Information Systems Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Mr. Right or Mr. Best: The Role of Information Under Preference Mismatch in Online Dating

Hongchuan Shen; , Chu (Ivy) Dang; , Xiaoquan (Michael) Zhang;

informs
https://pubsonline.informs.org/journal/isre

INFORMATION SYSTEMS RESEARCH
Vol. 35, No. 4, December 2024, pp. 2013–2029
ISSN 1047-7047 (print), ISSN 1526-5536 (online)

Mr. Right or Mr. Best: The Role of Information Under Preference Mismatch in Online Dating

Hongchuan Shen,[a] Chu (Ivy) Dang,[b] Xiaoquan (Michael) Zhang[c,d,*]

[a] Faculty of Business Administration, University of Macau, Macau, China; [b] Faculty of Business and Economics, The University of Hong Kong, Hong Kong; [c] Department of Decision Sciences and Managerial Economics, Business School, Chinese University of Hong Kong, Hong Kong; [d] Department of Management Science and Engineering, School of Economics and Management, Tsinghua University, Beijing 100084, China

*Corresponding author

Contact: hongchuanshen@um.edu.mo, https://orcid.org/0000-0002-4194-2366 (HS); ivydang@hku.hk, https://orcid.org/0000-0001-6349-6320 (C(I)D); zhang@cuhk.edu.hk, https://orcid.org/0000-0003-0690-2331 (X(M)Z)

Received: April 7, 2022
Revised: December 4, 2022; May 5, 2023
Accepted: July 19, 2023
Published Online in Articles in Advance: March 13, 2024

https://doi.org/10.1287/isre.2022.0233

Abstract. This paper examines the role of information in two-sided matching markets where preference mismatch is present. Two-sided markets are characterized by different preferences of the parties involved, and a match occurs when both sides show a preference for each other. In practice, however, there is often a preference mismatch. In this study, we use a large data set from an online dating website to provide empirical evidence for preference mismatch in the field. We also develop empirical models to investigate the impact of information under preference mismatch by analyzing variations in the amount of available information. Specifically, we compare *partial* and *complete* information contained in the users' short and long profiles, respectively. We find that more information about the other side does not necessarily improve the likelihood of a match. In fact, the side making the proposal has a better chance of matching if the decision is based on the information contained in the short profile rather than the long profile. This suggests that users are better off seeing partial rather than complete information about the candidates, a phenomenon we refer to as the "less information is more" effect. Our empirical analysis shows that this effect is driven by the mismatched preferences of the two sides. These results imply that there is an optimal amount of information that one side should possess about the other before making a proposal. Our study highlights the importance of optimal information design strategies to determine the appropriate amount of information that should be provided to each side. Our findings also offer managerial implications regarding information provision strategies for online platforms in general.

History: Pei-Yu Chen, Senior Editor; Xitong Li, Associate Editor.
Funding: This work was supported by the University of Macau [Grant SRG2023-00023-FBA] and the Research Grants Council of Hong Kong, University Grants Committee [Grants ECS 27504221, GRF 14500521, GRF 14501320, GRF 14503818, GRF 165052947, PDFS 2021-4H04, TRS:T31-604/18-N].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/isre.2022.0233.

Keywords: preference mismatch • matching platforms • two-sided markets • online dating • information disclosure • information design

1. Introduction

The rise of two-sided matching platforms such as Uber, Airbnb, Upwork, and Tinder has changed the way we commute, travel, work, and even date. These platforms facilitate value-creating matches while keeping search frictions low. Matching outcomes on these platforms depend on match-relevant information provided to the users, such as driver ratings on Uber, host information on Airbnb, or the age of a potential date on Tinder. Therefore, the success of these platforms depends on the role of information. What information and how much information should be provided? Conventional wisdom often suggests that providing more information to the disadvantaged side mitigates the problem of information asymmetry (Stiglitz 2000). This is what platforms like Airbnb do—they provide very detailed information to both sides; guests can check the location, photos, and reviews of an accommodation before booking it. Hosts can also check the booking histories, reviews, and self-introduction of the guest before accepting the booking request. In contrast, matching platforms like Uber deliberately withhold certain information about one side from the other side. For instance, passenger destination information is hidden from the driver before the driver accepts the ride request. Even platforms operating in the same domain can differ in their choices of how much information to show to their users. Online dating platforms OkCupid and Tinder are one example: OkCupid allows users to check more detailed information about a potential partner before making a match proposal, whereas Tinder allows users to view only limited information of a profile (e.g., the profile photo, a small bio, and the location) before swiping to like or dislike the profile.

So why do some matching platforms choose to show more information whereas others choose to withhold certain information? Findings from prior literature suggest that the answer depends on many factors such as market conditions (Halaburda et al. 2018, Fong 2023), the type of information (Bapna et al. 2016, Caldieraro et al. 2018, Horton 2019), the type of user (Tadelis and Zettelmeyer 2015), and whether the users are strategic (Bojd and Yoganarasimhan 2022). These works often model the preferences of the agents as exogenously given, and focus more on the outcome of their decisions (e.g., the match) given the preferences. To understand the process of the match, we need to focus on the preferences of the two sides and their selection decisions. Although earlier theoretical matching models like Becker's (1973) assume homogeneous preferences of agents on the same side, in reality, online matching platforms are characterized by users' heterogeneous preferences over potential partners (e.g., men and women on Tinder, freelancers and clients on Upwork, and hosts and guests on Airbnb). Notably, such heterogeneity in user preferences also impacts the role of information provided by the platform and will consequently influence the information provision strategy of the platform. For example, Tadelis and Zettelmeyer (2015) explain that given heterogeneous preferences over car quality rankings, sellers can benefit by disclosing more information (i.e., car quality) to bidders in wholesale automobile auctions. Prior research, including the examples we cited above, mainly considers the assumptions of the preferences within one side, that is, whether players from the same side share homogeneous or heterogeneous preferences over players or products from the other side. Less is studied about the preferences of players from different sides, and how they intertwine with the role of information.

To fill this gap, we focus on a defining characteristic of two-sided matching markets—that is, a match depends on the possibly different preferences of the two sides—and argue that the amount of information released depends on the extent to which the preferences of the two sides are mismatched. Specifically, in an empirical context of online dating, we find that when there exists *preference mismatch* between the two sides, seeking less match-relevant information about the other side leads to a better matching outcome. We call this the "less information is more" effect.

Our empirical framework benefits from the unique information disclosure feature on a dating platform. We leverage variations in information amount with *partial* information contained in the users' *short profiles* and *complete* information in their *long profiles*. We refer to users on the initiation side as the focal users and users on the other side as the candidates. Focal users who go to the site searching for a potential match can send messages to a group of candidates they are interested in (i.e., propose a match). If candidates who receive proposals are interested in the focal users, they can reply and exchange messages. By default, the candidates' short profiles are observed by the focal users. Unlike most other dating sites where users can only send messages to potential matches after browsing each candidate's long profile page (Hitsch et al. 2010a, Lee and Niederle 2015), the site we study allows users to contact the candidates immediately after reading their short profiles,[1] or, alternatively, they can send messages after browsing their long profiles. We leverage this feature to examine how matching outcomes are influenced by the amount of information obtained by the users.

Leveraging rich clickstreams of user search and matching records from the online dating website, we empirically demonstrate the existence of preference mismatch. First, we construct an attractiveness score for each candidate of a given focal user based on the estimation of mate preferences. This attractiveness score describes the subjective evaluation of a candidate for a given focal user. Second, we compare the attractiveness scores of candidates who received match proposals and those who replied when focal users' proposing decisions were based on (a) candidates' short profiles (i.e., *partial info* group) or (b) long profiles (i.e., *complete info* group). We find that among candidates who received match proposals, those who replied had lower attractiveness scores. This was true for both groups, suggesting that the focal users and the candidates had different opinions on their ideal match, as the candidates who replied were usually not the focal users' original top choices. This indicates a preference mismatch. Additionally, we found evidence of preference mismatch at the attribute level. Our data show that male and female users have different opinions on the ideal match's height, age, income, education level, etc. The prevalence and significance of preference mismatch are also reflected by the low response rate on dating platforms. For example, even on a lively site like OkCupid, only about 32% of the first messages get any response.[2] And the response rate is even lower on less popular dating websites such as AreYouInterested? (AYI.com; Hickey 2013).

Next, we look at the role of information under preference mismatch and find that the degree of preference mismatch between the focal users and the candidates in the complete info group is much greater than that between the focal users and the candidates in the partial info group. In essence, more information about the other side does indeed lead to stronger preference mismatch. This is likely because when more attributes are observed by the focal users, misaligned preferences in each attribute start to weigh in, hence, increasing the level of mismatch. Subsequently, the stronger level of preference mismatch will make focal users select candidates who are less likely to accept them (i.e., the "best" ones), ruling out potential candidates before unobservable attributes of the candidates, such as personality and hobbies, are even discovered. If they had a chance to be contacted and communicated with, they would be the "right" ones to be matched with. Our empirical test corroborates this. Specifically, we find that obtaining more information about the other side when proposing a match does not enhance the likelihood of matching. In fact, it is less likely for the focal users to get matched with a candidate if they propose based on information presented in the long profiles. This implies that the focal users are better off if they see only partial, rather than complete, information about the candidates. To provide more theoretical rationales, we build a stylized analytical model in Online Appendix F to supplement the empirical findings.

