Quant Mark Econ (2010) 8:393–427
DOI 10.1007/s11129-010-9088-6

What makes you click?—Mate preferences in online dating

Günter J. Hitsch · Ali Hortaçsu · Dan Ariely

Received: 28 September 2009 / Accepted: 12 April 2010 / Published online: 20 May 2010

Abstract We estimate mate preferences using a novel data set from an online dating service. The data set contains detailed information on user attributes and the decision to contact a potential mate after viewing his or her profile. This decision provides the basis for our preference estimation approach. A potential problem arises if the site users strategically shade their true preferences. We provide a simple test and a bias correction method for strategic behavior. The main findings are (i) There is no evidence for strategic behavior. (ii) Men and women have a strong preference for similarity along many (but not all) attributes. (iii) In particular, the site users display strong same-race preferences. Race preferences do not differ across users with different age,

Electronic supplementary material The online version of this article (doi:10.1007/s11129-010-9088-6) contains supplementary material, which is available to authorized users.

Note that previous versions of this paper ("What Makes You Click?—Mate Preferences and Matching Outcomes in Online Dating") were circulated between 2004 and 2006. Any previously reported results not contained in this paper or in the companion piece Hitsch et al. (2010) did not prove to be robust and were dropped from the final paper version.

G. J. Hitsch (✉)
Booth School of Business, University of Chicago,
5807 South Woodlawn Avenue, Chicago, IL 60637, USA
e-mail: guenter.hitsch@chicagobooth.edu

A. Hortaçsu
Department of Economics, University of Chicago,
1126 East 59th Street, Chicago, IL 60637, USA
e-mail: hortacsu@uchicago.edu

D. Ariely
Fuqua School of Business, Duke University, 1 Towerview Drive,
Box 90120, Durham, NC 27708-0120, USA
e-mail: dandan@duke.edu

income, or education levels in the case of women, and differ only slightly in the case of men. For men, but not for women, the revealed same-race preferences correspond to the same-race preference stated in the users' profile. (iv) There are gender differences in mate preferences; in particular, women have a stronger preference than men for income over physical attributes.

Keywords Mate preferences · Dating · Marriage

JEL Classification C78 · J12

1 Introduction

The measurement of mate preferences has a long history across fields as diverse as economics, sociology, psychology, anthropology, and evolutionary biology (see, e.g., Buss 1995; Etcoff 1999; Finkel and Baumeister 2010 for extensive surveys). We contribute to this literature using a novel data set obtained from an online dating service. We utilize detailed information on the partner search behavior of the site users to infer their revealed mate preferences. Our data allow us to estimate a rich preference specification that takes into account a large number of partner attributes, including detailed demographic and socioeconomic information, physical characteristics, and other information such as religion and political views.

Our data allow us to address three important questions the literature raises on mate preferences and marriage. The first question (see, e.g., Kalmijn 1998) is whether people prefer others who are similar to them or whether there is agreement among people regarding the desirability of mates. We call mate characteristics that are valued differently depending on own traits *horizontal attributes*, and characteristics that are agreed upon *vertical attributes*. Both types of preferences can lead to empirically observed assortative mating patterns (Becker 1973; Browning et al. 2008; Kalmijn 1998) and are thus indistinguishable using data on marriages only. For example, sorting on educational attainment (highly educated women date or marry highly educated men) may be the result of a preference for a mate with a similar education level. Alternatively, the same outcome can arise in equilibrium (as a stable matching) in a market in which all men and women prefer a highly educated partner over a less educated one.

The second question, which is closely related to the first question, specifically concerns the existence of same-race preferences. In the USA, marriages across ethnic groups are relatively rare, which could be due to same-race preferences or, alternatively, the fact that members of one ethnic group predominantly meet other members of the same ethnic group (Kalmijn 1998; Fisman et al. 2008; Hitsch et al. 2010). Our study provides new insights into the currently existing literature on estimating same-race preferences.

The third question concerns the existence of gender differences in mate preferences. This question is the subject of a vast literature, especially in psychology but also in other fields (e.g., Buss 1989; Eastwick and Finkel 2008; Fisman et al. 2006; Sprecher et al. 1994). Several theories, such as evolutionary psychology (Buss 1989; Buss and Schmitt 1993) and social structure theory (Eagly and Wood 1999), predict gender differences.

Beyond providing new insights into the three questions posed, our study makes data and methodological contributions to the literature on mate preferences. First, we base our results on a large and detailed data set. We utilize information on the attributes and mate search behavior of 6,485 users of an online dating service in two US metropolitan areas. As we show in Section 2, our sample is diverse and spans a wide variety of demographic and socioeconomic attributes. The detailed information on the users' traits allows us to estimate mate preferences over a much larger set of attributes than in the extant studies that we are aware of.

Second, most of the seminal studies in the mate preference literature (e.g. Buss 1989) are based on stated preference data, which may not reflect actual, revealed mate preferences. Eastwick and Finkel (2008) and Todd et al. (2007) provide some evidence for this concern. Stated preferences regarding a partner of a different race may be particularly unreliable, a suspicion our results in Section 5.2 confirm. Revealed preference methods rely on the researcher's ability to observe and model the choice context accurately. Online dating provides us with a near-ideal market environment that allows us to observe the participants' choice sets and their actual mate choices.[1] Specifically, our estimation approach utilizes the well-defined institutional environment of the dating site, where a user first views the posted "profile" of a potential mate and then decides whether to contact that mate by e-mail. This environment allows us to use a straightforward estimation strategy based on the assumption that a user contacts a partner if and only if the potential utility from a match with that partner exceeds a threshold value (a "minimum standard" for a mate). Using data from speed-dating experiments, research parallel to ours has pursued a similar approach to estimating revealed mate preferences (Kurzban and Weeden 2005; Fisman et al. 2006, 2008; Eastwick and Finkel 2008). Our work is also related to a literature that estimates mate preferences from observations on marriages using equilibrium restrictions of marriage market models (e.g., Choo and Siow 2006; Wong 2003). In contrast to this literature, our data allow us to observe the partner search process directly, providing us with information regarding the choice sets and partner choices. Our data also contain more detailed mate attribute information; for example, the US Census data do not include measures of physical traits or life-style variables.

[1]To be precise, we do not observe the site users' opportunities and mate choices outside the online dating environment, but control for the effect of these opportunities on mate choices using person-specific fixed effects.

Third, an important methodological concern is whether people act strategically in a way that shades their true preferences. This concern is relevant both for our approach based on online dating data and the speed-dating literature. For example, if people are afraid of rejection, they may not approach mates they deem as unattainable, and thus bias the empirical researcher's inference regarding their true mate preferences. In Section 4, we discuss two tests for the shading hypothesis, and reject that strategic behavior plays an important role in our data. We show how to correct for the potential bias in preference estimates due to strategic behavior. This approach can be utilized in environments where preference shading is an important concern.

One important question we cannot answer entirely is whether the estimated preferences apply to dating or other short-term relationships, or whether these preferences are also relevant for long-term relationships and marriages. In Hitsch et al. (2010), we discuss and present evidence that shows sorting patterns in short- and long-term relationships are similar, which suggests mate preferences do not substantially differ across short- and long-term relationships. Sprecher et al. (1994) present evidence that men and women are more "selective" with regard to long- versus short-term partners, in the sense of attaching larger weights to many mate attributes. However, they find some important gender differences, in particular the common finding that women place more emphasis on earnings potential relative to physical attractiveness compared to men, in the stated mate preferences for both a short- and long-term partner. Hence, some evidence exists that mate preferences inferred from online dating behavior may also apply to long-term relationships or marriages. The ultimate test of this claim, however, may require follow-up data on marriages that result from relationships initially formed on an online dating site.

The current paper is complementary to a related study, Hitsch et al. (2010), wherein we attempt to explain the sorting patterns we observe among the users of the dating site, and we extrapolate to sorting patterns observed in marriages. Hitsch et al. (2010) also provides preference estimates that, although not the main focus of that paper, serve as an input in the predictions of equilibrium sorting.[2] In contrast, the main focus of this paper is revealed mate preferences, which we estimate using a more general specification and discuss much more extensively. This paper also provides an extended description of the data and a full discussion of strategic behavior.

[2]Lee (2009) and Banerjee et al. (2009) follow Hitsch et al. (2010) by estimating a discrete choice preference model and simulating equilibrium match outcomes using the Gale–Shapley algorithm. They use, respectively, data from a South Korean matchmaking agency and a Bengali newspaper's matrimonial ads section. The data used by Lee (2009) allow her to follow the users of a matchmaking service through several stages of the dating process until an eventual marriage, and she adds a learning component to the choice model.

2 Data description

2.1 Data collection

We first give a brief overview of how online dating works and explain how we collected our data set.

Users who first join the dating site answer questions from a mandatory survey and create "profiles" of themselves.[3] A profile is a webpage that provides information about a user and can be viewed by the other members of the dating service. The users indicate various demographic, socioeconomic, and physical characteristics, such as their age, gender, education level, height, weight, and income. The users also answer a question about why they joined the service, for example, to find a partner for a long-term relationship or a partner for a "casual" relationship. In addition, the users provide information that relates to their personality, lifestyle, or views. For example, the site members indicate whether they are divorced, whether they have children, their religious and political views, and whether they are open to dating a partner with a different ethnic background. All this information is either numeric (such as age and weight) or an answer to a multiple-choice question, and hence quantifiable and usable for our statistical analysis. The users can also answer essay questions that provide more detailed information about their attitudes and personalities. Our analysis cannot use this information. Many users also include one or more photos in their profile. We used these photos to construct a measure of the users' physical attractiveness, as we will explain below.

After registering, the users can browse, search, and interact with the other members of the dating service. Typically, users start their search by indicating in a database query form a preferred age range and geographic location for their partners. The query returns a list of "short profiles" indicating the user name, age, a brief description, and, if available, a thumbnail version of the photo of a potential mate. By clicking on one of the short profiles, the searcher can view the full user profile, which contains socioeconomic and demographic information, a larger version of the profile photo (and possibly additional photos), and answers to several essay questions. Upon reviewing this detailed profile, the searcher decides whether to send an e-mail to the user. Our data contain a detailed, second-by-second account of all these user activities. In particular, we know if and when a user browses another user, views his or her photo(s), and sends an e-mail to another user.

In order to initiate a contact by e-mail, a user has to become a paying member of the dating service. Once the subscription fee is paid, there is no limit to the number of e-mails a user can send.

[3]Neither the names nor any contact information of the users were provided to us in order to protect the privacy of the users.

2.2 Sample description

Our full sample contains information on the attributes and online activities of about 22,000 users of an online dating service.[4] The users were located in Boston and San Diego, and we observe their activities over a three-and-a-half-month period in 2003. In Hitsch et al. (2010), we compare the distribution of socioeconomic and demographic attributes of the online dating sample with two population samples from the 2001 Current Population Survey (CPS). One CPS sample is representative of the general population in Boston and San Diego, the other is restricted to Internet users in the two metropolitan areas. We find that men are somewhat overrepresented in the online dating sample, and that, as expected, users of the dating service are younger than the population at large. Also, the site users are more educated and have higher incomes than the general population. Overall, however, no stark differences exist between the sample of online users and the population at large regarding socioeconomic and demographic characteristics.

To estimate mate preferences, we use data containing information on the users' browsing behavior and the decisions to send a first-contact e-mail to a potential mate. A first-contact e-mail is an initial, introductory e-mail that a user sends after viewing his or profile.

