\_game_i$	59.50	64.32	0	48	90	(0, 2,146)	24,653
$total\_game_i$	31.27	37.90	6	18	45	(1, 1,069)	24,653
num_pic_i	4.26	1.01	4	4	4	(0, 6)	22,669
pic_score_i	0.00	0.68	−0.52	−0.09	0.43	(−2.88, 3.29)	17,739
$gender_i$	(0) Female: 42.45%		(1) Male: 57.55%				24,653
$education_i$	(1) High school: 19.24%		(2) College: 78.12%		(3) Graduate: 2.64%		21,604

they gave each other, their match outcome, and their postmatch interactions. We describe these variables in detail below:

- $pref_{ijt}$: The preference ranking that user i *receives* from user j in game t; it can take values from one to four, with four indicating the highest preference and one the lowest. Users rank members of the opposite sex from one through four (as shown in Figure 1), with a rank of one indicating their highest preference and four indicating the lowest. We convert these rank orderings to preference rankings, such that a rank of one denotes a preference ranking of four, rank of two indicates a preference ranking of three, and so on. The transformed variable *pref* is easier to interpret because higher values of this variable correspond to more preference.
- $match_{ijt}$: A dummy variable indicating whether user i is matched with player j in game t. In each game, all players are uniquely matched with one other player from the opposite sex. So for woman (man) i in a game, this variable is equal to one for only one man (woman).
- $first_{ijt}$: A dummy variable indicating whether user i receives the first message from the matched partner (denoted by j) after game t. Because users cannot communicate with players that they have not been matched with, by default, this variable is zero if $match_{ijt} = 0$.
- $reply_{ijt}$: A dummy variable indicating whether user i receives a reply message from the matched partner j after game t, conditional on user i initiating the first message. By default, this variable is zero if $first_{jit} = 0$.

The summary statistics of these variables are shown in Table 2. The sample sizes of *pref* and *match* reflects the fact that there are 32 observations per game.[9] The distributions of *pref* and *match* are determined by the game structure, and their summary statistics are as expected. The sample size of $first_{ijt}$ reflects the fact that there are eight users matched with each other, and each of them can potentially initiate the first message. It is worth noting that the mean of $first_{ijt}$ is around 0.05 (of the 713,014 matches, only 39,377 messages were initiated). The observed number of first messages (39,377) defines the sample size of $reply_{ijt}$. The mean of $reply_{ijt}$ is around 0.08 (among 39,377 first messages only 3380 receive a reply). Interestingly, 76% of the conversations are initiated by men, which indicates that women are less likely to approach men after being matched. Furthermore, men receive replies to their messages 5% of the time, and women receive replies 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 with women (Kurzban and Weeden 2005, Fisman et al. 2006, Hitsch et al. 2010b).

Table 2. Summary Statistics of User-User-Level Data

Variables	Mean	SD	25th	50th	75th	(Min, max)	Size
$pref_{ijt}$	2.5	1.12	2	3	4	(1, 4)	3,008,560
$match_{ijt}$	0.25	0.43	0	0	0.5	(0, 1)	3,008,560
$first_{ijt}$	0.05	0.23	0	0	0	(0, 1)	713,014
$reply_{ijt}$	0.08	0.28	0	0	0	(0, 1)	39,377

4.3. User-Game-Level Data

We now describe user-game-level variables, that is, user-specific data that vary with each game:

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

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

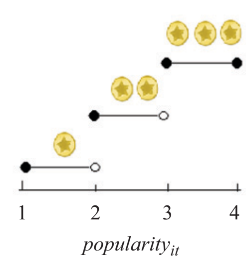
- $total\_game_{it}$: Total number of games that user i has played before game t. This is updated by one after each game played by user i.
- $star_{it}$: Indicates the user's star rating in game t; see Figure 1 for an example. The user's star rating is updated in real time after each game and is calculated as follows:

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

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

Figure 3 illustrates the relationship defined in Equation (2). Intuitively, an individual's star rating captures how popular or sought after she or he was in her or 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

Figure 3. (Color online) Pictorial Representation of the Star-Rating Rule



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

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

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

5. Descriptive Analysis

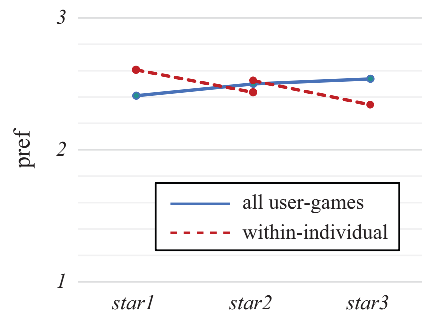
We now examine the relationship between a user's star rating and three measures of her or his demand—preference rankings received during the game and whether she or he receives a first message or reply message after the game—using simple model-free analyses. In this section, we focus on users who experienced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