This observation—the empirical demonstration of the information role under preference mismatch—has not been made before, to the best of our knowledge, and contributes to the existing literature on information disclosure and design in two-sided matching markets. Early theoretical works on matching usually focus on the equilibrium matching outcomes and put less emphasis on the searching process where preference mismatch usually happens (Gale and Shapley 1962, Roth 1982). In recent years, there has been a growing literature on search and matching (Chade et al. 2017) that has primarily focused on positive sorting. Most empirical works also tend to study final matching outcomes partly because they are limited by what is observed in data. Researchers usually only observe the outcome of the match and, hence, are not able to study preference mismatch, which requires observing the process of the match. It was only until recently, thanks to the availability of more granular data, that emerging empirical works have begun exploring search and selection process between the two sides (Bapna et al. 2016, Bruch et al. 2016). Preference mismatch is one of the prevalent phenomena that occur during such a process. We are among the first to empirically demonstrate the existence of preference mismatch and examine its interaction with information. It is also worth noting that the dating market is characterized by a high degree of heterogeneity in terms of both user characteristics and their preferences over potential partners. Therefore, online dating is an ideal context to study preference mismatch.

These findings generate managerial implications on the optimal information design and information provision strategies for matching platforms. To guide the application of our results to other matching contexts, we identify several boundary conditions. In short, our findings are more applicable to platforms that utilize decentralized matching mechanisms where user preferences play a role through the search process, and when communication cost is relatively low. The "less information is more" effect will be attenuated in contexts with lower mismatch cost and/or when the bargaining power between the two sides is highly imbalanced. Detailed discussions are provided in Section 6.5.

Lastly, we conduct a series of robustness analyses to rule out alternative explanations. One line of alternative explanations centers around the heterogeneity of users such as the heterogeneous goals of the focal users and the different levels of popularity or pickiness of the candidates. Another line of concerns focuses on user profiles, including the manipulation of short profiles and inference of long profiles based on short profiles. We rule out these alternative explanations in Section 6.

The rest of this paper is organized as follows. Section 2 reviews relevant literature. Then we introduce our empirical setting and data in Section 3. Next, we show empirical evidence of preference mismatch in Section 4, and test the role of information in Section 5. We discuss alternative explanations and boundary conditions in Section 6, and conclude in Section 7.

2. Literature

Most relevant to our research is a subset of literature in information design that focuses on the role of match-relevant information in two-sided markets. Findings from this literature are not conclusive. Some find that certain types of information are helpful; for instance, car quality information in wholesale automobile auctions (Tadelis and Zettelmeyer 2015), footprint traces of potential matches in online dating (Bapna et al. 2016), username and contribution amount on crowdfunding platforms (Burtch et al. 2016) and worker capacity information in online labor markets (Horton 2019) serve as useful signals to the market participants and/or help reduce information asymmetries. Other studies, in contrast, suggest complete information disclosure might be associated with suboptimal outcomes. Romanyuk and Smolin (2019) find that a platform's complete information disclosure may lead to market failure, which explains why Uber withholds the destination information from the drivers. Along this line of literature, other research shows that hiding or restricting information—like the number of choices, market thickness, and popularity information on dating sites (Halaburda et al. 2018, Bojd and Yoganarasimhan 2022, Fong 2023), winners' identities in business-to-business auction markets (Lu et al. 2019), worker type information in labor markets (Kanoria and Saban 2021), seller quality in online reputation systems (Shi et al. 2023), and borrower quality on peer-to-peer lending platforms (Caldieraro et al. 2018)—can lead to better market outcomes, such as more successful matches or higher platform revenues. In these works, the channel through which information influences the market lies in the preference differences on the same side. Our research, however, focuses on the role of information when there is preference discrepancy between different sides of the market.

Our work is also related to information designs in matching markets in general. The main challenge in the platform design of decentralized markets is creating an effective and efficient informative search process (Einav et al. 2016). It is well documented that information on search engines (Fradkin 2017), information contained in search results (Horton 2019), and choice set (Halaburda et al. 2018) all influence consumers' decision-making process. In addition to the content of information, there are also works studying how the order (Dinerstein et al. 2018), amount (Romanyuk and Smolin 2019), risk type (Kim et al. 2022), and timing (Fradkin et al. 2021) of information affect users. We focus on how the amount of information influences matching outcomes when preference mismatch exists. Conventional wisdom often suggests that users on platforms should try their best to obtain more information, but in this research we find circumstances where more information can actually hurt the users. These platform users have incentives not to obtain more information even if they face information asymmetry and the platform is willing to offer more information to them.

Our research also contributes to the economic literature on marriage and dating. Two major themes dominate this stream of research, with one focusing on the conditions of stable matching (Gale and Shapley 1962, Roth 1982) and the other focusing on explaining the positive sorting patterns observed in marriage (Becker 1973). Along this line of research, a subset of literature studies online dating, which is similar to our research context. Hitsch et al. (2010a) show that the matches are stable in online dating. They also measure mate preferences on this site and find no strategic behavior (Hitsch et al. 2010b). Lee and Niederle (2015) and Bapna et al. (2016) show that preference signaling increases an individual's match rate. Bruch et al. (2016) develop a discrete choice model to describe the multistage screening rules of users from an online dating site. Jung et al. (2019) explore the changes in user behavior when users use their mobile instead of the web version of the dating site. Fong (2023) studies how the information on market thickness affects the search and matching outcomes on a dating platform. Bojd and Yoganarasimhan (2022) find that popularity information on a centralized dating app can lead to strategic shading behavior of users. Our paper focuses on the misaligned preferences of the two sides of a dating website and how they impact the role of information. It is also important to distinguish the "beauty lies in the eyes of the beholder" effect from the preference mismatch effect. The former describes the diverging horizontal preferences of agents on the same side (Kanoria and Saban 2021, Du and Lei 2022), whereas the latter describes the preference discrepancy of a potential match that arises from the different sides of the market (Fisman et al. 2006).

3. Empirical Setting and Data

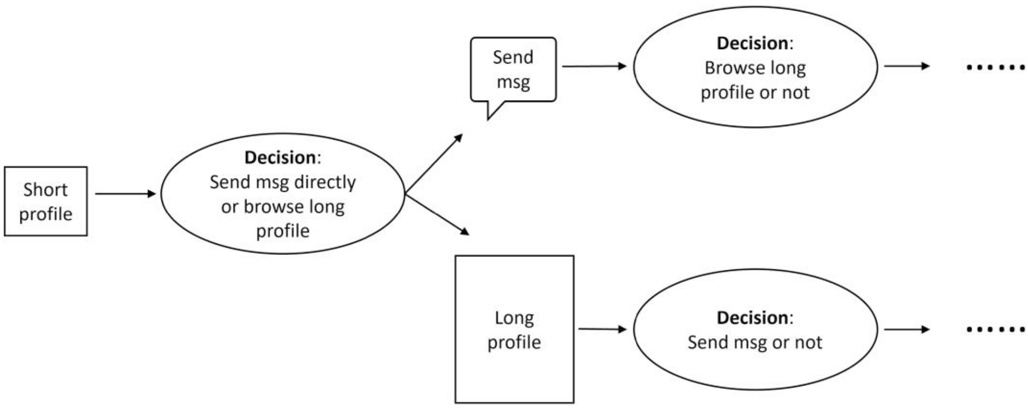
3.1. Empirical Setting

With a market size of USD 5.61 billion in 2021 and a worldwide user base of 300 million, online dating platforms are among the fastest-growing matching markets (*Economist* 2018). A survey in 2017 found that 39% heterosexual couples in the United States meet online, which has displaced friends (20%) as the main way of meeting potential partners (Rosenfeld et al. 2019). For homosexual couples in the United States, online dating has been the main way of meeting for over a decade.[3] In order to find an ideal partner, users usually spend quite some time and effort. Therefore, the information search process and designs of match-relevant information are particularly important for these types of platforms. In addition, the dating market is characterized by a high degree of heterogeneity in terms of both user characteristics and their preferences over potential partners. Therefore, online dating is an ideal context to study preference mismatch.

Our data come from one of the largest online dating sites in China, where single people search for potential partners to marry. Users can send messages to a group of candidates they are interested in (i.e., propose a match). If a candidate who receives a message is also interested in the focal user, she can reply and exchange messages with the focal user. Next, we describe the dating flow of the users on the site (Figure 1).

When a focal user (without loss of generality, a male) uses the online dating site to search for a potential partner, he will see a list of short profiles of candidates of the opposite gender (females).[4] At the time of our data collection (2011), the platform did not have personalized recommendations. The default landing page showed the same list of candidates for all users. The focal user can also search candidates based on certain attributes such as city and age. A short profile contains basic information about a candidate, including her nickname, age, education level, and home city. Based on this basic information, the focal user can (1) ignore this candidate, (2) contact the candidate immediately by sending a message, or (3) click on the candidate's photo and visit the candidate's long profile page. We show an example of the short profile in Online Appendix A.[5] On the long profile page, the focal user can observe more detailed information about the candidate, such as height, income, Chinese zodiac sign, astrological sign, religion, lifestyle, ownership of a house, and smoking habit. Most of the information on the long profile is required by the platform upon registration. In our data, more than 80% users filled out at least 90% of the long profile fields, and nearly all the users (99.8%) filled out at least 85% of the long profile fields.