The analysis of mate preferences is based on a sub-sample of 3,702 men and 2,783 women,[5] which we selected from the full sample based on two criteria. First, in their profiles, all users state their reason for joining the dating site. Because our interest is in mate preferences for a date or a long-term partner, we exclude users who state a preference for a casual relationship, and we also exclude users whose stated reasons for joining the site are difficult to interpret ("a higher power brought me here"). Users seeking a casual relationship account for 4% of all user activities, as measured by first-contact e-mails sent; all other excluded users account for 11% of all activities. We include users who want to "start a long-term relationship" (56% of observed activities) and those who are "just looking/curious" (21%), "like to make new friends" (5%), and who state that "a friend put me up on this" (3%). The latter three groups are likely to include users who want to sound less committal than those who state that they want to start a long-term relationship. Second, we only include single users (58% of activities), divorced users (32%), and those who describe themselves as "hopeful" (4%); we exclude the small number of married and engaged users. Table 1 provides summary statistics of the user attributes.

Applying the two selection criteria, we end up with 597,167 observations of user actions (browsed profiles) for men and 196,363 observations for women. The men sent a first-contact e-mail to 12.5% of all women whose profiles they viewed, whereas women sent a first-contact e-mail to 9% of all men whose

[4] Our sample includes only heterosexual users.

[5] In Hitsch et al. (2010) we employed a smaller sample of users because we randomly discarded some observations due to computer memory constraints.

Table 1 User attribute summary statistics

Variable	Men Percent	Women Percent
Age		
18–20	3.16	3.81
21–25	11.02	9.09
26–30	14.34	12.65
31–35	21.29	18.19
36–40	15.67	17.25
41–45	14.34	15.38
46–50	10.45	12.01
51–55	6.02	7.91
56–60	2.57	2.70
61–65	0.76	0.75
66–	0.38	0.25
Relationship status		
Single	67.21	57.91
Divorced	28.44	36.59
“Hopeful”	4.35	5.50
Reason for joining dating service		
Start a long term relationship	58.89	57.44
Make new friends	7.32	10.03
Just looking/curious	30.36	26.82
A friend put me up to this	3.43	5.72
Children		
Has children	30.96	42.16
No children	69.04	57.84
Does the profile include a photo?		
Has photo	51.16	50.79
No photo	48.84	49.21
Self-description of looks		
Very good looks	21.72	25.49
Above average looks	53.21	54.10
I look like anyone else walking down the street	23.82	19.77
Other	1.24	0.65
Height		
–5’2	0.41	17.11
5’3–5’4	0.70	24.62
5’5–5’6	3.94	28.79
5’7–5’8	13.37	19.81
5’9–5’10	25.31	7.84
5’11–6’0	28.80	1.80
6’1–6’2	20.07	
6’3–6’4	5.62	
6’5–	1.78	0.04
BMI		
–18	0.73	4.10
18–20	1.49	19.19
20–22	5.81	29.73
22–24	14.83	26.28
24–26	38.87	9.09
26–28	18.96	4.89
28–30	12.10	2.19
30–32	4.38	1.69
32–	2.84	2.84
Education level		
High school	5.65	5.14
Some college	21.53	26.74

Table 1 (continued)

Variable	Men Percent	Women Percent
2 year degree	5.29	3.56
College	33.77	31.20
In post-graduate program	3.21	3.67
Master's degree	15.40	16.64
Doctoral degree	4.92	3.06
Professional degree	5.19	6.25
Other	5.05	3.74
Annual income ($1,000)		
–12	1.57	3.41
15–25	1.27	1.11
15–25	2.78	2.52
25–35	5.43	6.00
35–50	9.56	12.40
50–75	13.72	12.11
75–100	9.27	4.06
100–150	6.51	1.40
150–200	2.19	0.40
200–250	0.95	0.04
250–	1.49	0.32
Only my accountant and the IRS know	42.52	51.55
Other	2.76	4.67
Ethnicity		
White/Caucasian	85.33	85.87
Black	3.57	1.62
Hispanic	4.62	4.53
Asian	1.59	4.24
Other	4.89	3.74
Religion		
Catholic	27.90	29.91
Jewish	4.00	4.06
Protestant	7.19	8.05
Muslim	0.54	0.40
Orthodox	0.92	0.93
Just Christian	17.15	19.34
Buddhist	0.62	0.86
Agnostic	2.65	1.58
I believe, but don't follow a particular religion	20.99	21.14
Religion is not a part of my life	12.34	8.52
Other	5.70	5.21
Political inclination[a]		
Conservative	13.07	11.21
Liberal	15.15	21.75
Other	71.77	67.04
Smoking behavior		
Non-smoker	88.47	87.17
Smoker	11.53	12.83
Alcohol consumption		
Does not drink	10.21	7.98
Drinks occasionally	88.49	90.76
Heavy drinker	1.30	1.26
Drug usage		
Does not use drugs	77.66	81.70
Uses drugs	22.34	18.30
No. obs.	3,702	2,783

[a]Users classified as "liberal" include those who indicate their political views as "Liberal. We need some changes," whereas "conservatives" include those who state that they are "Quite conservative, thank you." We classified users as "other" based on a wide variety of responses, such as "Liberal on social and conservative on fiscal issues," "Conservative on global and liberal on domestic issues," and "Who cares?" Correspondingly, liberals and conservatives according to our classification are likely to represent people with particularly strong political convictions

profiles they viewed. That women are more selective, in the sense that they approach fewer potential mates in their choice set than men, has been found in the speed-dating literature as well (e.g., Fisman et al. 2006; Kurzban and Weeden 2005). However, the primary difference in mate search behavior in our data concerns the much larger number of potential mates men browse.

A possible concern about our data is that the users might misrepresent some of their attributes. We cannot directly address this concern. However, we can compare the distribution of some reported characteristics with information from a more general sample. In Hitsch et al. (2010), we compare the reported weight and height of the users in our sample with information from the National Health and Examination Survey Anthropometric Tables. We find that the average weight reported by the women in our sample is somewhat lower than the average weight in the population, whereas the average weight reported by men is slightly higher. The stated height of both men and women is somewhat above the US average. Therefore, this comparison provides little evidence of misrepresentation.

2.3 Measuring physical attractiveness

Fifty-one percent of the men and women in our estimation sample post one or more photos online. To construct an attractiveness rating for these available photos, we recruited 100 subjects from the University of Chicago GSB Decision Research Lab mailing list. The subjects were University of Chicago undergraduate and graduate students in the 18–25 age group, with an equal fraction of male and female recruits.

Each subject received \$10 to rate, on a scale of 1 to 10, 400 male faces and 400 female faces displayed on a computer screen. We used each picture approximately 12 times across subjects. We randomized the ordering of the pictures across subjects to minimize bias due to boredom or fatigue.

Consistent with findings in a large literature in cognitive psychology, attractiveness ratings by independent observers appear to be positively correlated (for surveys of this literature, see Langlois et al. 2000, Etcoff 1999, and Buss 1995). We calculated Cronbach's alpha across 12 ratings per photo to be 0.80, which satisfies the reliability criterion (0.80) utilized in several studies that employed similar rating schemes.[6] To eliminate rater-specific mean and variance differences in rating choices, we followed Biddle and Hamermesh (1998) and standardized each photo rating by subtracting the mean rating given by the subject and dividing by the standard deviation of the subject's ratings. We then averaged this standardized rating across the subjects rating a particular photo.

Table 2 reports the results of regressions of (reported) annual income on the attractiveness ratings. Our results largely replicate the findings of Hamermesh and Biddle (1994) and Biddle and Hamermesh (1998), although

[6]Biddle and Hamermesh (1998) report a Cronbach alpha of 0.75.

Table 2 Earnings and physical attractiveness

Variable	Men		Women	
	(1)	(2)	(1)	(2)
Years of education	0.0809	0.0808	0.0762	0.0756
	(0.0055)	(0.0055)	(0.0065)	(0.0066)
Standardized photo rating	0.0988	0.0974	0.1244	0.1175
	(0.0230)	(0.0237)	(0.0246)	(0.0263)
Weight (lbs)		−0.0002		−0.0004
		(0.0006)		(0.0007)
Height (inches)		0.0140		0.0085
		(0.0054)		(0.0061)
No. obs.	1,665	1,665	1,136	1,136
R-squared	0.51	0.52	0.45	0.45

The dependent variable in each regression is the log of reported annual income. Each regression also includes indicator variables controlling for occupation, ethnicity, martial status, and the city (Boston or San Diego) where the user lives. We also included a "years in the workforce" variable, defined as the age of the user minus the years of education minus five. The square of this variable is also included. Standard errors are reported in parentheses

the cross-sectional rather than panel nature of our data makes arguing for a causal relationship between looks and earnings difficult. Nevertheless, the estimated correlations between attractiveness ratings and reported income are significant. The coefficient estimates on the standardized attractiveness score imply that a one standard deviation increase in a man's attractiveness score is related to a 10% increase in his earnings, whereas for a woman, the attractiveness premium is 12%. Interestingly, there also appears to be a significant height premium for men: a 1 in. increase is related to a 1.4% increase in earnings. For women, the corresponding height premium is smaller (0.9%) and not statistically significant. We find no important relationship between earnings and weight.

3 A modeling framework to analyze mate preferences

Out data contain records on the browsing behavior of each user and the decisions to send a first-contact e-mail. We interpret these decisions as choices concerning potential mates. Specifically, let $U_M(m, w)$ be the expected utility that male user m gets from a potential match with woman w. Similarly, let $U_W(w, m)$ be the expected utility that female user w gets from a potential match with m. Let $v_M(m)$ and $v_W(w)$ be the utility m and w get from the outside option of not responding to a specific profile. If m browses w's profile, he chooses to send an e-mail if and only if

$$U_M(m, w) \geq v_M(m). \tag{1}$$

A similar threshold-crossing rule, $U_W(w, m) \geq v_W(w)$, applies to the women in the market.

The threshold-crossing rule above arises naturally in two-sided search models of a dating or marriage market, such as Burdett and Coles (1997) and Adachi (2003). In these models, the thresholds $v_M(m)$ and $v_W(w)$ are reservation utilities that characterize optimal stopping rules and are determined optimally by searchers based on expectations of future gains from searching versus settling with the current best option. Unlike preferences, the reservation utilities are not model primitives but are endogenous objects in that they depend on the relative abundance or scarcity of different types of agents and the preferences of the agents in the market. For example, if very few women compared to men are in the market then women can be more selective in their search. Similarly, agents who possess attributes (e.g., attractiveness) in relatively short supply in the market will have higher reservation utilities.[7]

The stylized search models cited above do not describe exactly the actual partner search behavior in the online dating market we study.[8] However, these models capture basic mechanisms that apply to the workings of the market. The models capture the search process for a partner, and the notion that people have an understanding of their own "dating market value," which influences their threshold for a partner. This dating market value in turn depends on the searcher's own attributes, which searchers of the opposite sex value, as does competition from searchers of the same sex..

The threshold-crossing model above motivates the use of a discrete choice model to estimate mate preferences. We assume that mate preferences depend on observed own and partner attributes, x_m and x_w, and on an idiosyncratic preference shock: $U_M(m, w) = U_M(x_m, x_w; \theta_M) + \epsilon_{mw}$ (the utility function of women is defined analogously). Specifically, our preference model takes the following form:

$$U_M(x_m, x_w; \theta_M) = x_w' \theta_M^v + \iota(x_m, x_w)' \theta_M^h. \tag{2}$$

The parameter sub-vector θ_M^v accounts for vertical preference components, which only depend on the levels of the partner attributes x_w. The vector $\iota(x_m, x_w)$ includes interactions between the attributes of m and the potential mate w. These interaction terms account for the fact that m's preferences over w's attributes may also depend on his own attributes, a form of horizontal preferences. Some of the interactions are based on the difference between specific attributes in x_m and x_w; for example, we allow for preferences that depend on whether the age difference between w and m is at least 10 years, and we also allow this preference to differ depending on the age of m. Other interactions capture preferences that depend on the categorical values of some

[7] Reservation utilities of agents are also interdependent in that exogenous shocks to a subset of agents' reservation utilities will cause other agents to re-optimize their threshold-crossing rules. Adachi (2003) shows a strategic substitutability property in that more selective behavior (higher reservation utilities) by women (men) leads, in equilibrium, to less selective behavior (lower reservation utilities) by men (women).

