

Who’s Right and Who Writes: People, Profiles, Contacts, and Replies in Online Dating

Andrew T. Fiore¹, Lindsay Shaw Taylor², Xiaomeng Zhong¹, G.A. Mendelsohn², and Coye Cheshire¹

¹*School of Information
University of California, Berkeley*
{atf, zhongxm, coye}@ischool.berkeley.edu

²*Department of Psychology
University of California, Berkeley*
{lstaylor, jermend}@berkeley.edu

Abstract

In this analysis of profiles and messaging behavior on a major online dating service, we find that, consistent with predictions of evolutionary psychology, women as compared to men state more restrictive preferences for their ideal date. Furthermore, women contact and reply to others more selectively than men. Additionally, we identify connections among messaging behavior, textual self-descriptions in dating profiles, and relationship-relevant traits such as neuroticism.

1. Introduction

“Thousands of boys and girls who’ve never met plan weekends together, for now that punch-card dating’s here, can flings be far behind? And oh, it’s so right, baby. The Great God Computer has sent the word. Fate. Destiny. Go-go-go.”

— *Look Magazine, February 1966*

In the four decades since the advent of computer match-making, its modern successors — online dating systems — have been embraced by users around the world as an expedient way to find a date or meet a romantic partner. Almost 7 million adults in the U.S. alone have gone on a date with someone they met through an online dating Web site [21].

Researchers have studied behavior in computer-mediated communication (CMC) since its early days, but few studies have examined messaging behavior in a real-world online dating system. Online dating systems differ from most other forms of CMC in that they seek to help strangers find one another with the intent of meeting offline and perhaps forming a face-to-face relationship. Both the promise and the peril of online dating lies in the sheer number of potential partners a user encounters on a typical site: who contacts whom from a virtual catalog of potential mates, with detailed information about each person? Which qualities and behaviors are associated with the

likelihood of receiving a response from a contact? How do self-descriptions in profiles relate to personality traits and tendencies? In this paper we examine these questions using personal profiles, behavioral data, and questionnaire responses from users of a major U.S. online dating service.

2. Background

We begin by reviewing research relevant to contact and reciprocation in online dating. First, we describe social psychological studies of attraction and mate selection offline, which offer a theoretical and empirical foundation for the consideration of these phenomena when they begin online. Next, we provide an overview of the body of research that examines self-presentation and communication specifically in the context of online dating.

2.1. Psychology of attraction and mate selection

Social psychologists have adopted several distinct approaches to understanding mating behavior. These theoretical frameworks, which include evolutionary and sociocultural perspectives [8], seek to explain modal tendencies in the population, not to provide hard and fast rules for individual behavior. Research rooted in an evolutionary tradition suggests that characteristics that promote reproduction play a crucial role in attraction and mate selection. Specifically, healthy men with the resources to provide for a family should be highly attractive to women, while young, healthy women who are better able to bear children should be most attractive to men [5][17]. Markers of these characteristics, like occupation and height for men, or unwrinkled skin and good muscle tone for women [5], are therefore salient to people seeking mates offline and perhaps online as well.

Psychological factors such as personality traits, intelligence, and disposition toward intimate partners also affect partner choice and relationship satisfaction

[3][6][10]. On many of these characteristics, similarity to oneself has been shown to be attractive, and established couples often exhibit better-than-chance similarity on these dimensions [22]. The tendency of people to partner with mates similar to themselves — sometimes called “positive assortative mating,” or homophily — is evident with regard to physical attractiveness, socioeconomic status, ethnicity, attachment style, personality traits, and more [19][20][28][32]. Some assortment occurs on age as well, but on this dimension evolutionary psychology principles compete with the tendency toward assortative mating. Across cultures, men prefer younger women, who are more likely to be fertile, while women prefer older men, who are more likely to have the resources to support a family [5].

Finally, evolutionary psychology suggests that women will be pickier than men about finding a mate who satisfies their criteria for a partner, because women must make a much greater parental investment than men. Thus, it is important for women to pick a high-quality partner to conceive and support healthy children (a long-term strategy). Men, on the other hand, should be less choosy, because they can father many children without much parental involvement (a short-term strategy) [17][27].

2.2. Online dating

Online dating systems are typically Web sites that allow users to post personal profiles, view the profiles of others, and send private messages to them. Profiles usually include one or more photographs, basic demographics and categorical descriptors such as body type and style of humor, and a free-text self-description. Most sites also ask users to indicate the type of person they are seeking, at least in terms of basic demographic characteristics. Typically, sites allow users to browse the profiles of people in geographic proximity and to search other people’s profiles with a variety of detailed criteria. Some sites offer a “matching” feature, which suggests compatible partners according to a ranking algorithm [11].

2.3.1. People, profiles, and perceptions. Although online dating profiles contain a wealth of information, users sometimes find it difficult to form impressions of people from their profiles alone and to select partners from the many available options. Interviews with online dating users by Ellison and colleagues [9] found that users make substantial inferences from small cues, reading between the lines of a profile to glean the information they need to assess a potential date. Some theorists hold that given

the relative poverty of informational and social cues available through CMC, people fill in the blanks with what they would like to find, a sort of wishful thinking that leads to greater affinity than would be likely if they knew more [30][31].

