# Search Frictions and Sorting in Two-Sided Markets*

Hyesung Yoo† Song Yao‡ Ravi Bapna§ Jui Ramaprasad¶

February, 2025

**Abstract**

Search frictions and preferences substantially affect matching outcomes in two-sided markets. This study investigates these effects using data from a field experiment on an online dating platform. On the platform, a user typically incurs a cost to access a potential partner's preference regarding the focal user. In the experiment, however, treated users received this preference information for free, thereby reducing their search frictions. The findings reveal that lowering the frictions significantly decreases sorting, leading to matches between users with more dissimilar attributes. To explore the effects further, we develop a structural model that quantifies the roles of user preferences and search frictions in shaping matching outcomes. The results indicate that reducing search frictions not only reduces sorting but also enhances market efficiency, allowing users to connect with partners who more closely align with their preferences. This research highlights how platform designs that reduce frictions, such as by enabling free access to preference signaling, can promote diversity in matches and improve matching efficiency in two-sided markets, offering valuable insights for online platform development.

*We thank Rodrigo Belo, Tat Chan, Yulia Nevskaya, Raphael Thomadsen, and Maria Ana Vitorino for their comments. We also thank feedback by seminar participants at Arizona State University, Columbia University, INFORMS Marketing Science Conference 2020, KAIST, the Ohio State University, Santa Clara University, the University of Miami, the University of Minnesota, the University of Toronto, the University of Virginia, Washington University in St. Louis, the WISE Conference 2019, and the 2024 Payne Research Symposium at the University of Arizona.

†University of Toronto: hyesung.yoo@rotman.utoronto.ca

‡Washington University in St. Louis: songyao@wustl.edu

§University of Minnesota: rbapna@umn.edu

¶University of Maryland: jramapra@umd.edu

# 1 Introduction

Successful transactions on many two-sided platforms rely on participants being matched with one another. In these markets, participants have preferences regarding each other's characteristics. However, participants often do not observe the preferences of their potential matches and need to engage in costly searches to resolve the uncertainty. Consequently, to form a match, participants must not only rely on their preferences regarding the potential match but also assess its attainability to justify the costs incurred during the matching process. Accordingly, it may be rational for some agents to forgo potential matches that they prefer and instead pursue others whom they deem more attainable. To reduce such inefficiencies, many markets allow agents to signal their preferences to the other side of the market. For example, in US college admissions, many universities offer an "early decision" program, where high school students can apply to exactly one school before the regular application period and commit to matriculation if accepted, thus signaling to the school their strong preference. Similarly, in the economics Ph.D. job market, the American Economic Association (AEA) allows job applicants to signal their interest in receiving an interview with up to two potential employers.[1]

In this paper, we study the impact of search frictions and preference signaling on sorting, using data from an online dating platform. Sorting is a well-documented tendency for individuals to interact with others who are similar to themselves, along dimensions such as race, income, and education. This phenomenon is evident in various contexts, including workplace collaborations and household formations (Raquel and Rogerson, 2001), and has important implications for inequality. For instance, sorting in the labor market between workers and firms has been identified as an important driver of wage inequality (Card et al., 2018; Bagger and Lentz, 2019; Hong, 2024). Similarly, in the marriage market, sorting among couples has long-term consequences for economic development and inequality, particularly through its effects on the outcomes of children and the accumulation of human capital (Raquel and Rogerson, 2001; Raquel, 2003).

There are two major explanations for such sorting patterns (Kalmijn, 1998; Hitsch et al., 2010a): (1) The preference of an individual affects with whom they interact. For example, if people prefer those similar to themselves, sorting can be attributed to 'horizontal' preferences. Conversely, if

[1] See "AEA Guidance on the 2024-25 Economics Job Market Cycle.", https://www.aeaweb.org/news/member-announcements/2024-august-05, accessed on November 11, 2024.

preferences are 'vertical', meaning everyone ranks others based on the same criteria, then sorting will occur according to attributes aligned with these rankings. (2) Search frictions can influence with whom an individual interacts, regardless of the individual's preference. In online two-sided markets, search frictions typically take two common forms (Arnosti et al., 2021): (i) *screening* costs, which involve evaluating potential matches, and (ii) *communication* costs, which incur when an individual communicates with potential matches. Given these costs, individuals are more likely to balance their preferences with attainability, resulting in inefficient matches.

By providing information about the preferences of potential match candidates, signaling one's preference to potential matches can help avoid wasted effort, thereby mitigating inefficiencies caused by search frictions. In addition, preference signaling can influence sorting by directing users' pursuits toward more preferred candidates that align with their true preferences, rather than those who are merely more attainable.[2] For example, if people prefer similar (different) others, and if preference signaling helps users match with those they prefer more, this can lead to an increase (reduction) in sorting.

During the field experiment on the online dating platform that we study, each user in the treatment group was provided with a piece of information pertaining to whether a potential match candidate had $\ell iked$ the user–the information serving as a positive signal of the likelihood of a match. In contrast, the control group could only discover this signal if they first took action to $\ell ike$ the candidate. Thus, users in the control group could achieve two outcomes by $\ell iking$ a profile: (i) they could signal their interest, and (ii) they could find out if they were $\ell iked$ by the other user. We posit that $\ell iking$ a candidate is costly because it triggers anxiety due to the uncertainty about whether the candidate will $\ell ike$ them in reciprocation. Because the treatment group knows in advance who has $\ell iked$ them, they can $\ell ike$ those candidates without incurring the cost.

In our data, matches display sorting patterns, in part because users have lower probabilities of matching with candidates who are different from themselves. However, we find descriptive evidence suggesting that reducing search frictions through the treatment changes sorting patterns between matched couples. Specifically, when users matched with those who $\ell iked$ them, the couples in the treatment group displayed significantly greater differences in age, education, body type, and race

[2] According to Lee and Niederle (2015), people tend to respond more strongly to preference signals when the signal comes from a person who is considered more attractive than they are, which is also the conjecture in the AEA signaling mechanism (Coles et al., 2010).

compared to the control group's couples, while being more similar in desirability. We find that this is due to the $\ell ike$ signal encouraging users in the treatment group to reach out to those candidates with whom they would have a low likelihood of matching if the signal were non-existent.

To quantify how much of the sorting in matches observed on our platform is due to search frictions versus preferences, and to measure the extent to which the signaling reduces the influence of search frictions, we build a structural model of a two-stage decision process of a user searching for a partner. In the first stage, upon observing the characteristics of a candidate, the user first decides whether to $\ell ike$ the candidate's profile given the costs of $\ell iking$, the utility from a successful match, and their beliefs about the likelihood of a match. If the user decides to $\ell ike$ in the first stage, it would reveal whether the candidate has liked the user. In the second stage, the user decides whether to send a message. Because preferences and costs are interdependent in our model, it is empirically challenging to separately identify costs and utility parameters. Therefore, we rely on the user's treatment status for identification. The treatment status acts as an exclusion restriction that help identify costs and utility separately.

Based on the model estimates, we simulate who matches with whom under the default (control) setting, the treatment setting, and the frictionless environment using the Gale-Shapley algorithm. By comparing the matches in the control setting to the stable matches formed in a frictionless environment, we disentangle the relative impact of frictions and preferences on sorting. We find that frictions play a significant role in shaping the sorting patterns. For example, in man-optimal matches, frictions account for approximately 14% of sorting in age, 24% in years of schooling, 88% in desirability, and 26% in race for men. Additionally, we find that the treatment substantially reduces the impact of search frictions on sorting–resulting in a 74.14% reduction in sorting for age, 87.83% for years of education, 100% for desirability, and 92.16% for race.

We then examine the efficiency loss associated with search frictions. We first examine how much the treatment improves users' outcomes compared to the control setting, and then examine how much users gain when all the frictions are removed. To do this, we assign ordinal rankings to each matched partner based on estimated preference parameters and compare the average rankings achieved across different protocols. We find that reducing frictions by providing information about $\ell ikes$ through the treatment leads to a small but significant improvement. For example, in women-optimal matches, the treatment improves average partner rankings by 1.14 percentage points for men and 2.28 percentage points for women compared to the control setting. These results suggest that reducing search frictions through the provision of information via signaling significantly improves the match outcomes. When all frictions on the platform are removed, the average ranking of the partner is further improved, leading to a significant increase of 2.85 percentage points for men and 9.7 percentage points for women in terms of the highest achievable ranking.

In terms of net utility, which is the utility a user derives from a match minus the costs associated with $\ell iking$ and messaging, we find that, compared to the baseline control setting, the treatment increases net utility by 54% for men and 0.17% for women in men-optimal matches. When all frictions are removed, net utility increases by 104% for men, and 107% for women compared to the control setting.

The rest of this paper is structured as follows. The next section reviews the related literature and outlines this paper's contributions. Section 3 describes the institutional details of the dating platform and the details of the experimental design. Section 3.2 summarizes the data, and Section 4 presents descriptive evidence suggesting that reducing frictions through signaling may impact sorting patterns. In Section 5, we propose the structural model. Estimation details and results are discussed in Sections 6 and Section 7, respectively. Section 8 presents counterfactual exercises, and Section 9 concludes.

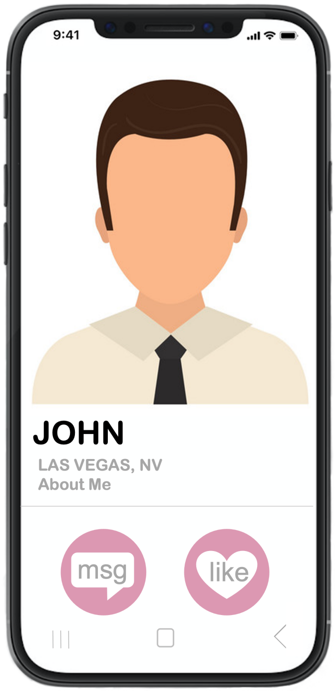
## 2 Related Literature

This paper closely relates to recent literature on frictions in online two-sided markets involving matching. Fradkin (2015) demonstrates the effects of search frictions on transactions on Airbnb, an online market for short-term real estate rentals. Horton (2014) shows that information frictions lead to inefficiencies in online labor markets. While these papers study frictions in two-sided online markets, our focus is on the impact of frictions on sorting in matches.

Our paper also relates to empirical work focusing on the estimation of mate preferences in romantic relationships (Wong, 2003; Choo and Siow, 2006; Flinn and Del Boca, 2012; Chan et al., 2015; Richards Shubik, 2015; Arcidiacono et al., 2016). While these papers use observed final match outcomes in their data to infer mate preferences, our data document each user's entire search process. The data enable us to estimate both preferences and costs based on users' actions at each stage of the decision process.

Finally, our paper advances the growing literature on speed-dating and online dating (Kurzban and Weeden, 2005; Fisman et al., 2006, 2008; Hitsch et al., 2010a,b; Lee, 2015; Lee and Niederle, 2015; Bapna et al., 2016; Halaburda et al., 2017; Fong, 2018; Shi and Viswanathan, 2023; Bruch and Newman, 2024; Huang et al., 2022; Jung et al., 2022). In particular, Lee (2015) shows that online dating promotes weaker sorting along geographic proximity but stronger sorting along education. Lee and Niederle (2015) find that signaling one's preference increases the success probability of a dating request. Bojd and Yoganarasimhan (2019) study the causal effect of popularity information in online dating and reveal evidence of strategic shading due to the fear of rejection. The findings of our research are qualitatively consistent with these studies on the effects of information signaling and users' fear of rejection. Using the same field experiment data as in this paper, Bapna et al. (2022) demonstrate how revealing "who likes you" affects user behavior in online dating. Our paper builds on previous literature and zooms in on the impacts of preferences and frictions on sorting. Furthermore, we quantify the extent to which signaling mitigates the impact of search frictions on sorting and efficiency losses.

The papers closest to ours are Hitsch et al. (2010a) and Banerjee et al. (2013). Using data from an online dating website, Hitsch et al. (2010a) show that matches predicted by the Gale-Shapley algorithm are similar to the actual matches, suggesting efficiency in the online dating market. They are also able to predict the sorting patterns observed in the matches, suggesting that sorting can arise even in the absence of search frictions and can be primarily driven by preferences and market mechanisms. Using a similar approach, Banerjee et al. (2013) study how caste preferences shape matching patterns, finding a strong preference for within-caste marriages. Compared to these two papers, our experiment introduces a source of exogenous variation in search frictions, which enables us to disentangle the contributions of frictions and preferences on the equilibrium sorting and efficiency.




*Note.* This figure is for illustrative purpose only. The actual app shows differently shaped/colored icons.

Figure 1: Profile of a Candidate Displayed to the Control Group

## 3 App Overview and the Experiment

The online dating platform underpinning our study is a typical "freemium" community: Most users sign up for a free account, which grants access to basic features (browsing profiles, *liking*, and sending messages), while some users pay a monthly subscription fee for a premium account that includes additional features, including the ability to know whether the candidate in the displayed profile has already *liked* the user. Our data sample only contains non-premium users. During the period of our study, there was no limit on the number of *likes* and messages that a user could send.

The platform is accessible through both the website and the mobile app. While the experiment involves randomly selected users from both channels, this paper focuses on mobile app users. In addition, we focus on users searching for partners through the "rapid matching" process, where a user looks at potential matches' profiles one by one, sequentially.[3] Below, we first elaborate on how the rapid matching process works for users in the *control group*, then describe how the treatment changes the process.

[3] In non-rapid matching, multiple profiles are displayed simultaneously, similar to an online retailer listing several comparable products all at once on its webpage.


(a) The candidate did not *like* you

(b) The candidate *liked* you

*Notes*: This figure illustrates how *liking* a profile reveals whether the candidate has *liked* the focal user. The *like* button turns red when the user chooses to *like*. If the candidate has already *liked* the focal user previously, both users receive a notification about the mutual *liking*. In addition to the notification, a heart icon appears in the upper right corner. If the candidate has not *liked* the focal user, neither a notification nor a heart icon will appear.

Figure 2: How *Liking* a Profile Reveals More Information

When a user in the control group opens the mobile app, a random profile is displayed to them. Figure 1 illustrates what the user sees, including the candidate's profile picture and characteristics such as age, race, and education level. Upon seeing the profile, the user can choose to *like* and/or send a message to the candidate. The user can choose to *like* by either clicking the *like* button or by swiping right. Similarly, the user can choose not to *like* by simply not clicking the *like* button or by swiping left.

More importantly, if a user *likes* a candidate and the candidate has already *liked* the user, both will receive a popup notification in their apps for their mutual *liking* (Figure 2b). Additionally, a heart icon appears next to the candidate's picture in the upper right corner, indicating the candidate has *liked* the focal user. Conversely, if neither a notification nor a heart icon appears upon the user choosing to *ℓike*, it implies that the candidate has not *ℓiked* the user, or has not yet seen the user (Figure 2a). The user is not able to distinguish between these two cases. In short, *ℓiking* a candidate reveals whether the candidate has *ℓiked* the user.

We posit that the action of *ℓiking* a profile entails a cost or disutility—the anxiety stemming from the uncertainty of reciprocity in *ℓiking*. Extensive research in psychology has established that people are often inhibited by fears of rejection from social overtures (Vorauer and Ratner, 1996; Vorauer et al., 2003), and the possibility of not being reciprocated in romantic advances can hurt one's ego (Baumeister et al., 1993). Because the information on mutual interest can be positive or negative, the uncertainty, particularly the potential rejection by the candidate, results in a cost or disutility when the user takes the action of *ℓiking*.