[8] Although especially Adachi (2003) pushes the realism of these models significantly forward by allowing agents to possess very general preferences.

own and partner attributes; for example, we allow for preferences that depend on the ethnicity or education level of both m and w. Because of the large number of mate attributes that we employ in the estimated model, we chose to focus on a preference specification in which the horizontal preference components are due to interactions of the observable attributes (plus the additive random utility term ϵ_{mw}); a generalization using a random coefficients specification with unobserved preference heterogeneity is possible but left for future research.

We assume ϵ_{mw} has the standard logistic distribution and is i.i.d. across all pairs of men and women. The reservation values $v_M(m)$ and $v_W(w)$ are estimated as person-specific fixed effects, denoted by c_m and c_w. The choice probabilities take the standard logit form:

$$\Pr\{m \text{ contacts } w | m \text{ browses } w\} = \frac{\exp\left(U_M(x_m, x_w; \theta_M) - c_m\right)}{1 + \exp\left(U_M(x_m, x_w; \theta_M) - c_m\right)}.$$

The preference parameters in this model can be estimated using a fixed effects logit estimator. A fixed effects estimation approach is required since $v_M(m)$ and $v_W(w)$ are functions of x_m and x_w.[9]

In Hitsch et al. (2010), we estimate a similar model, but using only a subset of the attribute information used here and stronger functional form assumptions. Nonetheless, the results in Hitsch et al. (2010) are largely in line with the more general preference estimates we present here.

4 Strategic behavior

The threshold rule (1) may not hold if sending a first-contact e-mail is costly or if there is a psychological cost of being rejected. In particular, a user may strategically refrain from contacting a potential mate if he or she believes there is only a small chance of going on an actual date. A priori, we expect that both factors—the cost of sending an e-mail and the cost of rejection—are less important in online than in offline dating.[10] Nonetheless, understanding how these factor may bias our inference of mate preferences and how we can test and control for strategic behavior is important. All analyses of mate

[9]We estimated the model in MATLAB using the KNITRO nonlinear optimization solver. Instead of concentrating out the fixed effects, we estimated all fixed effects directly along with the preference parameters. Using an analytic gradient and Hessian, convergence always occurred in less than 10 steps and in less than 120 s.

[10]The main cost associated with sending an e-mail is the cost of composing it. However, the marginal cost of producing yet another witty e-mail is likely to be small since one can easily personalize a polished form letter, or simply use a "copy and paste" approach. The fear of rejection should be mitigated by the anonymity provided by the dating site. Furthermore, rejection is common in online dating: in our data, 71% of men's and 56% of women's first-contact e-mails do not receive a reply.

preferences using a revealed preference approach, such as those based on speed dating or responses to newspaper personal ads, are subject to the same potential problem.

To understand how strategic behavior will bias the preference estimates, assume that a user incurs a cost k when contacting a potential mate, and cost r if he or she is rejected. Let $P_W(m, w)$ denote the (subjective) probability that w will accept m's invitation for a date. Then, *conditional on contacting* this mate, the utility m expects to get from a match with w is

$$EU^c_M(m, w) = -k + U_M(m, w)P_W(m, w) + (v_M(m) - r)(1 - P_W(m, w)).$$

Here, $P_W(m, w)$ is the probability that w will accept m's offer. Man m will then contact woman w if and only if $EU^c_M(m, w) \geq v_M(w)$, which simplifies to the threshold-crossing condition

$$U_M(m, w) - \frac{k + r}{P_W(m, w)} + r \geq v_M(m). \tag{3}$$

This equation illustrates the source of the potential bias in preference estimation. Both $U_M(m, w)$ and $P_W(m, w)$ depend on m's and w's attributes, x_m and x_w. Hence, differences in first-contact behavior related to differences in own and mate attributes reflect not only differences in the true preference for a mate but also differences in the expected cost of achieving a match. For example, suppose U_M increases in the looks of w while P_W decreases in w's looks. The left-hand side of Eq. 3 may then increase, decrease, or stay constant in the looks attribute of w. Note that as long as r is a function of x_m only, it will be absorbed into the fixed effect in our specification.

4.1 A first test for strategic behavior

Our first test is somewhat preliminary but easy to implement. The assumption that underlies this test is that physical attractiveness is a vertical attribute: everyone prefers to date a more attractive partner than a less attractive partner. Under this assumption, strategic behavior would "shade" the site users' decisions to contact a mate based on physical attractiveness. For example, a relatively unattractive user might be more likely to approach other unattractive mates rather than highly attractive mates even though the user prefers a more attractive partner. On the other hand, if the site users do not act strategically, their own physical attractiveness should not affect their first-contact decisions. A similar test based on attributes that are not likely to be vertical, such as age or education, will not yield conclusive results. For example, if we found evidence that the site users prefer partners with a similar education level, we could not be sure whether this finding was due to horizontal preferences or strategic behavior regarding education.

To implement our test, we first categorize all browsed mates according to the looks decile to which they belong. We then examine whether user i's own attractiveness rating affects his or her decision to contact a potential mate j with an attractiveness rating in decile d. Thus, we categorize all browsers i according to their own looks quintile, and estimate the following linear probability model:

$$email_{ij} = \sum_{d=1}^{10} \left(\beta_{1d} \mathbb{I}\{ad_j = d\} + \sum_{q=2}^{5} \beta_{qd} \mathbb{I}\{ad_j = d \text{ and } aq_i = q\} \right) + u_i + \epsilon_{ij}. \quad (4)$$

Here, *email* is the 0/1 decision to contact a potential mate upon browsing his or her profile, ad_j denotes the physical attractiveness decile of mate j, and aq_i denotes the attractiveness quintile of browser i. u_i is a browser-specific fixed effect that controls for i's threshold to contact a potential mate.

Table 3 shows the estimation results for model (4). Overall, the coefficients β_{1d} are increasing in the mate's looks decile d for both men and women. Hence, users in the lowest quintile of looks ratings do not refrain from contacting potential mates with higher attractiveness ratings. The coefficients

Table 3 Testing for strategic behavior: first-contact decisions based on own and partner looks

	Looks quintile of browser:									
	1		2		3		4		5	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
(a) First-contact behavior of men										
Looks decile of mate										
2	**−0.003**	0.019	**0.012**	0.025	**0.017**	0.024	**0.017**	0.024	**0.015**	0.026
3	**0.049**	0.019	**−0.022**	0.025	**−0.020**	0.024	**−0.010**	0.023	**−0.012**	0.026
4	**0.030**	0.020	**0.005**	0.025	**0.004**	0.025	**0.014**	0.024	**0.007**	0.026
5	**0.033**	0.019	**0.013**	0.024	**0.015**	0.023	**0.050**	0.023	**0.013**	0.025
6	**0.070**	0.018	**−0.005**	0.023	**−0.012**	0.023	**0.024**	0.022	**−0.021**	0.024
7	**0.062**	0.018	**0.018**	0.023	**0.015**	0.022	**0.053**	0.021	**0.030**	0.023
8	**0.066**	0.017	**0.015**	0.022	**0.033**	0.022	**0.050**	0.021	**0.024**	0.023
9	**0.098**	0.018	**−0.009**	0.023	**0.013**	0.022	**0.050**	0.021	**0.003**	0.023
10	**0.102**	0.019	**−0.003**	0.024	**0.029**	0.023	**0.075**	0.022	**0.045**	0.024
No. obs.	75,824									
(b) First-contact behavior of women										
Looks decile of mate										
2	**0.023**	0.025	**−0.035**	0.035	**−0.009**	0.034	**−0.034**	0.033	**−0.036**	0.035
3	**0.021**	0.024	**0.015**	0.034	**−0.021**	0.034	**−0.012**	0.032	**−0.014**	0.034
4	**0.026**	0.023	**−0.019**	0.032	**−0.010**	0.031	**−0.012**	0.030	**−0.013**	0.032
5	**0.051**	0.023	**−0.032**	0.032	**−0.022**	0.032	**−0.025**	0.031	**−0.028**	0.033
6	**0.042**	0.022	**−0.021**	0.031	**−0.024**	0.030	**−0.004**	0.029	**−0.014**	0.031
7	**0.065**	0.022	**−0.030**	0.030	**−0.039**	0.030	**−0.048**	0.029	**−0.028**	0.031
8	**0.053**	0.022	**−0.022**	0.030	**−0.014**	0.030	**−0.019**	0.029	**−0.009**	0.031
9	**0.050**	0.022	**−0.007**	0.031	**0.018**	0.030	**0.009**	0.029	**0.006**	0.031
10	**0.078**	0.022	**−0.007**	0.031	**0.001**	0.030	**0.015**	0.029	**0.024**	0.031
No. obs.	32,167									

The estimates for browsers in looks quintiles 2–5 are the differences in response rates relative to the response rates of browsers in looks quintile 1. Fixed effects are included for all browsers

β_{qd} measure the difference in first-contact probabilities between browsers in the first (lowest) looks quintile and browsers in quintile $2 \leq q \leq 5$. For men, we find some positive and statistically significant differences (at the 5% level) for browsers in looks quintile 4. All other coefficients for men and all estimated coefficients for women are not statistically different from zero. We also test the null hypothesis that $\beta_{2d} = \beta_{3d} = \beta_{4d} = \beta_{5d} = 0$ separately for each decile d and for men and women. For men, we reject this null hypothesis only for the top two looks deciles. For women, we cannot reject equality of the coefficients for any decile d.

We also estimate a version of model (4) separately for each quintile of browsers and without the interaction terms. Figure 1 depicts the estimated first-contact probabilities (note that Hitsch et al. 2010 contains the same graph; we replicate it here because it provides the easiest way to assess the evidence for or against strategic behavior). The increasing responses depicted in the graph and the tests discussed above provide little evidence that strategic behavior is of concern in our data.

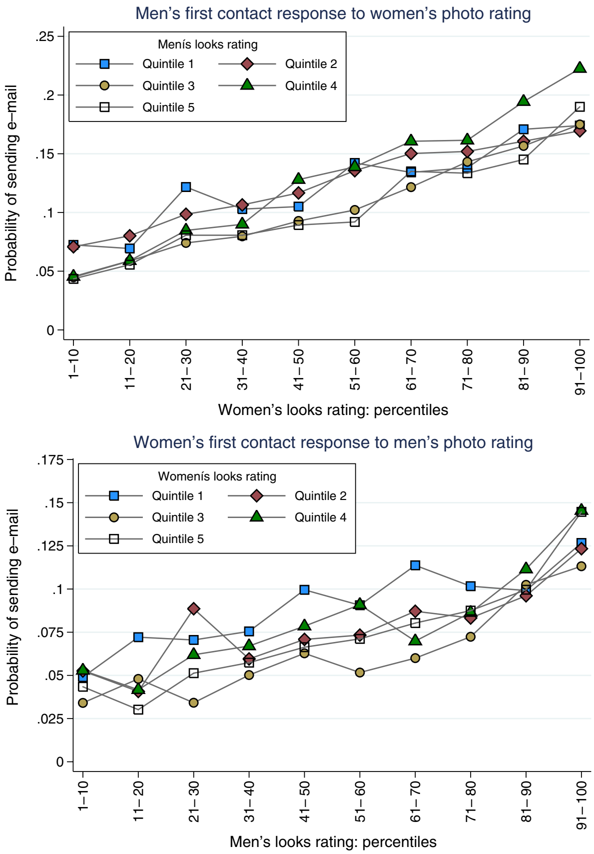

Fig. 1 Testing for strategic behavior. Note: The *vertical axis* plots the estimated mean probability of sending a first-contact e-mail to a browsed user, the *horizontal axis* indicates the photo rating of the browsed user. The estimated probabilities are based on separate regressions for each looks quintile of browsers. The estimates shown are for a sample of users ages 30–39

4.2 A second test and bias correction

In our second test for strategic behavior, we directly control for the probability of achieving a match (going on a date) that may influence a user's decision to send a first-contact e-mail. Compared to the original threshold rule (1), the threshold condition (3) involves the extra term $P_W(m, w)$. As in Heckman's (1979) two-step method for selection correction, we can estimate $P_W(m, w)$ directly from the data in a first step, and include it in our preference specification as a control. We estimate $P_W(m, w)$ using the information contained in our data on whether woman w sent a reply to m's first-contact e-mail. Here we assume that m's subjective perception of $P_W(m, w)$ coincides with the actual probability of receiving a reply (rational expectations). In principle, $P_W(m, w)$ could be estimated non-parametrically, but in practice, the large number of attributes makes nonparametric estimation infeasible. Thus, we use a binary logit specification to model the probability of receiving a reply conditional on own and partner attributes.[11] We then substitute the reciprocal of the predicted reply probability as an additional term into the first-contact specification:

$$\Pr\{m \text{ contacts } w|m \text{ browses } w\} = \frac{\exp\left(U_M(x_m, x_w; \theta_M) - c_m - \frac{k+r}{\hat{P}_W(m,w)}\right)}{1 + \exp\left(U_M(x_m, x_w; \theta_M) - c_m - \frac{k+r}{\hat{P}_W(m,w)}\right)}.$$

We also experimented with using $\hat{P}_W(m, w)$ directly as a covariate, but doing so did not change our results. To account for the propagation of estimation errors in the first- to the second-stage preference estimates, we calculate all standard errors using a bootstrap procedure.[12] Note that the coefficient on $1/\hat{P}_W(m, w)$ will be an estimate of $k + r$, the combined cost of rejection and e-mailing, which can be tested against the null hypothesis of zero.