Figure 1. The Dating Flow



Note. msg, Message.

When a candidate receives a message from the focal user, the focal user's short profile is attached to it. The candidate can (1) ignore the message, (2) reply, or (3) visit the focal user's long profile before deciding whether to reply. If the candidate does not reply, then no match occurs. The platform's main source of income is generated by charging for each message.[6] So the monetary cost of sending a message is higher than browsing a long profile, which is free. This also means that the monetary cost of sending messages with or without checking long profiles is the same. It is also worth noting that except for the messages they receive from the focal users, the candidates are unaware of any other actions executed by the focal users, such as whether the focal users have checked the candidates' short profiles or long profiles.

Because we do not observe users' activities outside this platform, we use the intensity of mutual communications as a proxy for matching, which is measured by the number of mutual messages exchanged between a focal user and a candidate.[7] The information structure of the platform allows us to examine how the amount of match-relevant information (i.e., the information in the short profiles versus information in the long profiles) influences the final matching outcomes.

3.2. Data

The data set was provided to us by the platform. The focal users in the data were a random sample of newly registered users, whose entire search and matching history between November 2011 and January 2012 was recorded as clickstream data. We have the interaction history of the focal users, including profile-checking records, messaging records, and the reactions from the candidates.[8]

3.2.1. Data Summary. Table 1 summarizes the raw data. In total, we observe 1,198,943 clicks from 33,504 focal users. Because we focus on the search and matching behaviors of the focal users, we will now describe the behavior of an average focal user. An average male user interacted with around 38 female candidates. The interactions included browsing candidates' long profiles, match proposals, and subsequent message communications (if any). Among all the candidates that the focal user interacted with, 45% of them were proposed to based on their short profiles. The total number of candidates that were proposed to, including proposals based on both short and long profiles, was around 20, but on average, only two female candidates replied. The matching rate of proposals sent by male focal users is around 9.8%. Here, a match is counted as successful if the focal user and a candidate exchanged messages at least once. The matching rate is 2.6% if we define a successful matching as the focal user and a candidate exchanging messages at least twice. Compared with male focal users, female focal users on the platform were more selective. They interacted with around 31 male candidates, and proposed to around 44% of them. The total number of proposals sent to male candidates was around 16. However, although female focal users proposed less frequently than their male counterparts, they enjoyed a much higher reply rate, with 4 out of 16 male candidates replying back. Consequently, the matching rate is higher for female users. This is consistent with the findings documented in Fong (2023), who also finds that female users on a popular dating app are more selective and get more matches than men. We summarize the attributes of focal users in Online Appendix B, Table A.1.

4. Empirical Evidence of Preference Mismatch

4.1. Preference Mismatch: Stylized Illustration

To explain the intuition behind the concept of preference mismatch, we first consider one example from our data. Based on the data, we find that a typical man's favorable height of a woman is 10 cm shorter than him,

Table 1. Data Summary

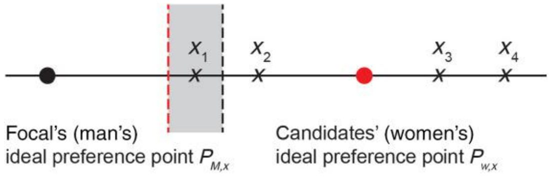
Variables	Male focal users	Female focal users
Time range	November 2011–January 2012	
Observations in clickstreams	1,198,943	
Provinces	34	
Focal users	33,504	
Total users	24,284	9,220
# cand. interacted	38	31
# cand. proposed short / # cand. interacted	45%	44%
# cand. proposed	20	16
# cand. proposed long	2.2	2.1
# cand. replied	2.0	3.9
# matches (mutual comm. ≥ 1)/# all proposals (%)	9.8	24.5
# matches (mutual comm. ≥ 2)/# all proposals (%)	2.6	5.5

Note. Cand., Candidate; comm., communication.

whereas a typical woman's favorable height of a man is around 20 cm taller than her. So preference mismatch exists in the interval between the height difference 10 cm and 20 cm. For example, Alice is 160 cm tall and Bob is 170 cm. If Bob suddenly grew to 175 cm, Alice would like him more because they now have a height difference of 15 cm compared with the previous 10 cm, which is closer to her ideal height difference. However, Bob would like Alice less, because the new height difference is further away from his ideal height difference of 10 cm. In this case, the 175 cm tall Bob would be less likely to accept a proposal from Alice than the 170 cm tall Bob. Therefore, mismatched preferences are reflected in the fact that Bob's 5 cm height increase makes him prefer Alice less, whereas it makes Alice prefer him more.

To formally define preference mismatch, we use a stylized illustration in Figure 2. Without loss of generality, we consider a man (i.e., the focal user) searching for female candidates. We denote the focal user's attribute as x_M, the candidate's attribute as x_W, and the difference between them as $x = x_W - x_M$. Let us say that attribute x denotes height. The focal male user is 10 cm taller than a female candidate, such that $x = -10$ cm. This relative height x is the same for both the male and the female users. We denote the focal user's ideal preference for x as $P_{M,x}$ and the female candidates' ideal preference for x as $P_{W,x}$.[9] The closer x is to $P_{M,x}$ ($P_{W,x}$), the higher the utility of the focal user (the candidates). Preference mismatch between the two sides exists when (1) the ideal attribute point differs between the two sides, $P_{M,x} \neq P_{W,x}$ and (2) the value of attribute x falls in between $P_{M,x}$ and $P_{W,x}$.

Figure 2. (Color online) Preference Mismatch: Stylized Illustration



We explain why this is the case. Suppose there are four candidates with relative attribute values x_1, x_2, x_3, and x_4.[10] The attributes of candidate 1 and candidate 2 fall within the interval between $P_{M,x}$ and $P_{W,x}$, whereas candidates 3 and 4 appear in the region outside the interval. Preference mismatch exists between the focal user and candidates 1 and 2. Given candidates 1 and 2, the focal user always prefers candidate 1 to candidate 2 because x_1 is closer to the focal user's ideal point $P_{M,x}$. On the candidates' side, however, candidate 2 actually prefers the focal user more than candidate 1 because x_2 is closer to $P_{W,x}$ than x_1. Translating this to our context, the focal user is more likely to propose to candidate 1 than candidate 2. But candidate 2 is more likely than candidate 1 to accept a proposal from the focal user. The misaligned preferences evidenced by the choice differences between the two sides are defined as *preference mismatch*. This mismatch happens only if candidates' attribute value lies in between $P_{M,x}$ and $P_{W,x}$; candidates 3 and 4, whose attribute values x_3 and x_4 fall outside the interval ($P_{M,x}$, $P_{W,x}$), have aligned preferences because they are on the same side of the focal user's and candidates' ideal points. It is also worth mentioning that, in reality, the focal user would probably not consider candidates 3 and 4 at all because x_3 and x_4 are too far away from the ideal preference point $P_{M,x}$. When a focal man's selection criterion is to propose to women with x falling between the black dot (i.e., the left dot in Figure 2) and the black dashed line (i.e., the right dashed line in Figure 2), and female candidates' selection criterion is to accept men with x falling in between the red dot (i.e., the right dot in Figure 2) and the red dashed line (i.e., the left dashed line in Figure 2), candidate 1 would be the only match for the focal user, because she is located in the "Mr./Ms. Right" region, the shaded area in Figure 2.

In Online Appendix F, we provide a theoretical model to formally model preference mismatch and its interaction with information. This model serves as a supplementary explanation to the focal mechanism.

4.2. Existence of Preference Mismatch

In this section, we show empirical evidence of preference mismatch. First, we estimate the mate preference of each side. Once we get the preference estimates, we construct an *attractiveness score* for each candidate of a given focal user, which captures the overall subjective evaluation of the candidate by the focal user. We then show evidence of preference mismatch at both the user level (using the attractiveness score) and the user attribute level. To avoid contamination of results across genders, in most analyses, we focus on the male focal users in our empirical analysis. We also conduct robustness checks with female focal users, and the results remain qualitatively the same.

4.2.1. Mate Preference Estimation.

The estimation procedure follows Hitsch et al. (2010a). Specifically, a focal user's preference for potential matches is revealed through his proposing decisions: whether he sends messages or not after browsing the long profile pages of the candidates. To be precise, we use a subsample of candidates whose long profiles were viewed by the focal user. Once we obtain the preference parameters, we compute the attractiveness score for every candidate including the ones whose long profiles were not browsed by the focal user.