The relationship between a user's star rating in a given game and the average preference ranking that she or he receives in that game is illustrated in Figure 4. The solid increasing line shows the relationship between the average preference rankings received for all user-game observations calculated for each star rating.[11] We see that in observations where users have higher star ratings, they also receive higher preference rankings. However, there is an obvious issue of correlated unobservables here, that is, users with higher star ratings are likely to be more attractive

Table 3. Summary Statistics of User-Game-Level Variables

Variables	Mean	SD	25th	50th	75th	(Min, max)	Size
$match\_level_{it}$	3.19	0.95	3	3	4	(1, 4)	752,140
$total\_game_{it}$	73.75	74.25	29	59	97	(0, 2,194)	752,140
$star_{it}$	2.00	0.10	2	2	2	(1, 3)	745,037

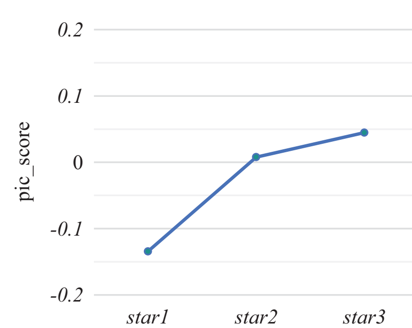
Figure 4. (Color online) Relationship Between Star Ratings and Average Preference Rankings Received



on other unobserved dimensions (e.g., physical attractiveness) as well. To examine whether this conjecture is true, we plot the average of users' *pic_score* for each star rating. As shown in Figure 5, users with higher star ratings also have higher physical attractiveness scores, on average. Thus, the effects shown by the solid line in Figure 4 cannot be interpreted as causal.

One way to cleanly capture the effect of star ratings is to look at the effect of star ratings *within* an individual, that is, if we compare preference rankings received by the same individual when she of he is shown with different star ratings, then our comparisons are less likely to be subject to endogeneity concerns. Consider an individual who was shown with a minimum of one star and a maximum of two stars and calculate two averages: (1) the average of preference rankings received in games where she or he is shown with one star, and (2) the average of preference rankings received in games where she or he is shown with two stars. We then perform an analogous exercise for users who were shown with a minimum of two stars and a maximum of three stars. These two comparisons are shown using dashed lines in Figure 4. As we can see, on average, the same set of users receive higher preference rankings when they are

Figure 5. (Color online) Relationship Between Star Ratings and Average Physical Attractiveness Score



shown with one star compared with two stars. Moreover, on average, the same set of users receive higher preference rankings when they are shown with two stars compared with three stars. This implies that higher star ratings lead to lower preference rankings; that is, users avoid those with higher stars. Note that the directions of the effect of star rating on preference rankings in solid line and dashed lines in Figure 4 are exactly opposite. This discrepancy implies that controlling for the endogeneity between star ratings and unobserved factors that affect user attractiveness is essential to deriving the causal impact of star ratings in our setting.

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

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

Figure 6. (Color online) Relationship Between Star Ratings and the Average Likelihood of Receiving the First Message

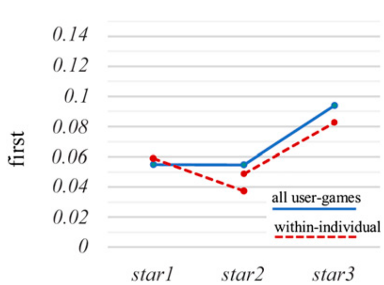
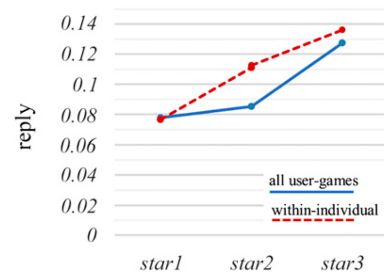


Figure 7. (Color online) Relationship Between Star Ratings and the Average Likelihood of Receiving a Reply Message



6. Effect of Star Ratings on Preference Rankings

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

6.1. Model Specification

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

$$pref^*_{ijt} = \beta_1 star1_{it} + \beta_2 star3_{it} + \gamma z_i + \eta_i + \epsilon_{ijt}, \qquad (3)$$