As we see in the sample selection literature, identification of the second-stage coefficients relies on the accuracy of the parametric model assumptions made (Goldberger 1983; Little 1985). Other studies have developed a number of semi-parametric estimation approaches (Heckman 1990; Newey et al. 1990) that rely on exclusion restrictions. Finding a variable that is excluded from $U_M(m, w)$ but shifts $P_W(m, w)$ is difficult in the context of online dating because the variable affecting $P_W(m, w)$ must be observable to m, but not in a way that affects m's utility from an eventual match with w. One factor possibly affecting $P_W(m, w)$ but not m's preferences is the time elapsed since w registered on the site. Users new to the site may initially receive a larger

[11] To be precise: The probability that m receives a reply from w is determined by the utility function $U_W(x_w, x_m)$, i.e. the preference of a woman with attributes x_w for a man with attributes x_m.

[12] We resample over individuals rather than individual choice instances to preserve within-person dependence structure.

number of e-mails from interested incumbent users than they will later. New users may also initially find organizing and replying to the e-mails they receive difficult. Therefore, the time elapsed since registration would likely affect the probability of receiving a reply from a new user but would not affect the utility of m from a date with w. However, the time elapsed since registration is not a valid excluded variable if it provides a signal of unobserved quality, in which case it will not be excluded from m's preferences.[13]

In Table I in the Appendix (Electronic Supplementary Material), we report the results of the first-stage estimates of reply probabilities for men and women. Most determinants of receiving a reply are intuitive and are in line with the preference results we discuss in the next section. We code the "days since registration" variable as an indicator for various lengths of time elapsed since a user registered. These indicators enter the reply probability significantly: men who registered recently are less likely to reply than men who registered more than 50 days ago. The same result also holds for women (although very new registrants are somewhat more likely to reply to e-mails than registrants who have been on the site between 4 and 50 days).

As we report in Table II in the Appendix, the coefficients in our main specification (Table 4) hardly change if we account for strategic behavior.[14] The point estimates of $k + r$, the combined cost of emailing and the cost of rejection, are not statistically significantly different from zero for both genders and are very small in size. We thus do not reject the null hypothesis that strategic behavior leading to "shading" of true mate preferences is of no concern in this market.

5 Preference estimates

Table 4 shows the preference estimates obtained from the fixed effects binary logit model. The table shows the preference coefficients for men, and the difference between women's and men's preference coefficients. We can thus directly assess if the difference between men's and women's preference coefficients is statistically significant. The table also shows the marginal effects of mate attributes on first-contact probabilities, which allows us to assess the quantitative significance of the different preference components. Note that the

[13]However, it is not clear whether a longer "time on market" in the context of dating should be considered a good or a bad signal of quality. A costly signalling story may suggest that "good" types can separate themselves from "bad" types by holding out longer. The opposite interpretation of time on market is possible if bad types reveal their unobserved quality during a date, are then rejected and hence stay longer in the market.

[14]The results reported in Table II are based on the predicted reply probabilities including the "days since registration" variable. The estimates based on the predicted reply probabilities without an excluded (from the first-contact decision) variable are similar.

Table 4 Preference parameter estimates

Variable	Model estimates				Marginal effects					
	Men		Women: differences		Men			Women		
	Est.	SE	Est.	SE	Est.	95% CI		Est.	95% CI	
Age										
Age < 30;										
5+ years younger	**−0.03**	0.036	**−0.28**	0.136	**−0.005**	−0.015	0.006	**−0.037**	−0.061	−0.007
5–10 years older	**−0.57**	0.032	**0.48**	0.062	**−0.072**	−0.078	−0.065	**−0.012**	−0.024	0.002
10+ years older	**−1.10**	0.035	**0.22**	0.074	**−0.116**	−0.120	−0.111	**−0.084**	−0.092	−0.075
Age ≥ 30 and < 40;										
10+ years younger	**−0.12**	0.022	**−0.74**	0.093	**−0.017**	−0.023	−0.011	**−0.083**	−0.094	−0.070
5–10 years younger	**0.14**	0.018	**−0.37**	0.047	**0.023**	0.017	0.029	**−0.028**	−0.037	−0.018
5–10 years older	**−0.43**	0.022	**0.30**	0.041	**−0.057**	−0.062	−0.052	**−0.016**	−0.024	−0.008
10+ years older	−**1.14**	0.028	**0.48**	0.056	**−0.119**	−0.122	−0.115	**−0.069**	−0.076	−0.061
Age ≥ 40 and < 50;										
10+ years younger	**−0.19**	0.024	**−0.81**	0.065	**−0.027**	−0.033	−0.021	**−0.092**	−0.098	−0.084
5–10 years younger	**0.15**	0.024	**−0.39**	0.049	**0.024**	0.016	0.032	**−0.029**	−0.038	−0.019
5–10 years older	**−0.44**	0.035	**0.45**	0.053	**−0.058**	−0.066	−0.050	**0.001**	−0.009	0.012
10+ years older	**−1.30**	0.062	**1.09**	0.086	**−0.128**	−0.135	−0.121	**−0.025**	−0.038	−0.012
Age ≥ 50;										
10+ years younger	**−0.53**	0.049	**−0.50**	0.110	**−0.067**	−0.077	−0.057	**−0.093**	−0.103	−0.081
5–10 years younger	**0.04**	0.051	**−0.33**	0.086	**0.007**	−0.008	0.023	**−0.034**	−0.048	−0.019
5+ years older	**−0.21**	0.102	**0.19**	0.129	**−0.030**	−0.055	−0.002	**−0.002**	−0.021	0.018
Marital status										
Single; divorced	**−0.05**	0.015	**−0.04**	0.035	**−0.007**	−0.012	−0.003	**−0.012**	−0.019	−0.004
Divorced; divorced	**0.09**	0.018	**0.07**	0.036	**0.014**	0.009	0.020	**0.023**	0.014	0.032
Reason for joining site										
"Longterm"; "Longterm"	**0.06**	0.012	**0.16**	0.029	**0.009**	0.006	0.013	**0.027**	0.021	0.034
Children										
Has children; has children	**0.23**	0.018	**−0.04**	0.035	**0.038**	0.031	0.044	**0.026**	0.018	0.035
No children; has children	**−0.26**	0.015	**−0.11**	0.037	**−0.036**	−0.040	−0.032	**−0.042**	−0.048	−0.035
Looks rating (percentile)										
11–20	**0.09**	0.040	**0.10**	0.073	**0.007**	0.001	0.013	**0.015**	0.005	0.025
21–30	**0.36**	0.037	**−0.09**	0.071	**0.031**	0.024	0.038	**0.021**	0.012	0.032
31–40	**0.47**	0.036	**−0.12**	0.069	**0.042**	0.035	0.050	**0.029**	0.018	0.040
41–50	**0.49**	0.036	**−0.16**	0.069	**0.045**	0.037	0.053	**0.027**	0.016	0.038
51–60	**0.68**	0.035	**−0.15**	0.067	**0.067**	0.059	0.076	**0.046**	0.035	0.059
61–70	**0.73**	0.035	**−0.14**	0.067	**0.073**	0.064	0.082	**0.053**	0.041	0.066
71–80	**0.83**	0.034	**−0.09**	0.067	**0.087**	0.078	0.096	**0.070**	0.057	0.085
81–90	**0.97**	0.034	**−0.16**	0.067	**0.107**	0.097	0.117	**0.079**	0.065	0.095
91–95	**1.16**	0.037	**−0.39**	0.077	**0.138**	0.125	0.150	**0.075**	0.059	0.092
96–100	**1.26**	0.038	**0.08**	0.076	**0.156**	0.142	0.169	**0.163**	0.140	0.187
Has photo	**−0.59**	0.038	**0.19**	0.076	**−0.106**	−0.117	−0.095	**−0.060**	−0.076	−0.042
Self-description of looks										
"Very good"	**0.56**	0.026	**−0.06**	0.062	**0.081**	0.072	0.089	**0.059**	0.045	0.075
"Above average"	**0.27**	0.024	**−0.10**	0.055	**0.035**	0.029	0.042	**0.017**	0.007	0.028
"Other"	**0.00**	0.141	**0.09**	0.253	**0.000**	−0.029	0.036	**0.010**	−0.027	0.060

Table 4 (continued)

