Preferences in Mate Choice

Preferences in the Early Stages of Mate Choice

Kevin Lewis, *University of California–San Diego*

Romantic partnership is often considered an optimal barometer of intergroup relations. To date, however, it has been challenging to distinguish the characteristics people prefer in a partner from the types of partners that are locally available. Online dating presents a new opportunity to address this puzzle. In this paper, I use behavioral data from a popular online dating site to answer three questions regarding preferences in the early stages of mate choice: First, to what extent do people prefer similarity versus status in a partner—and do these preferences vary by gender? Second, what is the relative importance of different types of preferences—and to what extent are apparent preferences for one characteristic merely a "by-product" of preferences for another characteristic with which the first is correlated? Third, do preferences vary at different moments of selection—and if so, how? These analyses not only provide a nuanced portrait of how interpersonal dynamics shape broader social structures—here, a network of romantic ties—but they recommend a future approach to mate choice that prioritizes processes over outcomes and more deeply engages the literature on gender, social networks, and symbolic boundaries.

Mate choice is a central topic in the study of inequality. In societies where romantic pairing involves intimacy and trust, mating patterns reflect the extent to which individuals from different backgrounds accept each other as equals. When romantic pairing leads to offspring, these patterns also tell us whether the status differences of today will be passed along to the children of tomorrow (Blossfeld 2009; Kalmijn 1998; Schwartz 2013).

Prior research has provided us with a detailed portrait of mating patterns in societies across the world. However, we still know comparatively little about how

Previous versions of this paper were presented at the 2011 International Sunbelt Social Network Conference in St. Pete Beach, FL, USA, and the 2011 Annual Meeting of the American Sociological Association in Las Vegas, NV, USA. This work benefited from feedback the author received at the Culture and Social Analysis Workshop at Harvard University; the Digital Seminar Series at the Harvard Business School; the Computational Social Science Initiative Seminar Series at the University of Massachusetts–Amherst; and the Departments of Sociology at UC San Diego, UC Berkeley, and Cornell University. The author is particularly indebted to Peter Marsden, Jocelyn Viterna, Marty Whyte, Michèle Lamont, Ken Lin, and Jen Lundquist, each of whom commented on full drafts of the paper; to Misiek Piskorski, with whom the author acquired the data set on which this research is based; and to three anonymous reviewers who provided thoughtful and constructive critiques. This research was supported by the Division of Research and Faculty Development at Harvard Business School. *Please address correspondence to Kevin Lewis, Department of Sociology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093–0533; E-mail: lewis@ucsd.edu*

Social Forces 95(1) 283–320, September 2016
doi: 10.1093/sf/sow036
Advance Access publication on 18 May 2016

these patterns are generated. In particular, because romantic choices are constrained by who is available—but data on romantic "opportunity structures" are rarely collected—it is challenging to distinguish the respective roles of *preference* and *availability* in the genesis of observed patterns (cf. Blau and Schwartz 1984). Consequently, a number of questions *about* romantic preferences have remained largely unanswered. Three are particularly important: First, to what extent do people prefer similarity versus status in a partner—and do these preferences vary by gender? Second, what is the relative importance of different types of preferences—and to what extent are apparent preferences for one characteristic merely a "by-product" of preferences for another characteristic with which the first is correlated? Third, do preferences vary at different moments of selection—and if so, how?

In this paper, I attempt to address these three questions. To do so, I join a growing body of scholarship that draws on data from online dating sites to better understand the process of mating (Anderson et al. 2014; Hitsch, Hortaçsu, and Ariely 2010; Kreager et al. 2014; Lin and Lundquist 2013; Skopek, Schulz, and Blossfeld 2011). Rather than relationship "outcomes" such as marriage or cohabitation, data from these sites represent the very early stages of romantic consideration and interest. However, because these data contain dynamic, directed information on interactions and multidimensional information on users—and because complete data on opportunity structures are available—online dating offers a unique vantage point for examining preferences. It is also increasingly worthy of study in its own right, as an unprecedented proportion of couples now meet online (Rosenfeld and Thomas 2012).

I begin by briefly reviewing the literature on mating patterns. Of the three general causes of these patterns—opportunities, third-party interference, and preferences—I explain why preferences have been particularly difficult to pinpoint and describe three important questions that have consequently remained unanswered. Next, I introduce my data and method. By focusing on users of the popular dating site OkCupid who live in New York City, I am able to approximate a population of locally available singles; focus on a region with particularly high site usage; and utilize exponential random graph models that can disentangle preferences from patterns (cf. Wimmer and Lewis 2010). I then present results from a series of statistical models. First, I test three hypotheses—matching, competition, and gender asymmetry—with respect to race, income, education, and religion. Second, I put the relative importance of demographic preferences in perspective and assess whether estimates of these preferences are impacted by the intersection of demographic and personal characteristics. Third, I show that site users display very different preferences when initiating than when reciprocating romantic interest—illustrating an important mechanism whereby social boundaries break down. I conclude by discussing the limitations and implications of these findings and pointing to a concrete avenue for further research.

Mating Patterns and Their Causes

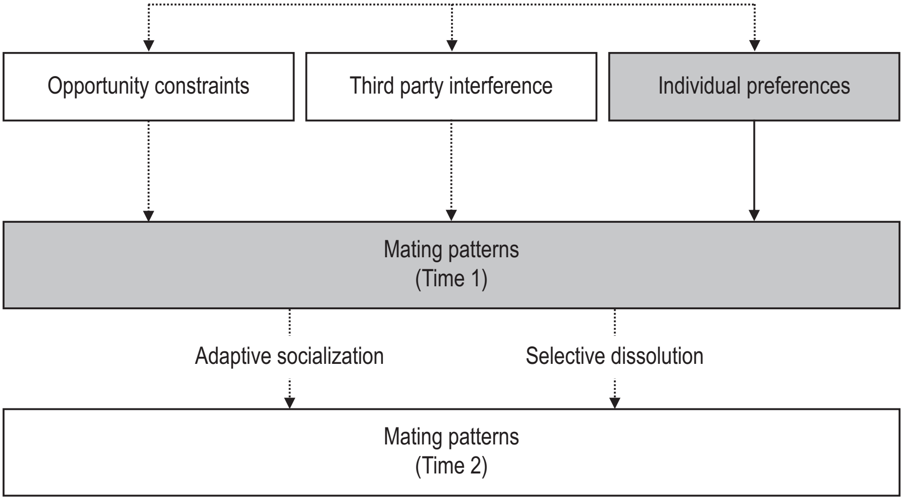
Given that homophily—or "birds of a feather flock together"—is a ubiquitous feature of social life (McPherson, Smith-Lovin, and Cook 2001), it is no surprise that mating patterns display this tendency as well. Most prior work has focused on marriage patterns across a variety of attributes and contexts. The findings from this research are unequivocal: Individuals display a marked tendency to select partners from the same social group (endogamy) or who are similar in terms of status (homogamy; see reviews in Blossfeld [2009]; Kalmijn [1998]; Schwartz [2013]).