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

- $star1_{it}$, $star3_{it}$: Indicator variables for the star rating of user i in game t, where $star2_{it}$ is considered as the base.
- z_i: Set of user-specific observables that can affect j's ranking of i, for example, gender of i.
- η_i: Set of unobservable (to the researcher) characteristics of user i that is visible to j and affects j's ranking of i. These could include the aspects of user i's physical attractiveness not captured in our laboratory study (e.g., other photos of the user), details in her or his bio description, employment details, her geographic location, etc.
- ϵ_{ijt}: Factors uncorrelated to the star rating of user i that can affect the preference ranking she or he receives from j in game t. We assume that ϵ_{ijt}'s have a logistic cumulative distribution. Three key sets of variables are subsumed in ϵ_{ijt}. First, it includes j's attributes (both observable z_j and unobservable η_j) because there is no correlation between j's and i's attributes. Second, it

includes all the attributes of the other three players of i's gender who i is being compared with, in game t.[13] The reason neither of the above two sets of variables affects our inference on star ratings is because the app adds users into a game randomly. Thus, there is no correlation between the attributes of users within a game.[14] Third, it can capture idiosyncratic factors that affect j's ranking of i within the game, for example, j's mood for going on a date with someone of i's type, etc.

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

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

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

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

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

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

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

$$\begin{aligned} LL(\beta_1, \beta_2, \gamma, \eta_i, \mu_k, \mu_{k+1}) = \sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{j=1}^{4}\sum_{k=1}^{4} \ln \\ \times \left[Pr\Big(pref_{ijt} = k \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_i, \mu_k, \mu_{k+1}\Big)^{I(pref_{ijt}=k)} \right], \end{aligned} \tag{7}$$

where N is the total number of users observed, and T_i is the total number of games played by user i. The unknown parameters in Equation (7) are $\beta_1, \beta_2, \gamma, \eta_i s$, μ_2, μ_3, μ_4.

6.2. Estimation

We are mainly interested in estimating the effect of star ratings (β_1 and β_2). The challenge comes from the potential correlation between η_i and $star_{it}$, that is, we expect that $E[star_{it} \cdot \eta_i] \neq 0$. We now discuss three estimation strategies that address this problem in varying degrees.

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

The second strategy is a pooled ordered logit model with control variables that includes user-specific variables (z_i) to control for the correlation between $star_{it}$'s and η_i. For example, controlling for users' physical attractiveness (pic_score_i) may reduce the bias in estimates of β_1 and β_2. However, this method is unable to control for the correlation between $star_{it}$'s and η_i.

Third is a fixed-effects ordered logit model, where we allow the user-specific unobservables η_i to be arbitrarily correlated with the star ratings. A naive approach to estimation with fixed-effects is to treat the η_i's as parameters and maximize the log-likelihood in Equation (7) directly. However, such a maximum likelihood estimator is inconsistent with large N and finite T because of the well-known incidental parameters problem (Neyman and Scott 1948). As a result, the estimates of β_1 and β_2 from this approach will be inconsistent too. Chamberlain (1980) provides an elegant solution to the incidental parameters problem by dichotomizing the ordered outcome variable. In Section 6.2.1, we describe how to apply the Chamberlain estimator to our setting, in Section 6.2.2, we describe how the Chamberlain estimators can be combined to form an efficient minimum distance (MD) estimator.

6.2.1. Chamberlain's Conditional Maximum Likelihood Estimator. The ordered outcome variable $pref_{ijt}$ can take K = 4 possible integer values, {1, 2, 3, 4}. Therefore, we can transform the random variable $pref_{ijt}$ into $K - 1 = 3$ possible binary variables $pref^k_{ijt}$, where

$$pref^k_{ijt} = I(pref_{ijt} \ge k), \quad \text{where} \quad k = 2,3,4. \tag{8}$$

For example, the binary variable $pref^4_{ijt}$ indicates whether user i received a preference ranking of four from user j in game t or not. Similarly, the binary variable $pref^3_{ijt}$ indicates whether user i receives a preference ranking of three or higher (i.e., three or four) from user j in game t or not. We can specify Chamberlain's conditional maximum likelihood (CML) estimator on each of these transformed binary variables. For each k, $pref^k_{ijt}$ is a binary logit variable such that

$$\begin{aligned} Pr(pref^k_{ijt} = 1 \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_i, \mu_k) \\ = 1 - \Lambda(\mu_k - \beta_1 star1_{it} - \beta_2 star3_{it} - \gamma z_i - \eta_i). \end{aligned} \tag{9}$$