Variable	Model estimates				Marginal effects					
	Men		Women: differences		Men			Women		
	Est.	SE	Est.	SE	Est.	95% CI		Est.	95% CI	
Height										
5'3–5'4	**0.07**	0.015			**0.009**	0.005	0.014			
5'5–5'6	**0.15**	0.017	**−0.37**	0.180	**0.021**	0.016	0.026	**−0.031**	−0.071	0.020
5'7–5'8	**0.12**	0.021	**−0.46**	0.177	**0.017**	0.010	0.023	**−0.046**	−0.081	0.000
5'9–5'10	**0.09**	0.029	**−0.32**	0.179	**0.013**	0.004	0.021	**−0.032**	−0.071	0.018
5'11–6'0	**0.24**	0.046	**−0.43**	0.185	**0.035**	0.021	0.050	**−0.027**	−0.068	0.025
6'1–6'2	**0.61**	0.134	**−0.78**	0.225	**0.102**	0.054	0.156	**−0.023**	−0.065	0.030
6'3–6'4			**−0.11**	0.185				**−0.016**	−0.060	0.041
6'5+			**−0.06**	0.194				**−0.009**	−0.057	0.052
Height difference; mate is										
5'+ shorter	**0.49**	0.031	**−2.66**	0.640	**0.064**	0.055	0.073	**−0.074**	−0.081	−0.049
2–5 in. shorter	**0.39**	0.023	**−1.07**	0.098	**0.049**	0.043	0.056	**−0.040**	−0.047	−0.031
2–5 in. taller	**−0.70**	0.038	**1.22**	0.063	**−0.058**	−0.063	−0.053	**0.050**	0.039	0.062
5+ in. taller	**−0.94**	0.126	**1.62**	0.141	**−0.071**	−0.082	−0.057	**0.071**	0.055	0.088
BMI										
18–20	**−0.06**	0.022			**−0.010**	−0.017	−0.003			
20–22	**−0.22**	0.022	**0.52**	0.132	**−0.036**	−0.043	−0.030	**0.031**	0.004	0.063
22–24	**−0.43**	0.024	**0.85**	0.128	**−0.066**	−0.072	−0.060	**0.045**	0.017	0.079
24–26	**−0.72**	0.031	**1.19**	0.130	**−0.101**	−0.107	−0.094	**0.052**	0.022	0.087
26–28	**−0.88**	0.041	**1.31**	0.135	**−0.117**	−0.124	−0.109	**0.047**	0.018	0.082
28–30	**−1.15**	0.056	**1.52**	0.143	**−0.140**	−0.148	−0.131	**0.040**	0.012	0.075
30–32	**−1.24**	0.075	**1.46**	0.157	**−0.147**	−0.157	−0.136	**0.021**	−0.005	0.054
32+	**−1.06**	0.067	**1.08**	0.166	**−0.133**	−0.143	−0.122	**0.002**	−0.023	0.033
BMI difference mate/browser										
Less than 2	**0.23**	0.018	**−0.49**	0.053	**0.032**	0.027	0.038	**−0.028**	−0.036	−0.018
More than 2	**−0.23**	0.035	**0.40**	0.047	**−0.027**	−0.035	−0.020	**0.021**	0.014	0.030
Education										
High school;										
Some college	**0.01**	0.063	**−0.18**	0.153	**0.002**	−0.016	0.021	**−0.020**	−0.049	0.015
College	**−0.10**	0.062	**−0.23**	0.149	**−0.015**	−0.032	0.003	**−0.039**	−0.063	−0.009
Graduate degree	**−0.17**	0.064	**−0.38**	0.156	**−0.025**	−0.041	−0.007	**−0.059**	−0.081	−0.032
Some college;										
High school	**−0.04**	0.036	**−0.15**	0.083	**−0.006**	−0.016	0.005	**−0.022**	−0.038	−0.005
College	**−0.12**	0.024	**−0.06**	0.049	**−0.017**	−0.024	−0.010	**−0.022**	−0.031	−0.012
Graduate degree	**−0.22**	0.026	**−0.04**	0.053	**0.031**	−0.037	−0.024	**−0.031**	−0.040	−0.021
College;										
High school	**−0.13**	0.030	**−0.16**	0.081	**−0.019**	−0.027	−0.011	**−0.034**	−0.049	−0.018
Some college	**−0.11**	0.019	**−0.09**	0.047	**−0.016**	−0.021	−0.010	**−0.024**	−0.034	−0.015
Graduate degree	**−0.11**	0.018	**0.12**	0.039	**−0.016**	−0.021	−0.011	**0.001**	−0.007	0.010
Graduate degree;										
High school	**−0.17**	0.036	**−0.36**	0.093	**−0.024**	−0.033	−0.014	**−0.057**	−0.071	−0.041
Some college	**−0.11**	0.023	**−0.38**	0.055	**−0.016**	−0.022	−0.009	**−0.053**	−0.062	−0.044
College	**0.00**	0.020	**−0.22**	0.041	**0.000**	−0.006	0.006	**−0.026**	−0.034	−0.018
Income (thousands of dollars)										
25–35	**0.17**	0.030	**−0.04**	0.137	**0.022**	0.014	0.030	**0.014**	−0.015	0.049
35–50	**0.20**	0.028	**0.03**	0.128	**0.027**	0.019	0.035	**0.028**	−0.001	0.063
50–75	**0.30**	0.029	**0.36**	0.126	**0.042**	0.033	0.050	**0.092**	0.054	0.136
75–100	**0.40**	0.037	**0.31**	0.133	**0.057**	0.046	0.069	**0.102**	0.061	0.149
100–150	**0.35**	0.049	**0.59**	0.138	**0.049**	0.034	0.064	**0.143**	0.097	0.196
150–200	**0.45**	0.071	**0.35**	0.153	**0.066**	0.043	0.090	**0.117**	0.072	0.170
200–250	**0.38**	0.079	**0.38**	0.171	**0.054**	0.031	0.080	**0.110**	0.061	0.168
250+			**1.10**	0.137				**0.177**	0.124	0.237

Table 4 (continued)

Variable	Model estimates: Men Est.	Men SE	Women: differences Est.	Women: differences SE	Marginal effects: Men Est.	Men 95% CI		Women Est.	Women 95% CI	
Income of mate										
25k+ less than browser	**−0.04**	0.022	**−0.10**	0.044	**−0.007**	−0.013	0.000	**−0.018**	−0.026	−0.009
25k+ more than browser	**−0.02**	0.020	**0.01**	0.050	**−0.003**	−0.009	0.003	**−0.002**	−0.013	0.010
"Only accountant knows"	**0.30**	0.023	**0.49**	0.119	**0.051**	0.043	0.059	**0.134**	0.089	0.183
"What, me work?"	**0.21**	0.031	**0.28**	0.136	**0.034**	0.024	0.045	**0.075**	0.032	0.124
Occupation										
Executive/managerial	**0.04**	0.026	**0.25**	0.095	**0.006**	−0.001	0.014	**0.040**	0.015	0.068
Administrative/clerical/secretarial	**0.06**	0.027	**0.29**	0.146	**0.009**	0.001	0.017	**0.049**	0.009	0.096
Financial/accounting	**0.02**	0.028	**0.16**	0.100	**0.003**	−0.005	0.012	**0.024**	0.000	0.052
Manufacturing	**0.14**	0.074	**0.14**	0.144	**0.022**	−0.001	0.046	**0.037**	0.004	0.076
Sales/marketing	**0.08**	0.025	**0.04**	0.096	**0.011**	0.004	0.019	**0.014**	−0.008	0.040
Technical/science/engineering/research/computer secretarial	**0.04**	0.028	**0.05**	0.095	**0.007**	−0.001	0.015	**0.012**	−0.010	0.037
Teacher/educator/professor	**0.07**	0.027	**0.01**	0.103	**0.011**	0.003	0.019	**0.010**	−0.014	0.037
Service/hospitality/food	**0.17**	0.041	**−0.43**	0.171	**0.026**	0.013	0.039	**−0.030**	−0.058	0.007
Legal/attorney	**0.12**	0.034	**0.46**	0.109	**0.017**	0.007	0.028	**0.086**	0.052	0.123
Law enforcement/Fire fighter	**0.06**	0.101	**0.46**	0.149	**0.009**	−0.019	0.041	**0.077**	0.042	0.116
Artistic/musical/writer	**0.21**	0.033	**−0.08**	0.112	**0.033**	0.022	0.043	**0.017**	−0.010	0.047
Health/medical/psychology/dental/nursing	**0.15**	0.025	**0.21**	0.100	**0.023**	0.015	0.031	**0.050**	0.023	0.082
Political/government/civil	**−0.07**	0.038	**0.20**	0.116	**−0.010**	−0.020	0.001	**0.017**	−0.010	0.048
Transportation	**0.12**	0.063	**−0.04**	0.135	**0.018**	0.000	0.038	**0.010**	−0.018	0.042
Entertainment/broadcasting/film	**0.08**	0.058	**0.22**	0.129	**0.012**	−0.005	0.030	**0.042**	0.010	0.079
Laborer/construction	**−0.21**	0.125	**0.18**	0.169	**−0.028**	−0.056	0.005	**−0.003**	−0.028	0.026
Military	**−0.00**	0.083	**0.47**	0.138	**0.000**	−0.023	0.024	**0.067**	0.033	0.105
Self employed	**0.12**	0.026	**0.05**	0.095	**0.018**	0.010	0.026	**0.022**	−0.001	0.048
Other	**0.17**	0.033	**−0.13**	0.116	**0.025**	0.015	0.036	**0.005**	−0.021	0.034
Same occupation	**0.08**	0.016	**0.05**	0.034	**0.012**	0.007	0.017	**0.018**	0.010	0.027
Race										
White;										
Black	**−0.88**	0.057	**0.12**	0.133	**−0.100**	−0.108	−0.091	**−0.075**	−0.091	−0.056
Hispanic	**−0.26**	0.024	**−0.30**	0.094	**−0.036**	−0.042	−0.030	**−0.060**	−0.074	−0.043
Asian	**−0.44**	0.029	**−1.12**	0.242	**−0.058**	−0.064	−0.051	**−0.118**	−0.131	−0.097
Other	**−0.15**	0.025	**−0.37**	0.079	**−0.022**	−0.029	−0.016	**−0.057**	−0.069	−0.043
Black;										
White	**−0.43**	0.218	**−1.35**	0.469	**−0.057**	−0.098	0.000	**−0.125**	−0.141	−0.090
Hispanic	**−0.50**	0.256	**−1.46**	0.941	**−0.064**	−0.109	0.000	**−0.130**	−0.150	−0.023
Asian	**−1.12**	0.296	**−0.47**	1.213	**−0.118**	−0.147	−0.069	**−0.119**	−0.151	0.117
Other	**−0.23**	0.256	**−0.73**	0.650	**−0.032**	−0.087	0.046	**−0.089**	−0.133	0.030

Table 4 (continued)

Variable	Model estimates				Marginal effects					
	Men		Women: differences		Men			Women		
	Est.	SE	Est.	SE	Est.	95% CI		Est.	95% CI	
Hispanic;										
White	**−0.43**	0.093	**−0.22**	0.250	**−0.057**	−0.076	−0.035	**−0.068**	−0.098	−0.024
Black	**−0.51**	0.239	**−0.39**	0.562	**−0.066**	−0.108	−0.006	**−0.086**	−0.128	0.013
Asian	**−0.81**	0.174			**−0.094**	−0.119	−0.061			
Other	**−0.05**	0.129	**−0.59**	0.408	**−0.007**	−0.042	0.033	**−0.066**	−0.111	0.017
Asian;										
White	**−0.61**	0.215	**0.52**	0.511	**−0.076**	−0.111	−0.027	**−0.011**	−0.091	0.139
Black	**−1.28**	0.788	**0.32**	1.199	**−0.127**	−0.174	0.044	**−0.089**	−0.143	0.139
Hispanic	**−0.23**	0.293	**−0.31**	0.669	**−0.033**	−0.094	0.058	**−0.058**	−0.123	0.103
Other	**0.14**	0.277	**−0.51**	0.635	**0.023**	−0.054	0.127	**−0.043**	−0.115	0.124
Religion										
Christian (non-Catholic);										
Catholic	**−0.09**	0.024	**−0.01**	0.054	**−0.013**	−0.020	−0.006	**−0.013**	−0.024	−0.001
Other religion	**−0.13**	0.022	**−0.09**	0.049	**−0.019**	−0.024	−0.012	**−0.026**	−0.035	−0.016
Not religious	**−0.16**	0.031	**−0.13**	0.063	**−0.023**	−0.032	−0.015	**−0.035**	−0.045	−0.023
Catholic;										
Christian	**−0.08**	0.024	**−0.07**	0.051	**−0.012**	−0.019	−0.005	**−0.019**	−0.029	−0.008
Other religion	**−0.15**	0.022	**−0.15**	0.048	**−0.022**	−0.028	−0.016	**−0.036**	−0.044	−0.027
Not religious	**−0.15**	0.030	**−0.13**	0.059	**−0.022**	−0.030	−0.014	**−0.033**	−0.043	−0.022
Other religion;										
Christian or Catholic	**−0.05**	0.017	**−0.16**	0.037	**−0.007**	−0.012	−0.002	**−0.025**	−0.032	−0.017
Not religious	**0.03**	0.027	**−0.12**	0.050	**0.004**	−0.004	0.012	**−0.012**	−0.022	−0.002
Not religious;										
Christian or Catholic	**−0.18**	0.038	**−0.02**	0.081	**−0.026**	−0.036	−0.015	**−0.024**	−0.039	−0.007
Other religion	**−0.09**	0.040	**−0.02**	0.085	**−0.014**	−0.025	−0.002	**−0.014**	−0.031	0.004
Political views										
Conservative;										
Liberal	**−0.24**	0.049	**−0.16**	0.112	**−0.033**	−0.045	−0.021	**−0.046**	−0.063	−0.025
Other	**−0.06**	0.041	**−0.08**	0.082	**−0.009**	−0.020	0.003	**−0.018**	−0.033	0.000
Liberal;										
Conservative	**−0.31**	0.050	**−0.10**	0.089	**−0.042**	−0.054	−0.030	**−0.046**	−0.059	−0.031
Other	**−0.08**	0.031	**−0.13**	0.057	**−0.012**	−0.020	−0.003	**−0.025**	−0.035	−0.014
Other;										
Conservative	**−0.05**	0.018	**0.06**	0.037	**−0.007**	−0.012	−0.002	**0.002**	−0.007	0.010
Liberal	**−0.04**	0.013	**0.01**	0.034	**−0.007**	−0.010	−0.003	**−0.004**	−0.012	0.004
Life-style										
Smoking										
Non-smoker; smoker	**−0.10**	0.016	**−0.25**	0.046	**−0.014**	−0.019	−0.010	**−0.040**	−0.048	−0.031
Smoker; smoker	**0.59**	0.045	**−0.14**	0.094	**0.107**	0.089	0.125	**0.068**	0.041	0.098
Drinking										
Do not drink;										
Drinks occasionally	**−0.18**	0.054	**0.14**	0.123	**−0.026**	−0.040	−0.011	**−0.006**	−0.031	0.024
Drinks heavily	**−0.22**	0.157	**0.01**	0.505	**−0.031**	−0.067	0.014	**−0.026**	−0.100	0.121
Drink occasionally; drinks heavily	**0.03**	0.046	**−0.09**	0.156	**0.004**	−0.009	0.019	**−0.008**	−0.041	0.033
Drug Use										
Do not use drugs; use drugs	**0.04**	0.015	**0.02**	0.033	**0.007**	0.002	0.012	**0.009**	0.002	0.017
Use drugs; use drugs	**0.06**	0.020	**0.05**	0.048	**0.009**	0.002	0.015	**0.014**	0.003	0.027