Prior research also provides a theoretical framework for understanding this pattern (figure 1). First, some degree of homogamy is explained by "supply side" constraints given that similar people tend to occupy similar social spaces and are particularly likely to meet one another (Blau and Schwartz 1984; Kalmijn and Flap 2001). Second, "third parties" such as friends, family members, and institutions may have incentives to encourage homogamous relationships (Bratter and Eschbach 2006; Clark-Ibáñez and Felmlee 2004). Third, the similarity between partners can be explained by individual preferences or cost/benefit calculations, that is, the "demand side" of the equation. These causes are interdependent and mutually reinforcing—preferences are influenced by social surroundings, while people tend to seek out those third parties and opportunities that help them realize their preferences—and in many circumstances, homogamy is further solidified following relationship creation: Partners can potentially become more similar over time, known as "adaptive socialization" (Oppenheimer 1988), and some kinds of heterogamous partnerships are more likely to terminate, called "selective dissolution" (Schwartz 2010).

Isolating the Role of Preferences

Recently, network researchers have made great strides toward teasing apart the above mechanisms. Kossinets and Watts (2009), for instance, analyze e-mail

Figure 1. Mating patterns and their causes, adopted from Kalmijn (1998)



Note: While individual preferences impact mating patterns through multiple causal pathways, the direct impact of preferences on patterns (shaded boxes and solid arrow) is the focus of this study.

exchanges in a large university community and conclude that even a modest degree of homophily (the preference for similarity) is reinforced by network structures: Similar people self-select into structurally proximate positions and are unusually likely to meet. Similarly, Goodreau, Kitts, and Morris (2009) quantify the contribution of sociality (the tendency for some people to form more or fewer friendships than others), triadic closure (the tendency for friends of friends to become friends), and homophily to the formation of friendships among high school students; and Wimmer and Lewis (2010), using picture postings on Facebook, provide a theoretical framework for understanding how racial homogeneity is generated and show that the importance of same-race preferences is exaggerated when alternative mechanisms of tie formation (e.g., ethnic homophily and reciprocity) are omitted from statistical models.

The main reason this progress has not extended to romantic ties has to do with the theoretical framework described above. Mating patterns are fundamentally constrained by who is available; but—unlike network data sets, which are often collected from relatively small, bounded settings—data on romantic partnerships (including marriage) tend not to include information on the kinds of local social contexts where couples actually meet (Kalmijn and Flap 2001). Consequently, most research assumes the composition of the population is regionally invariant (Harris and Ono 2005; Lichter, LeClere, and McLaughlin 1991) and generally does not consider single people, who constitute an integral part of romantic "opportunities" (Logan, Hoff, and Newton 2008). In short, without data on all relationships that *could* have formed, it is impossible to pinpoint the causes of those relationships that *did* form—including the role of preferences.

This absence of data on opportunity structures—combined with additional limitations of many data sets—has left three particularly important unresolved puzzles in the literature.

Matching, Competition, and Gender Asymmetry

First, prior research has struggled to adjudicate between alternative theories of preferences. The most common explanation for homogamy is the *matching* hypothesis: Partners tend to be similar because individuals prefer similarity in a partner. This preference stems from the ease of communication, mutual validation, and common understanding supported by shared experiences, values, and backgrounds (DiMaggio and Mohr 1985; Kalmijn 1994). While previous research has focused on categorical attributes such as racial background or religion, matching could also extend to ordinal attributes like income level or educational attainment (e.g., Blossfeld 2009)—not just due to the shared experiences with which these attributes are correlated, but because too large a difference between partners could threaten the status of the family unit (Oppenheimer 1977).

An alternative hypothesis is that attraction is governed less by matching—where individuals from the same background pursue one another—and more by status-based *competition*—where individuals from certain backgrounds are pursued by everyone (Edwards 1969; see also Kreager et al. 2014). Dominant explanations focus on socioeconomic resources: Insofar as economic well-being is shared among family members, individuals should try to maximize the collective earnings potential of the couple (Kalmijn 1994, 1998). Competition can also occur with respect to attractiveness (Stevens, Owens, and Schaefer 1990) and cultural resources (see DiMaggio and Mohr 1985). Unfortunately, it is often impossible to determine which process is at work: Both matching and competition generate an aggregate pattern of homogamy (Kalmijn 1998).

A final possibility is that whether one prefers similarity or status (or neither) in a partner may vary, in heterosexual mating, between men and women. Most research on *gender asymmetry* in preferences focuses on either evolved dispositions (e.g., Buss and Schmitt 1993) or else the degree of female participation in the workforce (e.g., Mare 1991). However, a growing body of literature also discusses stereotypes associated with men and women from different racial backgrounds, leading to gender-asymmetric processes of "competition" (e.g., Feliciano, Robnett, and Komaie 2009; McClintock 2010). Again, however, research has been limited by available data: Without "directed" data that distinguish male and female "choices," it has been challenging to assess *whose* preferences are driving observed patterns (cf. Lin and Lundquist 2013).

Multidimensionality and the By-Product Hypothesis

Second, most data sets on romantic partnerships contain only basic demographic variables, and multivariate analyses of mating patterns remain rare (Kalmijn 1991). Consequently, it has been challenging to assess the *relative* importance of different types of preferences or whether an apparent preference for one characteristic is a spurious *by-product* of a preference for a second characteristic with which the first is correlated (see Kalmijn and Vermunt 2007). In particular, a wide body of scholarship has highlighted the importance of shared beliefs, behaviors, and tastes for interpersonal affinity and exclusion (DiMaggio and Mohr 1985; Erickson 1996; Kalmijn 1991; Lamont and Lareau 1988). The extent to which such non-demographic or even "cultural" attributes are important for mating is both inherently interesting and consequential insofar as these attributes overlap with traditional dimensions of stratification (see Rivera 2012). To date, however, their relative importance to mate selection has remained a mystery.

Dynamic Variation in Preferences

A third puzzle involves dynamic variation in preferences. While most analyses of romantic partnerships focus on a single "stage" of formal relationship (e.g., marriage, cohabitation), several studies have compared patterns across multiple stages and suggested that differences in preferences may be one explanation (e.g., Blackwell and Lichter 2000; McClintock 2010; Schoen and Weinick 1993). More recently, a handful of scholars have considered whether romantic preferences may vary across micro-interactional *moments* by comparing patterns of first contacts and replies in online dating (Fiore and Donath 2005; Lin and Lundquist 2013; Skopek, Schulz, and Blossfeld 2011). Evidence on this question has remained inconclusive, however—in part because different researchers have focused on different attributes and in part because standard statistical tools are incapable of measuring both processes simultaneously (see below).