The focal user can also message the candidate they see in the profile. They may do so either without *ℓiking* or after they *ℓike* the candidate (i.e., after either Figures 2a or 2b appears). In both cases, a new candidate's profile will immediately appear after sending a message.

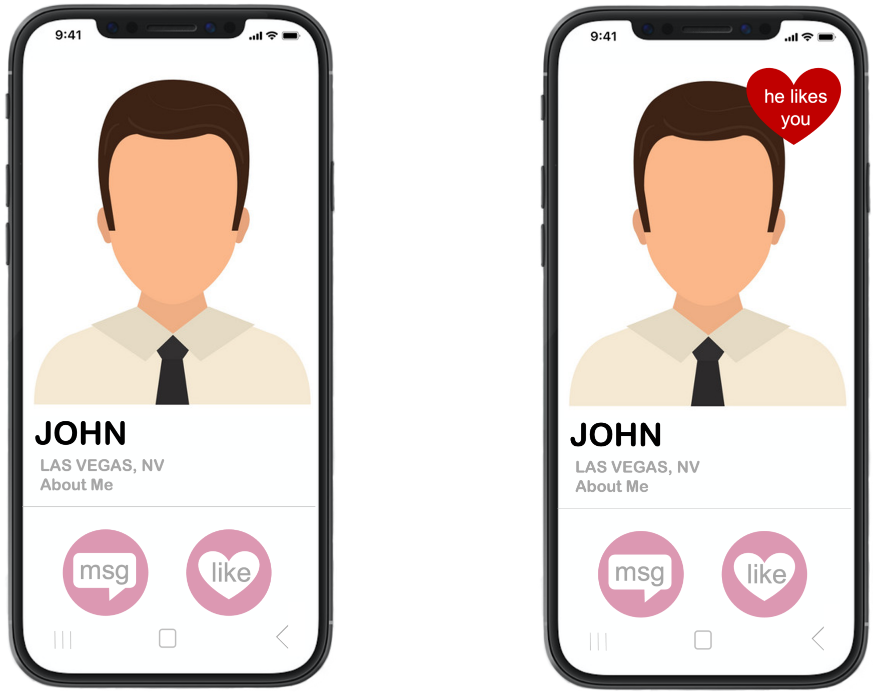
Finally, the user can always proceed to the next candidate's profile without messaging or *ℓiking* by clicking the "back" button in the lower right corner or swiping left.

### 3.1 The Experiment

In this section, we describe the design of the experiment. The experiment was conducted in 2016 and involved 100,000 randomly selected users who had newly registered on the platform (either via the website or the app) during the seven-day period prior to the experiment. These users accounted for less than 1% of the entire user population of the platform. The platform tracked these users' activities for three months. We refer to the three months as the pre-treatment (month 1), treatment (month 2), and post-treatment (month 3) periods, respectively. These 100,000 users were randomly and evenly divided into treatment and control groups. On the first day of the treatment period (i.e., month 2), the 50,000 treatment group users received the following email:

> *Hey username, you have been randomly selected to receive a superpower - for the next 30 days, we're giving you the ability to know whether someone has ℓiked you! Normally, this feature is restricted to paid premium users only. Enjoy!*[4]

[4]To disguise the identity of the platform, the messages presented in the paper are slightly modified from the actual messages sent to users.




(a) The candidate did not *like* you　　(b) The candidate *liked* you

Figure 3: Profile of a Candidate Displayed to the Treatment Group

The remaining 50,000 control group users received the following email:

> *Hey username, it's a good time to visit our platform! Enjoy!*

The treatment was endowed to users by the platform without requiring any further action on their part.

Figure 3 illustrates what a treated user saw when they opened the app. When a candidate's profile appears, the treated user can immediately see if the candidate has *liked* them. In particular, a heart icon appears in the upper right corner of the candidate's profile if the candidate has *liked* the focal user (Figure 3b); Otherwise, the heart icon is absent (Figure 3a). In short, treated users can tell if a candidate has *liked* them without incurring the psychological cost of uncertainty.

### 3.2 Final Sample and Balance Check

| | Control | | | Treatment | | | | |
|---|---|---|---|---|---|---|---|---|
| Variable | Mean | SD | Obs. | Mean | SD | Obs. | $t$-stat | $p$-value |
| **Panel A. Men** | | | | | | | | |
| Age | 31.508 | 9.278 | 5752 | 31.468 | 9.112 | 6055 | 0.2349 | 0.8143 |
| HighSchool | 0.128 | 0.334 | 1,002 | 0.117 | 0.321 | 1,055 | 0.7725 | 0.4399 |
| TwoYear | 0.175 | 0.380 | 1,002 | 0.177 | 0.382 | 1,055 | -0.1547 | 0.8770 |
| University | 0.554 | 0.497 | 1,002 | 0.537 | 0.499 | 1,055 | 0.7670 | 0.4432 |
| PostGrad | 0.144 | 0.351 | 1,002 | 0.169 | 0.375 | 1,055 | -1.5857 | 0.1130 |
| Skinny | 0.464 | 0.499 | 1,492 | 0.477 | 0.500 | 1,519 | -0.6678 | 0.5043 |
| Average | 0.424 | 0.494 | 1,492 | 0.409 | 0.492 | 1,519 | 0.7852 | 0.4324 |
| ExtraWeight | 0.093 | 0.291 | 1,492 | 0.090 | 0.287 | 1,519 | 0.2825 | 0.7775 |
| Overweight | 0.019 | 0.136 | 1,492 | 0.024 | 0.152 | 1,519 | -0.9382 | 0.3482 |
| Asian | 0.085 | 0.279 | 2,487 | 0.087 | 0.282 | 2,613 | -0.3073 | 0.7586 |
| White | 0.654 | 0.476 | 2,487 | 0.636 | 0.481 | 2,613 | 1.3123 | 0.1895 |
| Black | 0.079 | 0.270 | 2,487 | 0.089 | 0.285 | 2,613 | -1.2591 | 0.2080 |
| Hispanic | 0.103 | 0.303 | 2,487 | 0.114 | 0.317 | 2,613 | -1.2787 | 0.2011 |
| Other Race | 0.080 | 0.271 | 2,487 | 0.074 | 0.262 | 2,613 | 0.7717 | 0.4403 |
| Desirability | 2.011 | 0.812 | 5,559 | 1.988 | 0.822 | 5,873 | 1.5042 | 0.1326 |
| **Panel B. Women** | | | | | | | | |
| Age | 34.669 | 11.071 | 1,911 | 34.665 | 11.017 | 2,087 | 0.0119 | 0.9905 |
| HighSchool | 0.079 | 0.269 | 436 | 0.051 | 0.221 | 423 | 1.6313 | 0.1032 |
| TwoYear | 0.135 | 0.342 | 436 | 0.149 | 0.356 | 423 | -0.5709 | 0.5682 |
| University | 0.582 | 0.493 | 436 | 0.578 | 0.494 | 423 | 0.1264 | 0.8994 |
| PostGrad | 0.204 | 0.404 | 436 | 0.222 | 0.416 | 423 | -0.6469 | 0.5179 |
| Skinny | 0.308 | 0.462 | 615 | 0.318 | 0.466 | 641 | -0.3822 | 0.7024 |
| Average | 0.593 | 0.491 | 615 | 0.553 | 0.497 | 641 | 1.4444 | 0.1489 |
| ExtraWeight | 0.075 | 0.263 | 615 | 0.103 | 0.304 | 641 | -1.7517 | 0.0801 |
| Overweight | 0.024 | 0.154 | 615 | 0.027 | 0.161 | 641 | -0.2394 | 0.8108 |
| Asian | 0.118 | 0.323 | 972 | 0.140 | 0.347 | 1,054 | -1.4638 | 0.1434 |
| White | 0.616 | 0.486 | 972 | 0.605 | 0.489 | 1,054 | 0.5371 | 0.5912 |
| Black | 0.110 | 0.313 | 972 | 0.108 | 0.311 | 1,054 | 0.1386 | 0.8897 |
| Hispanic | 0.112 | 0.316 | 972 | 0.103 | 0.305 | 1,054 | 0.6329 | 0.5269 |
| Other Race | 0.043 | 0.203 | 972 | 0.044 | 0.204 | 1,054 | -0.0478 | 0.9619 |
| Desirability | 1.990 | 0.816 | 1,881 | 2.005 | 0.818 | 2,041 | -0.4117 | 0.6806 |

*Notes:* This table presents a comparison of the demographics of the control and treatment groups for users who were active during the treatment period, separately for each gender.

Table 1: User Demographics During the Treatment Period

For each of the 100,000 users in our experiment, we observe the following self-reported demographic variables: gender, sexual orientation, age, education level, race, and body type. Additionally, we observe each user's time-stamped actions (browsing, $\ell iking$, and messaging) over the three months, as long as they were active (browsed at least one profile). Furthermore, we observe time-stamped actions and demographics for all correspondent users (i.e., candidates) who interacted with the experimental users.[5] The data on correspondent users' *ℓiking* behavior allows us to observe who had *ℓiked* users in our experiment. We also observe whether a user was using a desktop or a mobile phone, has a premium account, and whether the account is valid (non-spammer/bot).

From the initial sample of users who were active during the treatment period, we limit our sample to heterosexual mobile app users who were searching for a partner using the *rapid matching* process. Our final sample consists of 8,142 treated and 7,663 control experimental users. To ensure randomization, we implement the balance check by testing for any differences in characteristics between the treatment and control groups during the treatment period. Table 1 summarizes the characteristics of the users by their treatment status, separately for men and women. We see that the control and treatment groups are statistically indistinguishable from each other in their characteristics. In Appendix A, we also report the test of pretreatment randomization, where we find no systematic difference between the treatment and control groups. One may also be concerned that the treated users with certain characteristics become more active due to the treatment (i.e., endogenous compliance), causing the treated and control groups to become systematically different. The results of the comparison in Table 1 mitigates such a concern. Furthermore, as an additional test to address the concern, we compare the number of sessions in which users were active by treatment status and gender in Appendix B. Again, we find no significant differences between the control and treatment groups for both men and women.

# 4 Summary Statistics and Descriptive Analysis

## 4.1 Impact of the Treatment on User Activities

We begin by showing the effect of the treatment on user activities. From the perspective of the focal user, there are two types of candidates: (1) candidates who had *ℓiked* them (henceforth "Likers"), and (2) candidates who did not *ℓike* them (henceforth "NotLikers"). Since the treatment allows

[5] In our data, less than 0.5% of the experimental users interacted with other members of the experiment, and less than 0.5% of the users in the treatment group interacted with other members in the treatment group. Hence, we do not have major concerns about contamination bias.

| | Men | | | | Women | | | |
|---|---|---|---|---|---|---|---|---|
| | control | treated | diff | t-stat | control | treated | diff | t-stat |
| | (1a) | (1b) | (1c) | (1d) | (2a) | (2b) | (2c) | (2d) |
| **Panel A. Matches with Likers** | | | | | | | | |
| *Likes sent* | | | | | | | | |
| Mean | 1.7808 | 2.0140 | 0.2333 | 2.5907 | 4.8817 | 6.4207 | 1.5390 | 3.3391 |
| Median | 0 | 1 | | | 1 | 2 | | |
| SD | 4.8778 | 4.9020 | | | 12.7792 | 16.0127 | | |
| *Initiated messages* | | | | | | | | |
| Mean | 0.7357 | 0.8912 | 0.1554 | 4.1160 | 0.9168 | 1.2372 | 0.3204 | 3.2550 |
| Median | 0 | 0 | | | 2.2510 | 3.7248 | | |
| SD | 2.0060 | 2.0925 | | | 2.2510 | 3.7248 | | |
| *Initiated messages that led to match* | | | | | | | | |
| Mean | 0.2498 | 0.2827 | 0.0329 | 2.0178 | 0.3182 | 0.4078 | 0.0896 | 2.0684 |
| Median | 0 | 0 | | | 0 | 0 | | |
| SD | 0.8683 | 0.9024 | | | 0.9166 | 1.6784 | | |
| **Panel B. Matches with NotLikers** | | | | | | | | |
| *Likes sent* | | | | | | | | |
| Mean | 104.9727 | 89.0842 | -15.8885 | -2.7703 | 20.3306 | 20.0748 | 0.2558 | 0.0915 |
| Median | 18 | 16 | | | 3 | 3 | | |
| SD | 332.4176 | 290.2234 | | | 83.0087 | 93.7785 | | |
| *Initiated messages* | | | | | | | | |
| Mean | 9.5570 | 9.5538 | -0.0033 | -0.0048 | 1.2988 | 1.2511 | | |
| Median | 1 | 1 | | | 0 | 0 | | |
| SD | 35.6085 | 38.7018 | | | 6.5044 | 5.6285 | | |
| *Initiated messages that led to match* | | | | | | | | |
| Mean | 0.6777 | 0.7207 | 0.0430 | 0.8456 | 0.2166 | 0.2348 | 0.0181 | 0.5678 |
| Median | 0 | 0 | | | 0 | 0 | | |
| SD | 2.5462 | 2.9579 | | | 0.9584 | 1.0539 | | |
| Obs. | 5,752 | 6,055 | | | 1,911 | 2,087 | | |

Notes: Initiated Matches are conditional on initiating a message.

Table 2: Summary Statistics of User Activities Toward Likers

users to know whether the candidate had *ℓiked* them without having to take any further action, it is reasonable to think that treated users would behave differently depending on whether a candidate is a Liker or a NotLiker.

In Table 2 Panel A, we present summary statistics of user activities toward Likers, separately for men and women. For men, the treatment increases the number of *ℓikes* and number of initiated messages by 13% and 17%, respectively. We see a similar pattern for women: the treatment increases the two metrics by 32% and 34%, respectively.

We also test whether the treatment has an impact on the number of successful matches achieved by our experimental users. While we do not observe whether users actually went on an offline date, we do observe the number of messages exchanged between each pair. Prior research by Bapna et al. (2016) and anecdotal evidence from the online dating industry have pointed out that exchange of three messages between potential couples is a good predictor of an actual online match, where phone numbers are exchanged or users ask the other out for an offline date. In fact, senior executives of the platform revealed that they strongly believe that this measure of a match is an accurate predictor of an offline date and hence use this metric as a measure of a successful match. Following Bapna et al. (2022), we take a more conservative stance and define a successful match as an exchange of at least four messages (2 messages sent and 2 received). Here we only consider "initiated" matches, where at least four messages were exchanged after the experimental user initiated a conversation. We find that the treatment increases men's initiated matches with Likers by 12% and increases women's initiated matches with Likers by 28%.

Panel B of Table 2 presents summary statistics of user activities towards NotLikers. Since treated users are unsure whether the candidate browsed their profile and decided not to $\ell ike$, or whether their profile was not browsed by the candidate, the effect of treatment on user behavior is ambiguous. Except for the reduction in $\ell ikes$ sent by men in the treatment group, the treatment does not lead to a significant difference in user activities nor the number of successful matches.[6]

### 4.2 Frictions and Sorting

Positive correlations in mate attributes have been widely documented and studied across multiple disciplines. In our experiment, the treatment reduces search friction by revealing information about $\ell ikes$ that the treated users have received, without requiring their costly actions. Consequently, by examining how the treatment affects sorting patterns among matched couples, we can obtain insights into the impact of search frictions on sorting.