Table 4 (continued)

Variable	Model estimates				Marginal effects			
	Men		Women: differences		Men		Women	
	Est.	SE	Est.	SE	Est.	95% CI	Est.	95% CI
First-contact probability for "median" partner					**0.187**		**0.155**	
Log-likelihood	−215,369.4							
No. obs.	597,167		196,363					
First-contact e-mails	74,518		17,599					
No. of individuals	3,702		2,783					

The table shows estimates of the preferences of men and the estimated differences in the preference coefficients between women and men. Whenever two attributes are shown (e.g. "Hispanic; White"), the first one refers to the browser and the second one refers to the mate. The model includes fixed effects for each browser. Details of how the marginal effects were calculated are provided at the beginning of Section 5. Note that the marginal effects reported for women are the actual levels, not the differences with respect to men

table displays the full marginal effects for women, not the difference between men's and women's marginal effects. To calculate the marginal effects, we first obtain the median of looks, height, BMI, income, and occupation for each gender in the sample. We then consider a mate who is characterized by the gender-specific median attributes and browses the profile of a potential partner who is also characterized by his or her gender-specific median attributes, and also has the same age, education, ethnicity, religious beliefs, and so forth as the browser. For each category of attributes, we calculate the marginal effect of an attribute as the difference in first-contact probabilities across two potential mates, where one mate has that specific attribute in the category under consideration and the other mate has the base attribute in the category (the mates are identical along all other attributes). For example, the marginal effect of being in the fifth decile of looks ratings is the difference in the first-contact probabilities for a mate in the fifth decile of looks ratings relative to a mate in the first decile of looks ratings. To evaluate the relative magnitude of the marginal effects, note that the "base" first contact probability is 0.187 if a median man browses a median woman, and 0.155 if a median woman browses a median man.

5.1 Physical attractiveness vs. earnings potential

Age We allow age preferences to depend on the age of the browser and the age of the potential mate. We estimate the age preferences separately for browsers of age 30 or younger, 30 to 39, 40 to 49, and 50 or older. In each category, age preferences depend on whether the potential mate is at least 10 years younger, 5 to 9 years younger, 5 to 9 years older, or at least 10 years older. Note that this specification of age preferences is very flexible because the site users state their age in 5-year categories, such as "age 31–35," and because 85% of all profile views occur for potential mates with an (absolute) age difference of less than 15 years.

Women show a slight preference for men their age over men who are 5 to 9 years younger (the average marginal effect across all age categories is −0.032), and a strong preference for men their age over men 10 or more years younger (the average marginal effect is −0.089). Men, on the other hand, show a slight preference for women who are 5 to 9 years younger,[15] but prefer women their age to women more than 10 years younger (the average marginal effect is −0.037). Women are typically indifferent between men their age and men who are 5 to 9 years older (only for women ages 30 to 39 is the effect slightly negative and statistically significant) but prefer men of a similar age to men who are older by 10 or more years (the average marginal effect is −0.059). In contrast, men consistently dislike older women: the average marginal effects are −0.054 with respect to women 5 to 9 years older, and −0.121 with respect to women 10 or more years older. The differences between the age coefficients across genders are statistically significant (with the exception of the preference for an older mate by users of age 50 or older).

Overall, men prefer women their own age or a few years younger; women, on the other hand, want to date a man their age or a few years older. These findings are consistent with Buss (1989). However, our results are inconsistent with the finding of Kenrick and Keefe (1992) that the preferences of men for younger women become more pronounced as men become older.[16]

Looks We find that 77.6% of all profile views in our sample occur for users who posted one or more photos of themselves. As Section 2.3 describes, we obtained a numerical looks measure for these users based on ratings in a laboratory environment. We classify the users into deciles based on their looks rating, and we also split the top decile in two halves.

As expected, both men and women prefer better-looking partners: the looks coefficients are increasing in the looks categories, and the implied marginal effects are large. For example, the probability that a man contacts a woman in the 81st to 90th percentile range of looks is 10.7% higher than the probability that he contacts a woman in the lowest decile of ratings; the corresponding increase in the contact probability is 7.9% for women. Men consistently

[15] The effect is slightly positive and statistically significant for men in the 30–39 and 40–49 age groups, and statistically insignificant otherwise.

[16] Most recent speed dating papers do not report age preferences, due to the small amount of variation in age among the students that comprise many of the analyzed samples. The exception is Kurzban and Weeden (2005), who consider only preferences over the age level, but not the age relative to a potential partner. Our results, however, show that the preference for a partner's age is strongly contingent on own age.

display stronger looks preferences than women. Quantitatively, however, the difference in preferences for a partner's looks is modest: the average difference in the marginal effects between men and women is 0.017, and only five out of all 10 differences are statistically significant.

We also find that both men and women respond to the self-descriptions of those site users who did not post a photo online. Users who describe themselves as very good-looking or above average have a higher chance of receiving a first-contact e-mail than users who state that they have average or "other" looks. We find no statistically significant differences in preferences based on self-described looks across the genders.

Our findings are consistent with numerous studies, based on both stated and revealed preferences, that document a preference for physical attractiveness in general and a larger importance of attractiveness in the preferences of men than of women (Buss 1989; Buss and Schmitt 1993; Eastwick and Finkel 2008; Fisman et al. 2006; Kurzban and Weeden 2005; Regan et al. 2000).

Body mass index (BMI)[17] Men and women differ strongly in their preferences for the weight of a potential mate. Women with a higher BMI have a lower chance of receiving a first-contact e-mail. The effect sizes are large: for example, a woman with a BMI of 24 to 26 has a roughly 10% lower chance of being approached than a women with a BMI of 20 or less. Women, on the other hand, prefer men with a BMI of 20 or above to men with a lower BMI. The point estimates of women's preferences peak for a BMI range of 24 to 26 (the marginal effect is 0.052), although the effect sizes are fairly similar over BMIs ranging from 20 to 30.

In addition to these preferences concerning the level of a partner's BMI, we also find that both men and women have preferences concerning the relative BMI of a potential mate. Men somewhat prefer women with a BMI that is lower than their own[18] and dislike women with a larger BMI. Women, on the other hand, somewhat prefer men with a BMI that is larger than their own but dislike men with a lower BMI.

Overall, BMIs have a smaller effect on women's preferences than on men's preferences. This finding is roughly consistent with the results of Tovée et al. (1998) and Maisey et al. (1999). In particular, Maisey et al. (1999) show that men's waist-chest ratios—more than their BMI ratings—influence the attractiveness ratings women give men.

Height Women prefer men who are taller than themselves (the marginal effect is 0.050 for men who are 2 to 5 in. taller, and 0.071 for men who are more than 5 in. taller), whereas men prefer women shorter than themselves.

[17] If weight is measured in pounds and height is measured in inches, the BMI is calculated as BMI $=$ $(\text{weight} \cdot 703)/\text{height}^2$.

[18] More precisely, we estimate preferences over BMI differences that are at least 2 in absolute value.

Likewise, women show no interest in men who are shorter than themselves, and men show no interest in women who are taller than themselves.

The absolute effect sizes of relative height are comparable across genders, although of opposite sign. For women, we find no evidence for a preference concerning the absolute height of a potential partner. Men appear to prefer women who are 5 ft 3 or taller, but the effect sizes are small (the exception is women who are 6 ft 1 or taller—but this group contains only 32 (0.45%) of all browsed women). We therefore conclude that the height preferences of men and women are almost exclusively horizontal, that is, relative to one's own height. These results are consistent with findings on the relative height of married men and women (Gillis and Avis 1980) and the probability of having children conditional on height (Pawlowski et al. 2000; Nettle 2002). These findings, however, do not provide direct evidence on height-related mate preferences.

Income Men and women prefer a high-income partners over low-income partners. This income preference is more pronounced for women. For example, men are 3.9% more likely to contact a woman with an income in the range \$150k to \$200k than a woman who earns \$35k to \$50k per year. For women, on the other hand, the corresponding difference in first-contact probabilities is 8.9%. For both men and women, we see the biggest increase in first-contact probabilities up to the \$50k to \$75k range, and smaller increases for higher income levels. However, some caution is in order when interpreting the latter results. Only 11.1% of men and 2.2% of women in our sample report an income above \$100k. We find a fair amount of measurement error in the preference coefficients for these income levels.

Our results also show that income preferences largely depend on the absolute and not the relative level of a mate's income. The only statistically significant relative income effect is the preference of women for men who earn at least \$25k less; the corresponding marginal effect, however, is small (-0.018).

Education Another predictor of the lifetime income of a mate is his or her education level. We classify all site users into four groups based on their completed years of education: ≤ 12 (high school), > 12, and < 16 (some college), 16 (college), and > 16 (graduate school).[19] We estimate education preferences separately for the four groups and allow these preferences to depend on the education group of a potential mate. Potential partners with the same education as the browser are the excluded group.

The results show that both men and women prefer a partner with a similar education level (21 of the 24 estimated coefficients are negative; the three positive ones are tiny and insignificant). The quantitative effect of education

[19] Note that according to this scheme, users with a two-year degree or similar education are subsumed in the "some college" group.

on first-contact probabilities is stronger for women than for men. For example, men with a graduate degree are equally likely to contact a woman with a college degree, 1.6% less likely to contact a women with "some college" education, and 2.4% less likely to contact a woman with a high school degree or less. Women with a graduate degree, on the other hand, are 2.6% less likely to contact a man with a college degree, 5.3% less likely to contact a man with "some college" education, and 5.7% less likely to contact a man with high school education (or less).

Interestingly, our results show that an apparent preference for similarity along education overwhelms any potential role of education as a signal of earnings potential. For example, both men and women with high school or "some college" education are less likely to contact a potential mate with a graduate degree compared to a partner with a similar education level.

Occupation Men show little preference for one occupation over another in their potential mates. The occupation that tends to draw the most attention relative to students, the omitted base occupation, is "Artistic/Musical/Writer." Even then, however, the corresponding marginal effect is small at only 0.033.