Online Dating

Similar to most social media, online dating sites enable their members to create personal profiles and interact in a variety of ways. Unlike other social media, dating sites' goal is to facilitate romantic connections between strangers. Online dating also offers a number of methodological advantages for studying preferences. First, it is possible to know who has an account at any given time—and therefore to control for the "opportunity structure" for interaction. Second, users typically report a variety of characteristics on their profiles. These include not just demographic traits, but other attributes that are typically unavailable to sociologists—from body type to smoking habits to whether a person has or likes pets. Third, all exchanges are digitally recorded, so it is possible to observe the actual process of interaction (rather than just its outcome).

Because of these advantages, a growing body of scholars is using data from online dating sites to study mate selection. Surprisingly, however, the general questions posed here have remained unanswered. First, prior work tends to focus on one or another specific dimension of mating, such as race (Lewis 2013), education (Skopek, Schulz, and Blossfeld 2011), or attractiveness (Kreager et al. 2014; Taylor et al. 2011)—preventing an assessment of the "by-product" hypothesis identified above. Second, prior work tends to focus on a single stage of interaction, such as profile views (Anderson et al. 2014) or first messages (Hitsch, Hortaçsu, and Ariely 2010), or compare results from multiple stages analyzed separately (Fiore and Donath 2005; Lin and Lundquist 2013)—treating different moments of the same underlying process as independent events. Finally, few of the aforementioned studies provide any meaningful information about the actual site being analyzed. This is problematic not only because it is impossible to assess the generalizability of results (given the staggering variety of dating sites that are now available), but also because most dating sites actively interfere with user behavior (e.g., by "recommending" matches)—and so it is impossible to tell whether or to what extent results might simply reflect site architecture (for a lengthier discussion, see Lewis [2015a]).

Data

Data for these analyses were acquired from OkCupid (www.okcupid.com), one of the most popular dating sites on the Internet (Rudder 2014). This site has several distinguishing features. First, membership is free. On one hand, this means users might be less "serious" about finding a mate. On the other hand, it eliminates a barrier to entry for users who cannot afford to join sites with large membership fees. Second, OkCupid is a "generalist" site, as opposed to the many "niche" sites that are now available. Users thus constitute a diverse swath of the population, although individuals who strongly prefer a certain type of partner may seek out a relevant niche site such that preferences on OkCupid are unusually inclusive.

Individual Data

My data set contains information on users who were active on the site between October 1 and December 15, 2010.[1] For this paper, I focus on users who self-identified as "straight," "single," and living in the United States. The majority of site users (84.4 percent) live in the United States, and we can expect that people in relationships might behave differently on a dating site than people who are single. The restriction to heterosexual users is an important limitation of this study. However, given potential variation in preferences by sexual orientation (see Lundquist and Lin 2015), it seemed unwarranted to include all users in the same analysis, and focusing on heterosexuals facilitates comparisons with prior research.

I additionally trimmed this sample in three ways. First, users indicate what they are "looking for" on the site (options include new friends, short-term dating, long-term dating, activity partners, long-distance pen pals, and casual sex). Given my focus on romantic (as opposed to platonic or physical) relationships, I restrict attention to users looking for "long-term dating," "short-term dating," or both (regardless of whether other options were selected).[2] Such individuals account for 65.4 percent of single, straight users in the United States.

Second, I restrict attention to users who joined the site between October 1 and November 30, 2010. This is to prevent artificial truncation of user interactions: If this restriction were not imposed, it would be impossible to determine whether a communication from user A to user B that occurred on October 1 (or any other date) was the initiation of a new exchange or the response to an unobserved communication B had previously sent. Omitting users who joined after November 30 similarly prevents tail-end truncation, as described below.[3]

Finally, I focus on users with a zip code beginning 10xxx—the area encompassing New York City. Such users account for 4.7 percent of the sample identified above—a higher percentage than any other two-digit zip code—and result in a final sample of $N = 7{,}671$ users. This restriction is practically necessary due to computational demands (cf. Hitsch et al. 2010; Kreager et al. 2014) and is akin to other studies of "complete" network data where it is necessary to impose some boundary beyond which ties are no longer considered (Marsden 1990; for alternative approaches, see, e.g., Anderson et al. [2014]; Lewis [2013]; Lin and Lundquist [2013]). The advantages of this approach are that it allows me to approximate a geographically dense set of users who plausibly consider one another as romantic prospects and to ground interpretation of results in this local cultural setting. The disadvantages are that messages exchanged with users outside this region are artificially censored;[4] the strength of preferences may be slightly underestimated, insofar as the cost of a "suboptimal" partner might be offset by the benefit of convenience; and preferences among these users may not be representative of other regions that are less urban, liberal, educated, or diverse. Descriptive statistics are presented in table 1.

Relational Data

The primary means by which users communicate is the site's internal e-mail system. My relational data consist of all messages exchanged among the sample

Table 1. Descriptive Statistics on Study Sample

	Male ($N = 3{,}146$)	Female ($N = 4{,}525$)
Demographic characteristics		
Race		
Asian	5.02	6.98
Black	4.67	5.46
Indian	2.03	1.24
Hispanic	6.58	4.95
White	52.80	55.65
Other	10.87	9.24
Unknown	18.02	16.49
Income		
0 to 30K	4.04	3.47
30 to 60K	4.77	3.34
60 to 100K	4.29	2.61
100 to 150K	2.96	0.84
150K and up	4.10	1.06
"Private"	20.92	22.25
Unknown	58.93	66.43
Education		
High school	3.31	1.79
2-year college	3.53	2.98
University	45.49	49.26
Master's	14.91	20.49
Law/Med/PhD	8.58	7.58
Other	2.32	0.88
Unknown	21.87	17.02
Religion		
Agnostic	8.74	8.80
Atheist	7.44	4.24
Catholic	11.44	13.22
Christian	13.76	16.95
Jewish	9.41	11.45
Other	9.15	7.87
Unknown	40.05	37.48
Age		
Mean	30.28	29.43
Standard deviation	8.99	8.41

(*Continued*)

Table 1. ***continued***

	Male ($N = 3{,}146$)	Female ($N = 4{,}525$)
Network properties		
Initiation messages		
Total sent	13168	3016
Average # sent per person	4.19	0.67
Initiation rate[a]	0.09	0.02
Response messages		
Total sent	1099	2058
Average # sent per person	0.35	0.45
Response rate[b]	36.44	15.63
Median response time (in days)	0.17	0.27

Note: All statistics are percentages unless otherwise indicated. The high proportion of missing data is an unavoidable limitation—but also accurately reflects the (lack of) information site users face when searching for a mate online.
[a]"Initiation rate" is defined as the total number of initiation messages sent from men to women or women to men as a percentage of the total possible messages that could have been sent, that is, the initiation "density" for each gender.
[b]"Response rate" is defined as the total number of response messages sent as a percentage of the total number of initiation messages received.

defined above, with the following restrictions. First, I look only at messages between a male and a female (99.5 percent of messages among this sample). Second, I include only the first time user A contacted user B, and—if applicable—B's first reply. Without message contents, lengthier exchanges are impossible to interpret: Two users who stop communicating could equally plausibly have lost interest or transitioned their interaction offline. Third, I consider only initiation messages sent between October 1 and November 30 and replies sent within two weeks. Because my messaging data extended only through mid-December, it was necessary to pick a cutoff point beyond which new initiations would no longer be considered but data on replies were still available. Two weeks seems a rather conservative window within which to expect a reply from an interested recipient; in fact, 98.7 percent of all replies met this criterion, and the median response time among men and women alike was less than seven hours (table 1).