We denote by $pref^k_i$ the entire history of preference rankings at level k received by user i over time, that is, $pref^k_i = \{pref^k_{i11}, pref^k_{i21}, pref^k_{i31}, pref^k_{i41}, \ldots, pref^k_{i1T_i}, pref^k_{i2T_i}, pref^k_{i3T_i}, pref^k_{i4T_i}\}$. Furthermore, let s^k_i be the sum of all the binary transformed preference rankings at level k received by user i such that $s_i^k = \sum_{t=1}^{T_i} \sum_{j=1}^{4} pref_{ijt}^k$. In other words, s_i^k shows the count of ones in the set of $pref_i^k$. Let B_i^k be the set of all possible vectors of length $4 \times T_i$ with s_i^k elements equal to one and $4 \times T_i - s_i^k$ elements equal to zero;[15] that is,

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

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

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

A key observation is that this conditional probability does not depend on η_i's (or the thresholds μ_k's or z_i's); that is, s_i^k is a sufficient statistic for η_i. Thus, we can now specify a conditional log-likelihood (CLL) that is independent of η_i's and μ_k's as shown below:

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

Because we can dichotomize $pref_{ijt}$ into three binary variables at each of the three cutoffs ($pref_{ijt}^4$, $pref_{ijt}^3$, and $pref_{ijt}^2$), the above CLL can be specified for each $pref_{ijt}^k$, where $k \in \{2,3,4\}$. Maximizing each of these CLLs gives us three separate but consistent estimates of β_1, β_2, which we denote by $\{\beta_1^k, \beta_2^k\}$, where $k \in \{2, 3, 4\}$. These are referred to as Chamberlain CML estimators.

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

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

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

where $\tilde{\beta}$ is the 6×1 matrix of Chamberlain estimators, M is the matrix of three stacked two-dimensional identity matrices, and $var(\tilde{\beta})$ is the variance-covariance matrix of the stacked Chamberlain estimates.

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

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

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

6.3. Identification

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

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

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

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

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

In order to be able to manipulate their star rating in any period t, users need to be aware of and be able to meaningfully change their popularity score ($popularity_{it}$) by manipulating their profile information. This is not feasible for a few reasons. First, as discussed in Section 4.1, users lack the ability to change many aspects of their profile in response to their star ratings. Although they can add few additional pictures and/or modify their bios, both of these are not very critical because they are not shown in the main screen of a game (see Figure 2). Therefore, although a user can change these in response to their star ratings, we do not believe that this was a frequent occurrence. Second, although users are aware of their star rating at any given point in time ($star_{it}$), they do not observe any of the ranks that they received in the past games or their popularity score ($popularity_{it}$) at any point in time. (They are simply shown the person they are matched with after each game; the rankings that they received from other players are never revealed to them.) Moreover, users were never informed of the threshold rule used by the platform to assign the star ratings. Although users may have correctly inferred that their star ratings are correlated with their prior rankings, they are unlikely to have inferred the exact rule. Finally, the marginal effect of the rankings received in a new game on the popularity score is vanishingly small as the number of games played increases (see Online Appendix B, Section B.2.2 for details). Thus, as user's gain experience, it is increasingly hard for them to move the needle on their popularity score (and their star rating).

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

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

6.4. Results

The results from the estimation exercise are presented in Table 4. As discussed in Section 6.2, we estimate three ordered logit models: (1) Model M1, a simple specification that includes only star ratings as the independent variable; (2) Model M2, a more elaborate model that also includes user-specific observables (z_i), and (3) Model M3, a fixed-effects model using MD estimator that controls for η_i's.

In the basic ordered logit model (Model M1), we see a positive and significant effect for higher star ratings; that is, one-star users receive lower preference rankings compared with two-star users, and two-star users receive lower preference rankings compared with three-star users. This result is consistent with Figure 4 (solid line). Next, we estimate Model M2, which controls for all the user-specific observables because a user's current star rating is likely to be positively correlated to user-specific observables such as physical attractiveness, age, education, etc.[18] However, the direction of the results remain unchanged. Nevertheless, without explicitly controlling for the endogeneity concerns discussed earlier ($E[star_{it}.\eta_i] \neq 0$), our estimates are likely to be biased. Therefore, we now focus on the results from the fixed-effects MD estimator (Model M3). Interestingly, here we find that the effect of star rating is *negative*—a user gets a worse preference ranking when she or he is shown with three stars as opposed to two stars. We do not find any significant effect of one star compared with two stars. In Section