User attributes in our data include age, education level (high school = $-4$, two-year college = $-2$, university = 0, postgraduate = 2), body type (skinny, average, extra weight, overweight), and race (Asian, Black, Hispanic, White, other). We also create a synthetic variable that measures the desirability of a user (Bruch and Newman, 2024; Shi and Viswanathan, 2023; Bapna et al., 2022). Specifically, for each user, we observe the number of unique candidates who sent messages to

[6] In Appendix C, we also examine whether the treatment effect diminishes a few days after it begins. We find that the treatment initially boosts $\ell iking$ and messaging behavior, but its effect diminishes over time, becoming insignificant for $\ell iking$ and weaker for messaging after the first week.

| Attribute | Manipulation | Obs. | Mean | SE | Diff | $t$ statistic | $p$-value |
|---|---|---|---|---|---|---|---|
| **Panel A. Matches with Likers** | | | | | | | |
| Age | Treatment | 2,563 | 4.5310 | 0.0940 | 0.4166 | 3.1464 | 0.0017 |
| Age | Control | 2,045 | 4.1144 | 0.0899 | | | |
| Education | Treatment | 534 | 1.9925 | 0.0886 | 0.2684 | 2.0715 | 0.0386 |
| Education | Control | 435 | 1.7241 | 0.0937 | | | |
| Race | Treatment | 1219 | 0.6136 | 0.0139 | 0.0430 | 1.9961 | 0.0459 |
| Race | Control | 906 | 0.5706 | 0.0164 | | | |
| Body type | Treatment | 798 | 0.6266 | 0.0171 | 0.0559 | 2.2121 | 0.0270 |
| Body type | Control | 708 | 0.5706 | 0.0186 | | | |
| Desirability | Treatment | 2,563 | 0.8837 | 0.0149 | -0.0473 | -2.1248 | 0.0337 |
| Desirability | Control | 2,045 | 0.9311 | 0.0166 | | | |
| **Panel B. Matches with NotLikers** | | | | | | | |
| Age | Treatment | 4,854 | 5.1376 | 0.0709 | 0.1321 | 1.2906 | 0.1969 |
| Age | Control | 4,312 | 5.0056 | 0.0735 | | | |
| Education | Treatment | 808 | 1.8775 | 0.0694 | 0.0497 | 0.4999 | 0.6172 |
| Education | Control | 778 | 1.8278 | 0.0712 | | | |
| Race | Treatment | 2179 | 0.5631 | 0.0106 | -0.0144 | -0.9145 | 0.3605 |
| Race | Control | 1813 | 0.5775 | 0.0116 | | | |
| Body type | Treatment | 1498 | 0.5868 | 0.0127 | -0.0521 | -2.8065 | 0.0050 |
| Body type | Control | 1282 | 0.6388 | 0.0134 | | | |
| Desirability | Treatment | 4,854 | 0.8869 | 0.0106 | -0.0152 | 0.0156 | -0.9752 |
| Desirability | Control | 4,312 | 0.9021 | 0.0115 | | | |

*Notes*. For race and body type, we report the z-statistic.

Table 3: Attribute Differences With Initiated Matches

them during the pretreatment period. We discretize these numbers across users into gender-specific tertiles and use the tertiles to measure each user's desirability (See Section 3.2).

To compare sorting patterns between the treatment and control groups, we construct an "attribute difference" measure to test whether the treatment leads to significantly different sorting patterns between matched men and women. Specifically, the attribute difference between a man $m$ and a woman $w$ is calculated as $\Delta = |X_m - X_w|$, where $X_m$ and $X_w$ are the attributes of $m$ and $w$, respectively. We compute attribute differences for age, education level, desirability, race, and body type. For categorical variables like race and body type, the attribute difference is an indicator variable that equals 1 when they differ between $m$ and $w$, and 0 otherwise.

Table 3 Panel A reports the mean attribute differences of couples that matched with *Likers*, separately for the control and treatment groups. Notably, the treatment group shows significantly larger attribute differences than the control group for age, education, race, and body type, while the opposite is true for desirability. For example, the age difference in the treatment group ($\Delta = 4.53$) is 0.4 years, or 10%, larger than that in the control group ($\Delta = 4.11$); approximately 61% of treatment

| dependent variable | message indicator | | | |
|---|---|---|---|---|
| | Men | | Women | |
| | (1) Likers | (2) NotLikers | (3) Likers | (4) NotLikers |
| equal desirability | -0.0687*** | -0.0043* | 0.0019** | 0.0018*** |
| | (0.0107) | (0.0026) | (0.0016) | (0.0004) |
| higher desirability | -0.1313*** | -0.0174** | 0.0091*** | 0.0039*** |
| | (0.0129) | (0.0037) | (0.0026) | (0.0008) |
| treated×lower desirability | 0.0180 | 0.0074 | 0.0022 | -0.0001 |
| | (0.0245) | (0.0052) | (0.0037) | (0.0008) |
| treated×equal desirability | 0.0363** | 0.0018 | 0.0092*** | -0.0005 |
| | (0.0163) | (0.0038) | (0.0032) | (0.0008) |
| treated×higher desirability | 0.0576*** | 0.0015 | 0.0126*** | -0.0010 |
| | (0.0103) | (0.0026) | (0.0033) | (0.0010) |
| constant | 0.2995*** | 0.0468 | 0.0175 | 0.0035 |
| | (0.0113) | (0.0034) | (0.0021) | (0.0005) |
| R-squared | 0.0131 | 0.0019 | 0.0022 | 0.0004 |
| Obs. | 38,433 | 2,711,317 | 166,986 | 993,829 |

*Notes*. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4: Messaging Behavior With Respect to Relative Desirability Levels

group users matched with partners of a different race, 4 percentage points more than that of the control group (57%); and the desirability difference ($\Delta = 0.88$) is 0.047 units lower than that of the control group ($\Delta = 0.93$). In contrast, Panel B reports differences in sorting between the treatment and control groups that matched with *NotLikers*. Because the user cannot distinguish whether a NotLiker chose not to $\ell ike$ them, or simply hasn't seen their profile yet, we do not find any consistent patterns in the impact of treatment on sorting.

### 4.3 Mechanism of the Treatment Effect on Sorting

In this section, we investigate the mechanism behind the treatment effect on sorting as shown earlier in Section 4.2. We conjecture that the treatment encourages users to reach out to candidates they would otherwise perceive as out of reach, and these candidates happen to be less similar to the users. Specifically, we examine if the treatment's effect on users' messaging behavior varies with the relative desirability of the user and the candidate. We first create a binary variable 'message indicator', which equals 1 if a user initiates a message to a candidate and 0 otherwise. We regress this binary variable on dummy variables–'lower desirability,' 'equal desirability', and 'higher desirability'–which take the value 1 if the candidate is less, equally, or more desirable than the user,

| dependent variable | message indicator | | | |
|---|---|---|---|---|
| | Men | | Women | |
| | (1) Likers | (2) NotLikers | (3) Likers | (4) NotLikers |
| medium match prob | -0.0119 | 0.0017 | -0.0205*** | -0.0032*** |
| | (0.0102) | (0.0019) | (0.0037) | (0.0009) |
| high match prob | -0.0270** | 0.0069** | -0.0283*** | -0.0054*** |
| | (0.0119) | (0.0032) | (0.0037) | (0.0009) |
| treated×low match prob | 0.0576*** | 0.0020 | 0.0212*** | -0.0017 |
| | (0.0137) | (0.0034) | (0.0059) | (0.0012) |
| treated×medium match prob | 0.0250** | 0.0026 | 0.0060** | -0.0004 |
| | (0.0126) | (0.0032) | (0.0030) | (0.0010) |
| treated×high match prob | 0.0268 | 0.0048 | 0.0051* | 0.0004 |
| | (0.0188) | (0.0047) | (0.0030) | (0.0007) |
| constant | 0.2466*** | 0.0372** | 0.0408*** | 0.0082*** |
| | (0.0093) | (0.0022) | (0.0036) | (0.0009) |
| $R$-squared | 0.0037 | 0.0004 | 0.0088 | 0.0006 |
| Obs. | 38,433 | 2,711,317 | 166,986 | 993,829 |

*Notes*. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Messaging Behavior With Respect to Ex Ante Match Probabilities

respectively, along with the interaction terms of these dummy variables with the treatment dummy variable.

Table 4 reports the regression results. From the table, we can see that, for both men and women, the treatment significantly increases the user's likelihood of messaging Likers (columns 1 and 3) only when the candidate is equally or more desirable than they are. This suggests that the treatment encourages users to message Likers who are at least as desirable as themselves. However, when messaging NotLikers (columns 2 and 4), the interaction terms are all insignificant, implying that the treatment does not affect messaging behavior if the user cannot tell if the candidate has liked them.

As a second test to confirm that the treatment encourages users to message out-of-reach candidates, we analyze how the treatment affects users' messaging behavior depending on an 'ex-ante' match probability. We define this probability as the likelihood of a successful match conditional on observed attributes of the user and the candidate. To obtain the ex-ante match probabilities, we regress the match indicator on the user's attributes, along with the positive and negative differences between attributes of the user and the candidate. The model is estimated with a logistic regression using only data from the control group to avoid potential treatment effects on the match probability.

| | Age | Education | Race | Body Type | Desirability |
|---|---|---|---|---|---|
| Men | -0.2439* | -0.2652* | -0.2573* | -0.0402* | -0.0594* |
| Women | -0.1155* | -0.1150* | 0.1465* | -0.0507* | -0.1644* |

*Notes.* This table reports Pearson correlation coefficients between user-candidate attribute differences and ex-ante match probabilities. * $p<0.05$

Table 6: Pearson Correlation of Attribute Gap and Ex Ante Match Probability

We then predict the ex-ante match probabilities for users in both groups, which we refer to as the ex-ante match probabilities.

Next, we regress the binary messaging indicator on tertiles of the ex-ante match probabilities (low, medium, high) and their interaction terms with the treatment variable. Table 5 reports the results of this regression. When men encounter Likers (column 1), two interaction terms, treatment×low match prob and treatment×medium match prob, are positive and statistically significant. It implies that the treatment increases men's likelihoods of messaging candidates with whom they have low and medium ex-ante match probabilities, and the effect size on the group with low-probability is more than double of that on the medium-probability group. For women (column 3), we observe a similar pattern. The treatment significantly increases the likelihood of messaging for candidates with low and medium match probabilities, with the effect size more than tripling for the low-probability group. For both men and women encountering NotLikers (columns 2 and 4), with the interaction terms being insignificant, we do not observe such effects of the treatment. In Appendix D, we also find consistent results when examining the treatment's effect on $\ell iking$ behavior with respect to different ex-ante match probabilities. These results suggest that the treatment encourages users to interact with candidates who would otherwise appear to be out of reach.

Finally, Table 6 reports Pearson correlation coefficients between user-candidate attribute differences and ex-ante match probabilities, separately for men and women. For both genders, nearly all attribute differences (except for race in female users) are negatively correlated with ex-ante match probabilities, suggesting that greater attribute differences reduce the likelihood of matching. These correlation coefficients, along with the results in Tables 4 and 5, suggest that the treatment encourages users to message Likers who are more different from themselves and perceived as out-of-reach based on observed attributes. In conclusion, our findings suggest that reducing friction with more information provision alters user behavior and sorting patterns.

# 5 Model

## 5.1 Model Overview

As documented in the previous section, the treatment reduces frictions through information provision, which in turn affects user behavior and sorting patterns. What remains unclear is the respective impact of frictions and user preferences on sorting and match outcomes, as well as the loss of efficiency due to frictions. To answer these questions, we need to compare matches formed in a market with frictions to those in a frictionless environment, where preferences alone shape matching outcomes. Accordingly, in the following sections, we develop and estimate a structural model, then use the model estimates to simulate and compare matches in settings with different levels of frictions.

Consider an online dating platform. Each period, male user $m$ ($m \in \{1, 2, ..., N_M\}$) and female user $w$ ($w \in \{1, 2, ..., N_W\}$) are searching for a partner. Time is discrete, and we assume that discounting over time is negligible (Hitsch et al., 2010a). In what follows, we describe the model from the perspective of *a male user in the control group*, noting that the model is symmetric for female users in the control group. The model of treated users is a straightforward extension of the control user model. We will explain the treated user model at the end of this section,

When male user $m$ first views female user $w$'s profile, he observes $w$'s characteristics $X_w$. If they successfully match, user $m$ derives a match utility, $U_M^{mw}$. Before a match can materialize, user $m$ goes through two decision stages: *$\ell$iking* stage and messaging stage. In the first stage, $m$ makes a discrete *$\ell$iking* decision, $d_{mw} \in \{0, 1\}$, where 1 denotes *$\ell$iking* $w$ and 0 denotes otherwise. If $m$ *$\ell$ikes* $w$, he incurs a *$\ell$iking* cost $c_{M,m}^{\ell ike}$. In the second stage, $m$ decides whether to initiate a message to $w$, choosing $\mu_{mw} \in \{0, 1\}$, where 1 denotes messaging $w$ and 0 otherwise. If $m$ sends the message, he incurs a messaging cost of $c_{M,m}^{msg}$, which is the time and effort spent on composing the message.

Let $\ell_{wm}$ denote whether $w$ has *$\ell$iked* user $m$ when $m$ views $w$'s profile. $\ell_{wm}$ equals 1 if $w$ has *$\ell$iked* $m$ and 0 otherwise. For the *control* group, $m$ does not directly observe the true value of $\ell_{wm}$ without him choosing to *$\ell$ike* $w$ first. Hence $m$'s information regarding $\ell_{wm}$ depends on $d_{mw}$, his *$\ell$ike* decision, as well as the characteristics of both users, $X_m$ and $X_w$. We denote $m$'s information about $\ell_{wm}$ as follows:

$$\widetilde{\ell}_{wm}(d_{mw}|X_m, X_w) = \begin{cases} \mathbb{E}[\ell_{wm}|X_m, X_w] & \text{if } d_{mw} = 0 \\ \ell_{wm} & \text{if } d_{mw} = 1. \end{cases} \tag{1}$$

While $\ell_{wm}$ takes the value of 1 or 0, $\mathbb{E}[\ell_{wm}|X_m, X_w]$ takes a value between 0 and 1 and can be interpreted as the probability that $w$ *likes* $m$ given their respective characteristics.

User $m$'s decisions to *like* or message $w$ depend on the following factors: (1) $m$'s utility from matching with $w$, denoted as $U_M^{mw}$, (2) $m$'s reservation value of remaining single and continuing the search, denoted as $V_M(m)$, and (3) $m$'s belief about the match probability, denoted as $P(X_m, X_w, d_{mw}, \mu_{mw}, \widetilde{\ell}_{wm})$, a function of his own and $w$'s characteristics, his *liking* and messaging decisions, and his information about $\ell_{wm}$ at the point in time during his decision-making. To simplify the exposition, we henceforth denote $P(X_m, X_w, d_{mw}, \mu_{mw}, \widetilde{\ell}_{wm})$ as $P_{d_{mw},\mu_{mw},\widetilde{\ell}_{wm}}$. In particular, we sometimes update the subscripts of $P_{d_{mw},\mu_{mw},\widetilde{\ell}_{wm}}$ to reflect the then-current actions or information of user $m$. For example, $P_{1,0,\ell_{wm}}$ represents that user $m$ *likes* but does not message $w$ ($d_{mw} = 1, \mu_{mw} = 0$), and he knows the exact value of $\ell_{wm}$ being 1 or 0 after his action $d_{mw} = 1$.

Next, in a backward induction manner, we first elaborate on user $m$'s messaging decision in the second stage. We then describe $m$'s liking decision in the first stage.

## 5.2 Second Stage: User $m$'s Messaging Decision

User $m$'s decision to initiate a message in the second stage depends on whether he *liked* $w$ in the first stage. Therefore we need to consider the two cases separately.