It is important to acknowledge that the assumption that replies represent mutual interest, while analytically necessary (see also Lin and Lundquist 2013; Skopek, Schulz, and Blossfeld 2011), may not always be true in practice. For instance, some users may simply reply to everyone; other users may reply out of appreciation for an especially nice message or to express a "polite rejection." However, it is generally true that site users signal uninterest by not replying (80 percent of initiation messages are unanswered); because my models control for the baseline tendency to reciprocate (see below), results will be biased only if *patterns* of "uninterested" replies vary systematically from patterns of interested replies; and insofar as we can expect that uninterested replies are *shorter*, on average, than interested replies, I found no systematic differences in the patterning of replies by length in another analysis of the same data set (see Lewis 2013).

Generalizability and Site Interference

One critical question is whether individuals in these analyses are different from the broader population of singles. We know from recent surveys that online dating is most common among college-educated urbanites and suburbanites in their mid-twenties through mid-forties (Smith and Duggan 2013). However, once they limited consideration to single Internet users (i.e., people "at-risk" of online dating), Sautter, Tippett, and Morgan (2010) found that no socio-demographic variables significantly influence the likelihood of Internet dating. Comparing the specific sample in these analyses to all unmarried, Internet-using adults in the New York Metropolitan Area, the sample contains more women, fewer blacks and Hispanics, and more whites and persons from an "other" racial background. The sample is also younger and more educated than we would expect by chance—though the income distribution of the two populations is similar.[5]

Second, I can claim to isolate preferences only if all exogenous influences have been eliminated—including interference from the dating site. Fortunately, OkCupid's approach to matching is unusual in two ways: It is *transparent*, and it is *user driven*. Specifically, at any time, users may answer questions about (1) their own personality, (2) their ideal partner's personality, and (3) how much each characteristic matters. The site then uses these responses to calculate a compatibility score for every two users (www.okcupid.com/help/match-percentages); every time user A encounters user B on the website (including when B appears in A's search results), their compatibility score is displayed. In short, both primary ingredients of "matchmaking" on OkCupid—how users find one another (primarily through custom searches) and how compatible the site believes two people are (represented by compatibility scores)—depend on users' explicit preferences; the remaining steps to communication (actually viewing a person's profile and deciding to contact that person) are driven solely by individual volition.[6]

Method

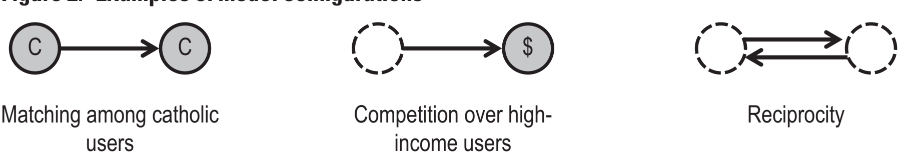
Exponential random graph (ERG) modeling is a powerful network analytic technique that is increasingly used to understand how patterns in relational data were generated (e.g., Goodreau, Kitts, and Morris 2009; Wimmer and Lewis 2010).[7] In ERG models, the possible ties among actors—here, messages among dating site users—are regarded as random variables. In this case, I consider only interactions between a male and a female; that is, same-gender messages are prohibited. If **Y** is the matrix of all such variables and **y** is the matrix of observed ties, then these models have the following form:

$$\Pr(Y = y) = \left(\frac{1}{\kappa}\right)\exp\left\{\sum_{A} \eta_A g_A(y)\right\}$$

The analyst begins by positing any number of "tie-generating mechanisms," or reasons why one user might contact another. These reasons correspond to the possible effects in a model, and each is represented by a configuration A. We can think of configurations as small network "substructures" that might be observed (figure 2). For instance, we may speculate that Catholic users tend to contact one another (matching); that all users tend to contact high-income users (competition); or that users tend to reciprocate messages (reciprocity). Above, $g_A(\mathbf{y})$ is the *count* of how many times configuration A appears in network $\mathbf{y}$ (respectively, the number of times a Catholic user messages another Catholic user; the number of times a high-income user receives a message; and the number of reciprocated messages). Each configuration is associated with a parameter η_A representing whether A occurs more frequently (positive parameter) or less frequently (negative parameter) than we would expect by chance, conditional on other effects in the model; and κ is a normalizing constant. The formula therefore represents the probability of observing the empirical network that actually was observed as a function of the various possible micromechanisms that might have generated it.[8]

One advantage of ERG models for this paper is that they are able to statistically disentangle the relative contribution of matching, competition, and gender asymmetry to mating patterns—all while controlling for a variety of confounding mechanisms. Additionally, by explicitly modeling the reciprocation process (as opposed to examining patterns of initiation and reply separately), these models can integrate both stages of interaction under a single umbrella; measure the strength of reciprocity itself as a generative mechanism; and precisely quantify how initiations and replies differ. The presence of these reciprocity effects—which entail a complex dependence structure among observations—means that we cannot directly estimate the maximum likelihood values for the η parameters (in short, because of the difficulties calculating κ). Instead, I use a simulation-based technique called Markov chain Monte Carlo maximum likelihood estimation (Robins et al. 2007). This approach proceeds by simulating a distribution of random graphs from some starting set of parameter estimates; comparing this distribution of graphs against the observed graph with respect to all effects in the model and refining parameter estimates accordingly; and repeating this process until parameter estimates stabilize (i.e., until they produce a distribution of graphs in which the observed graph is typical for all effects in the model). At this point, the model is considered to have converged.[9]

Figure 2. Examples of model configurations



Note: Shaded nodes represent site users with a specified attribute (C = Catholic, $ = high-income); unshaded nodes represent anyone. For instance, the middle configuration above refers to any instance in which someone messages a high-income user, regardless of who that "someone" is.

Results

First, I assess the roles of matching, competition, and gender asymmetry for four attributes on which prior literature has focused: race, income, education, and religion (Kalmijn 1998; Schwartz 2013). Second, I document the relative importance of matching based on each of these dimensions as well as based on non-demographic characteristics. Finally, I show how dating site users' preferences vary depending on whether they are expressing or reciprocating interest.