In terms of how people select among the many available profiles, Fiore and colleagues [13] found that the attractiveness of the photo in an online dating profile was the most important predictor of the profile’s overall attractiveness. Although the appeal of the free-text self-description also contributed to judgments of a profile’s attractiveness, overall, individuals seem to prefer profiles with attractive pictures to profiles with attractive text [13][26].

Lastly, Alterovitz and Mendelsohn [1] examined whether predictions of evolutionary psychology would hold throughout the lifespan in online dating profiles. Indeed, they found that men on average sought women younger than themselves, no matter their own age; women through middle age sought older men, but older women sought somewhat younger men. Furthermore, men provided more information about their social status, and women of all ages were pickier than men in their stated preferences for a mate.

2.3.2. Communication among online dating users.

Who contacts whom? One study of communication patterns in online dating demonstrated positive assortment on a wide variety of characteristics among communicating pairs. Online daters contacted people like themselves in terms of ethnicity, religion, being previously married, smoking habits, and drinking habits more than would be predicted by chance [12].

However, Hitsch and colleagues [18] found that online dating users on one site did not demonstrate assortment by physical attractiveness. Rather, they were more likely to contact attractive people regardless of their own attractiveness, though attractive message recipients responded more readily to suitors who were also attractive [18]. Similarly, Shaw Taylor and colleagues showed in a lab study that although participants with high self-worth preferred attractive online dating profiles *more strongly* than did participants with low self-worth, all preferred more attractive profiles to less attractive profiles [26].

3. Research questions

In this work, we address several research questions regarding both stated preferences for an ideal date and actual patterns of communication in a real online dating system. (1) What is the relationship between the sociodemographic characteristics of initiators and recipients of online dating messages?

(2) How do men and women differ in their self-descriptions? How are these self-descriptions associated with relationship-relevant psychosocial constructs? (3) What characteristics are associated with contacting and being contacted by others? With replying to another's initial message?

4. Methodology

To address our research questions, we analyzed several overlapping data sets obtained through a collaboration with a major U.S. online dating Web site. This dating site allows users to browse and search the profiles of others, and it also provides a matching mechanism. Its private messaging feature works much like Web-based email. The data we use in this work derive from profiles and private messaging activity on the dating site in the 9-month period from early April 2008 to early January 2009. We extracted most of the data from log files and users' profiles, anonymized it, and compiled it in a data warehouse. We also administered a questionnaire to 1,100 users of the dating site to glean additional information that was not available from the dating site itself.

We used data organized in two ways: by person, so that each observation is one individual, and by pairwise contact, so that each observation is one pair of people in which one sent a message to the other, who might or might not have replied. (We hesitate to call them "dyads" since most contacts were unreciprocated.) The site we studied is primarily oriented toward heterosexuals, so we considered only heterosexual contacts in this work.

Table 1 summarizes the variables describing people and pairwise contacts. In the following sections, we discuss the variables and their organization into various data sets in more detail.

4.1. Variables describing individuals

We gathered three types of information about individuals: messaging statistics, which we obtained for all users of the site during the data-collection period; profile characteristics, which we obtained from the online dating profiles of 11,600 users; and questionnaire responses, including a series of psychometric scales, which we obtained for 1,100 users. All of these variables are described briefly in Table 1, but a few points deserve elaboration.

4.1.1. Messaging statistics. The primary quantity we consider in computing messaging statistics is not the number of messages sent or received but the number of *distinct people* whom a person contacts or is

contacted by. This places the focus not on how many messages a pair exchanges, but rather on who initiated the contact. In particular, one key variable in many of our analyses is *popularity*. A person's popularity is indexed by the average number of people who initiate contact with him or her per day that he or she is active on the site. We believe, following Fiore and Donath [12], that this measure serves as a reasonable proxy for overall attractiveness as well, as we expect that more attractive people will, on average, receive more unsolicited attention than less attractive people. The proportion of people who reply to one's initial contacts is another potential proxy for attractiveness.

Since users of the site were free to join and leave during the data-collection period, we sought to compute relevant statistics as per-day averages rather than absolute counts. To do this, we normalized the number of contacts initiated and received by the number of days on which a user was active. Since we were not able to obtain information about when users merely logged in, we counted active days as those on which a user sent at least one message. This metric is an approximation which, if anything, *undercounts* the number of days on which a user actually logged in, since a user could log in without sending a message but not the reverse. Nonetheless, we think it is a reasonable choice, since we applied it consistently and believe that the approximation error is not systematically greater for any particular group of users.

4.1.2. Profile characteristics. An important but potentially confusing quantity is the proportion of sought preferences specified in a profile. Most online dating profiles, including the ones on the site we studied, allow each user to specify the characteristics he or she is seeking in a partner using categorical descriptors. This permits the user to select, for example, one or more preferred ethnicities from a list of possible ethnicities. However, for any of the characteristics, the user may also choose "Any" to indicate that any of the possibilities is acceptable. The variable *sought preferences specified*, then, indicates for how many characteristics a person stated a preference rather than choosing "Any."