We denote by U_{ij} the latent utility of focal user i from matching with candidate j. As shown in Equation (1), it is a function of user characteristics, including continuous attributes such as age and height, which we denote by x, and also discrete attributes such as gender and lifestyle, which we denote by d. The term d_{im} (d_{in}) is the indicator for category m (n); $|x_i - x_j|_{+/-}$ denotes the attribute difference if the difference is positive/negative; and η_i controls the focal user's individual fixed effects. To model the proposing decision and avoid the potential incidental problem, we employ a linear probability regression with fixed effects (Equation (2)) instead of a fixed effect logit model. We assume users are not strategic when proposing a match.[11] Given that there are a large number of observed user attributes, we perform a lasso regression to identify attributes that are of statistical and economic significance and include them in the mate preference regression.[12]

$$U_{ij} = I\{x_j - x_i > 0\} \cdot \beta + (|x_j - x_i|_+^T) \cdot \gamma^+ + (|x_j - x_i|_-^T) \cdot \gamma^- + \sum_{m,n=1}^{N} I\{d_{im} = 1 \text{ and } d_{jn} = 1\} \cdot \theta^{mn} + \eta_i + \epsilon_{ij}, \quad (1)$$

$$Prob(i \text{ proposes to } j \,|\, \text{long profile of } j \text{ browsed}) = \overline{U}_{ij}. \quad (2)$$

Table 2 presents the estimation result. Overall, male focal users look for potential partners who are from the same city, around 10 cm shorter than themselves, and of similar age and education level. The effect of income is very interesting. Male focal users look for someone with a lower income level than themselves (as evident by the negative coefficient of $I\{income_j - income_i > 0\}$), but not too low (as evident by the negative coefficient of $|income_j - income_i|_-$). However, if a female candidate has an income level that is sufficiently high, the focal users will disregard their preference for candidates with lower income levels (as evident by the positive coefficient of $|income_j - income_i|_+$). Divorced focal users prefer to propose to candidates who are also divorced. Focal users who do not have children prefer candidates who do not have children either. Focal users also favor candidates with Chinese zodiac signs that best match theirs.[13] Those who own a house also prefer candidates who own a house, and those who do not own a house avoid candidates who do not own a house. The effect of lifestyle is not so informative because 70% of users leave

Table 2. Mate Preference Estimation

	(1)	
	Propose a match	
Age: elder	−0.0122***	(0.0018)
Age difference (+)	−0.0016***	(0.0003)
Age difference (−)	−0.0032***	(0.0002)
Degree: higher	0.0030	(0.0020)
Degree difference (+)	−0.0076***	(0.0011)
Degree difference (−)	−0.0019**	(0.0006)
From the same city	0.0082**	(0.0025)
Has photo	0.0017	(0.0012)
Height: higher than baseline	0.0035*	(0.0016)
Height difference (+)	−0.0029***	(0.0002)
Height difference (−)	−0.0033***	(0.0002)
Income: higher	−0.0093***	(0.0020)
Income difference (+)	0.0032***	(0.0007)
Income difference (−)	−0.0067***	(0.0007)
Marriage: both divorced	0.0175***	(0.0047)
Marriage: either single	−0.0052	(0.0029)
Have children: both	−0.0021	(0.0034)
Have children: neither	0.0223***	(0.0028)
Zodiac sign: best match	0.0069***	(0.0012)
Zodiac sign: worst match	−0.0001	(0.0014)
Religious: both	0.0107*	(0.0053)
Religious: neither	−0.0004	(0.0015)
Own house: both	0.0131***	(0.0032)
Own house: neither	−0.0114***	(0.0024)
Lifestyle: irregular	0.0003	(0.0011)
Lifestyle: unknown	0.0110***	(0.0011)
Provide looks rating	−0.0012	(0.0010)
Log # words in self-intro	−0.0009*	(0.0004)
Focal Fixed Effects	Yes	
Adjusted R^2	0.005	
Observations	484,742	

Notes. The proposing decision is conditional on viewing long profiles. The baseline of the height variable is that the female is 10 cm shorter than the male. Other omitted baseline categories are Marriage: both single; Have children: either; Zodiac sign: others; Religious: either; Own house: either; Lifestyle: regular. Standard errors are clustered at the focal user level.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

this information empty. Interestingly, focal users like candidates with shorter self-introductions.

4.2.2. Candidate Attractiveness Scores. Based on the mate preference estimates, we are able to construct attractiveness scores of candidates by summing over candidate attributes weighted by their coefficients estimated by Equation (2): $Q_{ij} = \sum_k \gamma_k attribute_{ijk}$. We then standardize the scores by dividing them by the maximum score from the candidate of the same focal user. This way, candidates of different focal users are comparable on the same scale, and all the scores are between zero and one. Candidates with an attractiveness score equal to one are those who are most favored by the focal users.

4.2.3. Preference Mismatch: Evidence from Attractiveness Score Distributions. With candidate attractiveness scores constructed, we can examine the mutual choices of the two sides. If the preferences of the two sides are aligned, that is, those who are favored by the focal users also favor the focal users back, then candidates with high attractiveness scores should be the ones that are most likely to reply. If the preferences are mismatched, then more attractive candidates will be less likely to reply.

Figure 3 shows the attractiveness score distributions of candidates who received match proposals and those who also replied when focal users' proposing decisions were based on (a) candidates' short profiles (i.e., partial info group) and (b) long profiles (i.e., complete info group). For both groups, the attractiveness score distributions of candidates who replied shift toward the left compared with all candidates who received match proposals. The two-sample Kolmogorov–Smirnov (KS) test scores of the distribution shifts in Figure 3, (a) and (b), are KS = 0.083 ($p < 0.001$) and KS = 0.16 ($p < 0.001$), respectively. That is, candidates who replied back were relatively lower in attractiveness score compared with all the candidates who received match proposals; the candidates who replied were usually not the focal users' original top choices. This demonstrates that the focal users and the candidates have different opinions on what their ideal matches are. It also echoes the earlier stylized illustration in that the replying candidates are not those who are near the focal users' ideal preference point.

4.2.4. Preference Mismatch: Evidence from Attribute-Level Preferences. We also find empirical evidence of preference mismatch at the attribute level. Specifically, we conduct regression analyses to test the preferences of the focal users and the candidates on various attributes. Here we take height as an example. In Table 3, we regress focal users' match proposal decisions on the relative height differences between the candidates and the focal users while controlling for the candidate attributes and the fixed effects of the focal users. The height difference in the table is the height of the female minus the height of the male. This definition holds in both columns. For male users, they favor females who are around 10 cm shorter than them, whereas females favor men who are 19 cm taller than them. The results reveal a mismatched preference in height difference ranging from −25 to −11 cm, where the coefficients of male focal users and female focal users in this region are of opposite signs. Female users prefer height differences in this region, whereas male users do not. One may argue that the focal females in Table 3 may not be the same group of females that the focal male users interacted with. To alleviate this concern, we conducted a robustness check where we used a subsample of female focal users who were also the candidates of the focal male users (see Table A.2 in Online Appendix C).[14] The results draw