Table 4. Ordered Logit Estimates of the Effect of Star Rating on Preference Rankings Received

	(M1) Ordered logit	(M2) Ordered logit	(M3) FE ordered logit
$star1_{it}$	−0.14452***	−0.12991***	0.02852
	(0.02315)	(0.02312)	(0.01804)
$star3_{it}$	0.06063***	0.06863***	−0.05101***
	(0.01560)	(0.01893)	(0.01464)
Controls (z_i)		✓	
Fixed effects (η_i)			✓
μ_2	−1.09924***	−1.11212***	
	(0.00203)	(0.01197)	
μ_3	−0.00053	−0.00893	
	(0.00188)	(0.01192)	
μ_4	1.09828***	1.09239***	
	(0.00205)	(0.01192)	
Individuals	24,393	16,461	3,494
Observations	2,980,148	2,339,168	630,160

Notes. Standard errors (in parentheses) are clustered at the user level. Controls (z_i) in Model M2 include age_i, $college_i$, $graduate_i$, pic_score_i, num_pic_i, $employment_i$, and bio_i. FE, fixed effect.
***$p < 0.01$.

8, we present a battery of robustness checks to confirm the validity of these empirical findings.

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

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

7. Effect of Star Ratings on Messaging Behavior

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

7.1. Model and Estimation

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

$$first_{ijt} = \begin{cases} 1 & \text{if} \quad first^*_{ijt} > 0 \\ 0 & \text{else} \end{cases}, \quad reply_{ijt} = \begin{cases} 1 & \text{if} \quad reply^*_{ijt} > 0 \\ 0 & \text{otherwise} \end{cases}. \tag{14}$$

These latent variables are defined as

$$first^*_{ijt} = \beta^f_1 star1_{it} + \beta^f_2 star3_{it} + \gamma^f z_i + \eta^f_i + \epsilon^f_{ijt}, \tag{15}$$

$$reply^*_{ijt} = \beta^r_1 star1_{it} + \beta^r_2 star3_{it} + \gamma^r z_i + \eta^r_i + \epsilon^r_{ijt}, \tag{16}$$

where the interpretations of $\{\beta^f_1, \beta^f_2, \gamma^f, \eta^f_i, \epsilon^f_{ijt}\}$ and $\{\beta^r_1, \beta^r_2, \gamma^r, \eta^r_i, \epsilon^r_{ijt}\}$ are similar to those in Section 6.1. Furthermore, following the same arguments, we allow for η^f_i and η^r_i to be arbitrarily correlated to $star1_{it}$ and $star3_{it}$. Assuming that ϵ_{ijt}'s are independent and identically distributed and drawn from a logistic distribution, the probability that user i receives a first message from user j (conditional on i and j being matched in game t) is

$$Pr(first_{ijt} = 1 \mid match_{ijt} = 1, X_{it}, \eta^f_i) = \frac{exp(\beta^f_1 star1_{it} + \beta^f_2 star3_{it} + \gamma^f z_i + \eta^f_i)}{1 + exp(\beta^f_1 star1_{it} + \beta^f_2 star3_{it} + \gamma^f z_i + \eta^f_i)}.$$

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

$$Pr(reply_{ijt} = 1 \mid match_{ijt} = 1, first_{jit} = 1, X_{it}, \eta^r_i) = \frac{exp(\beta^r_1 star1_{it} + \beta^r_2 star3_{it} + \gamma^r z_i + \eta^r_i)}{1 + exp(\beta^r_1 star1_{it} + \beta^r_2 star3_{it} + \gamma^r z_i + \eta^r_i)}.$$

As in the case of the ordered logit model, we can use these probabilities to specify two CLLs that are independent of η_i's and then maximize the two CLLs to derive consistent estimates of $\{\beta^f_1, \beta^f_2\}$ and $\{\beta^r_1, \beta^r_2\}$. Because these steps are very similar to those described in Section 6.2, we relegate the details to Online Appendix C.