We also consider the textual self-description each person provides in his or her profile. These statements are rich sources of information about their authors, and an appealing self-description has been shown to correlate with overall profile attractiveness [13]. To analyze the 11,600 profiles we obtained, we turned to the Linguistic Inquiry and Word Count (LIWC) software, a well-validated tool for identifying the percentage of words in a passage of text that fall into various syntactical and semantic

Table 1. Overview of variables with brief descriptions.

VARIABLES DESCRIBING INDIVIDUALS	
MESSAGING STATISTICS — AVAILABLE FOR ALL USERS OF THE DATING SITE	
Messages sent per day	<i>Average number of messages sent per day active on site</i>
Messages received per day	<i>Average number of messages received per day active on site</i>
Contacts initiated per day	<i>Average number of people this person contacted per day active on site</i>
Contacts received per day (Popularity)	<i>Average number of people who contacted this person per day active on site</i>
Reply proportion	<i>Proportion of initial contacts from others to which this person replied</i>
Replied-to proportion	<i>Proportion of initial contacts from this person to which others replied</i>
PROFILE CHARACTERISTICS — AVAILABLE FOR 11,600 USERS OF THE DATING SITE	
Gender, age, ethnicity (own and sought)	<i>Own demographics and those of the type of person sought</i>
Sought preferences specified	<i>Proportion of ideal date’s attributes for which this person stated a preference</i>
Self-description word count	<i>Number of words in free-text self-description on profile</i>
Self-description % words in categories: home, work, money, sex, emotion	<i>Percentage of words in self-description associated with each category according to the Linguistic Inquiry and Word Count software (LIWC) [25]</i>
QUESTIONNAIRE RESPONSES — AVAILABLE FOR 1,100 USERS OF THE DATING SITE	
Interest in serious relationship	<i>Interest in serious dating, marriage, civil union (Likert-type scale)</i>
Big Five personality traits: Neuroticism, Openness, Extraversion, Conscientiousness, Agreeableness	<i>Personality traits assessed with the Ten-Item Personality Inventory [16]</i>
Romantic relationship disposition: Anxiety, Avoidance	<i>Adult attachment style dimensions assessed with the Experiences in Close Relationships–Revised instrument [15]</i>
Dispositional trust and caution	<i>Assessed with Yamagishi’s instrument [33]</i>

VARIABLES DESCRIBING PAIRWISE CONTACTS	
Reciprocated	<i>Whether the recipient of the initial contact ever replied</i>
Messages from initiator and recipient	<i>Number of messages each party sent</i>
Distance	<i>How far apart in miles initiator and recipient were (by postal code)</i>
Duration	<i>How long the interaction lasted</i>
Time until first reply	<i>Hours from initial contact until first reply, if any</i>
Simultaneous interactions	<i>How many other people the initiator and recipient each contacted and were contacted by in the week following the initial contact</i>

categories [25]. For this analysis, we considered those categories evolutionary psychology suggests should be most relevant to mate selection: home, work, money, and sex. We also included emotion-related words and tentative language, since they pertain to intimate relationships, and men and women are known to use them differently [4][24].

4.1.3. Questionnaire responses. Our questionnaire, answered by approximately 1,100 users of the site as part of a larger study of online dating behavior, includes well-validated psychometric constructs that provide useful information about the disposition of respondents in social situations and relationships.

In particular, we included the Ten-Item Personality Inventory, a short measure of the Big Five personality traits [16]; the Experiences in Close Relationships–Revised instrument, which assesses adult attachment style [15]; and a general trust and caution scale developed by Yamagishi [33].

The Big Five personality traits have been shown to be related to both the number and quality of one’s personal relationships [2]. Adult attachment style concerns a person’s behavior in close relationships, particularly romantic but also familial. Those who score low in anxiety and avoidance are said to have a “secure” attachment style in such relationships, neither worrying about abandonment nor avoiding intimacy, whereas individuals who score high on one

or both dimensions tend to have difficulty forming and maintaining close relationships [15]. The general trust scale measures default expectations of others on two dimensions, trust and caution. General trust is negatively correlated with general caution, which deals with one's tendency to act in a guarded or careful manner with others. High trusters have a lower threshold for interacting with others in risky and uncertain environments. On the other hand, low trusters tend to prefer the security of known, stable relationships over the potential rewards of new opportunities [33].

4.2. Variables describing pairwise contacts

The second part of Table 1 provides an overview of the variables in our data that describe messaging contact among pairs of users. Note that the binary variable *reciprocated* indicates whether the recipient ever replied within the data collection period. (Any replies that came after data collection ended would represent a negligible proportion given the volume of data and the length of the collection period.)