#### Case 1: $m$ *liked* $w$ in the first stage

If $m$ *liked* $w$ in the first stage ($d_{mw} = 1$), his expected utility from sending a message is given by

$$EU^{mw,stage2}_{\mu_{mw}=1|d_{mw}=1,\ell_{wm}} = \overline{EU}^{mw,stage2}_{\mu_{mw}=1|d_{mw}=1,\ell_{wm}} + \varepsilon^{msg}_{mw} \tag{2}$$

$$\text{where} \quad \overline{EU}^{mw,stage2}_{\mu_{mw}=1|d_{mw}=1,\ell_{wm}} = U_M^{mw} \cdot P_{1,1,\ell_{wm}} + V_M(m) \cdot (1 - P_{1,1,\ell_{wm}}) - c^{msg}_{M,m}. \tag{3}$$

On the one hand, with probability $P_{1,1,\ell_{wm}}$, $w$ accepts $m$'s match offer and $m$ receives utility $U_M^{mw}$. On the other hand, with probability $1 - P_{1,1,\ell_{wm}}$, $m$ does not match with $w$, and he receives the reservation value $V_M(m)$ for remaining single and continuing the search. Furthermore, $c_{M,m}^{msg}$ is the cost of messaging, and $\varepsilon_{mw}^{msg}$ is an error term observed by $m$ (but unobserved by the researcher) that affects $m$'s decision to send a message.

Conversely, the expected utility of not sending a message is given by

$$EU^{mw,stage2}_{\mu_{mw}=0|d_{mw}=1,\ell_{wm}} = \overline{EU}^{mw,stage2}_{\mu_{mw}=0|d_{mw}=1,\ell_{wm}} + \varepsilon^{nmsg}_{mw} \tag{4}$$

$$\text{where} \quad \overline{EU}^{mw,stage2}_{\mu_{mw}=0|d_{mw}=1,\ell_{wm}} = U_M^{mw} \cdot P_{1,0,\ell_{wm}} + V_M(m) \cdot (1 - P_{1,0,\ell_{wm}}) \tag{5}$$

and $\varepsilon_{mw}^{nmsg}$ is an error term that affects $m$'s decision to not send a message. Note that even if $m$ does not send a message to $w$, he may match with $w$ with positive probability $P_{1,0,\ell_{wm}}$ due to the signaling effect of $\ell iking$ $w$ in the first stage. Assuming that error terms $\varepsilon_{mw}^{\text{msg}}$ and $\varepsilon_{mw}^{\text{nmsg}}$ follow i.i.d Type I EV distribution, the probability of $m$ sending a message to $w$ conditional on $d_{mw} = 1$ is

$$\Pr(\mu_{mw} = 1|d_{mw} = 1) = \frac{\exp\left(\overline{EU}^{mw,stage2}_{\mu_{mw}=1|d_{mw}=1,\ell_{wm}} - \overline{EU}^{mw,stage2}_{\mu_{mw}=0|d_{mw}=1,\ell_{wm}}\right)}{1 + \exp\left(\overline{EU}^{mw,stage2}_{\mu_{mw}=1|d_{mw}=1,\ell_{wm}} - \overline{EU}^{mw,stage2}_{\mu_{mw}=0|d_{mw}=1,\ell_{wm}}\right)} \tag{6}$$

**Case 2: $m$ did not $\ell ike$ $w$ in the first stage**

User $m$'s expected utility from sending a message when he did not $\ell ike$ $w$ in the first stage ($d_{mw} = 0$), is given by

$$EU^{mw,stage2}_{\mu_{mw}=1|d_{mw}=0,\mathbb{E}[\ell_{wm}]} = \overline{EU}^{mw,stage2}_{\mu_{mw}=1|d_{mw}=0,\mathbb{E}[\ell_{wm}]} + \varepsilon^{msg}_{mw} \tag{7}$$

$$\text{where} \quad \overline{EU}^{mw,stage2}_{\mu_{mw}=1|d_{mw}=0,\mathbb{E}[\ell_{wm}]} = U_M^{mw} \cdot P_{0,1,\mathbb{E}[\ell_{wm}]} + V_M(m) \cdot \left(1 - P_{0,1,\mathbb{E}[\ell_{wm}]}\right) - c^{msg}_{M,m}. \tag{8}$$

We assume that if $m$ neither $\ell ikes$ nor messages $w$, the match probability is zero, i.e., $P_{0,0,\cdot} = 0$. Then $m$'s expected utility from not sending a message will be

$$EU^{mw,stage2}_{\mu_{mw}=0|d_{mw}=0,\mathbb{E}[\ell_{wm}]} = V_M(m) + \varepsilon^{nmsg}_{mw}. \tag{9}$$

Again, assuming that error terms $\varepsilon_{mw}^{\text{msg}}$ and $\varepsilon_{mw}^{\text{nmsg}}$ follow i.i.d Type I EV distribution, the probability of $m$ messaging $w$ conditional on $d_{mw} = 0$ can be written as

$$\Pr(\mu_{mw} = 1 | d_{mw} = 0) = \frac{\exp\left(\overline{EU}^{mw,stage2}_{\mu_{mw}=1|d_{mw}=0,\mathbb{E}[\ell_{wm}]} - V_M(m)\right)}{1 + \exp\left(\overline{EU}^{mw,stage2}_{\mu_{mw}=1|d_{mw}=0,\mathbb{E}[\ell_{wm}]} - V_M(m)\right)}. \tag{10}$$

### 5.3 First Stage: User $m$'s $\ell iking$ Decision

In the first stage, $m$'s decision to $\ell ike$ depends on his expectation of whether he will choose to message in the second stage. Let us define $\mathcal{U}_{d_{mw},\widetilde{\ell}_{wm}(d_{mw})}$ as the expectation of the second stage (expected) utility conditional on the probability of sending a message in the second stage:

$$\begin{aligned}\mathcal{U}_{d_{mw},\widetilde{\ell}_{wm}(d_{mw})} =& \Pr(\mu_{mw} = 0 | d_{mw}) \cdot EU^{mw,stage2}_{\mu_{mw}=0|d_{mw},\widetilde{\ell}_{wm}(d_{mw})} \\ &+ \Pr(\mu_{mw} = 1 | d_{mw}) \cdot EU^{mw,stage2}_{\mu_{mw}=1|d_{mw},\widetilde{\ell}_{wm}(d_{mw})}.\end{aligned} \tag{11}$$

If $m$ $\ell ikes$ $w$ in the first stage, the true value of $\ell_{wm}$ reveals. Ex ante, $\ell_{wm}$ equals 1 with probability $\mathbb{E}[\ell_{wm}]$ and equals 0 with probability $1 - \mathbb{E}[\ell_{wm}]$. Accordingly, $m$'s (expected) utility from $\ell iking$ is an expectation over the true value of $\ell_{wm}$, and is given by

$$\begin{aligned} EU^{mw,stage1}_{d_{mw}=1} &= \overline{EU}^{mw,stage1}_{d_{mw}=1} + \varepsilon^{\ell ike}_{mw} \\ \text{where} \quad \overline{EU}^{mw,stage1}_{d_{mw}=1} &= \mathbb{E}[\ell_{wm}] \cdot \mathcal{U}_{1,1} + (1 - \mathbb{E}[\ell_{wm}]) \cdot \mathcal{U}_{1,0} - c^{\ell ike}_{M,m}.\end{aligned} \tag{12}$$

$\varepsilon^{\ell ike}_{mw}$ is an error term that affects $m$'s decision to $\ell ike$, and $c^{\ell ike}_{M,m}$ is the cost of $\ell iking$. Conversely, if $m$ does not $\ell ike$ $w$, the true value of $\ell_{wm}$ will not be revealed, and $m$'s utility will be given by

$$EU^{mw,stage1}_{d=0} = \overline{EU}^{mw,stage1}_{d=0} + \varepsilon^{n\ell ike}_{mw} \tag{13}$$

$$\text{where} \quad \overline{EU}^{mw,stage1}_{d=0} = \mathcal{U}_{0,\mathbb{E}[\ell_{wm}]} \tag{14}$$

and $\varepsilon^{n\ell ike}_{mw}$ is an error term affecting $m$'s decision to not $\ell ike$.

Assuming that error terms $\varepsilon^{\ell ike}_{mw}$ and $\varepsilon^{n\ell ike}_{mw}$ follow i.i.d Type I EV distribution, the probability that $m$ chooses to $\ell ike$ $w$ can then be written as

$$\Pr(d_{mw}=1) = \frac{\exp\left(\overline{EU}^{mw,stage1}_{d_{mw}=1} - \overline{EU}^{mw,stage1}_{d_{mw}=0}\right)}{1+\exp\left(\overline{EU}^{mw,stage1}_{d_{mw}=1} - \overline{EU}^{mw,stage1}_{d_{mw}=0}\right)}. \tag{15}$$

### 5.4 Model of Treated Users

The model described above corresponds to the decision process of users in the control group. The model for the treatment group is a straightforward adaption. Specifically, because treated user $m$'s knowledge about whether $w$ has *ℓiked* him is directly endowed to him and does not depend on his *ℓike* action $d_{mw}$, Equation (1) becomes $\widetilde{\ell}_{wm}(d_{mw}|X_m, X_w) = \ell_{wm}$. As a result, all aspects of the model for treated users remain the same as in the model for control users, except that occurrences of $\mathbb{E}[\ell_{wm}|X_m, X_w]$ are replaced with $\ell_{wm}$.

We further note that the cost of *ℓiking* stems from a psychological burden, arising from the anxiety associated with the possibility of not receiving a reciprocal *ℓike*. Because a treated user knows upfront whether the candidate has *ℓiked* them, the level of the *ℓiking* cost differs between the control and treatment groups. Although the cost of *ℓiking* shares a common notation in the model formulation, the empirical specification allows its levels to differ depending on whether a user is in the treatment group. The details of this specification will be discussed in the next section.

## 6 Estimation and Identification

To estimate the model based on the data, we first specify functional forms for utility, match probability, and the costs of *ℓiking* and messaging. We then discuss the estimation strategy.

### 6.1 Utility Specification

Let $X_M = (x_m, x_m^d)$ and $X_W = (x_w, x_w^d)$ be $m$'s and $w$'s observed characteristics, respectively. $x_m$ and $x_w$ are vectors that contain variables of continuous values, and $x_m^d$ and $x_w^d$ are sets of categorical variables. User $m$'s utility from matching with $w$ is specified as

$$\begin{aligned} U_M^{mw}(X_m, X_w; \Theta_M) =& x_w'\beta_M + (|x_w - x_m|_+')\beta_M^+ + (|x_w - x_m|_-')\beta_M^- \\ &+ \sum_{r,s=1}^{N} \mathbb{1}\{x_{mr}^d = 1 \text{ and } x_{ws}^d = 1\} \cdot \beta_{M,rs}^d + \epsilon_{mw}, \end{aligned} \tag{16}$$

where $|x_w - x_m|_+ = \max(x_w - x_m, 0)$ and $|x_w - x_m|_- = \max(x_m - x_w, 0)$. $\epsilon_{mw}$ is an idiosyncratic preference shock which follows standard normal distribution, denoted as $F_\epsilon(\epsilon)$. The set of preference parameters to be estimated is $\Theta_M = (\beta_M, \beta_M^+, \beta_M^-, \beta_M^d)$.

We note that whether $w$ has liked $m$ (i.e., $\ell_{wm}$ being 1 or 0) does not enter the match utility function. Reciprocity of liking–where people tend to like those who express liking for them–is well-established in non-romantic contexts (e.g., Kenny, 1994; Montoya and Insko, 2008). However, the validity of these findings in romantic contexts is unclear due to mixed findings.[7] Accordingly, we let the match utility be independent of $\ell_{wm}$. On the other hand, research has demonstrated that people are reluctant to initiate romantic overtures without some indication of mutual interest (Vorauer et al., 2003), and are less likely to pursue a potential partner when the probability of rejection is high (Bernstein et al., 1983). Therefore, we posit that reciprocity of liking affects users' *expected* match utility by increasing the probability of a successful match.

## 6.2 Match Probability

We obtain users' beliefs about the match probability, $P_{d_{mw},\mu_{mw},\widetilde{\ell}_{wm}}$, as follows:

- Conditional on $m$'s actions $\alpha = (d_{mw}, \mu_{mw})$, we first construct a binary variable, $Match^{\alpha}_{mw}$, that equals 1 if $m$ and $w$ match. For example, $Match^{(1,0)}_{mw} = 1$ indicates $m$ and $w$ match upon $m$ *liking* but not messaging $w$.

- We then regress this match indicator on the user's attributes, the differences between the user's and the candidate's attributes, and an indicator of whether $w$ has *liked* $m$:

$$\begin{aligned} Match^{\alpha}_{mw} =& x_m'\nu_W + (|x_m - x_w|_+')\nu_W^+ + (|x_m - x_w|_-')\nu_W^- \\ &+ \sum_{r,s=1}^{N} \mathbb{1}\{x^d_{wr} = 1 \text{ and } x^d_{ms} = 1\} \cdot \nu^d_{W,rs} + \psi^a_W \cdot \ell_{wm} + \eta_{wm}. \end{aligned} \tag{17}$$

  $\Upsilon_W = \{\upsilon_W, \upsilon_W^+, \upsilon_W^-, \upsilon_W^d\}$ and $\psi^a_W$ are the parameters to be estimated. The subscript $W$ indicates that these parameters reflect women's preferences to accept $m$'s match offer. $\psi^a_W$ captures the impact of $w$ *liking* $m$ on the match probability, and $\eta_{wm}$ is an error term.

[7] Walster et al. (1973) argue that unconcealed romantic interest can appear desperate and unappealing, and Eastwick and Finkel (2009) suggest that romantic liking may only be reciprocated when it is selective.

- We use the predicted values of the regression as the match probabilities. We employ the bootstrapping approach to account for the uncertainty in the match probabilities predicted from the regression.

There is a potential endogeneity issue with $\ell_{wm}$. Unobserved components that affect the match between $m$ and $w$, may be correlated with $\ell_{wm}$, leading to biased estimates of $\psi^a$. To address this issue, we use an instrumental variable, *ℓike*-rate, to isolate the causal effect of *ℓike* on match probabilities. The *ℓike*-rate is defined as the ratio of profiles *ℓiked* to profiles browsed by the candidate. This ratio reflects $w$'s average tendency of *ℓking* a profile and is correlated with whether $w$ *ℓikes* $m$. However, it influences $w$'s decision to accept $m$'s match offer only through her specific decision of *ℓiking* $m$.

Table 7 reports parameter estimates of $\psi^a$ from Equation 6.2, estimated using OLS and IV regressions, for the values of $\alpha = (1, 0)$ and $\alpha = (\cdot, 1)$, which correspond to matching after $m$ *ℓiking*–only (but not messaging), and messaging $w$ (regardless of *ℓiking* or not), respectively. Columns (1) and (2) show estimates with $Match_{mw}^{(1,0)}$ as the outcome variable, while Columns (3) and (4) use $Match_{mw}^{(\cdot,0)}$ as the outcome variable. Panel A presents results for men. In Columns (1) and (2), the IV estimate, after addressing selection, is smaller than the OLS estimate. The first-stage partial $F$-statistic of 13,328.5 suggests that the instrument has strong explanatory power. Conversely, Columns (3) and (4) of Panel A show that the IV estimate is larger than the OLS estimate, with a first-stage partial $F$-statistic of 4,350.38. Panel B provides analogous results for women, revealing similar patterns across genders.