7.2. Results

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

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

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

Table 5. Effect of Star Rating on Messages Received

	First message		Reply message	
	(M4) Logit	(M5) Logit FE	(M6) Logit	(M7) Logit FE
$star1_{it}$	0.14989	0.51448***	−0.05065	−0.09053
	(0.13078)	(0.12034)	(0.22566)	(0.31932)
$star3_{it}$	0.63824***	0.73056***	0.46113***	0.40377**
	(0.09112)	(0.07482)	(0.15023)	(0.18095)
Controls (z_i)	✓		✓	
Controls (z_j)	✓	✓	✓	✓
Fixed effects (η_i)		✓		✓
Constant	−4.17057***		−2.04840***	
	(0.08135)		(0.30083)	
Individuals	16,364	1,797	3,446	385
Observations	436,652	83,693	25,062	6,566

Notes. Standard errors (in parentheses) are clustered at the user level. Controls (z_i) include $gender_i$, age_i, $college_i$, $graduate_i$, pic_score_i, num_pic_i, $employment_i$, and bio_i. Controls (z_j) include age_j, $college_j$, $graduate_j$, pic_score_j, num_pic_j, $employment_j$, bio_j, $star1_{jt}$, and $star3_{jt}$. FE, fixed effect.

$^{**}p < 0.05$; $^{***}p < 0.01$.

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

8. Robustness Checks

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

8.1. Effect of Stars on Preference Rankings—Linear Model

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

8.2. Estimation Sample

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

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

8.3. Within-Game Correlation

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

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

8.4. Star Configuration in a Game

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

9. Discussion of Mechanism

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

9.1. Players' Ranking and Messaging Strategy

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

9.1.1. Ranking Strategy During the Game. Let EU_{ijt} denote the expected utility that user j gets conditional on being matched with i, such that

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

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

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

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

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

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

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

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

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

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

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

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

9.2. Strategic Shading

We now formally define strategic shading.

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

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

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

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

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

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

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

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

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

The above inequality is based on Assumption 1, which asserts that users' ranking behavior during the game reflects their true preferences; that is, preference rankings reflect users' underlying expected utilities. Because we know from Inequality (20) that $U(star_{it} = 3) > U(star_{it} = 2)$, Inequality (21) can be explained only by rejection concerns, that is, due to perceived positive probability of rejection $\mathcal{P} < 1$ and nonzero cost of being rejected ($C > 0$). Thus, the negative effect of star ratings during the game can therefore be directly attributed to rejection concerns.

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

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

However, we do find that users' prior success in postmatch conversation shapes their ranking strategy. We stratified rank givers into two groups based on their prior conversation history as *successful* and *unsuccessful*. Successful rank givers are defined as those who have had more successful conversations with their past matches compared with the median user. Here a successful conversation from a user's perspective is defined as one where she or he either received a first message from the matched partner or received a reply to a message that she or he initiated. We find that the negative effect of popularity (or three-star rating) comes mainly from the unsuccessful rank givers (see Online Appendix E for the details of the model and the table of results). Indeed, we see that users who have been successful in engaging in postmatch conversations actually give higher preference rankings to three-star users. This suggests that the strategic shading mainly stems from users who have not had much success in postmatch conversations in the past, and therefore avoid popular users. Moreover, the personal nature of dating can give rise to significant psychological costs of rejection ($C > 0$). If users suffer from being rejected, then they will shade away from popular people whom they perceive as less likely to reciprocate in the postmatch conversation stage.[21]

9.3. Alternative Mechanisms

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

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

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

10. Conclusion

In this paper, we examine the effect of a user's popularity on her or his demand in a mobile dating app at different stages of the matching process and the drivers of these effects. Specifically, we document the causal impact of a user's star rating on the preference rankings that she receives during a game and her likelihood of receiving messages after a game. We show that, everything else being constant, compared with two-star users, three-star users receive lower preference rankings during the game but receive more reply messages after the game. We then link the heterogeneity across outcomes to the perceived severity of postmatch rejection concerns and establish strategic shading as the underlying mechanism for the negative effect of popularity during the game.

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

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

Acknowledgments

The authors thank an anonymous firm for providing the data and the UW-Foster High Performance Computing Laboratory for providing computing resources. They also thank Anja Lambrecht, Simha Mummamaleni, and Omid Rafieian for detailed comments, which improved this paper. Thanks are also due to seminar participants at the IESE Business School, University of California at Irvine, University of Maryland, University of Memphis, and University of Texas at Dallas, and participants at the 2018 Conference on Information Systems and Technology, the 2020 UT Dallas Frontiers of Research in Marketing Science Conference, and the 2021 Theory and Practice in Marketing Conference.

Endnotes

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

[2] Although our setting constitutes a centralized market, there remain significant postmatch rejection concerns. In this aspect, our setting is unlike standard centralized matches where matches are binding, for example, residents and hospitals in the NRMP (Roth and Sotomayor 1990).

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

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

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

[6] Similarly, a women-proposing Gale–Shapley deferred acceptance algorithm is woman optimal; that is, none of the women can do better using a different algorithm.