4.3. Data sets

Not all variables of interest were available in every data set, so we created some data sets with many observations and relatively few variables and others with fewer observations but many variables. Each data set constitutes a sample of the population of people or pairwise contacts on the dating site. When possible, the samples were chosen randomly, but when the variables of interest were available only for some observations, in some cases we included all for which the data were available, making a non-random sample.

We used three data sets organized by person — that is, where each observation is one individual:

1. **Many People:** 676,500 users of the dating site who were active during the data collection period. Includes age, gender, and messaging statistics. (All active users whose age and gender could be determined by our system, a non-random sample.)
2. **Profile People:** 11,600 users for whom we obtained full online dating profiles in addition to messaging statistics. (Random sample from several U.S. cities.)
3. **Questionnaire People:** 1,100 users of the online dating site who completed an online questionnaire. (Self-selected sample of those recruited by online advertisement in several U.S. cities.)

We also used four data sets in which each observation is one pairwise contact:

1. **Many Contacts:** 6.3 million pairwise contacts among 905,500 distinct people in which one sent a message to the other, who might or might not have reciprocated. Includes gender of each person and number, order, and timing of messages sent by each. (All contacts for which these data were available.)
2. **Location Contacts:** 300,000 pairwise contacts among 288,200 distinct people in the continental United States in which one sent a message to the other, who might or might not have reciprocated. Includes the variables in Many Contacts plus the geographic location of each person. (Random sample of contacts with location data available.)
3. **Reciprocated Contacts:** 200,000 pairwise contacts among 197,785 distinct people in which one sent a message to the other and got a response. Like Many Contacts, but reciprocated. (Random sample of reciprocated contacts.)
4. **Profile Contacts:** 73,300 pairwise contacts of 11,600 distinct people in which one person sent a message to the other, who might or might not have reciprocated. Includes everything in Location Contacts plus descriptors from each person's online dating profile. (All pairwise contacts among Profile People.)

5. People and profiles

In this section, we present descriptive statistics about the users of the online dating site we studied. First, we discuss the demographics of these users and the types of people they were seeking to meet, as specified in their profiles. Next, we examine their free-text self-descriptions. Finally, we describe typical messaging behavior and present two regression models to predict how many new contacts men and women receive on average per day. Throughout this section and the next, we report the results of parametric tests of means, but in most cases we use medians to describe distributions in order to provide a clearer picture of central tendency.

5.1. Profile demographics: self and sought

Men outnumbered women in the data we examined. In our largest set of online dating users, the Many People data set, approximately 42.1 percent were female and 57.9 percent male. The median age was 42 for men and 41 for women, but our data included substantial numbers of people in their 20s, 30s, 40s, 50s, 60s, and 70s. In this data set, 77.5 percent of users who specified an ethnicity said that they were Caucasian, 9.3 percent Hispanic or Latino, 7.4 percent African-American, and 1.6 percent Asian.

A small number reported another ethnicity or mixed ethnicity.

On their profiles, men said they were seeking women who were from a median 3 years older than themselves to a median 11 years younger. Women sought men who were from a median 7 years older than themselves to a median 5 years younger. These ranges support the contention of evolutionary psychology that men will seek younger women and women will seek older men. Furthermore, consistent with the notion that women will be pickier in selecting a mate, women's age ranges were significantly smaller than men's, a median 12 years compared to 14 years for men, $t(10,350) = -19.8, p < 0.0001$. Figure 1 shows how these median age ranges vary with a user's own age — both men and women tended to prefer increasingly younger people relative to themselves as they got older.

Moreover, men were less picky in specifying the type of person they were seeking. Men expressed a specific preference for the characteristics of the person they sought on an average of 9.1 characteristics, whereas women expressed a preference on an average of 11.9 characteristics, $t(9,542) = -30.3, p < 0.0001$. This difference was consistent across age. The most striking characteristic on which women were pickier than men was ethnicity. Women were more than twice as likely as men to specify in their profiles that they were seeking someone of their own ethnicity, whereas men were more likely to specify that any ethnicity was acceptable. (See [14] for additional analyses of the relationship between ethnicity preference and gender.)

5.2. Textual self-descriptions

Most self-descriptions were short passages, with 90 percent of them between 40 and 305 words. Men's self-descriptions were 11 percent shorter than women's, a median of 106 words for men versus a median of 118 words for women $t(9,767) = 3.9, p < 0.001$. Self-descriptions by those in their 30s, 40s, and 50s were 13 percent longer than those by younger or older people, $t(4,538) = 7.9, p < 0.0001$.