The occupation of a male partner, on the other hand, strongly influences women's preferences in some instances. For example, relative to their likelihood of approaching students, women are 8.6% more likely to approach a man described as "Legal/Attorney," 7.7% more likely to approach men in the occupation "Law enforcement/Fire fighter," 6.7% more likely to approach men in the military, and 5% more likely to approach men in the health profession. Although some of these occupation effects may reflect a preference for earnings potential not captured by the stated income level, others (such as the preferences for men in law enforcement or fire fighters) are unlikely to be income related.

Both men and women also prefer a partner with the same occupation, although the quantitative significance of this effect is small.

5.2 Preferences for race or ethnicity

Few studies report evidence on race preferences. Obviously, people may be unwilling to *state* that they would not want to date or marry a partner of a different race, which is a likely reason for the relative dearth of evidence on race preferences. Hence, a revealed preference approach is of particular importance to assess mate preferences for race or ethnicity. We present evidence below comparing revealed and stated race preferences that support this point. The most extensive extant study on race preferences obtained from revealed partner choices is Fisman et al. (2008). Compared to their study, we base our results on a more representative and larger sample; this data aspect is particularly important as, due to their small sample size, Fisman et al. control only for physical attractiveness apart from race in their preference specification.

We classify all the site users in our sample as white, black, Hispanic, Asian, or "other," based on the ethnicity they state in their profile. White users include those who state "I'm White/Caucasian" and those who indicate their ethnicity is French, German, Danish, and so on. Asian users are those who state "I'm Asian;" we do not include in this group users who indicate they are East Indian. The group "other" includes users who state they are Aborigine, African, Arabic, East Indian, Native American, Polynesian, or some "Other ethnic variance." Of all users making first-contact decisions, 4.4% belong to this group. We separately estimate race preferences for white, black, Hispanic, and Asian site users. The race preferences depend on which ethnic group a potential mate belongs to; mates of the same ethnicity as the browser are the excluded category.

Same-race preferences Both men and women have same-race preferences: 30 of the 31 estimated race coefficients are negative. Some of the implied marginal effects are large: the chance of a white man contacting a black woman is 10% lower than the chance of him contacting a white woman, and the chance of a white woman contacting an Asian man is 12% lower than the chance of her contacting a white man. We estimate the race coefficients precisely for the white members of the dating site. Due to the smaller number of observations, however, we estimate the race coefficients less precisely for black, Hispanic, or Asian members (six out of 12 estimated coefficients for non-white men and three out of 11 coefficients for non-white women are statistically significant). Therefore, judging whether differences in race preferences across ethnic groups exist is difficult, although the point estimates of the marginal race effects are of comparable magnitude for the different ethnicities. Also, selection may exist in the sense that non-white members with strong same-race preference join a dating site that specifically caters to their own ethnicity. Thus, our results are unlikely to provide conclusive evidence indicating whether and how same-race preferences differ across ethnic groups.

Examining gender differences among the white members of the dating site, we find that—with the exception of preferences for a black partner—the race coefficients of women are more negative, and the differences between the coefficients of men and women are statistically significant. For non-white site members, nine out of the 11 estimated race coefficients are more negative for women than for men, although only one of these differences is statistically significant. These findings suggest that women have more pronounced same-race preferences than men.

Differences in race preferences across users with different attributes Do race preferences differ across site users with different observable attributes? In order to study this question, we create a categorical variable indicating whether a potential mate is of a different ethnicity than the browser, and interact this variable with various observed browser characteristics. In order to reduce the number of estimated coefficients, we do not separately study race preferences for different ethnicities, and we also do not allow for race preferences to

Table 5 Same-race preferences and marginal effects across different user groups

Variable	Race interactions				Interactions marginal effects					
	Men		Women		Men			Women		
	Est.	SE	Est.	SE	Est.	95% CI		Est.	95% CI	
(a)										
Mate: different ethnicity	**−0.375**	0.103	**−0.856**	0.342	**−0.051**	−0.073	−0.025	**−0.086**	−0.121	−0.023
Interaction of "different ethnicity" with:										
Age	**0.000**	0.002	**0.004**	0.005	**0.000**	−0.001	0.001	**0.001**	−0.001	0.002
Years of education	**−0.001**	0.006	**−0.003**	0.020	**0.000**	−0.002	0.002	**0.000**	−0.006	0.005
Income	**0.001**	0.000	**0.002**	0.003	**0.000**	0.000	0.000	**0.000**	0.000	0.001
Log-likelihood	−166,817.7		−48,685.3							
(b)										
Mate: different ethnicity	**−0.301**	0.061	**−0.608**	0.184	**−0.042**	−0.056	−0.026	**−0.067**	−0.093	−0.031
Interaction of "different ethnicity" with:										
Age 30–50	**0.036**	0.038	**0.082**	0.117	**0.006**	−0.006	0.018	**0.011**	−0.019	0.047
Age 50–	**−0.202**	0.072	**0.073**	0.206	**−0.029**	−0.047	−0.009	**0.010**	−0.040	0.075
Education: some college	**−0.184**	0.059	**−0.102**	0.192	**−0.027**	−0.042	−0.010	**−0.013**	−0.055	0.041
Education: college	**−0.076**	0.055	**−0.252**	0.193	**−0.011**	−0.027	0.005	**−0.031**	−0.068	0.018
Education: graduate school	**−0.116**	0.057	**−0.039**	0.197	**−0.017**	−0.032	−0.001	**−0.005**	−0.050	0.053
Income > $70k	**0.106**	0.035	**0.237**	0.215	**0.017**	0.006	0.028	**0.035**	−0.024	0.110
Log-likelihood	−166,800.7		−48,683.2							

(c)										
Mate: different ethnicity	**−0.282**	0.072	**−0.717**	0.238	**−0.039**	−0.056	−0.021	**−0.075**	−0.105	−0.031
Interaction of "different ethnicity" with:										
Age 30–50	**0.046**	0.038	**0.102**	0.117	**0.007**	−0.004	0.019	**0.014**	−0.017	0.050
Age 50–	**−0.197**	0.072	**0.091**	0.207	**−0.028**	−0.046	−0.008	**0.013**	−0.038	0.079
Education: some college	**−0.185**	0.059	**−0.114**	0.192	**−0.027**	−0.042	−0.010	**−0.015**	−0.056	0.039
Education: college	**−0.076**	0.055	**−0.262**	0.194	**−0.011**	−0.027	0.005	**−0.032**	−0.069	0.017
Education: graduate school	**−0.133**	0.057	**−0.089**	0.199	**−0.019**	−0.035	−0.003	**−0.012**	−0.055	0.045
Income > $70k	**0.108**	0.035	**0.228**	0.215	**0.017**	0.006	0.028	**0.033**	−0.024	0.108
Political orientation: conservative	**−0.124**	0.047	**−0.304**	0.184	**−0.018**	−0.031	−0.005	**−0.037**	−0.071	0.008
Political orientation: liberal	**0.119**	0.048	**0.165**	0.119	**0.019**	0.004	0.035	**0.023**	−0.009	0.061
Religion: Christian	**−0.027**	0.044	**0.077**	0.170	**−0.004**	−0.017	0.009	**0.011**	−0.032	0.063
Religion: other	**−0.024**	0.046	**0.181**	0.173	**−0.004**	−0.017	0.010	**0.026**	−0.020	0.083
Log-likelihood	−166,792.6		−48,679.3							
(d)										
Mate: different ethnicity	**−0.053**	0.166	**−0.840**	0.427	**−0.008**	−0.051	0.045	**−0.083**	−0.124	0.000
Interaction of "different ethnicity" with:										
Stated pref. for same ethnicity	**−0.966**	0.173	**−0.185**	0.436	**−0.106**	−0.127	−0.077	**−0.023**	−0.096	0.110
Stated pref. for different ethnicity	**1.011**	0.239	**1.478**	0.622	**0.200**	0.096	0.314	**0.293**	0.038	0.577
Stated pref. for "different species"	**−0.191**	0.167	**0.347**	0.431	**−0.027**	−0.066	0.022	**0.052**	−0.055	0.224
Log-likelihood	−166,662.5		−48,669.8							
No. obs.	597,167		196,363							
No. of individuals	3,702		2,783							

The first column shows the preferences of men, and the second column shows the preferences of women. The estimated models also include all other variables that are not related to race as reported in Table 4. The model includes fixed effects for each browser depend on the ethnicity of the browser. We report the results in Table 5 (although not reported, the estimated models also include all other variables reported in Table 4). All specifications allow for a main effect if a potential mate is of a different ethnicity than the browser. Our discussion focuses on preference differences across site users with different attributes conditional on gender. Therefore, we report the estimated preference coefficients for men and women separately, not the difference between women's and men's coefficients as in Table 4.

In model (a), we interact the "different ethnicities" dummy with the browser's age, education level (in years of completed schooling), and income. The main effect of "different ethnicities" is negative, just as we found before. However, all interactions, apart from the income interaction for men, which is small, are statistically insignificant. Hence, the results from this model specification reject the notion that race preferences differ across either men or women who differ in age, education, or income. In model (b), we interact the "different ethnicities" indicator with categorical variables indicating a specific age range of the browser (age 30 to 49, age 50+), a specific education level (some college, college, or graduate school), and an income level of $70k or above. For women, we find no evidence for statistically significant differences in race preferences across the different attribute categories. For men, on the other hand, we find some differences. Men of age 50+ have somewhat stronger same-race preferences than younger men. Men with "some college" or graduate education, but not college-educated men, have somewhat stronger same-race preferences than men with no college education; thus, no clear evidence for a *systematic* relationship between race preferences and education for men exists. Men with an income of $70k or above have weaker same-race preferences than men with less income. Although these estimated differences are statistically significant, Table 5 shows that the corresponding marginal effects are small. Therefore, based on our estimates, men's race preferences differ little across demographic attributes.

In model (c), we add additional interactions related to the political and religious views of the site users. First, we interact the "different ethnicities" indicator with categorical variables classifying the users as "conservative" (12.3% of the browsers in the sample) or "liberal" (18% of browsers). The omitted category contains users whose political views are not easily classifiable or who do not declare any political views. About 70% of all browsers in our sample are part of this category; correspondingly, the users we classify as "conservative" or "liberal" may be those who hold unusually strong political views on either side of the political spectrum, and we cannot simply equate them with Republicans or Democrats or other well-defined political groups in the USA. Second, we examine interactions with categorical variables classifying users as "Christian" or of some "other religion;" site users who declare themselves to be agnostic or state that religion plays no role in their life constitute the omitted category. We find that men classified as "conservative" have stronger same-race preferences, whereas men classified as "liberal" have weaker same-race preferences. However, the corresponding marginal effects are small. For women, we find no evidence for differences in race preferences related to their political views.[20] We also do not find evidence that race preferences differ by religion, both for men and women.

While there are statistically significant differences in race preferences across men with different attributes and no corresponding statistically significant differences for women, we cannot reject that the preference estimates for men and women are equal. The reason is that the race preference coefficients for women, in particular, are estimated imprecisely. Consequently, a more cautious interpretation of the findings above is that while race preferences differ slightly across men with different observed attributes, the data provide insufficient information to judge if there are similar differences in women's race preferences or not.

Revealed versus stated race preferences In model (d) (Table 5) we examine whether revealed race preferences and the users' stated preferences for dating a mate of a different ethnicity coincide. In our sample of browsers, 80.3% of men and 54.7% of women state that the ethnic background of their partner "doesn't matter." On the other hand, 17% of men and 41.6% of women state that they prefer a partner whose ethnicity is "the same as mine." Furthermore, 1.1% of men and 2.3% of women prefer a partner with a different ethnicity, whereas 1.7% of men and 1.5% of women prefer "a different species." To investigate whether revealed and stated preferences are related, we estimate the preferences for a partner of a different ethnicity separately for different groups of site users, defined by their answer to what ethnic background they seek in a partner. For men in the omitted group of users with no stated ethnicity preference, our estimates show no evidence for same-race preferences. On the other hand, men who want to date a partner of the same ethnicity strongly discriminate against potential mates with a different ethnic background, and men who state they want to date someone with a different ethnicity indeed strongly favor partners with a different ethnicity. Women who declare no preference for the ethnicity of a partner, however, nonetheless reveal strong same-race preferences. In fact, the difference in same-race preferences between women who do and those who do not declare a preference for a partner of the same race is not statistically significant. On the other hand, women who state that they seek a partner of a different ethnicity also behave accordingly. For users without and for users with a stated preference for a partner of the same ethnicity, we can reject equality of men's and women's preference estimates at the 10% significance level.