Using the estimates from the IV regression, we predict match probabilities, $P_{d_{mw},\mu_{mw},\widetilde{\ell}_{wm}}$, by replacing $\ell_{wm}$ with $m$'s beliefs about it, $\widetilde{\ell}_{wm}(d_{mw})$, to reflect his beliefs about the match probability based on his knowledge about $\ell_{wm}$ that depends on his decision at each stage.

### 6.3 Costs

Separately identifying costs and preferences is difficult in models where the two components are interdependent. In our context, the interdependence arises because *ℓiking* and messaging $w$ could reflect either strong preferences for a profile's characteristics or low costs of *ℓiking* and messaging. Correspondingly, we rely on an exclusion restriction to separately identify preferences from costs.

| dependent variable | Match from *ℓiking* only | | Match from sending a message | |
|---|---|---|---|---|
| | (1) OLS | (2) IV | (3) OLS | (4) IV |
| **Panel A. Men** | | | | |
| $\psi^{(1,0)}$ | 0.0766*** | 0.0356*** | 0.2522*** | 0.2804*** |
| se | (0.0003) | (0.0022) | (0.0030) | (0.0162) |
| first-stage $F$ | | 13,328.5 | | 4,350.38 |
| **Panel B. Women** | | | | |
| $\psi^{(\cdot,1)}$ | 0.1597*** | 0.1001*** | 0.1501*** | 0.1979*** |
| se | (0.0020) | (0.0045) | (0.0089) | (0.0264) |
| first-stage $F$ | | 11,536.3 | | 1,189.76 |

*Notes*. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 7: Effect of *ℓike* on match probability

When at least one covariate enters the cost function but not the (expected) utility function, that covariate can serve as an exclusion restriction for identification (Chen and Yao, 2017). Once the cost and utility are empirically separated, the identification of model parameters follows from the usual econometric theory for discrete choice models.

We specify the cost of *ℓiking* as

$$c_{M,m}^{\ell ike} = \omega_M^{\ell ike} \times (1 - Treatment_m \times \ell_{wm}) \times \frac{\exp(K_m' \Lambda_M)}{1 + \exp(K_m' \Lambda_M)}. \tag{18}$$

where $\omega_M^{\ell ike}$ is a scalar coefficient. $K_m$ is a vector of user characteristics affecting the cost of *ℓiking*. Namely, we include a constant as well as indicators for older age (above 50), higher education level (university or higher), being white (due to the predominance of white users in our data), larger body size (extra weight and overweight), and higher desirability (the third tertile of observed desirability levels). $Treatment_m$ is a binary variable indicating whether $m$ is in the treatment group. Note that the cost of *ℓiking* is zero when $Treatment_m = 1$ and $\ell_{wm} = 1$. This is because when $m$ is in the treatment group and observes that $w$ has *ℓiked* him upfront, he does not incur the psychological cost of *ℓiking* her. $\omega_M^{\ell ike}$ and $\Lambda_M$ are the parameters to be estimated.

The identification of the parameters comes from an exclusion restriction–user $m$'s treatment status. The treatment is an exogenous variable that affects the user's *ℓiking* decision through its influence on his knowledge of the true value of $\ell_{wm}$ without influencing the utility of matching (see the discussion in Section 6.1). One may be concerned that because the treatment influences the match probability, it affects the expected utility of matching and thus invalidates it as an exclusion restriction. However, since the match probability is obtained separately using the predicted values of Equation (6.2) (as described in Section 6.2), the treatment status does not affect the expected utility either. In short, $Treatment_m$ enters the *ℓiking* cost function but not the expected utility function, serving as a valid exclusion restriction for identification.

The cost of messaging, which is the time and effort to compose a message, is specified as

$$c_{M,m}^{msg} = \omega_M^{msg} \times \frac{\exp(K_m' \Phi_M)}{1 + \exp(K_m' \Phi_M)}. \tag{19}$$

Contrary to *ℓiking*, sending a message does not change one's knowledge about $\ell_{wm}$. Therefore, $Treatment_m$ does not enter the cost of messaging. However, recall that the user's messaging decision depends on his *ℓiking* decision in the first stage (see Section 5.2). Because the treatment exogenously varies *ℓiking* decisions, it in turn exogenously influences one's messaging decision. Consequently, conditional on the identification of utility and *ℓiking* cost, the messaging cost parameters, $\omega_M^{msg}$ and $\Phi_M$, are identified by the exogenous variations in the messaging decision and $K_m$.

### 6.4 Reservation Value

To estimate $V_M(m)$ and $V_W(w)$, the reservation values of remaining single and continuing the search, we assume that $V_M(m)$ and $V_W(w)$ remain constant across different profiles a user browses. We first group users with similar characteristics using the $K$-means clustering method, an unsupervised machine learning algorithm for classification. We partition users into six groups for each gender (three for control and three for treatment groups) based on their observed characteristics. We then estimate the reservation value at the group level.

### 6.5 Likelihood

To estimate the coefficients of utility and costs, we maximize the joint likelihood of $m$'s decisions at the *ℓiking* stage and messaging stage for each profile he encounters. Let $\Pr(\ell ike_{mw} = 1)$ denote the probability that $m$ *ℓikes* $w$ in the first stage, which depends on the error term of the utility, $\epsilon_{mw}$, as well as the error terms affecting his messaging decisions in the second stage, $\varepsilon_{mw}^{\text{msg}}$, and $\varepsilon_{mw}^{\text{nmsg}}$. We assume that $\epsilon_{mw}$ follows a standard normal distribution, denoted as $F(\epsilon)$, while $\varepsilon_{mw}^{\text{msg}}$ and

$\varepsilon_{mw}^{\text{nmsg}}$ follow i.i.d. Type I EV distributions with location 0 and scale 1, denoted as $G(\varepsilon)$. Then the probability that $m$ *ℓikes* $w$, integrated over the distributions of the error terms, is:

$$\Pr(\ell ike_{mw} = 1) = \iiint \Pr\left(d_{mw} = 1 \mid \epsilon_{mw}, \varepsilon_{mw}^{\text{msg}}, \varepsilon_{mw}^{\text{nmsg}}\right) f(\epsilon_{mw}) g(\varepsilon_{mw}^{\text{msg}}) g(\varepsilon_{mw}^{\text{nmsg}})\, d\epsilon_{mw}\, d\varepsilon_{mw}^{\text{msg}}\, d\varepsilon_{mw}^{\text{nmsg}}. \tag{20}$$

Next, let $\Pr(msg_{mw} = 1 \mid \ell ike_{mw})$ be the probability that $m$ messages $w$ in the second stage, conditional on his first stage decision. This probability depends on the utility error term, $\epsilon_{mw}$, and is integrated as:

$$\Pr(msg_{mw} = 1 \mid like_{mw}) = \int \Pr\left(\mu_{mw} = 1 \mid d_{mw}, \epsilon_{mw}\right) f_{\epsilon}(\epsilon_{mw})\, d\epsilon_{mw}. \tag{21}$$

Given the probabilities or *ℓiking* and messaging, the joint likelihood function is given by

$$\begin{aligned} L = \textstyle\prod_{m=1}^{N_M} \prod_{w=1}^{J_m} & \left[\Pr(\ell ike_{mw} = 1) \cdot \Pr(msg_{mw} = 1 \mid \ell ike_{mw} = 1)^{\delta_{mw}} \left(1 - \Pr(msg_{mw} = 1 \mid \ell ike_{mw} = 1)\right)^{1-\delta_{mw}}\right]^{\vartheta_{mw}} \\ & \times \left[(1 - \Pr(\ell ike_{mw} = 1)) \cdot \Pr(msg_{mw} = 1 \mid \ell ike_{mw} = 0)^{\delta_{mw}} \left(1 - \Pr(msg_{mw} = 1 \mid \ell ike_{mw} = 0)\right)^{1-\delta_{mw}}\right]^{1-\vartheta_{mw}}. \end{aligned} \tag{22}$$

$J_m$ is the total number of profiles that $m$ encounters, $\theta_{mw}$ indicates the decision made at the first stage ($\vartheta_{mw}$ =1 if $d_{mw} = 1$, $\vartheta_{mw} = 0$ otherwise), and $\delta_{mw}$ indicates the decision at the second stage ($\delta_{mw} = 1$ if $\mu_{mw} = 1$, $\delta_{mw} = 0$ otherwise).

## 7 Estimation Results

Table 8 reports the maximum likelihood estimates of preference parameters, separately for men and women. Our estimation results align with prior research (Kurzban and Weeden (2005); Fisman et al. (2006); Hitsch et al. (2010a); Hitsch et al. (2010b)). Both genders prefer younger partners, but men prefer women who are younger than themselves, while women generally avoid men who are younger than they are. Men prefer candidates with a university education but are averse to women with master's degrees. Women, on the other hand, dislike high school and two-year college graduates and slightly prefer men with master's degrees. Both genders prefer slimmer body types to heavier ones, with men favoring being larger than their partners and women strongly disliking men who are smaller than themselves. We also find that desirability is a key factor for both genders: men strongly prefer more desirable women, while women have a strong aversion to less desirable

| | Preferences of Men | | Preferences of Women | |
|---|---|---|---|---|
| | Coefficients | SE | Coefficients | SE |
| Age | -0.0078* | 0.0046 | -0.4771*** | 0.0277 |
| Age Difference (+) | 0.0892*** | 0.0140 | -0.3888*** | 0.1435 |
| Age Difference (−) | 0.0149 | 0.0244 | 0.0546 | 0.0554 |
| HighSchool | -0.0570*** | 0.0167 | -0.0457*** | 0.0178 |
| TwoYear | 0.0070 | 0.0217 | -0.0455* | 0.0247 |
| University | 0.0983** | 0.0444 | 0.0836 | 0.0640 |
| Masters | -0.1362*** | 0.0422 | 0.0201* | 0.0109 |
| Law | -0.0075 | 0.0067 | 0.0007 | 0.0021 |
| Medical | -0.0045 | 0.0075 | 0.0008 | 0.0031 |
| PhD | -0.0141 | 0.0090 | 0.0103 | 0.0078 |
| Education Difference (+) | -0.1176** | 0.0557 | -0.3708*** | 0.1041 |
| Education Difference (−) | 0.1694*** | 0.0410 | 0.2267 | 0.1886 |
| Skinny | 0.5565*** | 0.1200 | 0.1991* | 0.1049 |
| Average | 0.0753 | 0.0454 | -0.0050 | 0.0170 |
| ExtraWeight | -0.1183 | 0.0425 | -0.0898** | 0.0441 |
| Overweight | -0.0887*** | 0.0271 | -0.0160* | 0.0086 |
| BMI Difference (+) | 0.1268*** | 0.0443 | 0.1068 | 0.1302 |
| BMI Difference (−) | -0.4461*** | 0.0819 | -0.2150*** | 0.0560 |
| Desirability | 0.8912*** | 0.0539 | 1.0862 | 0.5344 |
| Popularity Difference (+) | -0.0164 | 0.1116 | -0.8291** | 0.3836 |
| Popularity Difference (−) | 0.2209*** | 0.0628 | 0.5809 | 0.3242 |
| Asian; mate White | 0.0457*** | 0.0152 | 0.0723* | 0.0380 |
| Asian; mate Black | -0.0081** | 0.0038 | -0.0019 | 0.0014 |
| Asian; mate Hispanic | -0.0041 | 0.0034 | 0.0017 | 0.0025 |
| Asian; mate other | 0.0599** | 0.0245 | 0.0130 | 0.0102 |
| White; mate Asian | -0.0529*** | 0.0206 | -0.0368** | 0.0184 |
| White; mate Black | -0.2164*** | 0.0503 | -0.0675* | 0.0353 |
| White; mate Hispanic | 0.0017 | 0.0160 | -0.0265* | 0.0114 |
| White; mate other | 0.0619*** | 0.0214 | -0.1076* | 0.0598 |
| Black; mate Asian | 0.0031 | 0.0047 | -0.0085* | 0.0043 |
| Black; mate White | -0.0528*** | 0.0195 | 0.0013 | 0.0059 |
| Black; mate Hispanic | 0.0227*** | 0.0071 | -0.0022 | 0.0037 |
| Black; mate other | 0.0378* | 0.0222 | -0.0641* | 0.0312 |
| Hispanic; mate Asian | -0.0148* | 0.0076 | -0.0280* | 0.0154 |
| Hispanic; mate White | 0.0005 | 0.0148 | 0.0262 | 0.0189 |
| Hispanic; mate Black | -0.0353*** | 0.0103 | -0.0205* | 0.0108 |
| Hispanic; mate other | 0.0804*** | 0.0204 | -0.0399* | 0.0205 |
| Other; mate Asian | 0.1489*** | 0.0373 | -0.0822** | 0.0414 |
| Other; mate White | 0.1314** | 0.0485 | 0.2598* | 0.1435 |
| Other; mate Black | -0.3167*** | 0.0638 | -0.0371* | 0.0194 |
| Other; mate Hispanic | 0.1200*** | 0.0394 | -0.0113* | 0.0066 |

*Notes.* To account for the standard errors of the match probabilities, we employ bootstrapping, and report the means and standard deviations of the parameter estimates across the 100 bootstrap replications. *** p<0.01, ** p<0.05, * p<0.1

Table 8: Preference Parameter Estimates

| | Men | | Women | |
|---|---|---|---|---|
| | Coefficients | SE | Coefficients | SE |
| **A. Cost of Liking** | | | | |
| $\omega_M^{\ell ike}$ | 2.030*** | 0.0136 | 2.5978*** | 0.1695 |
| $\lambda_M^0$ | -2.7167*** | 0.0751 | 0.6826*** | 0.1210 |
| $\lambda_M^{old}$ | 2.2910*** | 0.3771 | -0.0613*** | 0.0234 |
| $\lambda_M^{high\,edu}$ | 0.1412** | 0.0704 | 0.0193 | 0.0857 |
| $\lambda_M^{white}$ | 1.4085*** | 0.4372 | 0.1478*** | 0.0519 |
| $\lambda_M^{large}$ | -0.5428*** | 0.1009 | -0.0299** | 0.01366 |
| $\lambda_M^{desirable}$ | -0.1342** | 0.0646 | -0.0650*** | 0.01140 |
| **B. Cost of Messaging** | | | | |
| $\omega_M^{msg}$ | 4.2819*** | 0.0211 | 6.8352*** | 0.5427 |
| $\phi_M^0$ | 1.7881*** | 0.0330 | 0.6138*** | 0.0860 |
| $\phi_M^{old}$ | 0.4880*** | 0.0847 | 0.0323 | 0.0831 |
| $\phi_M^{high\,edu}$ | -0.0248 | 0.0535 | 0.2350*** | 0.0892 |
| $\phi_M^{white}$ | 0.5105*** | 0.0084 | 0.0866*** | 0.1520 |
| $\phi_M^{large}$ | -0.1118 | 0.0746 | 0.0559 | 0.0937 |
| $\phi_M^{desirable}$ | -1.8330*** | 0.0118 | -0.0600** | 0.0261 |
| Log-Likelihood | -700,254.752 | 1,743.053 | -597,710.927 | 1,549.821 |

*Notes.* This table reports the coefficients for the costs of liking ($\omega$) and messaging ($\phi$). Superscripts indicate the following demographic characteristics: *ℓike* for the cost of liking, *msg* for the cost of messaging, 0 for baseline, *old* for age over 50, *high edu* for higher education, *white* for white race, *large* for larger body size, and *desirable* for higher desirability (third tertile). To account for the standard errors of the match probabilities, we employ bootstrapping, and report the means and standard deviations of the parameter estimates across the 100 bootstrap replications. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Cost Parameter Estimates

men. Finally, although both genders show ambiguous preferences for candidates of different races, women generally show a stronger aversion to interracial matches.