Women tended to use a greater percentage of words determined by the Linguistic Inquiry and Word Count software to be related to home and sex than did men, while men used a greater percentage of words related to work. Women used more words related to affect in general and positive emotion specifically. (All differences were significant at

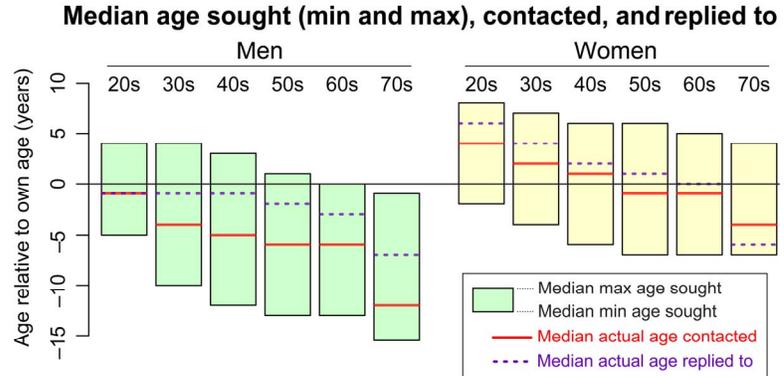


Figure 1. Ages sought, contacted, and replied to by age and gender.

$p < 0.001$; pairwise comparisons and t statistics available from the authors upon request.)

The use of certain word categories was also associated with scores on the psychometric scales we administered in the questionnaire completed by 548 women and 566 men on the dating site. Women who used more negative emotion words, such as “hate,” in their self-descriptions had lower general trust ($r = -0.13, p < 0.01$), higher general caution ($r = 0.12, p < 0.01$), and higher attachment anxiety ($r = 0.14, p < 0.01$). Similarly, for men, the use of more positive emotion words, such as “love” or “nice,” was associated with higher levels of general caution ($r = 0.13, p < 0.01$) and attachment anxiety ($r = 0.12, p < 0.01$). Those who used more tentative qualifiers, such as “maybe” and “perhaps,” showed lower levels of general trust ($r = -0.13, p < 0.01$) and higher levels of attachment anxiety ($r = 0.10, p < 0.05$). Such linguistic markers of psychological traits and dispositions might serve as important cues “given off” by profile writers that readers can use to better evaluate potential dates.

Moreover, though these results should be considered preliminary because of the small sample size relative to the large space of possible self-descriptive text and the small magnitude of the correlations, the above findings suggest that it might be possible to use automated tools such as LIWC to estimate a person's scores on relationship-relevant dimensions such as attachment anxiety and avoidance, extraversion, neuroticism, and general trust and caution, without having to administer surveys. Indeed, researchers have already had some success in inferring personality traits from somewhat longer self-descriptive passages [7].

6. Characterizing communications

Making initial contact with a potential date is the crucial first step toward a face-to-face meeting. Since

meeting in person is the goal for most online daters, geographical proximity is important. In the Location Contacts data set, more than 75 percent of pairwise contacts were between people within 100 miles of each other (median 28.2 miles). Although most were local, some contacts spanned the United States. In the following sections, we first discuss the rates of initial contact and subsequent reciprocation by age and gender. Then we present binary logistic regression models to predict when reciprocation occurs.

6.1. Initial contact

Rates of initial contact differed sharply by gender. Men in the Many People data set initiated a median 1 contact per day compared with 0.875 for women, $t(343,085) = 23.3$, $p < 0.0001$. Given this difference combined with the greater number of men on the site, women tended to be contacted much more often than men, a median 2 times per day, compared to 0.5 for men, $t(307,103) = -178.2$, $p < 0.0001$. Indeed, among the pairwise contacts in the Many Contacts data set, 77.1 percent consisted of a man initiating contact with a woman. Only 22.9 percent of contacts consisted of a woman contacting a man.

Older men and younger men were contacted equally often, but older women were contacted less often than younger women ($r = -0.186$, $p < 0.001$). Both men and women tended to initiate contact with people toward the middle of the age range they said they were seeking on their profile (Figure 1). Furthermore, they contacted people of the same ethnicity as themselves more often than chance would predict given the actual distributions of ethnicities among men and women in the sample (a finding which we report in more detail in another work based on these data [14]).

Among the 1,100 people who completed our psychometric questionnaire, men and women who scored higher on general caution contacted others more often, $r = 0.11$, $p < 0.001$, as did those people high in neuroticism, $r = 0.12$, $p < 0.0001$, perhaps as a way to control with whom they correspond or as a strategy to evaluate a larger pool of candidates before meeting. Furthermore, men high in general caution were contacted by others less often, $r = -0.10$, $p < 0.05$, though it is unclear what cues they were giving off that might lead to lower levels of contact.

Finally, more popular men and women — those who were contacted more often per day — initiated contact with others slightly less often, $r = -0.11$ for men and $r = -0.05$ for women, $p < 0.0001$. This suggests that those to whom potential dates present themselves need not pursue others as actively.

6.2. Reciprocation

Men contacted by women replied 26.4 percent of the time, while women contacted by men replied just 15.9 percent of the time. These differences perhaps reflect the relative rarity of a woman initiating contact with a man, but they also suggest that women are pickier in their communications.

Older men's initial contacts garnered replies more often than did those of younger men — the correlation between men's age and the fraction of their contacts that were answered was $r = 0.13$ ($p < 0.0001$). Older women, perhaps to compensate for being contacted less often, replied more often than younger women to men who contacted them ($r = 0.18$, $p < 0.001$). Also, more popular people replied less often than did their less popular peers, $r = -0.20$, $p < 0.0001$ for men, $r = -0.30$, $p < 0.0001$ for women. This finding mirrors that of Hitsch and colleagues [18].