Control Terms

Table 2 presents results from model 1—a baseline, "control model" of preferences. All terms in this model are included in all subsequent models. The negative

Table 2. Parameter Estimates and Standard Errors from Control Model

	Model 1	
	Coef	SE
Structural effects		
Density	–9.001***	.076
Reciprocity	6.784***	.032
Activity spread	–3.175***	.038
Gender/age effects		
Female-receiver	2.717***	.105
Age-sender	.106***	.002
Age-receiver	.105***	.004
Age × female-sender	–.138***	.005
Age × female-receiver	–.201***	.005
Age-absolute difference	–.155***	.004
Age-absolute difference^2	–.002***	3e-4
Propinquity effect		
Same 3-digit zip code	.238***	.012
Site usage effects		
Account duration-sender	–.020***	.001
Account duration-receiver	–.004***	.001
Login time-sender	2e-5***	5e-7
Login time-receiver	2e-6**	6e-7
Has picture(s)-sender	–.035**	.012
Has picture(s)-receiver	.096***	.015
Account overlap	.031***	.001

Note: All models presented in this paper are exponential random graph models of messaging behavior among heterosexual dating site users in New York ($N = 7{,}671$). Ties are prohibited between two users of the same gender and also between two users whose membership on the dating site did not overlap.
**$p < .01$ ***$p < .001$ (two-tailed tests)

"density" coefficient indicates that the baseline likelihood for two users to message each other is far below 50 percent. The positive "reciprocity" coefficient indicates that the log-odds of user A messaging user B increase tremendously if B has contacted A first. The final structural effect, "activity spread," captures unmodeled heterogeneity in the tendency for some users to send more or fewer messages overall (Robins, Pattison, and Wang 2009). In essence, this is a higher-order "star" configuration that represents the *additional* tendency for a node to accumulate ties depending on the quantity of ties it already has. The negative coefficient implies that users who send very many messages are rare and the variance of the quantity of messages sent is relatively small—as opposed to a positive coefficient, which could reflect a core-periphery structure (Snijders et al. 2006).

In the next effects, we see that women are more likely to receive messages than men; men are more likely to send and receive messages as they get older, while women are *less* likely than men to send and receive messages as they get older; and dating site users tend to contact other users who are similar in terms of age. Two users who share the same three-digit zip code are more likely to contact each other than two users who do not. Users who spend more time online are more likely to send and receive messages, but users who have had an account longer are less likely to send and receive messages. Finally, two users are more likely to contact each other the longer their membership periods overlapped.

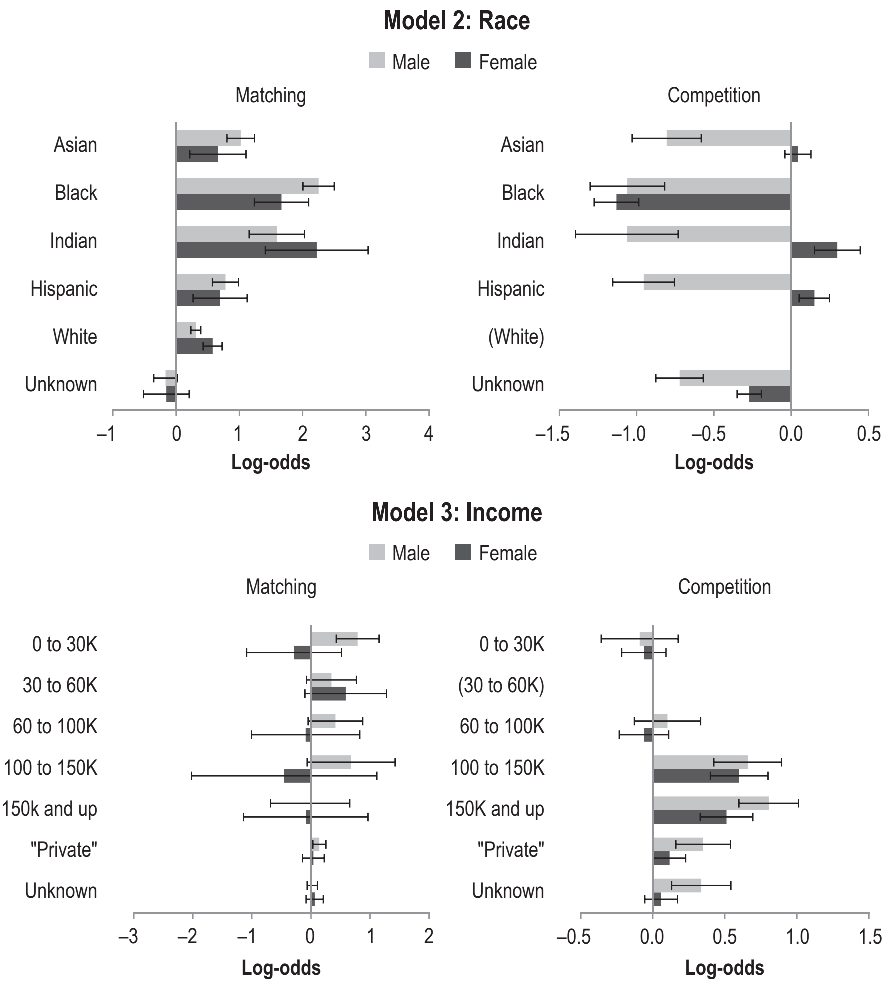
Similarity, Status, and Gender

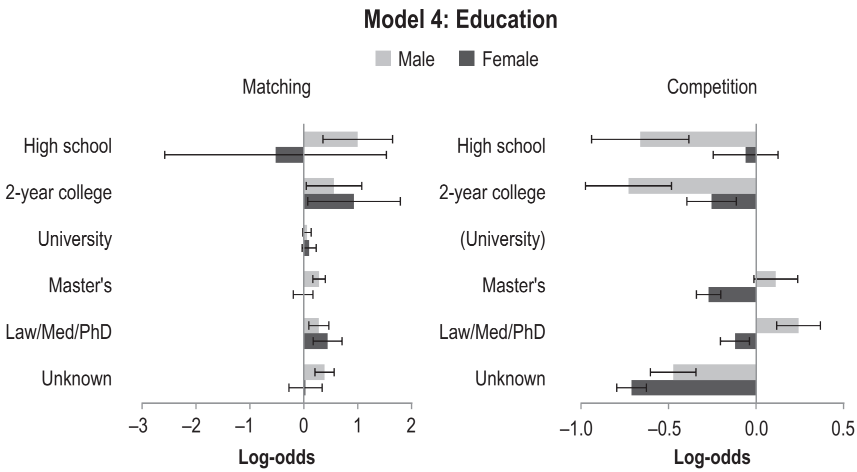
Figure 3 presents results from four models of preferences—one each for race, income, education, and religion. Each model includes four sets of terms: similarity-based matching (male and female) and status-based competition (male and female). I focus on the five largest categories for each variable (except income, where I aggregated available brackets into five bins); and for income, I include the extra category of "private." Also included in each model (but not pictured) are all terms from the control model; matching and competition effects for users from a background "other" than the five presented here; and "sociality" effects capturing the tendency for users from each category to send more or fewer messages overall.[10]