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

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

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

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

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

[12] We discuss other modeling frameworks such as rank-ordered logit or regression discontinuity design in Online Appendix A and explain why they are not appropriate for our setting.

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

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

$\| g_j - g_i \| \le 500$. Because we already control for user i's location (g_i) through η_i, the remaining δ is random noise.

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

[16] For individuals who have played a large number of games (large T_i) and have a large number of positive values of $pref_{ijt}^k$ (large s_i^k), calculating all combinations of outcomes can lead to numerical overflow and computational issues. For example, if user i plays 100 games ($T_i = 100$) and receives one preference ranking of four in each game, then $s_i^4 = 100$ and $\binom{4 \times 100}{100} = 2.24e + 96$. Therefore, we limit our empirical analysis to users' first 100 games. Of the 3,494 users who experience a star change, only 352 (10%) users play more than 100 games. The consistency of the estimates is not affected if we choose a subset of games for players who have played a large number of games.

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

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

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

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

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

References

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

Abdulkadiroğlu A, Pathak PA, Roth AE (2005) The New York City high school match. *Amer. Econom. Rev*. 95(2):364–367.

Anderson M, Vogles EA, Turner E (2020) The virtues and downsides of online dating. Accessed March 15, 2021, https://www.pewresearch.org/internet/2020/02/06/the-virtues-and-downsides-of-online-dating/.

Arnosti N, Johari R, Kanoria Y (2021) Managing congestion in matching markets. *Manufacturing Service Oper. Management*, ePub ahead of print January 19, https://doi.org/10.1287/msom.2020.0927.

Baetschmann G, Staub KE, Winkelmann R (2015) Consistent estimation of the fixed effects ordered logit model. *J. Roy. Statist. Soc. Series A* 178(3):685–703.

Bapna R, Ramaprasad J, Shmueli G, Umyarov A (2016) One-way mirrors in online dating: A randomized field experiment. *Management Sci*. 62(11):3100–3122.

Biddle JE, Hamermesh DS (1998) Beauty, productivity, and discrimination: Lawyers' looks and lucre. *J. Labor Econom*. 16(1):172–201.

Cai H, Chen Y, Fang H (2009) Observational learning: Evidence from a randomized natural field experiment. *Amer. Econom. Rev*. 99(3):864–882.

Cameron JJ, Stinson DA, Wood JV (2013) The bold and the bashful: Self-esteem, gender, and relationship initiation. *Soc. Psych. Personality Sci*. 4(6):685–691.

Chamberlain G (1980) Analysis of covariance with qualitative data. *Rev. Econom. Stud*. 47(1):225–238.

Che YK, Koh Y (2016) Decentralized college admissions. *J. Political Econom*. 124(5):1295–1338.

Chen Y, Wang Q, Xie J (2011) Online social interactions: A natural experiment on word of mouth vs. observational learning. *J. Marketing Res*. 48(2):238–254.

Chevalier JA, Mayzlin D (2006) The effect of word of mouth on sales: Online book reviews. *J. Marketing Res*. 43(3):345–354.

Choo E, Siow A (2006) Who marries whom and why. *J. Political Econom*. 114(1):175–201.

Das M, Van Soest A (1999) A panel data model for subjective information on household income growth. *J. Econom. Behav. Organ*. 40(4):409–426.

Dewan S, Ho YJ, Ramaprasad J (2017) Popularity or proximity: Characterizing the nature of social influence in an online music community. *Inform. Systems Res*. 28(1):117–136.

Duan W, Gu B, Whinston AB (2009) Informational cascades and software adoption on the internet: An empirical investigation. *MIS Quart*. 33(1):23–48.

Eastwick PW, Finkel EJ (2008) Sex differences in mate preferences revisited: Do people know what they initially desire in a romantic partner? *J. Personality Soc. Psych*. 94(2):245–264.

Fisman R, Iyengar SS, Kamenica E, Simonson I (2006) Gender differences in mate selection: Evidence from a speed dating experiment. *Quart. J. Econom*. 121(2):673–697.

Fisman R, Iyengar SS, Kamenica E, Simonson I (2008) Racial preferences in dating. *Rev. Econom. Stud*. 75(1):117–132.

Fong J (2020) Search, selectivity, and market thickness in two-sided markets: Evidence from online dating. Working paper, University of Michigan Ross School of Business.

Gale D, Shapley LS (1962) College admissions and the stability of marriage. *Amer. Math. Monthly* 69(1):9–15.