Figure 3. (Color online) Candidate Attractiveness Score Distributions

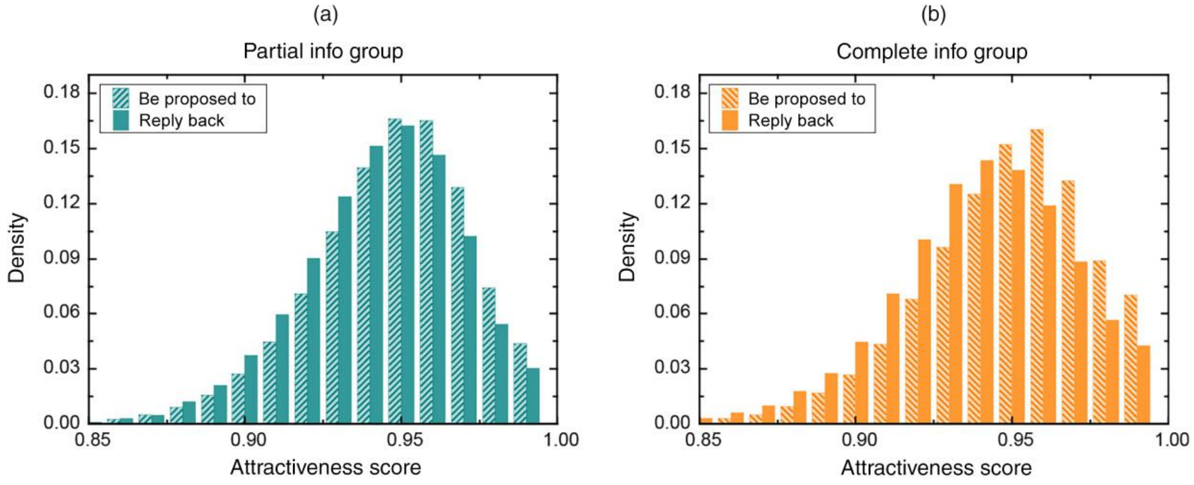


Table 3. Preference Mismatch: Height

	Propose a match			
	(1)		(2)	
Height difference (cm)	Male focal users		Female focal users	
−25	−0.0492***	(0.0062)	0.0041	(0.0099)
−24	−0.0414***	(0.0063)	0.0001	(0.0095)
−23	−0.0425***	(0.0043)	0.0142	(0.0075)
−22	−0.0394***	(0.0043)	0.0160*	(0.0068)
−21	−0.0333***	(0.0041)	0.0218**	(0.0070)
−20	−0.0217***	(0.0033)	0.0221***	(0.0049)
−19	−0.0232***	(0.0035)	0.0295***	(0.0055)
−18	−0.0146***	(0.0030)	0.0232***	(0.0044)
−17	−0.0136***	(0.0027)	0.0219***	(0.0045)
−16	−0.0109***	(0.0028)	0.0222***	(0.0044)
−15	−0.0098***	(0.0024)	0.0166***	(0.0040)
−14	−0.0032	(0.0026)	0.0184***	(0.0041)
−13	−0.0051*	(0.0024)	0.0184***	(0.0038)
−12	0.0032	(0.0022)	0.0100**	(0.0038)
−11	−0.0007	(0.0025)	0.0100*	(0.0040)
−9	0.0034	(0.0024)	0.0005	(0.0042)
−8	−0.0010	(0.0023)	−0.0112**	(0.0038)
−7	−0.0006	(0.0024)	−0.0059	(0.0044)
−6	−0.0026	(0.0028)	−0.0167***	(0.0048)
−5	−0.0016	(0.0025)	−0.0282***	(0.0046)
−4	−0.0071*	(0.0030)	−0.0334***	(0.0054)
−3	−0.0101***	(0.0030)	−0.0335***	(0.0055)
−2	−0.0161***	(0.0031)	−0.0411***	(0.0061)
−1	−0.0217***	(0.0036)	−0.0495***	(0.0067)
0	−0.0357***	(0.0026)	−0.0397***	(0.0049)
Candidate attributes	Yes		Yes	
Focal fixed effects	Yes		Yes	
Adjusted R^2	0.005		0.017	
Observations	484,742		156,605	

Notes. The proposing decision is conditional on viewing long profiles. The height difference in the table is the height of the female minus the height of the male. The baseline of the height difference is that the female is 10 cm shorter than the male. The controlled candidate attributes are the same set of attributes as in the mate preference estimation. Standard errors are clustered at the focal user level.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

similar conclusions on the existence of preference mismatch in height.

Apart from height, preference mismatch also exists in other attributes such as age (Table A.3, Online Appendix C), income (Table A.4, Online Appendix C), etc. We include this evidence in Online Appendix C. In fact, preference mismatch is very common in life and not only exists in marriage and dating markets, but also prevails in many other matching contexts such as school choice, student–teacher matching, and job searches.

5. The Role of Information Under Preference Mismatch

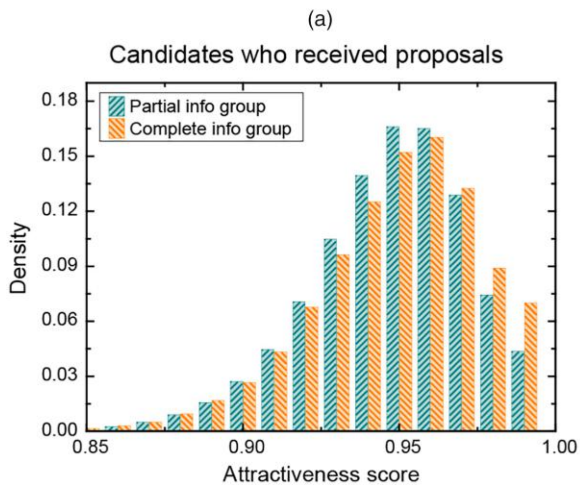
Given the existence of preference mismatch, should the online dating platform provide more or less information to their users? In product marketplaces, the situation is much simpler because only one side makes decisions; given complete information disclosure of product information, the consumers choose the best alternative among all options, which ends the decision journey. In contrast, in two-sided matching markets, including the online dating context that we study, the decision consumers make at the beginning does not necessarily lead to the final match, because they also need to wait for the other side's decision. With the existence of preference mismatch, knowing more information about the other side is not necessarily better. The intuition is as follows: if the focal user knows more information about the candidates, he will select candidates who are closer to his ideal preference (i.e., those who are Ms. Best), and candidates who are in the Ms. Right region will be ruled out. This is because, more, compared with less, information about the other side leads to stronger preference mismatch. Therefore, we expect that the degree of preference mismatch between the focal user and the candidates in the complete info group is stronger than that between the focal user and the candidates in the partial info group. And given the same focal user, making a decision to propose based on information contained in the short profile versus the long profile results in a better matching outcome (i.e., the "less information is more" effect).

5.1. Testing the Role of Information

To test whether the degree of preference mismatch is stronger in the complete info group than in the partial info group, we compare the attractiveness score distributions of the two groups. In Figure 4(a), we plot the attractiveness score distributions of candidates who received match proposals. Overall, the complete info group has a higher attractiveness score compared with the partial info group (KS = 0.046, $p < 0.001$). This is as expected because the more information that is provided, the more favored candidates the focal users will propose to. The result implies that the partial info group candidates would be less likely to receive match proposals if their complete information were revealed to the focal users in the beginning. Taking a look at the candidates' side, among those who replied, the partial info group candidates were actually more attractive to the focal users than those from the complete info group (KS = 0.06, $p < 0.001$; see Figure 4(b)). This means that the degree of preference mismatch between the focal users and the candidates in the complete info group is much stronger than that between the focal users and the candidates in the partial info group. In sum, as expected, more (versus less) information about the other side does lead to stronger preference mismatch. When more attributes are observed by the focal users, misaligned preferences in each attribute begin to weigh in, hence, increasing the level of mismatch.

Taking advantage of the information structure of the site, we examine whether more match-relevant

Figure 4. (Color online) Attractiveness Score Distributions of Candidates Who Received Proposals



information about the candidates helps the focal user find a match. We regress matching outcomes ($Match_{ij}$) on whether a focal user i obtained more information about a candidate j, conditional on the focal user proposing a match (i.e., sending a message to the candidate). The variations in the information amount (partial versus complete) come from the attribute information contained in the short versus long profile.

The term $Match_{ij}$ is the variable that captures whether the focal user i and the candidate j are successfully matched. Because we do not observe their offline activities, nor do we know the content of their messages, we adopt the platform's definition and use the number of messages exchanged between the two sides as a proxy for the likelihood of matching.[15] We define $Match_{ij}$ as a binary variable where $Match_{ij} = 1$ if the frequency of message exchange is over a certain cutoff number, and otherwise $Match_{ij} = 0$. In Table 4, we report the results when the cutoff number equals one. We also change the cutoff number to two, three, and four (see Tables A.5 and A.6 in Online Appendix D) and treat the dependent variable as a continuous variable and as a count variable (see Table A.7 and A.8 in Online Appendix D). The similar results hold.

Table 4 summarizes the results. We run the same set of models for both the male focal users (columns (1) to (3)) and the female focal users (columns (4) to (6)). We test whether searching for more information about a candidate brings a better outcome for a focal user, conditional on proposing a match to the candidate (i.e., sending a message to the candidate). First, we regress the matching outcomes on the action of browsing candidates' long profiles, controlling for focal user fixed effects (columns (1) and (4)). The results show that the candidates that a focal user proposed to based on the information shown on the candidates' short profiles are more likely to reply compared with the candidates that the focal user proposed to after browsing their long profiles.

The effect might be driven by the popularity of the candidates or candidates' profile photos. For example, maybe candidates that a focal user proposed to without

Table 4. The Role of Information on Matching Outcomes

	Male focal users			Female focal users		
	(1)	(2)	(3)	(4)	(5)	(6)
Obtaining more information	−0.0539***	−0.0482***	−0.0486***	−0.0947***	−0.0727***	−0.0737***
	(0.0020)	(0.0020)	(0.0020)	(0.0051)	(0.0056)	(0.0056)
Choice environment (cand.)			−0.0050***			−0.0027
			(0.0009)			(0.0016)
Choice environment (focal)			−0.0004***			−0.0013***
			(0.0001)			(0.0002)
Candidate fixed effects	No	Yes	Yes	No	Yes	Yes
Focal fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.003	0.326	0.326	0.005	0.369	0.370
Observations	479,987	340,933	340,933	144,162	73,854	73,854

Notes. Standard errors are clustered at the focal user level for columns (1) and (4) and at both the focal user and candidate levels for the other columns.
***$p < 0.001$.

browsing their long profiles are also the ones who are more likely to reply, or maybe the focal user's proposals are solely based on candidates' profile photos and those who look less attractive are more likely to reply.[16] To alleviate these concerns, we control for candidate fixed effects in columns (2) and (5) of Table 4. The results rule out popularity or photos of candidates as a major concern. As another robustness check, we replace candidate fixed effects with demographic differences between the candidates and the focal user (see Table A.9 in Online Appendix D).

Furthermore, matching outcomes may be influenced by the dynamic choice environment of the two sides. To proxy for the time-varying choice environment of each user, we count the number of ongoing communications a user has at time t when she or he is making messaging decisions. We denote the relevant variables as *Choice environment (cand.)* and *Choice environment (focal)*, and include them in the regression model as control variables (columns (3) and (6)).