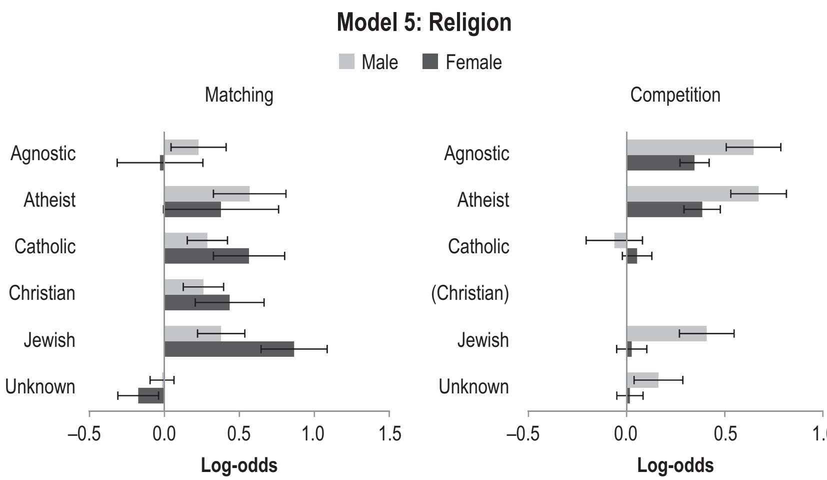
Race Men and women alike from all five backgrounds prefer racial similarity in a partner (model 2). This preference is strongest for blacks and Indians, followed by Asians and Hispanics, and finally by whites—an ordering that is similar to Blackwell and Lichter's (2000) study of racial endogamy in marriage and cohabitation. (While whites tend to have the lowest rates of intermarriage in the United States—see, e.g., Qian and Lichter [2007]—they also constitute the majority of the population, and so they have the fewest opportunities available for out-marriage. Consequently, other studies that control for the racial composition of the population similarly find the lowest degree of endogamy among whites.) Competition also plays a substantial role net of the effects of matching. While white men (the reference category) receive far more messages than men from all other backgrounds, white women receive far more messages than black women—and significantly *fewer* messages than Indian and Hispanic women. In this powerfully gendered hierarchy, it seems the critical divide for women is between white and non-white partners (cf. Lin and Lundquist 2013). New York City men, on the other hand, may seek Indian and Hispanic women for their "exoticism" and local scarcity (they are the smallest of the five categories considered here), while avoiding black women because they defy idealized notions of femininity (cf. Feliciano, Robnett, and Komaie 2009).

Income Prior research typically interprets socioeconomic similarity between partners as evidence of competition. In model 3, we see that matching also plays a role: Men who make \$0 to \$30,000 significantly prefer women in the same income bracket. This is plausibly due to a "floor" effect where all men prefer women who make less than they do—an option uniquely unavailable for men in the lowest category (cf. Verbrugge 1977). Shifting attention to competition, it is generally true that the higher one's income, the more messages one receives,

Figure 3. Parameter estimates and 95 percent confidence intervals from models of matching, status competition, and gender asymmetry among dating site users







Note: For competition terms, the reference category is in parentheses. Matching effects refer to the tendency to *send* messages to others from the *same* background (e.g., the tendency for Asian men to send messages to Asian women). Competition effects refer to the tendency to *receive* messages from others from *any* background (e.g., the tendency for Christian women to receive messages from all men). In addition to matching and competition effects, also included in each model (but not presented above) are all effects from the control model (model 1); effects capturing the tendency for men and women from each background to send more or fewer messages overall ("sociality"); and, in the case of race, education, and religion, matching and competition effects for men and women from a background "other" than the five presented above. Full model results are presented in appendix A.

though the most salient distinction appears to be between men and women who make over $100,000 annually and less than $100,000 annually. This pattern might also arise from the distinct economic composition of New York City. Results are otherwise difficult to interpret given the small proportion of users who report their income, resulting in imprecise estimates. Nonetheless, the presence of significant competition effects for both men and women is notable (given that emphasis is commonly placed on *male* economic resources), as is the presence of one significant instance of income-based matching (combined with a handful of other positive coefficients of borderline significance, especially for men).

Education Excepting men with a university degree, men at all other levels of education significantly prefer women who share their educational background (model 4). This tendency is strongest among high school–educated men, perhaps again due to a "floor" effect. On the other hand, only women with a two-year college degree and with a JD/MD/PhD prefer educational similarity in a partner—perhaps due to a comparable "ceiling" effect that prevents the latter from dating upward (cf. Skopek, Schulz, and Blossfeld 2011). Turning to competition effects, more educated men are generally more likely to be contacted by women, regardless of their own attainment level. Meanwhile, the important thing for women in terms of attracting a potential mate is having a bachelor's degree—and *only* a bachelor's degree: Women with a two-year college degree, a master's degree, or a JD/MD/PhD are significantly *less* likely to receive messages than women with a university education. (High school–educated women also receive fewer messages than college-educated women, though this difference is not statistically significant.) Mating preferences with respect to education are thus powerfully shaped by gender (see also Blossfeld 2009): While highly educated men are desired by everyone—*especially* highly educated women—highly educated women are desired *only* by highly educated men and avoided by everyone else.