Hansen SL (1977) Dating choices of high school students. *Family Coordinator* 26(2):133–138.

Hitsch GJ, Hortaçsu A, Ariely D (2010a) Matching and sorting in online dating. *Amer. Econom. Rev*. 100(1):130–163.

Hitsch GJ, Hortaçsu A, Ariely D (2010b) What makes you click? Mate preferences in online dating. *Quant. Marketing Econom*. 8(4):393–427.

Hole AR, Dickerson A, Munford L (2011) A review of estimators for the fixed-effects ordered logit model. *United Kingdom Stata Users Group Meeting* 2011, vol. 5 (Stata Users Group).

IBISWorld (2019) Dating services industry in the US—Market research report. Accessed March 15, 2021, https://www.ibisworld.com/industry-trends/market-research-reports/other-services-except-public-administration/personal-laundry/dating-services.html.

Jung J, Bapna R, Ramaprasad J, Umyarov A (2019) Love unshackled: Identifying the effect of mobile app adoption in online dating. *MIS Quart*. 43(1):47–72.

Kurzban R, Weeden J (2005) HurryDate: Mate preferences in action. *Evolution Human Behav*. 26(3):227–244.

Lee S (2016) Effect of online dating on assortative mating: Evidence from South Korea. *J. Appl. Econometrics* 31(6):1120–1139.

Ludden J (2016) Do You Like Me? NPR (February 11). Accessed March 15, 2021, https://www.npr.org/sections/thetwo-way/2016/02/11/466342716/do-you-like-me-swiping-leads-to-spike-in-online-dating-for-young-adults.

Moretti E (2011) Social learning and peer effects in consumption: Evidence from movie sales. *Rev. Econom. Stud*. 78(1):356–393.

Neyman J, Scott EL (1948) Consistent estimates based on partially consistent observations. *Econometrica* 16(1):1–32.

Niederle M, Yariv L (2009) Decentralized matching with aligned preferences. Technical report, National Bureau of Economic Research, Cambridge, MA.

Niederle M, Roth AE, Sönmez T (2008) Matching and market design. Derlauf S, Blume L, eds. *The New Palgrave Dictionary of Economics*, 2nd ed. (Palgrave Macmillan, Hampshire, UK).

Roth AE (1982) The economics of matching: Stability and incentives. *Math. Oper. Res*. 7(4):617–628.

Roth AE (2008) Deferred acceptance algorithms: History, theory, practice, and open questions. *Internat. J. Game Theory* 36(3–4):537–569.

Roth AE, Peranson E (1999) The redesign of the matching market for American physicians: Some engineering aspects of economic design. *Amer. Econom. Rev*. 89(4):748–780.

Roth AE, Sotomayor MAO (1990) *Two-Sided Matching* (Cambridge University Press, Cambridge, UK).

Roth AE, Xing X (1997) Turnaround time and bottlenecks in market clearing: Decentralized matching in the market for clinical psychologists. *J. Political Econom*. 105(2):284–329.

Safronova V (2018) An inside look at your favorite dating sites. *New York Times* (April 11), https://www.nytimes.com/2018/04/11/style/match-shaadi-league-farmersonly-dating-apps.html.

Salganik MJ, Dodds PS, Watts DJ (2006) Experimental study of inequality and unpredictability in an artificial cultural market. *Science* 311(5762):854–856.

Shaw Taylor L, Fiore AT, Mendelsohn G, Cheshire C (2011) "Out of my league": A real-world test of the matching hypothesis. *Personality Soc. Psych. Bull*. 37(7):942–954.

Sorensen AT (2007) Bestseller lists and product variety. *J. Indust. Econom*. 55(4):715–738.

Sun M (2012) How does the variance of product ratings matter? *Management Sci*. 58(4):696–707.

Toch T, Aldeman C (2009) Matchmaking: Enabling mandatory public school choice in New York and Boston. Report, Education Sector, Washington, DC.

Tucker C, Zhang J (2011) How does popularity information affect choices? A field experiment. *Management Sci*. 57(5):828–842.

Wong LY (2003) Structural estimation of marriage models. *J. Labor Econom*. 21(3):699–727.

Yoganarasimhan H (2012) Impact of social network structure on content propagation: A study using YouTube data. *Quant. Marketing Econom*. 10(1):111–150.

Yoganarasimhan H (2013) The value of reputation in an online freelance marketplace. *Marketing Sci*. 32(6):860–891.

Zhang J (2010) The sound of silence: Observational learning in the US kidney market. *Marketing Sci*. 29(2):315–335.