Table 9 Panel A reports estimates of the cost of *ℓiking*. As expected, the scalar coefficient of $\omega_M^{\ell ike}$ is positive and significant for both men and women, suggesting that users incur a significant cost when *ℓiking* a profile. For men, older age, higher education, and being white increase the cost of *ℓiking*, while being larger and more desirable reduce it. For women, being white also raises the cost of *ℓiking*, while older age, larger body size, and higher desirability lower it. Panel B reports estimates for the cost of messaging. The scalar coefficient on $\omega_M^{msg}$ is positive and significant for both genders, with a larger effect for women. For men, older age and being white increase the cost of messaging, while higher desirability lowers it. For women, higher education and being white increase the cost, while higher desirability reduces it.

# 8 Matching Patterns under Different Levels of Frictions

We quantify the impact of frictions and preferences on sorting, and measure the extent to which the preference signaling reduces the influence of search frictions. Specifically, we compare the equilibrium matches across the following matching protocols: (1) the default information setting of the platform, i.e., the control condition where a user does not know if the candidate has *ℓiked* them without *ℓiking* the candidate first, (2) the information setting of the focal experiment's treatment condition where a treated user is provided with a signal that reduces search frictions, and (3) a frictionless environment where *ℓiking* and messaging are costless. Note that in protocol (2), the focal users are treated, while candidates they browse remain in the control setting. We do not examine the impact of treatment under the two-sided treatment condition, where both focal users and candidates are treated, as this would require estimating match probabilities with both sides treated, which is not feasible with the available data. By comparing matches in (1) and (2), we assess the efficiency gains from lowering search costs through the treatment, providing managerial insights into whether revealing more user preference information on profile pages would improve matches. Comparing (1) and (3) allows us to quantify the relative impact of frictions and preferences on sorting, and the impact of frictions on efficiency. Finally, comparing the differences between (1) and (2) with those between (1) and (3) allows us to assess the extent to which signaling mitigates the impact of search frictions.

We use the users who are part of the experiment—11,807 men and 3,998 women—as our pool of individuals attempting to find a partner, and simulate matches among them. We simulate equilibrium matches in a frictionless environment using the Gale-Shapley deferred-acceptance algorithm (Gale and Shapley, 1962). Predicted matches under the control and treatment settings are obtained by introducing frictions into the deferred-acceptance algorithm using the cost components estimated from the structural model. This approach is similar to Banerjee et al. (2013), who introduce ad-hoc constraints to the deferred-acceptance algorithm to account for search frictions.

Before we proceed to describe how we simulate equilibrium matches, we first construct rankings for each man over the entire set of women, which will then be used in the deferred-acceptance algorithm. We begin by calculating the utility each man derives from matching with each woman, using estimated preference parameters from the model. Specifically, for each $m$ and $w$ we compute the predicted utility:

$$\hat{U}_{M,mw} = U_M^{mw}(X_m, X_w; \hat{\Theta}_M) \tag{23}$$

These predicted utilities are then transformed into ordinal rankings. For each user $m$, the ranking $R_m(w)$ of woman $w$ is assigned as:

$$R_m(w) = n \qquad \text{if} \qquad \begin{cases} \hat{U}_{M,mw'} > \hat{U}_{M,mw} > \hat{U}_{M,mw''} \\ \text{and } R_m(w') = n-1 \text{ and } R_m(w'') = n+1 \end{cases} \tag{24}$$

where $n$ is an integer.

## 8.1 Deferred-Acceptance Algorithm

### 8.1.1 No Frictions

The *man-optimal* stable matching, in which each man is paired with the best possible partner, is obtained using the deferred-acceptance algorithm, which is implemented as follows: Each man proposes to his top-ranked woman, as long as $\hat{U}_{M,mw} \geq V_M^{GS}(m)$, where $V_M^{GS}(m)$ denotes the reservation value in a frictionless environment. Ties are broken randomly.

1. Each woman selects her highest-ranked man among the offers she receives, provided $\hat{U}_{W,mw} \geq V_W^{GS}(w)$.

2. Unmatched men then propose to their next-highest ranked woman.

3. If a woman receives an offer from a man who is ranked higher than her current accepted offer, the woman releases the old offer and keeps the new offer. The released man proposes to the next woman in his ranked list.

4. This process continues until all men have proposed to every woman, with $\hat{U}_{M,mw} \geq V_M^{GS}(m)$.

To obtain reservation values, $V_M^{GS}(m)$, in the absence of frictions, we use the following procedure:

1. Using the estimated parameters from the structural model, calculate $m$'s utility, $U_M^{mw}$, and costs, $c_{M,mw}^{like}$ and $c_{M,mw}^{msg}$, for each candidate $w$. Then compute $m$'s average utility and average costs across all candidates, i.e., $\bar{U}_M^{mw}$, $\bar{c}_{M,mw}^{like}$ and $\bar{c}_{M,mw}^{msg}$.

2. Regress $V_M(m)$ on $\bar{U}_M^{mw}$, $\bar{c}^{like}_{M,mw}$, $\bar{c}^{msg}_{M,mw}$, and their respective squared terms.

3. Obtain the fitted value $\hat{V}_M(m)$ after setting $\bar{c}^{like}_{M,mw}$ and $\bar{c}^{msg}_{M,mw}$ to zero.

4. Use $\hat{V}_M(m)$ as $V_M^{GS}(m)$.

#### 8.1.2 Incorporating Frictions

We next describe how we incorporate frictions into the Gale-Shapley algorithm in the other two settings: (1) the default setting of the platform, and (2) the one-side treatment setting, where focal users receive the treatment while the candidates do not. Since we use only the pool of experimental users (assigned to either the treatment or control group) to form the two sides of the market, we need to first set up the *ℓikes* that focal users received from candidates, which will then enable them to search through those candidates. To that end, we simulate the $\ell_{wm}$ that a user $m$ receives from a candidate $w$, for each pair of $m$ and $w$. Specifically, the simulation of $\ell_{wm}$ and matching in an environment with frictions is executed as follows:

Step 0. Construct $\ell_{wm}$ that $w$ sends to $m$ for each $w$ and $m$:

(a) For each $w$, using estimated preference parameters, construct the expected utility from *ℓiking* ($\ell_{wm} = 1$) and not *ℓiking* ($\ell_{wm} = 0$), given by equations (13) and (14), respectively.

(b) $w$ chooses $\ell_{wm} = 1$ if and only if the expected utility from *ℓiking* exceeds that from not *ℓiking*.

Step 1. Given $\ell_{wm}$, $m$ decides whether to *ℓike* $w$ followed by the decision to send a message to $w$

(a) Using estimated preference parameters, construct expected utilities from *ℓiking* ($d_{mw} = 1$) and not *ℓiking* ($d_{mw} = 0$), given as equations (13) and (14), respectively.

(b) $m$ chooses $d_{mw} = 1$ if and only if the expected utility from *ℓiking* exceeds that from not *ℓiking*.

(c) If $d_{mw} = 1$, the expected utilities from messaging and not messaging are given by equations (3) and (5), respectively. Conversely, if $d_{mw} = 0$, they are given by equations (8) and (9), respectively.

| | Control (CT) | | | Treated (TR) | | | Gale-Shapley (GS) | | | GS - CT |
|---|---|---|---|---|---|---|---|---|---|---|
| | Mean | SD | Obs. | Mean | SD | Obs. | Mean | SD | Obs. | 95% CI |
| **Panel A. Men-Optimal** | | | | | | | | | | |
| Age | 3.953 | 1.146 | 2,287 | 4.162 | 1.125 | 2,488 | 4.606 | 1.323 | 2,514 | [ 0.582, 0.723 ] |
| Education | 1.284 | 1.189 | 293 | 1.545 | 1.141 | 317 | 1.694 | 1.289 | 334 | [ 0.214, 0.605 ] |
| Desirability | 1.029 | 0.566 | 2,287 | 1.025 | 0.551 | 2,488 | 0.904 | 0.476 | 2,514 | [ 0.095, 0.154 ] |
| BodyType | 0.508 | 0.281 | 418 | 0.543 | 0.268 | 469 | 0.538 | 0.328 | 472 | [-0.010, 0.071 ] |
| Race | 0.429 | 0.386 | 408 | 0.577 | 0.381 | 427 | 0.579 | 0.414 | 471 | [ 0.098, 0.204 ] |
| **Panel B. Women-Optimal** | | | | | | | | | | |
| Age | 5.104 | 2.258 | 2,520 | 5.274 | 2.113 | 2,628 | 5.272 | 2.509 | 2,739 | [ 0.038, 0.297 ] |
| Education | 1.517 | 1.343 | 334 | 1.591 | 1.338 | 335 | 1.636 | 1.411 | 356 | [-0.087, 0.325 ] |
| Desirability | 0.889 | 0.512 | 2,520 | 0.868 | 0.462 | 2,628 | 0.857 | 0.544 | 2,739 | [ 0.004, 0.061 ] |
| BodyType | 0.477 | 0.393 | 448 | 0.512 | 0.512 | 490 | 0.565 | 0.419 | 520 | [ 0.037, 0.140 ] |
| Race | 0.632 | 0.404 | 475 | 0.544 | 0.396 | 512 | 0.546 | 0.423 | 517 | [ 0.034, 0.138 ] |

*Notes:* This table reports the average of means, standard deviations, and number of observations across 100 simulations.

Table 10: Attribute Correlations in Predicted Matches

(d) $m$ chooses to send a message if and only if the expected utility from sending a message is greater than that of not sending one.

Step 2. Compute equilibrium matches:

(a) Define $I_m$ as the set of all profiles that $m$ messaged.

(b) All men first propose to their most highly-ranked woman within the set $I_m$. Ties are broken randomly.

(c) Woman selects the most highly ranked man, as long as $\hat{U}_{W,wm} \geq \hat{V}_W(w)$.

(d) A man who has not been chosen proposes to the next best woman in his set $I_m$.

(e) If a woman receives a new offer that is preferable to the one she is currently holding, she releases the previous offer. The released man then has to propose to the next woman on his list within the set $I_m$.

(f) Steps 2 (a) through (e) continue until each man exhausts the list of women in his set $I_m$.

## 8.2 Frictions and Sorting

Table 10 reports the mean attribute differences between couples for predicted matches under the assumption of the default of the platform (i.e., the control condition or CT), one-sided treatment (TR), and frictionless (GS) environments. We include all matches (Likers and NotLikers) and report the 95% confidence intervals for the difference between the frictionless and control protocols (GS-CT). While the treatment setting generally shows larger attribute differences compared to the control setting, the overall pattern is ambiguous due to the inclusion of NotLikers. However, when all frictions are removed, we observe significant increases in attribute differences for age, education, body type, and race, alongside a decrease in desirability difference compared to the control setting.

For example, in man-optimal matches, the age difference is 3.95 years in the control setting and 4.61 years in the frictionless setting, suggesting that 14% of the positive sorting in age is due to frictions, while 86% (=3.9531/4.6055) is attributed to preferences. For years of education, 24% of the sorting is attributed to frictions and 76% to preferences. For desirability, 88% of the sorting is due to preferences, while 12% is due to frictions. While not statistically significant, 6% of the sorting in body type is due to frictions and 94% due to preferences. Finally, 26% of the positive sorting in race is due to frictions and 74% due to preferences.

We observe similar patterns in woman-optimal matches. The age difference between couples is 5.10 years in the control setting and 5.27 years in the frictionless setting, with approximately 3% of the sorting in age attributed to frictions. While not statistically significant, the difference in years of schooling is 1.52 years in the control and 1.64 years in the frictionless setting, with 7% of the sorting due to frictions. For desirability, 3.7% of the sorting is due to frictions, and 96.3% to preferences. Interestingly, in the frictionless setting, women show more sorting with partners of the same race, with 86% of positive sorting due to preferences and 14% due to frictions.

Table 11 reports the mean attribute differences of couples *who matched with Likers* under the control (CT) and treatment (TR) settings, along with the 95% confidence intervals for their differences (TR-CT). Panel A reports the results for men. Consistent with the patterns we observe in our data, attribute differences in the treatment setting are significantly larger compared to those of the control setting for age, education, and race, and is significantly smaller for desirability. For matches with Likers, the treatment increases the age gap by 0.41 years. Given that the total effect of search friction on the age difference in matches with Likers is 0.65 years (=4.6055−3.9531), the treatment reduces the impact of search friction on sorting by 74% (=0.41/0.65). For years of education, the treatment reduces the impact of search frictions on sorting by 88%, and for desirability, the treatment fully removes the impact of search frictions. For race, the treatment reduces the impact of friction by 92%. We do not find significant differences for body type.

| | Control (CT) | | | Treated (TR) | | | TR-CT |
|---|---|---|---|---|---|---|---|
| | Mean | SD | Obs. | Mean | SD | Obs. | 95% CI |
| **Panel A. Men-Optimal** | | | | | | | |
| Age | 4.045 | 1.418 | 1,177.8 | 4.461 | 1.397 | 1,445.9 | ( 0.307, 0.524 ) |
| Education | 1.258 | 1.279 | 149.15 | 1.641 | 1.270 | 180.72 | ( 0.105, 0.660 ) |
| Desirability | 0.970 | 0.569 | 1,177.8 | 0.878 | 0.546 | 1,445.9 | ( 0.050, 0.135 ) |
| BodyType | 0.459 | 0.331 | 217.55 | 0.430 | 0.315 | 286.6 | (-0.028, 0.086 ) |
| Race | 0.471 | 0.411 | 211.19 | 0.571 | 0.408 | 277.97 | ( 0.027, 0.174 ) |
| **Panel B. Women-Optimal** | | | | | | | |
| Age | 5.120 | 2.403 | 1,464 | 5.476 | 2.363 | 1,494.5 | ( 0.184, 0.528 ) |
| Education | 1.488 | 1.340 | 186.27 | 1.630 | 1.351 | 192.4 | (-0.130, 0.415 ) |
| Desirability | 0.872 | 0.533 | 1,464 | 0.830 | 0.498 | 1,494.5 | ( 0.005, 0.079 ) |
| BodyType | 0.471 | 0.409 | 249.32 | 0.544 | 0.403 | 263.4 | ( 0.002, 0.143 ) |
| Race | 0.645 | 0.414 | 257.4 | 0.595 | 0.407 | 271.58 | (-0.020, 0.120 ) |

*Notes:* This table reports mean attribute differences for matches achieved with Likers across 100 simulations.

Table 11: Attribute Difference in Predicted Matches with Likers

Panel B reports results for women. The treatment setting shows significantly larger attribute differences for age and body type, but significantly smaller differences for desirability. The treatment increases the education level difference compared to the control, but this difference is not statistically significant. The treatment fully removes the impact of search frictions on age and desirability, and reduces the impact on body type by 77%.

### 8.3 Efficiency

We quantify the departure from efficiency caused by frictions on the platform, and study whether reducing frictions through signaling improves efficiency and makes users better off. Let $R_m^{s1}$ and $R_m^{s2}$ represent the ranks of $m$'s matched partner in two different settings, where $s1$ and $s2$ are elements of the set $\{CT, TR, GS\}$. The difference in rankings for $m$ between the two settings is $\Delta R_m^{s1-s2} = R_m^{s1} - R_m^{s2}$, and its average across users is denoted as $\Delta \bar{R}^{s1-s2}$. A positive difference indicates that setting $s1$ leads to a better match allocation than setting $s2$.