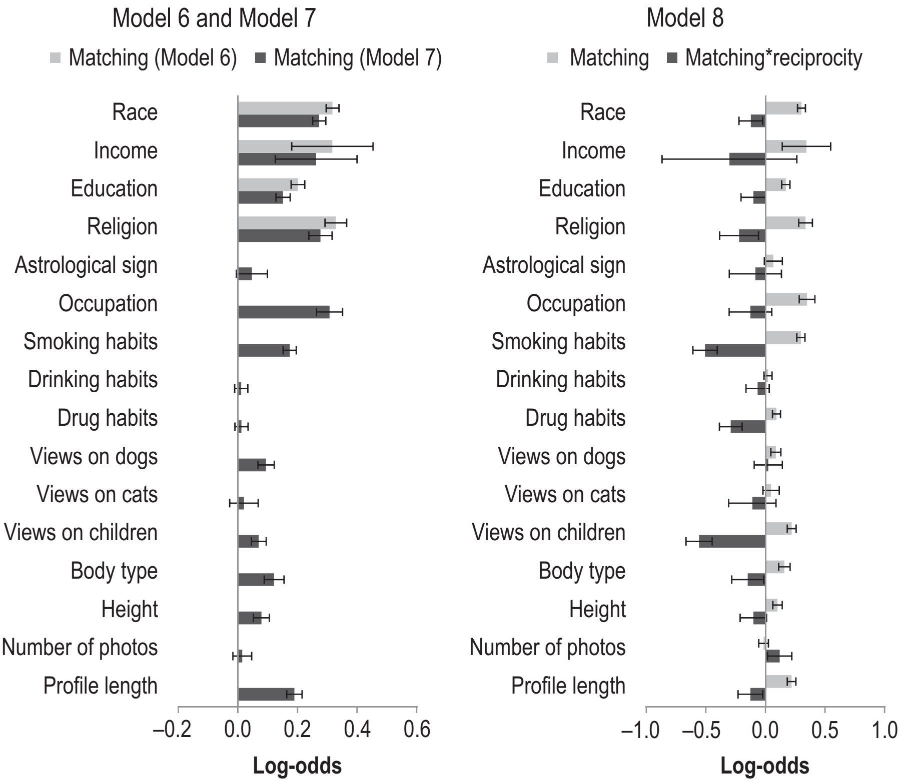
Religion Men and women from nearly all religious categories significantly prefer mates from the same background (model 5)—the only exception being agnostic women (and, marginally, atheist women), who as a consequence of their beliefs may feel open to religious and irreligious suitors alike. This preference is stronger for atheist and agnostic men compared to women and for theistic (and especially Jewish) women compared to men. Moving to competition effects, agnostics and atheists are more likely to receive messages than are users from all other religious affiliations. This is surprising given the moral boundaries surrounding atheists in the United States (Edgell, Gerteis, and Hartmann 2006), but could be a product of the (relatively young, educated, and areligious) population of OkCupid in historically liberal New York City. Women are also considerably more likely to contact Jewish men than Christian men, while Jewish women share no such success compared to Christian women—perhaps due to cultural stereotypes.

To date, homogeneity in mating patterns has been viewed as equally strong evidence for matching and competition. Here, it is apparent that both mechanisms are important—but not uniformly so. Matching effects are indeed most consistent for the categorical attributes of race and religion, but site users seek similarity with respect to many ordinal characteristics as well (e.g., men at most education levels and at least one income bracket). The strength of this tendency also varies considerably across categories (e.g., black and Indian users prefer similarity much more than do Asian, Hispanic, or white users); and in some instances (e.g., agnostic women), site users do not prefer similarity at all. Regarding competition, women prefer men the more education and income they report, while men prefer women with high incomes. But male preferences regarding education are impossible to interpret if we assume men simply maximize the collective status of the partnership; and substantial "competition" exists with respect to racial background and religious affiliation as well—especially over white men, Indian and Hispanic women, Jewish men, and male and female atheists and agnostics.

Multiple Dimensions of Preferences

In figure 4, I adopt a multivariate approach to mate selection. While this horizon forces me, for computational reasons, to omit the depth of models 2–5, models 6–8 enable us to examine the relative importance of matching based on a number of diverse attributes. In other words, by removing all competition effects and including a single parameter for each attribute that does not differentiate between male and female matching, these results tell us less about detailed underlying preferences than about overt, crosscutting patterns of homogeneity (cf. Blau and Schwartz 1984).

Figure 4. Parameter estimates and 95 percent confidence intervals from models of demographic and non-demographic matching and reciprocity among dating site users



Note: Also included in each model (but not presented above) are all effects from the control model (model 1). Full model results are presented in appendix B.

Model 6 includes matching effects for all four variables described above. Interestingly, once we take a multivariate approach—where each dimension of segregation is measured net of the others—matching based on race, income, and religion are all approximately equal in magnitude and somewhat greater than matching based on education. This finding is noteworthy, first, because education is commonly used as a proxy for income; and second, because prior research on social networks in the United States has typically found the greatest segregation according to race (McPherson, Smith-Lovin, and Cook 2001; although see Wimmer and Lewis 2010). It could be that prior work has exaggerated the degree of racial homogamy by ignoring local opportunity structures (see Harris and Ono 2005). The measurement of matching according to education and income, in particular, will also depend on how exactly the brackets surrounding "in-group" categories are drawn.

In model 7, I expand model 6 by incorporating a variety of additional attributes that are reported on user profiles. These attributes are diverse in nature—some having to do with attitudes (e.g., toward pets or children), others with behaviors (e.g., smoking and drinking), and still others with physical appearance (e.g., height and body type). Results convey two important points. First, we know from model 6 that users strongly prefer mates with whom they share demographic characteristics. However, two users who work in the same occupation are *more* likely to contact each other than are two users who share the same religious affiliation, racial background, or income bracket; and two users who have a similar profile length or who share the same smoking habits are more likely to contact each other than are two users with the same educational attainment level. Also significantly likely to message each other are users with the same self-described body type ($\eta = .12$), users with the same views about dogs ($\eta = .09$), users with a similar height ($\eta = .08$), and users with the same views about children ($\eta = .07$).

The second point has to do with the effects of including these new terms on our absolute measurement of demographic matching. Moving from model 6 to model 7, the coefficient for sharing the same racial background decreases by 14 percent; for sharing the same religious affiliation by 16 percent; for sharing the same income bracket by 17 percent; and for sharing the same educational attainment by 25 percent. In other words, some degree of demographic preferences can indeed be explained by attitudinal, behavioral, and physical preferences; and the omission of the latter has potentially led to a slight overestimation of the importance of demographic boundaries *per se* in prior work. Much of this effect may be due, in particular, to the important role that occupations play as "badges" summarizing a variety of features of one's lifestyle and values (Kalmijn 1994). Nonetheless, the coefficients for demographic matching remain among the largest in the model—indicating that there is still a substantial amount of demographic segregation that is irreducible to the non-demographic factors measured here.

Boundary-Crossing and Reciprocity

In model 8, I go one step further by taking a closer look at the tendency to reciprocate messages. Specifically, I add an additional term for each matching effect capturing the interaction between that effect and the general tendency to reciprocate messages. So while a positive matching coefficient again reflects a baseline preference for similarity, a positive (negative) reciprocity coefficient indicates that this preference is *stronger* (*weaker*) for replies than for initiations.