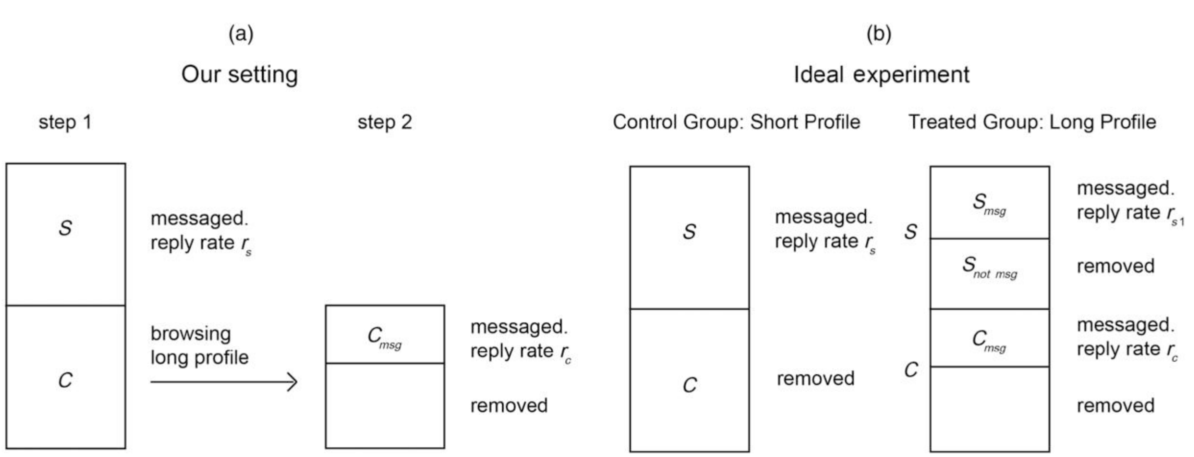
Across the three sets of regression models in Table 4, including the robustness checks we mentioned earlier, the results consistently corroborate the "less information is more" effect. It suggests that a focal user has a higher chance of getting matched if he only sees partial instead of complete information of the candidates.

5.1.1. Possible Selection Issue: Our Setting vs. the Ideal Experiment. One may argue that because the short versus long profiles are not exogenously given to the users, results in Table 4 may suffer from the issue of self-selection. In this section, we compare our setting to a hypothetical ideal experiment and discuss the extent to which such a selection might influence the results. We prove that the conclusion drawn from the ideal experiment should be qualitatively the same as the conclusion drawn from our empirical context, conditional on the users' decision rules and behaviors in our context.

First, we describe in detail where the selection might come from in our setting (Figure 5(a)). Upon seeing the short profiles of a list of candidates, a focal user forms a judgment on the *expected attractiveness* of each candidate based on the partial information contained in each candidate's short profile. The focal user sends messages directly to a group of candidates (denoted as S) with higher expected attractiveness, and checks the long profiles of candidates (denoted as C) with lower expected attractiveness. This decision rule is supported by the fact that the cost of sending a message, which includes both time and paying a fee to the platform, is much higher than browsing a long profile. After checking the long profiles of candidates C, the focal user can then evaluate the attractiveness scores of the candidates based on their complete information. The focal user will send messages to some of them (denoted by C_{msg}) with higher attractiveness scores.[17] The regression test in Table 4 essentially compares the message reply decision of those in S and those in C_{msg}. There is a selection concern because the focal user endogenously chooses to browse the long profiles of C and send messages to C_{msg}. Additionally, there is a group of candidates to which the focal user neither browses the long profiles nor sends a message. Because these candidates are the lowest in expected attractiveness and are immediately rejected, we do not consider these candidates in the following discussion. In our data, we observe the proportion of candidates who replied, and we denote the reply rates by r_s and r_c for group S and C_{msg}, respectively. Results in Table 4 show that $r_s > r_c$.

We now describe a hypothetical experiment setting where profiles with different amounts of information (long profile versus short profile) are randomly shown to the focal user, and the focal user makes proposal decisions based on the information he observes (Figure 5(b)). In this scenario, candidates' characteristics should be the same in the short profile (control group) and long profile (treated group) conditions.

Figure 5. Our Setting vs. Ideal Experiment



For the control group, similar to our setting, candidates in S still receive match proposals. However, candidates in C are removed from consideration because there is no option for the focal user to browse the candidates' long profiles. Candidates in C have lower expected attractiveness, and thus do not satisfy the focal user's match proposal criteria. Therefore, the reply rate of the control group $r_{control}$ is still equal to r_s, which is the reply rate of the partial info group in our empirical setting.

Next, we derive the reply rate r_{treat} of the treated group in the ideal experiment. A similar pool of candidates is in this group, which we also denote by S and C. Consistent with the empirical setting, when the long profiles of candidates in C are revealed, the focal user will send messages to candidates C_{msg}. The reply rate of these candidates is r_c. Candidates in S are those who would receive match proposals if their short profiles were revealed to the focal user. We indicate in Section 4.2.3 that the subjective attractiveness rankings of candidates in C_{msg} are higher than the rankings of candidates in S when their long profiles are revealed. So for the treated group, given that C_{msg} all received messages and their attractiveness scores are higher than those in S, only a fraction of candidates in S will receive the match proposals. We denote those candidates with higher attractiveness scores who receive proposals by S_{msg} and those with lower attractiveness scores who do not receive proposals by $S_{not\ msg}$. In order to compare the reply rates of r_{treat} and $r_{control}$, we assume candidates in $S_{not\ msg}$ are also given match proposals. We then have the equation $S \times r_s = S_{msg} \times r_{s1} + S_{not\ msg} \times r_{s2}$, because candidates' reply decisions do not depend on whether the focal user observes their short or long profiles. Here, r_{s1} is the reply rate of candidates S_{msg}, and r_{s2} is the reply rate of candidates $S_{not\ msg}$ if they were proposed to.

To compare r_{s1} and r_{s2}, we need to borrow another observation we find in our data (Section 4.2.3), that candidates of a higher attractiveness score have a lower reply rate (i.e., the existence of preference mismatch). As noted earlier, attractiveness scores are computed based on the complete information of the candidates, so $r_{s1} < r_{s2}$. Given $S \times r_s = S_{msg} \times r_{s1} + S_{not\ msg} \times r_{s2}$, we have $r_{s1} < r_s < r_{s2}$.

Now we have shown that $r_c < r_s$ and $r_{s1} < r_s$. Finally, we can show the reply rate of the treated group: $r_{treat} = \frac{S_{msg} \times r_{s1} + C_{msg} \times r_c}{S_{msg} + C_{msg}} < \frac{S_{msg} \times r_s + C_{msg} \times r_s}{S_{msg} + C_{msg}} = r_s = r_{control}$. To summarize, if we were to have the opportunity to conduct an ideal experiment where we could randomly manipulate the length of the user profile, the conclusion drawn from the ideal experiment would be the *same* as the conclusion drawn from our current empirical analysis: the matching outcome is better if less information is shown to the focal user.

To completely alleviate this concern, we conduct online experiments and simulations to exogenously provide different amounts of information to the focal users. The results replicate the "less information is more" effect. Detailed experimental and simulation procedures can be found in Online Appendices E and G, respectively.

5.2. The Lasting Effect of Preference Mismatch

Table 4 shows that preference mismatch impacts the overall matching outcomes measured by the total number of message communications. To understand the effects of preference mismatch on each round of communications, we look at the reply decisions of focal users and candidates separately. Specifically, Table 5 examines the focal user and the candidate's message reply decisions on whether the candidate is from the partial info or the complete info group. The results show that, conditional on a focal user sending the first contact message, preference mismatch not only affects the candidate's first reply decision (column (1)) and the focal user's first reply decision (column (2)), but also affects their subsequent reply decisions (columns (3) and (4)). The stylized theoretical model in Online Appendix F can also be easily extended to account for the multiround reply decisions and make predictions that are consistent with the empirical evidence in Table 5.

6. Discussions

6.1. Heterogeneous Goals

One potential concern is that different focal users may have different goals, which may cause a self-selection problem. For example, those who are not looking for

Table 5. Multiround Message Reply Decisions

	(1) Candidate's 1st reply	(2) Focal's 1st reply	(3) Candidate's 2nd reply	(4) Focal's 2nd reply
If partial info group	0.0551***	0.1079***	0.0869***	0.0935***
	(0.0019)	(0.0106)	(0.0151)	(0.0221)
Candidate attributes	Yes	Yes	Yes	Yes
Focal fixed effects	Yes	Yes	Yes	Yes
Adjusted R^2	0.026	0.010	0.010	0.003
Observations	479,987	47,350	22,854	12,384

Notes. Here the controlled candidate attributes are the same set of attributes as in the mate preference estimation. Standard errors are clustered at the focal user level.
***$p < 0.001$.

marriage or a serious relationship may be more attractive and may be more likely to send a message without reading long profiles. Nevertheless, our data show that most focal users had both the following behaviors: sending a message after reading the short profiles of some candidates, and sending a message after reading the long profiles of some candidates. The information role is identified using the within focal-user variations in the profile-checking behaviors. User heterogeneity such as goal differences is controlled by individual fixed effects.