In Panel A of Table 12, we report the means and standard deviations of predicted average ranking differences across protocols for men-optimal and women-optimal equilibria. In men-optimal matches, the average ranking difference between the treated and control settings is 66.22 for men and 9.47 for women. To interpret the magnitude of this difference, we can express the ranking differences as a percentage of the highest achievable ranking ($100 \times \Delta R_m^{s1-s2}/N_w$), which results in

| | Men | | | | Women | | | |
|---|---|---|---|---|---|---|---|---|
| | Mean | Median | SD | 95% CI | Mean | Median | SD | 95% CI |
| ***Panel A. Changes in Partner Ranking*** | | | | | | | | |
| **Men-Optimal** | | | | | | | | |
| $\Delta \bar{R}^{TR-CT}$ | 66.22 | 61.99 | 64.9 | (65.05, 67.39) | 9.47 | 6.67 | 102.6 | (6.291, 12.65) |
| $\Delta \bar{R}^{GS-CT}$ | 95.80 | 103.32 | 115.9 | (93.71, 97.89) | 96.59 | 102.42 | 70.5 | (94.41, 98.78) |
| **Women-Optimal** | | | | | | | | |
| $\Delta \bar{R}^{TR-CT}$ | 45.71 | 26.93 | 82.6 | (44.22, 47.20) | 269.19 | 157.53 | 260.3 | (261.1, 277.3) |
| $\Delta \bar{R}^{GS-CT}$ | 113.8 | 110.1 | 99.6 | (112.0, 115.6) | 1,145.4 | 1,120.4 | 341.4 | (1,135, 1,156) |
| ***Panel B. Changes in Net Utility*** | | | | | | | | |
| **Men-Optimal** | | | | | | | | |
| $\%\Delta \bar{U}^{TR-CT}$ | 53.57 | 62.59 | 37.59 | (52.41, 54.74) | 0.170 | 0.033 | 17.45 | (-0.371, 0.711) |
| $\%\Delta \bar{U}^{GS-CT}$ | 104.1 | 103.3 | 3.26 | (104.0, 104.2) | 107.4 | 103.2 | 66.61 | (105.3, 109.5) |
| **Women-Optimal** | | | | | | | | |
| $\%\Delta \bar{U}^{TR-CT}$ | 0.286 | 0.247 | 5.203 | (0.192, 0.380) | 76.69 | 79.36 | 13.19 | (76.28, 77.10) |
| $\%\Delta \bar{U}^{GS-CT}$ | 107.0 | 105.6 | 7.21 | (106.9, 107.1) | 107.5 | 107.1 | 3.47 | (107.4, 107.6) |

*Notes:* This table reports the average of means, medians, and standard deviations for changes in partner ranking (Panel A) and percentage changes in utility (net of costs) (Panel B) across 100 simulations.

Table 12: Changes in Efficiency

a ranking improvement of 1.65 percentage points for men and 0.08 percentage points for women. Despite the small magnitude, this suggests that offering treatment as a free feature benefits users by improving the ranking of their matched partner. When all frictions are removed, men's partner rankings improve by 95.8, and women's by 96.59 compared to the control setting, or 2.4 percentage points for men and 0.82 percentage points for women. Similar patterns are observed in women-optimal matches, with the average difference between the treatment and control group matches being 45.7 for men and 269.2 for women (improvements of 1.14 percentage points for men and 2.28 for women). The difference between the Gale-Shapley matches and the control matches is 113.8 for men and 1,145.4 for women (improvement of 2.85 percentage points for men and 9.7 percentage points for women).

In a frictionless environment, matches are determined solely by preferences, whereas in the control and treatment settings, users incur costs for *ℓiking* and messaging. To assess improved outcomes for users in terms of overall utility, we compare the utility users receive, net of the costs associated with *ℓiking* and messaging. Specifically, we subtract the total costs incurred by a user during the entire process of the deferred-acceptance algorithm from the utility that a user obtains from a match, and we compare the percent change of this value between (1) the treatment and control environments, denoted as $\%\Delta\bar{U}^{TR-CT}$, which captures the effect of the treatment on net utility; and (2) the percent change in net utility between the frictionless and control environments, denoted as $\%\Delta\bar{U}^{GS-CT}$, which assesses the impact of introducing costs relative to a frictionless setting.

In men-optimal matches, the average percent change in net utility between the treatment and control settings is 54% for men and 0.17% for women, suggesting that the treatment leads to higher utility for users, even after accounting for the costs of $\ell iking$ and messaging. When all frictions are removed, the percent change in utility between the frictionless and control environments is 104% for men and 107% for women. In women-optimal matches, the average percent change in net utility between the treatment and control settings is 0.29% for men and 76% for women. Similarly, the percent change in utility between the frictionless and control environments is 107% for men and 108% for women.

## 9 Conclusion

With agents on both sides of two-sided markets having private preferences, finding a match based on mutual agreement requires extensive costly search. This paper studies the impact of search frictions on match formation in two-sided markets and explores how signaling preferences can mitigate the impact of these frictions on sorting. Using data from an online dating platform, we investigate how signaling a higher likelihood of a match can reduce the impact of frictions on sorting and efficiency, while also decomposing the impact of preferences and frictions on sorting.

Our findings can provide important managerial implications for the pricing of premium features, such as how much users are willing to pay to receive a signal about the preferences of the other side. In addition, our findings shed light on what type of information should be displayed on users' profiles. Information that is helpful in gauging the preferences of other users can greatly improve the consumer experience. Our findings are not limited to online dating contexts; they are broadly applicable to other two-sided market contexts where matches may not form due to imperfect information about each other's preferences.

Our paper also provides insights into how online platform design can contribute to diversity, given that one-third of US marriages now originate from online encounters. However, due to data limitations, we cannot quantify the long-term effects of reducing frictions in online dating, such as the impact on marriages and the accumulation of human capital through children's education. Future research should aim to address this gap, as understanding these long-term effects can offer solutions to alleviating persistent social inequality.

**Declarations: Funding and Competing Interests**

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. The authors have no funding to report.

## References

Arcidiacono, P., A. Beauchamp, and M. McElroy (2016): "Terms of endearment: An equilibrium model of sex and matching," *Quantitative Economics*, 7, 117–156.

Arnosti, N., R. Johari, and Y. Kanoria (2021): "Managing Congestion in Matching Markets," *Manufacturing Service Operations Management*, 23, 547–730.

Bagger, J. and R. Lentz (2019): "An Empirical Model of Wage Dispersion with Sorting," *The Review of Economic Studies*, 86, 153–190.

Banerjee, A., E. Duflo, M. Ghatak, and J. Lafortune (2013): "Marry for What? Caste and Mate Selection in Modern India," *American Economic Journal: Microeconomics*, 5, 33–72.

Bapna, R., E. I. McFowland, P. Mojumder, J. Ramaprasad, and A. Umyarov (2022): "So, Who Likes You? Evidence from a Randomized Field Experiment," *Management Science*, 69, 3759–4361.

Bapna, R., J. Ramaprasad, G. Shmueli, and A. Umyarov (2016): "One-Way Mirrors in Online Dating: A Randomized Field Experiment," *Management Science*, 62, 3085–3391.

Baumeister, R. F., S. R. Wotman, and A. M. Stillwell (1993): "Unrequited Love: On Heartbreak, Anger, Guilt, Scriptlessness, and Humiliation," *Journal of Personality and Social Psychology*, 64, 377–394.

Bernstein, W. M., B. O. Stephenson, M. L. Snyder, and R. A. Wicklund (1983): "Causal Ambiguity and Heterosexual Affiliation," *Journal of Experimental Social Psychology*, 19, 78–92.

Bojd, B. and H. Yoganarasimhan (2019): "Star-Cursed Lovers: Role of Popularity Information in Online Dating," *Working Paper*.

Bruch, E. and M. Newman (2024): "Aspirational Pursuit of Mates in Online Dating Markets." *Science Advances*, 4, 9815.

Card, D., A. R. Cardoso, J. Heining, and P. Kline (2018): "Firms and Labor Market Inequality: Evidence and Some Theory," *Journal of Labor Economics*, 36, S13–S70.

Chan, T. Y., B. H. Hamilton, and N. W. Papageroge (2015): "Health, Risky Behavior and the Value of Medical Innovation for Infectious Disease," *Review of Economic Studies*, 83, 1456–1510.

Chen, Y. and S. Yao (2017): "Sequential Search with Refinement: Model and Application with Click-Stream Data," *Management Science*, 63, 4345–4365.

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

Coles, P., J. Cawley, P. B. Levine, M. Niederle, A. E. Roth, and J. J. Siegfried (2010): "The Job Market for New Economists: A Market Design Perspective," *Journal of Economic Perspectives*, 24, 187–206.

Eastwick, P. W. and E. J. Finkel (2009): "Reciprocity of Liking," *Encyclopedia of Human Relationships*, 3, 1333–1336.

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

——— (2008): "Racial Preferences in Dating," *Review of Economic Studies*, 75, 117–132.

Flinn, C. J. and D. Del Boca (2012): "Endogenous household interaction," *Journal of Econometrics*, 166, 49–65.

Fong, J. (2018): "Search, Selectivity, and Market Thickness in Two-Sided Markets," *working paper*.

Fradkin, A. (2015): "Search, Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb," *Working Paper*.

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

Halaburda, H., M. J. Piskorski, and P. Yildirim (2017): "Competing by Restricting Choice: The Case of Matching Platforms," *Management Science*, 64, 3574–3594.

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

——— (2010b): "What makes you click?-Mate preferences in online dating," *Quantitative Marketing and Economics*, 8, 393–427.

Hong, L. (2024): "Coworker Sorting, Learning, and Inequality," SSRN No. 4291004, 86, https://doi.org/10.2139/ssrn.4291004.

Horton, J. J. (2014): "Misdirected Search Effort in a Matching Market: Causes, Consequences and a Partial Solution," *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, 357–357.

Huang, N., G. Burtch, Y. He, and Y. Hong (2022): "Managing Congestion in a Matching Market via Demand Information Disclosure," *Information Systems Research*, 33, 1119–1516.

Jung, J., R. Bapna, J. Ramaprasad, and A. Umyarov (2022): "Love Unshackled: Identifying the Effect of Mobile App Adoption in Online Dating," *MIS Quarterly*, 33, 47–70.

Kalmijn, M. (1998): "Intermarriage and Homogamy: Causes, Patterns, Trends," *Annual Review of Sociology*, 24, 395–421.

Kenny, D. A. (1994): "Interpersonal Perception: A Social Relations Analysis." *Guilford Press.*

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

Lee, S. (2015): "Effect of Online Dating on Assortative Mating: Evidence from South Korea," *Journal of Applied Econometrics*, 31, 1120–1139.

Lee, S. and M. Niederle (2015): "Propose with a rose? Signaling in internet dating markets," *Experimental Economics*, 18, 731–755.

Montoya, R. M. and C. A. Insko (2008): "Toward a More Complete Understanding of the Reciprocity of Liking Effect." *European Journal of Social Psychology*, 38, 477–498.

Raquel, F. (2003): "Household formation, inequality, and the macroeconomy," *Journal of the European Economic Association*, 1, 683–697.

Raquel, F. and R. Rogerson (2001): "Sorting and Long-Run Inequality," *Quarterly Journal of Economics*, 116, 1305–1341.

Richards Shubik, S. (2015): "Peer effects in sexual initiation: Separating demand and supply mechanisms," *Quantitative Economics*, 14, 663–702.

Shi, L. and S. Viswanathan (2023): "Optional Verification and Signaling in Online Matching Markets: Evidence from a Randomized Field Experiment." *Information Systems Research*, 4, 1321–1814.

Vorauer, J. D., J. J. Cameron, J. G. Holmes, and D. G. Pearce (2003): "Invisible Overtures: Fears of Rejection and the Signal Amplification Bias." *Journal of Personality and Social Psychology*, 84, 793–812.

Vorauer, J. D. and R. K. Ratner (1996): "Who's Going to Make the First Move? Pluralistic Ignorance as an Impediment to Relationship Formation," *Journal of Social and Personal Relationships*, 13, 483–506.

Walster, E., G. W. Walster, J. Piliavin, and L. Schmidt (1973): ""Playing Hard to Get": Understanding an Elusive Phenomenon." *Journal of Personality and Social Psychology*, 26, 113–121.

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

# Appendix

## A Randomization Check

| Variable | Control Mean | Control SD | Control Obs. | Treatment Mean | Treatment SD | Treatment Obs. | $t$-stat | $p$-value |
|---|---|---|---|---|---|---|---|---|
| **Panel A. Men** | | | | | | | | |
| Age | 30.594 | 10.646 | 30,869 | 30.653 | 10.592 | 30,653 | 0.0822 | 0.9345 |
| HighSchool | 0.146 | 0.353 | 6,261 | 0.145 | 0.352 | 6,166 | 0.2595 | 0.7953 |
| TwoYear | 0.176 | 0.381 | 6,261 | 0.179 | 0.383 | 6,166 | -0.4428 | 0.6579 |
| University | 0.535 | 0.499 | 6,261 | 0.529 | 0.499 | 6,166 | 0.6552 | 0.5123 |
| PostGrad | 0.143 | 0.350 | 6,261 | 0.147 | 0.354 | 6,166 | -0.7076 | 0.4792 |
| Thin | 0.165 | 0.372 | 5,634 | 0.161 | 0.367 | 5,582 | 0.7036 | 0.4817 |
| Average | 0.656 | 0.475 | 5,634 | 0.655 | 0.475 | 5,582 | 0.1175 | 0.9065 |
| ExtraWeight | 0.141 | 0.348 | 5,634 | 0.147 | 0.354 | 5,582 | -0.8739 | 0.3822 |
| Overweight | 0.038 | 0.190 | 5,634 | 0.038 | 0.191 | 5,582 | -0.0476 | 0.9620 |
| Asian | 0.103 | 0.304 | 8,146 | 0.102 | 0.302 | 8,085 | 0.2782 | 0.7808 |
| White | 0.670 | 0.470 | 8,146 | 0.664 | 0.472 | 8,085 | 0.8379 | 0.4021 |
| Black | 0.106 | 0.308 | 8,146 | 0.113 | 0.316 | 8,085 | -1.2996 | 0.1938 |
| Hispanic | 0.105 | 0.307 | 8,146 | 0.107 | 0.309 | 8,085 | -0.3943 | 0.6934 |
| Other Race | 0.085 | 0.279 | 8,146 | 0.083 | 0.275 | 8,085 | 0.5351 | 0.5926 |
| **Panel B. Women** | | | | | | | | |
| Age | 30.907 | 10.724 | 19,131 | 30.789 | 10.763 | 19,347 | 1.0901 | 0.2757 |
| HighSchool | 0.135 | 0.342 | 3,329 | 0.134 | 0.341 | 3,380 | 0.1027 | 0.9182 |
| TwoYear | 0.179 | 0.383 | 3,329 | 0.172 | 0.377 | 3,380 | 0.7365 | 0.4614 |
| University | 0.531 | 0.499 | 3,329 | 0.541 | 0.498 | 3,380 | -0.8236 | 0.4102 |
| PostGrad | 0.155 | 0.362 | 3,329 | 0.152 | 0.359 | 3,380 | 0.2651 | 0.7909 |
| Thin | 0.274 | 0.446 | 2,498 | 0.270 | 0.444 | 2,514 | 0.3286 | 0.7425 |
| Average | 0.532 | 0.499 | 2,498 | 0.516 | 0.500 | 2,514 | 1.1703 | 0.2420 |
| ExtraWeight | 0.141 | 0.348 | 2,498 | 0.159 | 0.366 | 2,514 | -1.8446 | 0.0652 |
| Overweight | 0.053 | 0.225 | 2,498 | 0.056 | 0.229 | 2,514 | -0.3199 | 0.7491 |
| Asian | 0.127 | 0.333 | 4,548 | 0.131 | 0.338 | 4,764 | -0.6200 | 0.5352 |
| White | 0.655 | 0.476 | 4,548 | 0.647 | 0.478 | 4,764 | 0.7304 | 0.4652 |
| Black | 0.132 | 0.339 | 4,548 | 0.130 | 0.336 | 4,764 | 0.3164 | 0.7517 |
| Hispanic | 0.106 | 0.308 | 4,548 | 0.108 | 0.310 | 4,764 | -0.1986 | 0.8426 |
| Other Race | 0.056 | 0.229 | 4,548 | 0.054 | 0.226 | 4,764 | 0.3119 | 0.7551 |

*Notes:* This table presents a pretreatment comparison of the demographics between the control and treatment groups, separately for each gender.