Among the women who completed the psychometric questionnaire, those with higher levels of caution replied less often to men who contacted them, $r = -0.12$, $p < 0.01$, as did women higher in neuroticism, $r = -0.11$, $p < 0.01$. Men's reply rates did not vary with their responses to the questionnaire.

Some characteristics that we expected to correlate with the rate of reciprocation did not. In particular, the distance between the two parties in pairwise contacts was not associated with the reply rate, probably because the initiator already chooses based on distance, so the recipient does not need to. Also, the proportions of words in the self-description related to home, work, sex, money, and emotion were not substantially associated with rates of contacting or replying.

6.3. Predicting reciprocation

The success of a user in online dating depends on his or her ability to garner a response from a potential date. In this section, we present two binary logistic regression models to predict whether the recipient of an initial contact will reply, one for men contacting women and one for women contacting men. Table 2 shows the maximum likelihood estimates of the parameters for these two models.

Both male and female initiators were more likely to receive replies when they were popular but less likely to be answered when the recipients of their messages were popular. For men contacting women, the interaction between his popularity and that of his recipient was also associated with a slightly higher probability of reply, implying that a reply is more likely when the two people have similar levels of

Table 2. Binary logistic regression models to predict whether an online dater who initiates contact will receive a response. The first model is for men initiating contact with women, and the second is for women initiating contact with men. The columns give estimates of the regression coefficients, standard errors for the coefficients, and exponentiated coefficients, which correspond to odds ratios for each term. All binary variables are coded with 1 for “yes” and 0 for “no.” The Z statistics and exact *p* values are available from the authors upon request.

Outcome variable Model term	Men contacting women Does woman reply? (binary)		Women contacting men Does man reply? (binary)	
	$\hat{\beta}$ (S.E.)	$e^{\hat{\beta}}$	$\hat{\beta}$ (S.E.)	$e^{\hat{\beta}}$
	Initiator popularity	0.082 (.019) ***	1.086	0.036 (.005) ***
Recipient popularity	-0.138 (.003) ***	0.871	-0.292 (.016) ***	0.747
Recipient simultaneous contacts initiated	0.046 (.004) ***	1.047	0.023 (.003) ***	1.023
Recipient simultaneous contacts received	0.006 (.001) ***	1.006	0.021 (.005) ***	1.021
Age difference (male – female)	-0.026 (.002) ***	0.975	0.020 (.004) ***	0.980
Init.’s age in recip.’s range (binary)	0.221 (.028) ***	1.248	0.352 (.049) ***	1.422
Init.’s ethnicity preferred by recip. (binary)	0.604 (.048) ***	1.829	0.303 (.119) *	1.354
Number of preferences stated by initiator	0.002 (.003)	1.002	0.014 (.004) ***	1.015
Number of preferences stated by recipient	0.012 (.003) ***	1.012	0.020 (.004) ***	1.020
Initiator popularity × Recipient popularity	0.012 (.001) ***	1.012	< 0.001 (.002)	1.000
(Constant)	-1.372 (.069) ***		-1.277 (.142) ***	
Number of observations	46,339		14,339	
Percent replied	21.6		36.6	
χ^2	5,877		1,221	
Degrees of freedom	10		10	

*** *p* < .001 ** *p* < .01 * *p* < .05

popularity, all else being equal. This interaction was not significant for women contacting men.

Furthermore, for both men and women, the number of other people the recipient contacted or was contacted by around the time of the initiator’s message was positively associated with the chance that the recipient would reply. We interpret this to mean that activity begets more activity among recipients choosing whether to reply.

We entered age difference as the man’s age minus the woman’s, a quantity which was positive when he was older and negative when she was older. When women contacted men older relative to themselves, they got replies slightly more often than they did from relatively younger men. The odds ratio suggests that the chance of the man replying would go up 2 percent per year of his age relative to hers, all other things being equal. This was unsurprising, as evolutionary psychology would predict and our own data suggest that older women receive replies less often than younger ones. For men contacting women, though, the result was unexpected — men who were

older relative to the women they contacted had a slightly *lower* chance of getting a reply, about 2.5 percent less per year of his age relative to hers.

It seems that what matters more is not age difference *per se* but rather whether the initiator’s age is within the recipient’s preferred age range, as stated on his or her profile. Indeed, the odds ratio for this term in the model suggests that a man contacted by a woman would be 42 percent more likely to reply when her age was in his preferred range, with all other factors held equal. Similarly, a woman contacted by a man would be 25 percent more likely to reply when his age was in her preferred range, all else being equal.

The chance of receiving a reply also improved when the initiator’s ethnicity was one of those preferred by the recipient, as specified in his or her profile. When a woman was contacted by men whose ethnicity she preferred, the odds ratio suggests that she would be 83 percent more likely to reply than she would be with a man of an ethnicity she did not favor, all else being equal. Similarly, a man contacted by a woman would be 35 percent more likely to reply

if her ethnicity was one he favored. The difference in magnitude between men and women reinforces the more restrictive ethnicity preferences expressed by women (section 5.1).