To further address the concern, we run our main regression of information role using subsamples of different focal users based on their heterogeneous proposing behaviors. The assumption is that the heterogeneous proposing behaviors are the results of goal differences (if any). Specifically, we categorize focal users into different groups based on their heterogeneity in the proposing method (Table 6, panel A) and proposal frequency (Table 6, panel B). More precisely, if a focal user proposes based on long profiles more than 50% of the time, then he belongs to the group of "more with complete info" (columns (1), (3), (5), (7) in panel A). Otherwise, he belongs to the "more with partial info" group (columns (2), (4), (6), (8) in panel A). In panel B, we median split the focal users into "top half" (columns (1), (3), (5), (7) in panel B) and "bottom half" (columns (2), (4), (6), and (8) in panel B) based on how frequently they propose throughout the observation window on the platform. Results in Table 6 replicate our main findings of the information role across heterogeneous focal user groups, ruling out goal differences as an alternative explanation.

6.2. Candidate Popularity and Pickiness

Another alternative explanation for the "less information is more" effect might be that candidates whose long profiles are searched also turn out to be the ones who are more popular or who have a higher reservation value (i.e., they are pickier). Hence, the more popular and pickier candidates are less likely to reply. Our previous regression results controlled for the candidate fixed effects, so any unobserved heterogeneity across candidates, such as different levels of popularity or reservation value, were already accounted for.

To further reduce the concern of this alternative explanation, we test whether candidates in the complete info group are indeed more popular. First, we construct a popularity measure for each candidate, which is proxied by the total number of focal users who proposed to the candidate. For a given focal user, we compute the average popularity of candidates who are in the partial info group and those who are in the complete info group. Using a logit regression (Table 7, column (1)), we show that the average popularity of the candidates in the complete info group is actually lower, rather than higher, compared with the candidates in the partial info group. For each candidate, we also compute her probability of being in the complete info group.[18] We then regress this probability on the candidate's popularity.

Table 6. The Role of Information on Matching Outcomes in Different Groups

Panel A. More likely to propose based on complete info vs. partial info								
	Male focal users				Female focal users			
	(1) More with complete info	(2) More with partial info	(3) More with complete info	(4) More with partial info	(5) More with complete info	(6) More with partial info	(7) More with complete info	(8) More with partial info
Obtaining more info	−0.0635***	−0.0421***	−0.0585***	−0.0370***	−0.1104***	−0.0746***	−0.0910***	−0.0559***
	(0.0028)	(0.0027)	(0.0029)	(0.0035)	(0.0050)	(0.0087)	(0.0088)	(0.0109)
Candidate fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Focal fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.005	0.002	0.313	0.347	0.007	0.003	0.373	0.369
Observations	239,927	240,060	139,005	139,257	73,787	70,375	26,906	26,960
Panel B. Proposing more frequently vs. less frequently								
	(1) Top half	(2) Bottom half	(3) Top half	(4) Bottom half	(5) Top half	(6) Bottom half	(7) Top half	(8) Bottom half
Obtaining more information	−0.0509***	−0.0578***	−0.0475***	−0.0510***	−0.0877***	−0.1042***	−0.0720***	−0.0676***
	(0.0030)	(0.0024)	(0.0032)	(0.0032)	(0.0077)	(0.0055)	(0.0093)	(0.0098)
Candidate fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
Focal fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.003	0.003	0.338	0.320	0.004	0.005	0.358	0.382
Observations	239,712	240,275	139,817	139,997	72,117	72,045	27,066	27,055

Notes. Standard errors are clustered at the focal level for columns (1), (2), (5), and (6) and at both the focal and candidate levels for columns (3), (4), (7), and (8).
$^{***}p < 0.001$.

Table 7. Effect of Candidate Popularity

	(1)		(2)	
	Complete info group		Probability of being in complete info group	
Average popularity	−0.1171***	(0.0038)		
Candidate popularity			−0.0528***	(0.0006)
Constant	0.6267***	(0.0193)	0.6208***	(0.0011)
Pseudo/adjusted R^2	0.0384		0.086	
Observations	47,552		330,137	

Note. Standard errors are in parentheses.
$^{***}p < 0.001$.

The result again shows that candidates with a lower popularity score are more likely to be in the complete info group (Table 7, column (2)). Together these results rule out candidate popularity as an alternative explanation.

6.3. Manipulation of Short Profiles

Another concern is the manipulation of short profiles as a whole because candidates who are eager to find a match may polish their short profiles to make them very attractive. They are also more likely to reply given their desire to be matched sooner rather than later. However, we argue that this is unlikely to affect the results of the analysis for two reasons. First, there is not much room to manipulate or polish a short profile. It contains only a few pieces of information about the user, including nickname, profile photo, age, and home city. Apart from the photo, all other fields are required, and the users cannot design their short profiles. Second, the only flexible factor in the short profile is the profile photo, which we have discussed in Section 5.1. Related to this, users may learn about the usefulness of profiles overtime. We test and exclude consumer learning as an alternative explanation in Online Appendix H.

6.4. Inferring Long Profiles from Short Profiles

Now we discuss a case where some information in the long profiles can be inferred from the short profiles. The current premise of the results is based on the different amounts of information that are uncovered in the short profiles; more information in the long profiles results in a higher level of preference mismatch. Now let us assume some information in the long profiles can be inferred from the short profiles. In this scenario, the difference in the amount of information uncovered from the two profiles will be smaller compared with the benchmark case where no information can be inferred from the short profiles. The smaller difference will consequently lead to a lower level of preference mismatch. This would mean that our current result of the role of information under preference mismatch is underestimated if people can infer some information from the short profiles. That is, the "less information is more" effect should be even more pronounced in the benchmark case where no inference is being made.

6.5. Boundary Conditions

In order to apply our results to other two-sided matching settings where preference mismatch might exist, we need to acknowledge the possible boundary conditions.

First, different platforms have different matching mechanisms. A centralized matching platform like Uber assigns drivers to passengers in real time. Passengers care more about time efficiency than their preference over other dimensions, such as car type or a driver's demographics. Given that drivers do not know the destination of a ride before accepting the ride order, their preferences over destination or passengers usually do not matter. Therefore, the effect of preference mismatch should be negligible in these types of settings. Our results are more applicable to decentralized matching platforms, like Tinder, Upwork, and Airbnb, where user preferences play a key role in the search process. For example, on job-matching platforms, freelancers and clients need to search for each other before the final match is made, and they may have different preferences over hourly rate, job difficulty, and work quality; thus, preference mismatch will play an important role.

Second, users on various matching platforms also differ in mismatch cost, the disutility of not matching with candidates at the ideal point. When mismatch cost is relatively low (e.g., when you want to find a part-time job online), the "Mr./Ms. Right" region in our stylized illustration (Figure 2) should be bigger compared with settings where mismatch cost is relatively high (e.g., in online dating or the marriage market), meaning there are more potential matches. Additionally, with lower mismatch cost, there will be fewer candidates who will be ruled out by the presence of more information. Together, we expect that our results should be attenuated in contexts with lower mismatch cost.

Third, our results are also subject to the bargaining power between the two sides. If the two sides have highly imbalanced bargaining power, then the final matching outcome will depend more on the preference of the side with the stronger bargaining power. The effect of preference mismatch will be much attenuated.

Last, we discuss how communication cost affects our results. Assuming zero communication cost, the best strategy is to send proposals to all candidates without browsing their profiles and communicate with them repeatedly. After all the user characteristics including the unobservables, such as personality, are revealed, focal users can finally make their optimal matching decisions. However, in reality, communication cost is not negligible. To save cost, users should obtain some information about the other side to exclude candidates who are far from the matching criteria. There is a trade-off between communication cost and preference mismatch, implying an optimal amount of information needed before proposing a match. For example, on Airbnb, communication cost is relatively high, especially when deadlines are imminent and guests and hosts may not always be available online. To expedite the matching process in such scenarios, guests may need more upfront information to reduce communication time. When communication cost is extremely high that it dominates the preference mismatch effect, the "less information is more" effect will diminish.

7. Conclusion

In this paper, we study the role of information under preference mismatch in two-sided matching markets. Given the ubiquitous existence of preference mismatch and the pivotal role of information, it is important to understand their interactions.

We focus on an empirical context of online dating, which is characterized by a high degree of heterogeneity in terms of both user characteristics and their preferences of potential partners. We find that platform users—who in many situations are on the disadvantaged side of information asymmetry—will be better off when obtaining less information about the other side. This effect is driven by the mismatched preferences between the two sides. We show evidence of preference mismatch by using the attractiveness scores of candidates. We also find that preference mismatch exists at the attribute level. More information about the other side leads to stronger preference mismatch. Subsequently, the stronger level of preference mismatch will lead to focal users' selecting candidates who are less likely to accept them (i.e., the "best" ones), thus ruling out potential candidates (i.e., the "right" ones) too early. The findings imply that there exists an optimal amount of information that one side should know about the other side prior to a matching proposal.