Table A.1: Pretreatment Randomization Check

## B Number of Sessions

| | Men | | | | Women | | | |
|---|---|---|---|---|---|---|---|---|
| | control (1a) | treated (1b) | diff (1c) | t-stat (1d) | control (2a) | treated (2b) | diff (2c) | t-stat (2d) |
| **Panel A. Sessions with 1 hour gap** | | | | | | | | |
| Mean | 11.821 | 11.626 | -0.195 | -0.569 | 7.514 | 8.072 | 0.558 | 1.410 |
| SD | 18.836 | 18.394 | | | 11.750 | 13.150 | | |
| **Panel B. Sessions with 3 hour gap** | | | | | | | | |
| Mean | 9.512 | 9.398 | -0.115 | -0.502 | 6.385 | 6.772 | 0.388 | 1.372 |
| SD | 12.520 | 12.308 | | | 8.592 | 9.229 | | |
| Obs. | 5,752 | 6,055 | | | 1,911 | 2,087 | | |

Notes: New sessions are based on periods of no activity.

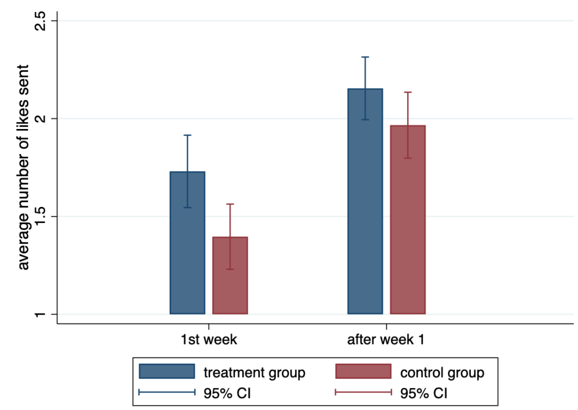
Table B: Number of Sessions During the Treatment Period

We compare the number of active sessions by treatment status. We define sessions based on periods of inactivity. Specifically, in Table B, Panel A, a new session begins after 1 hour of inactivity by a user. We observe no significant differences between users in the control and treatment groups, for both genders. We repeat the analysis with a new definition of a session starting after 3 hours of inactivity (Panel B), and find similar results.

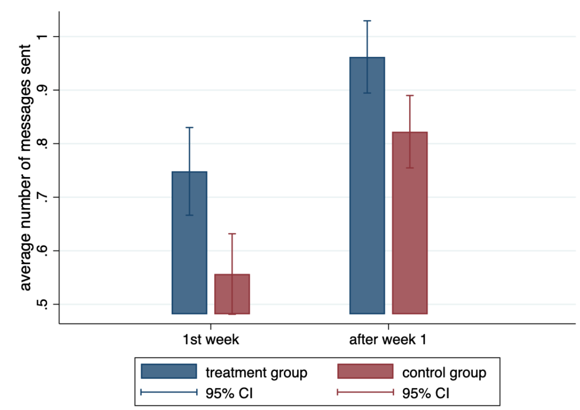
# C Treatment Effect Over Time

Knowing who *ℓiked* you may initially increase user engagement due to the excitement it generates. However, as the novelty of this feature wears off, the behavior of treated users may change a few days after exposure to the treatment. To investigate this potential novelty effect, we compare the behavior of the treated users over time with that of users in the control group.

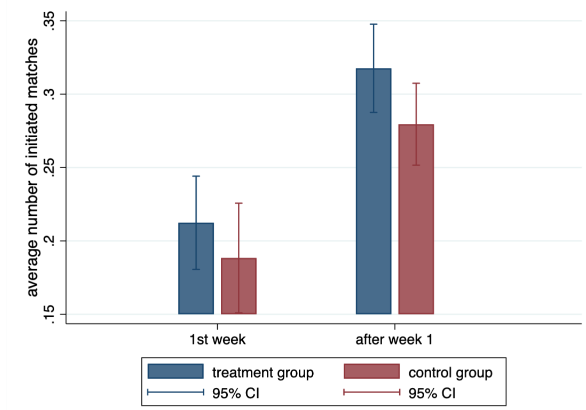
Figure C.1, subfigures (a) and (b) present the average number of *ℓikes* and messages sent by men to Likers over two periods: (1) the first week of treatment and (2) the subsequent weeks (weeks 2-4). For *ℓiking* behavior, the treatment effect is significant in the first period, but this difference diminishes and becomes insignificant in the following period. Regarding messaging, while a significant difference is observed in both periods, the effect diminishes over time, with the difference being smaller in the second period. In contrast, for initiated matches (subfigure (c)), no significant differences are observed between the treatment and control groups in either period. For NotLikers (subfigures (d) – (f)), no notable differences in activity are observed between the treatment and control groups over time. A similar pattern to that of men is observed for women, across both Likers and NotLikers (Figure C.2).



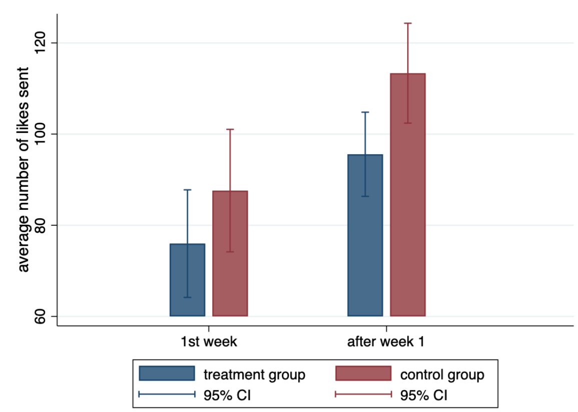

(a) $\ell ikes$ Sent to Likers Over Time



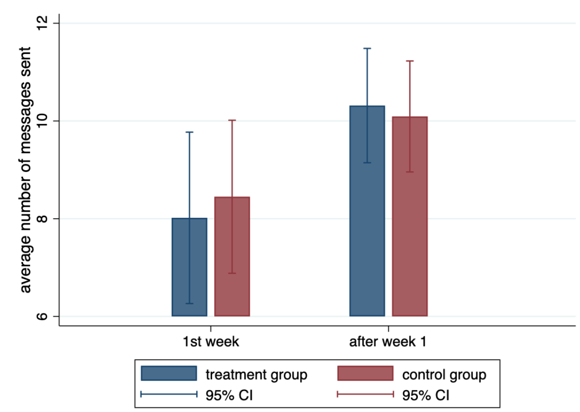

(b) Messages Sent to Likers Over Time



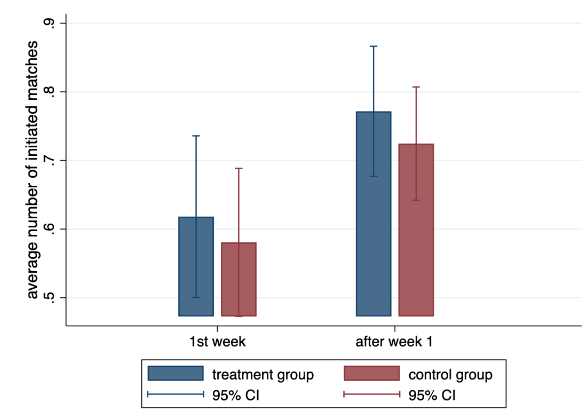

(c) Matches Initiated to Likers Over Time




(d) $\ell ikes$ Sent to NotLikers Over Time

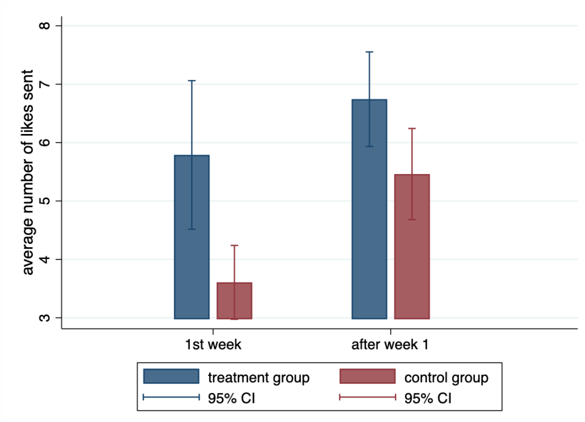



(e) Messages Sent to NotLikers Over Time

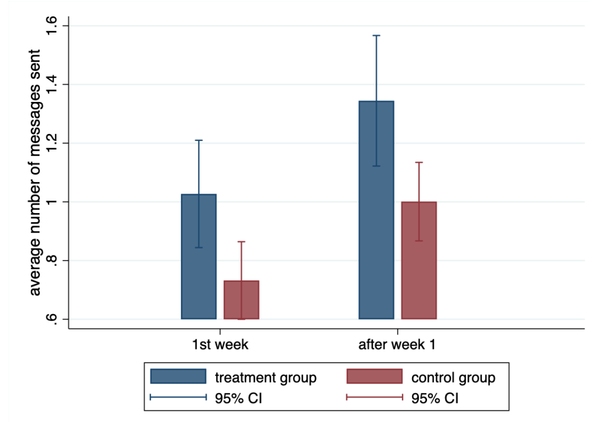



(f) Matches Initiated to NotLikers Over Time

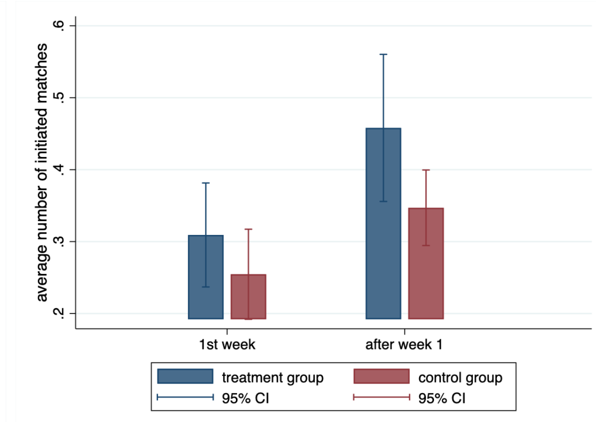
Figure C.1: Changes in User Behavior Over Time – Men



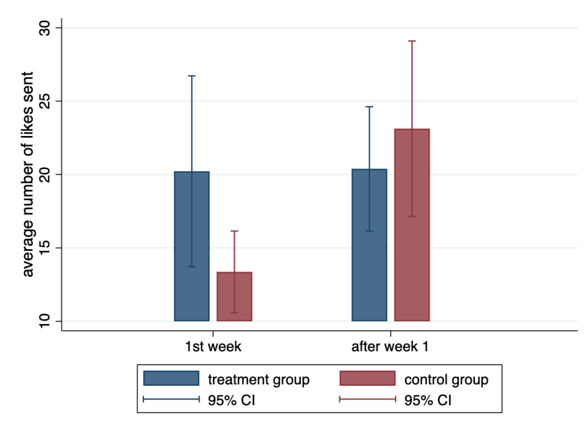

(a) $\ell ikes$ Sent to Likers Over Time



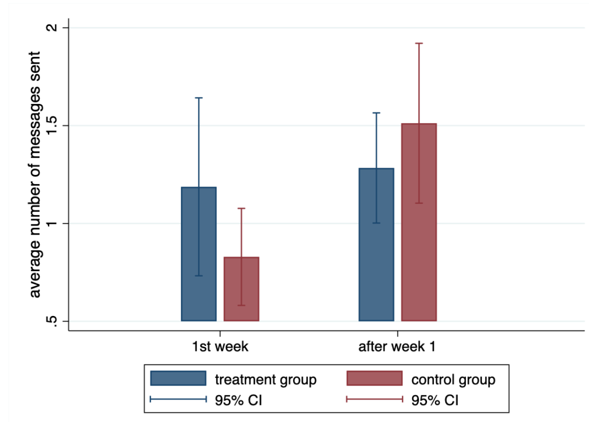

(b) Messages Sent to Likers Over Time



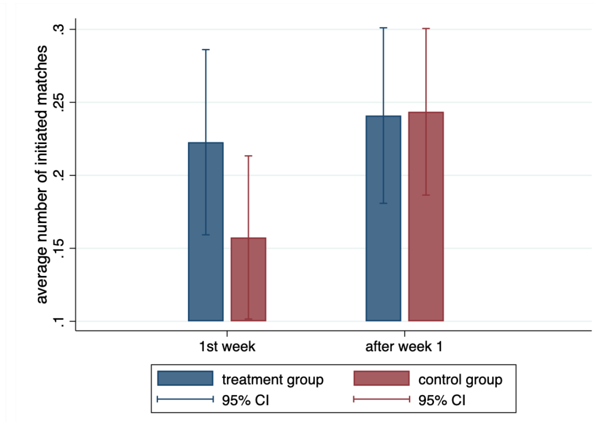

(c) Matches Initiated to Likers Over Time




(d) $\ell ikes$ Sent to NotLikers Over Time




(e) Messages Sent to NotLikers Over Time




(f) Matches Initiated to NotLikers Over Time

Figure C.2: Changes in User Behavior Over Time – Women

# D Liking Behavior With Respect Relative Match Probability

| dependent variable | *like* indicator | | | |
|---|---|---|---|---|
| | Men | | Women | |
| | Liker | NotLiker | Liker | NotLiker |
| medium match prob | -0.0597*** | 0.0129 | -0.0281** | -0.0328** |
| | (0.0136) | (0.0147) | (0.0121) | (0.0140) |
| high match prob | -0.163*** | -0.0924*** | -0.0599*** | -0.0405*** |
| | (0.0171) | (0.0178) | (0.0139) | (0.0150) |
| treated×low match prob | 0.0661*** | -0.0288 | 0.0681*** | -0.0102 |
| | (0.0172) | (0.0240) | (0.0184) | (0.0190) |
| treated×medium match prob | 0.0684*** | -0.0417** | 0.0410*** | 0.00538 |
| | (0.0174) | (0.0198) | (0.0115) | (0.00906) |
| treated×high match prob | 0.0123 | -0.0209 | 0.0279** | -0.00571 |
| | (0.0299) | (0.0188) | (0.0120) | (0.00853) |
| constant | 0.661*** | 0.463*** | 0.155*** | 0.106*** |
| | (0.0128) | (0.0173) | (0.0122) | (0.0142) |
| Obs. | 38,433 | 2,711,317 | 166,986 | 993,829 |
| $R$-squared | 0.030 | 0.009 | 0.011 | 0.004 |

*Notes*. This table presents the treatment's effect on *liking* behavior with respect to different ex-ante match probabilities. Robust standard errors in parentheses.
*** p<0.01, ** p<0.05, * p<0.1

Table C: Liking Behavior With Respect Relative Match Probability