The number of characteristics of an ideal date that online daters specified in their profiles — which we might consider a measure of pickiness — was also significantly associated with the chance of a reply. Across both genders, recipients who were pickier in this manner were more likely to reply, perhaps because indicating preferences in one's profile makes it less likely that one will be contacted by ineligible people in the first place. Women with more stated preferences also received replies to their messages more often than did women who were less specific about their ideal date.

6.3.1. How late is too late to reply? In a reciprocated contact, how long it takes the recipient of the initial message to reply could affect the progression of the conversation: too quick, and the recipient might appear overeager; too slow, and he or she might seem uninterested. In the Reciprocated Dyads sample of 182,000 reciprocated contacts, the median time from the initial message to the first reply was 16.1 hours after a woman initiated contact with a man and 19.2 hours after a man initiated contact with a woman.

We built a simple logistic regression model to predict whether the initiator would send a follow-up *after* the first reply — that is, the third message in the exchange — based on the number of days that passed between the initial message and the reply. Indeed, this model suggested that the chance of the initiator following up dropped about 0.7 percent per day that the recipient of the initial message waited to reply. After a month, the model predicts that the chance of the initiator responding again would drop by nearly 18 percent. Furthermore, the data suggested that there was no such thing as too quick a reply.

7. Discussion

Women are pickier than men in online dating: their preferences for age and ethnicity are stricter than men's, and they initiate contact and reply to others at a lower rate than men. Men cast a wider net, stating fewer and less restrictive preferences for their ideal date, contacting a relatively large number of women, and discriminating less in their replies. These findings align with the long-substantiated predictions of evolutionary psychology, though our research was not intended as a test of evolutionary theory. In our view, the modal tendencies described by this tradition provide useful direction, but evolutionary arguments are not the only explanation for these phenomena [8].

We also found evidence for positive assortment on ethnicity — both men and women contacted people of the same ethnicity more often than chance would predict — and on popularity, a proxy for attractiveness, in that women replied more often to men whose popularity was close to their own.

Furthermore, we demonstrated a connection between the use of word categories in self-descriptive text and relationship-relevant psychosocial constructs such as attachment style, personality traits, and general trust and caution. Although the word categories themselves were not associated with contact and reply behavior, some of the psychosocial constructs were correlated with these behaviors. Taken together, these findings suggest a promising future direction: the automated estimation of relationship-relevant qualities of a person solely from self-descriptions.

7.1. Implications for daters and designers

For online daters seeking to improve their odds of finding a mate, the message is clear: choose wisely and, if possible, be female. Women in our study were contacted much more often than men and so had their pick of whom they replied to. Those who respect others' preferences for age and ethnicity when they choose whom to contact are more likely to get a reply. Also, more popular people reply less often, so slightly less attractive targets might be a safer bet.

For designers of online dating systems, these findings suggest interesting new design directions. If popular people aren't as likely to reply, perhaps sites should direct users toward targets who are slightly less popular but more likely to respond — a trade-off many users might willingly accept.

8. References

- [1] Alterovitz, S.S.-R., & G.A. Mendelsohn. (2009). Partner Preferences Across the Life Span: Online Dating by Older Adults. In *Psychology and Aging* 24 (2): pp. 513–517.
- [2] Asendorpf, J.B., & S. Wilpers. (1998). Personality Effects on Social Relationships. In *Journal of Personality and Social Psychology* 74 (6): pp. 1531–1544.
- [3] Botwin, M.D., Buss, D.M., & T.K. Shackelford. (1997). Personality and Mate Preferences: Five Factors In Mate Selection and Marital Satisfaction. In *Journal of Personality* 65 (1): pp. 107–136.
- [4] Brody, L.R. (1993). On understanding gender differences in the expression of emotion: Gender roles, socialization, and language. In (S.L. Ablon, D. Brown, E.J. Khantzian, and J.E. Mack, Eds.) *Human feelings: explorations in affect development and meaning*. New York: Routledge.