7.1. Managerial Implications

Our study provides insights into how the amount of information available to each side affects matching outcomes on two-sided platforms and offers guidance on information design strategies. Specifically, online dating platforms should consider providing only a partial set of information to users prior to proposing a match. For example, apps like Tinder display only a limited amount of information on each user, such as profile photos, brief bios, and location, before allowing the user to like or dislike the profile. Our research suggests that this limited approach is not just for brevity and that limiting information improves the probability of matching, thereby increasing the value of the platform. This could be a reason why Tinder outperforms competitors like OkCupid in terms of matching effectiveness.[19] In terms of user profile design, we recommend that short profiles encompass attributes with little or no preference mismatch. This will enable both parties to make effective proposing and reply decisions without missing out on potential partners. Online dating platforms should also measure the mismatch cost of each user attribute on both sides to inform their information design strategies. Additionally, our findings are not confined to dating websites and can be extended to other matching platforms, such as Airbnb and Upwork, where misaligned preferences can exist between the two sides. However, it is important to recognize that the optimal amount of information provided on platforms like Airbnb and Upwork should be greater than that on online dating platforms because of their relatively higher communication costs. Beyond information design, an additional practical challenge for online dating sites is dating algorithm design. Only 21% of adults in the United States believe that the algorithms used by dating sites are effective in finding matches.[20] This indicates that dating algorithms have ample room for improvement by considering preference mismatch, as suggested by our research.

7.2. Possible Future Extensions

Our research raises many interesting questions that can serve as avenues for future research. First, we specified the type of information that user profiles should contain. An interesting question for future study is, if users can design their own profiles, what effects would it have on the optimal matching outcomes if users could choose what information to disclose or hide? Second, our findings suggest that females should initially receive less information than males, given their higher selectivity. Further examination of the effect of user heterogeneity on optimal information design strategies is warranted. The interplay between recommendation algorithms and preference mismatch and their impact on platform information design strategies is another area that carries potential for exploration: should dating websites recommend the "best" or "right" matches to their users? Future investigations might also scrutinize how preference mismatch integrates with advanced machine learning algorithms utilized in dating and other forms of matching platforms. Last, this study is founded under a rational framework where uniqueness seeking, social conformity, or superstitions may impact one's rational preferences and thus the effect of preference mismatch. Future research can explore the effects of these psychological factors.

Acknowledgments
The authors are grateful to the review team for their insightful guidance, and thankful to Jinzhao Du, Jin Li, Xi Li, and the participants at seminars for comments and suggestions. Hongchuan Shen and Chu (Ivy) Dang contributed equally to this work.

Endnotes
[1] There is an envelope icon for users to click to send messages.

[2] See https://theblog.okcupid.com/online-dating-advice-optimum-message-length-8a2887c3d6ca.

[3] See https://www.businessofapps.com/data/dating-app-market/.

[4] The site is designed for heterosexual adults.

[5] The firm unfortunately did not provide us with screenshots of short and long profiles in 2011. We provide a screenshot of the current website in Online Appendix A. It shows the difference in information amount contained in short and long profiles, though the attributes may not be exactly the same as in the 2011 version.

[6] The platform charges outgoing but not incoming messages (those sent by the focal users but not those sent in reply by the candidates). The users can spend "red beans," the virtual currency on the platform, to send messages. The red beans can be acquired by either completing simple tasks (e.g., logging into the platform) or by directly spending money to purchase them. The monetary cost of message sending is relatively low given the income level at that time.

[7] This measure is used by the platform as a key indicator of the effectiveness of matching.

[8] We observe the action records of messaging, but because of the company's privacy policy, we do not observe the content of the messages.

[9] To be more specific, $P_{W,x}$ is the average ideal preference for x of female candidates.

[10] In the remainder of the section, we will use "attribute" to denote "relative attribute" when there is no confusion.

[11] This dating website does not provide candidate popularity information. So strategic shading behavior due to rejection concerns is unlikely to happen here (Bojd and Yoganarasimhan 2022). Our setting is more similar to the decentralized dating website studied by Hitsch et al. (2010a), where there was no evidence of strategic behavior of users.

[12] In the lasso regression, the dependent variable is the total number of messages received by the focal user, and the covariates are all the observable attributes of the same user.

[13] The best-matched zodiac signs are well known in Chinese culture according to the Chinese zodiac astrology known as Sheng Xiao or Shu Xiang.

[14] This ensures female candidates who were selected by the male focal users in column (1) of Table A.2 (Online Appendix C) are the same group of female focal users in column (2) of Table A.2.

[15] One indirect support of this definition comes from the findings in Hitsch et al. (2010a), where multiple rounds of email communications were found to be a strong indicator of the final match. We do, however, acknowledge the limitation of the current definition because we were not able to observe offline interactions or long-term outcomes such as marriage because of the privacy protections of the platform.

[16] As mentioned in the empirical part, the platform did not share the profile photos of their users with us. We know only whether a user had a profile photo or not, and their self-reported looks rating, which most candidates did not provide.

[17] Note that "expected attractiveness" is formed based on the partial information in the short profile, and the "attractiveness score" is based on the long profile. Expected attractiveness and attractiveness score can be different values because different amounts of information are used to derive them. Also, both of the values are subjective measures, meaning different focal users may have different preferences toward user attributes.

[18] For example, if a candidate received one proposal that was sent from a user who had seen only her short profile and two other proposals sent from other focal users who had seen her long profile, then the probability of her being in the complete info group is 2/3.

[19] See https://beyondages.com/okcupid-vs-tinder/.

[20] See https://www.pewresearch.org/fact-tank/2023/02/02/key-findings-about-online-dating-in-the-u-s/ft_2023-02-02_key-findings-online-dating_08/.

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

Becker GS (1973) A theory of marriage: Part I. *J. Political Econom.* 81(4):813–846.

Bojd B, Yoganarasimhan H (2022) Star-cursed lovers: Role of popularity information in online dating. *Marketing Sci.* 41(1):73–92.

Bruch E, Feinberg F, Lee KY (2016) Extracting multistage screening rules from online dating activity data. *Proc. Natl. Acad. Sci. USA* 113(38):10530–10535.

Burtch G, Ghose A, Wattal S (2016) Secret admirers: An empirical examination of information hiding and contribution dynamics in online crowdfunding. *Inform. Systems Res.* 27(3):478–496.

Caldieraro F, Zhang JZ, Cunha M Jr, Shulman JD (2018) Strategic information transmission in peer-to-peer lending markets. *J. Marketing* 82(2):42–63.

Chade H, Eeckhout J, Smith L (2017) Sorting through search and matching models in economics. *J. Econom. Literature* 55(2):493–544.

Dinerstein M, Einav L, Levin J, Sundaresan N (2018) Consumer price search and platform design in internet commerce. *Amer. Econom. Rev.* 108(7):1820–1859.

Du J, Lei Y (2022) Information design of matching platforms when user preferences are bidimensional. *Production Oper. Management* 31(8):3320–3336.

Economist (2018) The irresistible rise of internet dating (August 17), https://www.economist.com/graphic-detail/2018/08/17/the-irresistible-rise-of-internet-dating.

Einav L, Farronato C, Levin J (2016) Peer-to-peer markets. *Annual Rev. Econom.* 8:615–635.

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

Fong J (2023) Search, selectivity, and market thickness in two-sided markets: Evidence from online dating. Preprint, submitted June 29, https://dx.doi.org/10.2139/ssrn.3458373.

Fradkin A (2017) Search, matching, and the role of digital marketplace design in enabling trade: Evidence from Airbnb. Preprint, submitted March 23, https://dx.doi.org/10.2139/ssrn.2939084.

Fradkin A, Grewal E, Holtz D (2021) Reciprocity and unveiling in two-sided reputation systems: Evidence from an experiment on Airbnb. *Marketing Sci.* 40(6):1013–1029.

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

Halaburda H, Piskorski MJ, Yıldırım P (2018) Competing by restricting choice: The case of matching platforms. *Management Sci.* 64(8):3574–3594.

Hickey W (2013) Here's how many messages men have to send to women on a dating site to be sure of getting a response. *Bus.*

Insider (July 17), https://www.businessinsider.com/online-dating-message-statistics-2013-7.

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

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

Horton JJ (2019) Buyer uncertainty about seller capacity: Causes, consequences, and a partial solution. *Management Sci.* 65(8):3518–3540.

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

Kanoria Y, Saban D (2021) Facilitating the search for partners on matching platforms. *Management Sci.* 67(10):5990–6029.

Kim K, Park J, Pan Y, Zhang K, Zhang X (2022) Risk disclosure in crowdfunding. *Inform. Systems Res.* 33(3):1023–1041.

Lee S, Niederle M (2015) Propose with a rose? Signaling in Internet dating markets. *Experiment. Econom.* 18(4):731–755.

Lu Y, Gupta A, Ketter W, Van Heck E (2019) Information transparency in business-to-business auction markets: The role of winner identity disclosure. *Management Sci.* 65(9):4261–4279.

Romanyuk G, Smolin A (2019) Cream skimming and information design in matching markets. *Amer. Econom. J. Microeconom.* 11(2):250–276.

Rosenfeld MJ, Thomas RJ, Hausen S (2019) Disintermediating your friends: How online dating in the United States displaces other ways of meeting. *Proc. Natl. Acad. Sci. USA* 116(36):17753–17758.

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

Shi Z, Srinivasan K, Zhang K (2023) Design of platform reputation systems: Optimal information disclosure. *Marketing Sci.* 42(3):500–520.

Stiglitz JE (2000) The contributions of the economics of information to twentieth century economics. *Quart. J. Econom.* 115(4):1441–1478.

Tadelis S, Zettelmeyer F (2015) Information disclosure as a matching mechanism: Theory and evidence from a field experiment. *Amer. Econom. Rev.* 105(2):886–905.