In summary, we find relatively few men with a stated preference for a partner of the same ethnicity in our sample (17%), and we find that only this

[20]Using data from speed dating events, Eastwick et al. (2009) find that among whites, the relative preference for a white partner over a black partner is stronger for conservative than for liberal speed dating participants, while relatively conservative blacks have a stronger relative preference for a white than for a black partner compared to liberal blacks. Their study, however, does not report gender differences.

group of men actually discriminates against women of a different ethnicity (the estimated marginal effect is −0.106). On the other hand, we find that both women who have no declared race preference (54.7%) and those who do (41.6%) have an equally strong same-race preference (the corresponding marginal effect is −0.083).

5.3 Preferences for similarity

So far, we have found that the site users have a preference for a partner who has a similar education level or is of the same ethnicity. Here, we document preferences for similarity along several other attributes.

Regarding their marital status, single users avoid divorced users, whereas divorced users have a preference for a partner who is also divorced. Users who seek a long-term relationship prefer a partner who states the same goal. Users with children prefer a partner who also has children, whereas users without children avoid such mates.

Regarding religion, we find that non-Catholic Christians prefer other non-Catholic Christians over Catholics, members with a different religion, and non-religious members. Similarly, Catholics prefer other Catholics over non-Catholic Christians, members of other religions, and non-religious members. Finally, non-religious members tend to avoid Christian site members or members of other religions. Hence, most site members prefer to date a partner with similar beliefs. The effect sizes are small to moderate and somewhat more pronounced for women.

Similarity in political views also matters: among the site members in our sample, conservatives dislike liberals and vice versa (but note again that only about 30% of the users classify themselves as conservative or liberal and hence may represent people with unusually strong political views).

Non-smokers do not want to date smokers, and smokers have a strong preference for partners who smoke.

6 Summary and conclusions

We can divide our substantive results regarding revealed mate preferences into three main groups related to the main questions posed in the introduction.

Determining whether preferences are horizontal or vertical is important to understand the sources of the observed assortative mating patterns among dating or married couples (e.g. Hitsch et al. 2010). We find that vertical preferences alone cannot easily characterize mate preferences; horizontal preferences, and preferences for similarity in particular, play an important role. Although physical attractiveness and income are largely vertical attributes, preferences concerning a partner's age, education, race, and height reflect a strong preference for similarity. Our finding that preferences for education have important horizontal components qualifies some existing theories of mate selection. To the extent that education provides information on an individual's future earnings potential (in addition to the information provided by current income), horizontal preferences for education are not fully consistent with theories of mate selection based on earnings capacity (e.g. Buss 1989). We also find a preference for similarity along religion, political views, and "complementary" habits, such as smoking behavior.

Perhaps our most novel results concern same-race preferences, where the extant knowledge is comparatively scarce. We find that both men and women have same-race preferences (the same-race preference estimates of Fisman et al. (2008) are not statistically significant, which is likely due to their small sample size) and that the same-race preferences of women are more pronounced than the same-race preferences of men. The same-race preferences of women do not differ by age, education, or income. For men, we find some differences in same-race preferences across demographics, but the quantitative significance of these effects is small. The absence of a clear relationship between same-race preferences and education, in particular, is one of our most surprising results.

We also find that there are differences between the stated and revealed preferences for a partner's race or ethnicity. However, these differences between stated and revealed preferences are not the same across men and women. For men, we find that the revealed preference for a partner of a different ethnicity coincides with the corresponding stated preference. Women, however, reveal preferences that do not reflect their stated preferences: we estimate statistically and quantitatively significant same-race preferences for women both with and without a declared same-ethnicity preference, and the difference in these same-race effects is not statistically significant. These findings underline the importance of utilizing revealed preference methods when studying mate preferences over race or ethnicity.

Finally, we provide detailed evidence on gender differences in revealed mate preferences. Compared to previous revealed preference studies, we employ a much larger sample and control for a large number of own and partner attributes, which is necessary to guard against omitted variable bias. Our results are overall consistent with many of the previous findings in psychology and economics (e.g., Buss 1995; Eastwick and Finkel 2008; Fisman et al. 2006; Kurzban and Weeden 2005; Regan et al. 2000). In particular, women have a stronger preference for income relative to physical attributes, such as facial attractiveness, height, or body mass index. These results are consistent with predictions from evolutionary psychology (Buss 1989; Buss and Schmitt 1993) and the competing social structure theory (Eagly and Wood 1999).

An important methodological contribution of our paper is the discussion of strategic behavior. Both a simple test for strategic behavior and a more elaborate bias correction method do not provide evidence that the site users strategically "shade" their true preferences. To the extent that this result holds more generally, not just in the context of the particular dating site studied in this paper, it provides additional support for the validity of other revealed preference estimates, in particular those obtained from speed-dating experiments.

Acknowledgements We thank Babur De los Santos, Chris Olivola, Tim Miller, and David Wood for their excellent research assistance. We are grateful to Elizabeth Bruch, Jean-Pierre Dubé, Eli Finkel, Emir Kamenica, Derek Neal, Peter Rossi, Betsey Stevenson, and Utku Ünver for comments and suggestions. Seminar participants at the 2006 AEA meetings, Boston College, the Caltech 2008 Matching Conference, the Choice Symposium in Estes Park, the Conference on Marriage and Matching at New York University 2006, the ELSE Laboratory Experiments and the Field (LEaF) Conference, Northwestern University, the 2007 SESP Preconference in Chicago, SITE 2007, the University of Pennsylvania, the 2004 QME Conference, UC Berkeley, UCLA, the University of Chicago, UCL, the University of Naples Federico II, the University of Toronto, Stanford GSB, and Yale University provided valuable comments. This research was supported by the Kilts Center of Marketing (Hitsch), a John M. Olin Junior Faculty Fellowship, and the National Science Foundation, SES-0449625 (Hortaçsu).

References

Adachi, H. (2003). A search model of two-sided matching under nontransferable utility. *Journal of Economic Theory, 113*, 182–198.

Banerjee, A. V., Duflo, E., Ghatak, M., & Lafortune, J. (2009). Marry for what? Caste and mate selection in modern India. Manuscript (MIT)

Becker, G. S. (1973). A theory of marriage: Part I. *Journal of Political Economy, 81*(4), 813–846.

Biddle, J. E., & Hamermesh, D. S. (1998). Beauty, productivity, and discrimination: Lawyers' looks and lucre. *Journal of Labor Economics, 16*(1), 172–201.

Browning, M., Chiappori, P.-A., Weiss, Y. (2008). The economics of the family. http://www.tau.ac.il/~weiss/fam_econ/.

Burdett, K., & Coles, M. G. (1997). Marriage and class. *Quarterly Journal of Economics, 112*(1), 141–168.

Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and Brain Sciences, 12*, 1–49.

Buss, D. M. (1995). *The evolution of desire: Strategies of human mating*. New York: Basic Books.

Buss, D. M., & Schmitt, D. P. (1993). Sexual strategies theory: An evolutionary perspective on human mating. *Psychological Review, 100*(2), 204–232.

Choo, E., & Siow, A. (2006). Who marries whom and why. *Journal of Political Economy, 114* (1), 175–201.

Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior. *American Psychologist, 54*(6), 408–423.

Eastwick, P. W., Finkel, E. J. (2008). Sex differences in mate preferences revisited: Do people know what they initially desire in a romantic partner? *Journal of Personality and Social Psychology, 94*(2), 245–264.

Eastwick, P. W., & Finkel, E. J., Richeson, J. A., Son, D., & Finkel, E. J. (2009). Is love colorblind? Political orientation and interracial romantic desire. *Personality and Social Psychology Bulletin, 35*(9), 1258–1268.

Etcoff, N. (1999). *Survival of the prettiest: The science of beauty*. New York: Doubleday Books.

Finkel, E. J., & R. F. Baumeister (2010). Attraction and rejection. In R. F. Baumeister & E. J. Finkel (Eds.), *Advanced social psychology*. New York: Oxford University Press.

Fisman, R., Iyengar, S. S., Kamenica, E., & Simonson, I. (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *Quarterly Journal of Economics, 2*, 673–697.

Fisman, R., Iyengar, S. S., Kamenica, E., & Simonson, I. (2008). Racial preferences in dating. *Review of Economic Studies, 75*, 117–132.

Gillis, J. S., & Avis, W. E. (1980). The male-taller norm in mate selection. *Personality and Social Psychology Bulletin, 6*(3), 396–401.

Goldberger, A. S. (1983). Abnormal selection bias. In S. Karlin, T. Amemiya, & L. A. Goodman (Eds.), *Studies in econometrics, time series, and multivariate statistics*. New York: Academic.

Hamermesh, D. S., & Biddle, J. E. (1994). Beauty and the labor market. *American Economic Review, 84*(5), 1174–1194.

Heckman, J. J. (1979). Sample selection bias as specification error. *Econometrica, 47*(1), 153–162.
Heckman, J. J. (1990). Varieties of selection bias. *American Economic Review, 80*(2), 313–318.
Hitsch, G. J., Hortaçsu, A., & Ariely, D. (2010). Matching and sorting in online dating. *American Economic Review, 100*(1), 130–163.
Kalmijn, M. (1998). Intermarriage and homogamy: Causes, patterns, trends. *Annual Review of Sociology, 24*, 395–421.
Kenrick, D. T., & Keefe, R. C. (1992). Age preferences in mates reflect sex differences in human reproductive strategies. *Behavioral and Brain Sciences, 15*, 75–133.
Kurzban, R., & Weeden, J. (2005). HurryDate: Mate preferences in action. *Evolution and Human Behavior, 26*, 227–244.
Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin, 126*(3), 390–423.
Lee, S. (2009). Marriage and online mate search services: Evidence from South Korea. manuscript (University of Maryland).
Little, R. J. A. (1985). A note about models for selectivity bias. *Econometrica, 53*(6), 1469–1474.
Maisey, D. S., Vale, E. L. E., Cornelissen, P. L., & Tovée, M. J. (1999). Characteristics of male attractiveness for women. *Lancet, 353*, 1500.
Nettle, D. (2002). Women's height, reproductive success and the evolution of sexual dimorphism in modern humans. *Proceedings of the Royal Society B, 269*, 1919–1923.
Newey, W. K., Powell, J. L., & Walker, J. R. (1990). Semiparametric estimation of selection models: Some empirical results. *American Economic Review, 80*(2), 324–328.
Pawlowski, B., Dunbar, R. I. M., & Lipowicz, A. (2000). Tall men have more reproductive success. *Nature, 403*, 156.
Regan, P. C., Levin, L., Sprecher, S., Christopher, F. S., & Gate, R. (2000). Partner preferences: What characteristics do men and women desire in their short-term sexual and long-run romantic partners? *Journal of Psychology & Human Sexuality, 12*(3), 1–21.
Sprecher, S., Sullivan, Q., & Hatfield, E. (1994). Mate selection preferences: Gender differences examined in a national sample. *Journal of Personality and Social Psychology, 66*(6), 1074–1080.
Todd, P. M., Penke, L., Fasolo, B., & Lenton, A. P. (2007). Different cognitive processes underlie human mate choices and mate preferences. *Proceedings of the National Academy of Sciences, 104*(38), 15011–15016.
Tovée, M. J., Reinhardt, S., Emery, J. L., & Cornelissen, P. L. (1998). Optimum body-mass index and maximum sexual attractiveness. *Lancet, 352*, 548.
Wong, L. Y. (2003). Structural estimation of marriage models. *Journal of Labor Economics, 21*(3), 699–727.