- [5] Buss, D.M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. In *Behavioral and Brain Sciences* 12: pp. 1–49.
- [6] Buss, D.M., & M. Barnes. (1986). Preferences in Human Mate Selection. In *Journal of Personality and Social Psychology* 50 (3): pp. 559–570.
- [7] Chung, C.K. & J.W. Pennebaker. (2008). Revealing dimensions of thinking in open-ended self-descriptions: An automated meaning extraction method for natural language. In *Journal of Research in Personality* 42: 96–132.
- [8] Eagly, A.H., & W. Wood. (1999). The Origins of Sex Differences in Human Behavior: Evolved Dispositions Versus Social Roles. In *American Psychologist* 54 (6): pp. 408–423.
- [9] Ellison, N., Heino, R., & J. Gibbs. (2006). “Managing Impressions Online: Self-Presentation Processes in the Online Dating Environment.” In *Journal of Computer-Mediated Communication* 11 (2).
- [10] Feeney, J.A. (1994). Attachment style, communication patterns, and satisfaction across the life cycle of marriage. In *Personal Relationships* 1: pp. 333–348.
- [11] Fiore, A.T., & J. Donath. (2004). Online Personals: An Overview. Short paper, ACM Computer-Human Interaction 2004.
- [12] Fiore, A.T., & J. Donath. (2005). Homophily in Online Dating: When Do You Like Someone Like Yourself? Short paper, ACM Computer-Human Interaction 2005.
- [13] Fiore, A.T., Shaw Taylor, L., Mendelsohn, G.A., & M.A. Hearst. (2008). Assessing Attractiveness in Online Dating Profiles. In Proc. ACM Computer-Human Interaction 2008.
- [14] Fiore, A.T., Shaw Taylor, L., Zhong, X., Mendelsohn, G.A., & C. Cheshire. (2010). Whom we (say we) want: Stated and actual preferences in online dating. Poster presented at the Eleventh Annual Meeting of the Society for Personality and Social Psychology, Las Vegas, NV.
- [15] Fraley, R. C., Waller, N. G., & K.A. Brennan. (2000). An item-response theory analysis of self-report measures of adult attachment. In *Journal of Personality and Social Psychology* 78: pp. 350–365.
- [16] Gosling, S. D., Rentfrow, P. J., & W.B. Swann, Jr. (2003). A Very Brief Measure of the Big Five Personality Domains. *Journal of Research in Personality* 37: pp. 504–528.
- [17] Greer, A.E., & D.M. Buss. (1994). Tactics for Promoting Sexual Encounters. In *Journal of Sex Research* 31 (3): pp. 185–201.
- [18] Hitsch, G.J., Hortaçsu, A., & D. Ariely. (2006). What Makes You Click: An Empirical Analysis of Online Dating. Working paper, retrieved Sept. 2009 from: <http://home.uchicago.edu/~hortacsu/onlinedating.pdf>
- [19] Klohnen, E.C., & G.A. Mendelsohn. (1998). Partner Selection for Personality Characteristics: A Couple-Centered Approach. In *Personality and Social Psychology Bulletin* 24 (3): pp. 268–278.
- [20] Luo, S., & E.C. Klohnen. (2005). Assortative Mating and Marital Quality in Newlyweds: A Couple-Centered Approach. In *Journal of Personality and Social Psychology* 88 (2): pp. 304–326.
- [21] Madden, M., & A. Lenhart. (2006). Online Dating. Pew Internet and American Life Project.
- [22] Montoya, R.M., Horton, R.S., & J. Kirchner. (2008). Is actual similarity necessary for attraction? A meta-analysis of actual and perceived similarity. In *Journal of Social and Personal Relationships* 25 (6): pp. 889–922.
- [23] Norton, M., Frost, J., & D. Ariely. (2007). Less is more: The lure of ambiguity, or why familiarity breeds contempt. In *Journal of Personality and Social Psychology* 92 (1), pp. 97–105.
- [24] Palomares, N. A. (2009). Women are sort of more tentative than men, aren't they? How men and women use tentative language differently, similarly, and counter-stereotypically as a function of gender salience. In *Communication Research* 36 (4): pp. 538–560.
- [25] Pennebaker, J.W., Francis, M.E., & R.J. Booth. (2001). *Linguistic Inquiry and Word Count (LIWC): LIWC2001*. Mahwah: Lawrence Erlbaum Associates.
- [26] Shaw Taylor, L., Fiore, A. T., Cheshire, C., Mendelsohn, G. A., & X. Zhong. (2009). Self-Worth and Romantic Partner Evaluation and Selection. Poster presented at Association for Psychology Science 21st Annual Conference, San Francisco, CA.
- [27] Trivers, R.L. (1972). Parental investment and sexual selection. In (B. Campbell, ed.) *Sexual Selection and the Descent of Man*. New York: Aldine de Gruyter, pp. 136–179.
- [28] Vandenberg, S.G. (1972). Assortative mating, or who marries whom? In *Behavioral Genetics* 2 (2–3): pp. 127–157.
- [29] Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction: A relational perspective. In *Communication Research* 19 (1): pp. 52–91.
- [30] Walther, J.B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. In *Communication Research* 23 (1): pp. 3–44.
- [31] Walther, J.B., Slovacek, C., and L.C. Tidwell. (2001). Is a picture worth a thousand words? Photographic images in long term and short term virtual teams. In *Communication Research* 28 (1): pp. 105–134.
- [32] Watson, D., Klohnen, E.C., Casillas, A., Simms, E.N., Haig, J., & D.S. Berry. (2004). Match Makers and Deal Breakers: Analyses of Assortative Mating in Newlywed Couples. In *Journal of Personality* 72 (5): pp. 1029–1068.
- [33] Yamagishi, T. (2001). Trust as a Form of Social Intelligence. In (Karen Cook, Ed.) *Trust in Society*. New York: Russell Sage Foundation, pp. 185–201.