Results from this model are striking. While in general, reciprocity coefficients are less likely to be statistically significant (because the number of reciprocated ties is comparatively small), an overarching pattern is clear: 15 out of 16 matching coefficients are positive, while 14 out of 16 reciprocity coefficients are *negative*. Further, reciprocity coefficients are often statistically significant in precisely those cases where the boundary for an initial message is the strongest—such as matching with respect to race, religion, smoking habits, and views about children. Put differently, dating site users are more open to intergroup interaction when they reply to someone than when they initiate contact with someone (cf. Lewis 2013). Negative, significant reciprocity coefficients also appear for matching with respect to body type, drug habits, and profile length.[11]

As with all interaction effects, it is important to remember that each reciprocity coefficient must be interpreted in conjunction with the two other effects to which it relates: 1) the baseline reciprocity effect (a control term, not pictured) and 2) the specific matching effect beneath which it appears. First, this means that *any* two people are more likely to reply to each other than to initiate with each other—because the baseline reciprocity effect is so large ($\eta = 7.53$). Second, it is not necessarily the case that users prefer dissimilarity in replies in an absolute sense. Rather, this depends on the combined value of the matching effect and the matching × reciprocity effect, which is sometimes positive (e.g., for shared occupation) and sometimes negative (e.g., for shared views about children) and not necessarily significantly different from zero. For instance, the log-odds of user A messaging user B increase by .16 (matching effect) if (all else equal) they share the same body type. If A and B share the same body type *and* B has already contacted A, the log-odds of A messaging B increase by 7.54 (baseline reciprocity effect [7.53] + matching effect [.16] + matching × reciprocity effect [–.15]). And if A and B do *not* share the same body type and B has already contacted A, the increase in log-odds is not much different: 7.53 (baseline reciprocity effect). In other words, matching with respect to body type exists only with respect to initial messages, not replies.

In previous models, we saw that site users self-segregate according to a variety of demographic, behavioral, attitudinal, and physical traits. However, once we decompose an exchange into its smallest constituent parts, it is clear that the process of reciprocating a message is fundamentally different from the process of extending one in a way that is conducive to boundary erosion—and in fact, the most robust social divisions are also the most fragile. In short, someone who initiates contact on an online dating site—even (and especially) if it is someone one would not have otherwise considered—now appears in an entirely different light: Instead of an anonymous face in a sea of potential partners, she is a distinct, interested (and possibly the only) face in one's inbox—and also someone who has made the bold move of crossing an otherwise salient boundary. Such a person must at least be worth a look.[12]

Discussion

Patterns in romantic relationships are generated by multiple, interrelated factors. To date, however, there has remained a gap between our theoretical understanding of these causes and our empirical understanding of their importance. In this paper, I utilized network analytic methods and data from OkCupid to address three puzzles regarding preferences in the early stages of mate choice. First, I found that preferences for similarity as well as status contribute toward observed patterns of interaction—preferences that are not always symmetrical between men and women. Second, I showed that site users seek interpersonal compatibility (in terms of shared values, activities, and looks) as much as demographic similarity; and while some degree of demographic matching is a "by-product" of matching on these commonalities, much of these effects are independent. Finally, I showed that site users are more open to intergroup interaction in replies than in initiations—suggesting that "preferences" themselves are influenced by basic interactional dynamics.

Earlier, I noted that several other scholars have been attracted to the methodological advantages of online dating for studying mate choice. Their research provides some opportunity for triangulating results. Most closely paralleling my own findings, both Hitsch, Hortaçsu, and Ariely (2010) and Lin and Lundquist (2013) identify that both matching ("homophily") and competition ("hierarchy") play critical roles in user interaction—in contrast to other studies that have focused primarily on the former (e.g., Skopek, Schulz, and Blossfeld 2011) or the latter (e.g., Kreager et al. 2014). Fiore and Donath (2005) and Hitsch, Hortaçsu, and Ariely (2010) document matching according to several of the non-demographic variables examined here, including attitudes toward children, height, body type, and smoking behavior (although Fiore and Donath do not employ a multivariate framework and Hitsch, Hortaçsu, and Ariely find surprisingly little evidence for occupational matching). Finally, several other studies (in particular, Lin and Lundquist 2013; see also Kreager et al. 2014) have drawn attention to the importance of variation in preferences by *gender* and/or *stage of interaction*, although only Fiore and Donath (2005) compare initiations and replies across a wide variety of attributes, and—though they do not use statistical models—also find evidence that the preference for similarity is generally much smaller in replies. Unfortunately, insofar as my findings depart from prior work, it is often impossible to assess whether these differences result from focusing on a different stage of interaction (views vs. initiations vs. replies), sampling from different time periods or geographic regions, employing different statistical tools, or examining different (and typically unidentified) dating sites—each potentially with its own population idiosyncrasies, behavioral norms, and interface constraints (Lewis 2015b).

Several limitations of my analyses should also be kept in mind. In particular, my data do not contain information on two critical determinants of online interaction: user *attractiveness* and the *content* of users' profiles. If perceptions of attractiveness are correlated with any of the central variables of this analysis, this could account for apparent variation in preferences (e.g., Kreager et al. 2014). My findings will also be biased to the extent to which some users openly advertise or otherwise signal strong preferences in their profiles, preemptively dissuading certain others from contacting them. As noted earlier, the absence of data on message contents creates similar obstacles to interpretation—for example, some replies may be "polite rejections." Finally, because these analyses are based on aggregate patterns (any given user, for instance, might personally send very few messages), it is challenging to identify preferences on an individual level. We know, for instance, that there is evidence for pronounced matching on the basis of racial background, but whether this reflects a smaller proportion of individuals with strong preferences or a larger proportion of individuals with weaker preferences remains an open question for future research.

Nonetheless, these findings have a variety of implications for stratification and homogamy. In particular, because these data document the earliest stages of romantic "search" (Rosenfeld and Thomas 2012)—before prospective partners even meet, much less marry—it is clear that one explanation for contemporary homogamy is that patterns of romantic *interest* are heavily segregated from the beginning (see also Skopek, Schulz, and Blossfeld 2011). Similarly, when certain types of individuals are marginalized at this stage—particularly at the intersection of multiple attributes, such as black women with "too little" (*or* "too much") education—this severely constrains their opportunities to pursue and enjoy the various long-term benefits that partnership affords (e.g., Waite 1995). The consequences of this exclusion could extend to relationship satisfaction as much as participation, insofar as such individuals may be more likely to "settle" for a suboptimal partner; and despite speculation regarding the transformative possibilities of online dating (see Barraket and Henry-Waring 2008), the ease of searching for and filtering out potential mates on the basis of ascribed characteristics may exacerbate rather than ameliorate such inequality.

My aim has been deliberately broad rather than deep: Rather than providing an in-depth exploration of mating dynamics with respect to one variable (e.g., Anderson et al. 2014; Lin and Lundquist 2013; Skopek, Schulz, and Blossfeld 2011), I have focused on general mechanisms and overarching patterns across a wide variety of attributes, although grounded in a concrete empirical setting. While I believe such an approach is important given basic, unanswered questions in the literature, this step points naturally to a concrete direction for future research. Because we have only just begun to disentangle preferences from the other causes of mating patterns, *explanations* for these preferences are still largely underdeveloped. Why do Jewish women, for instance, prefer similarity in a partner more than Jewish men—and why should religious similarity matter more or less for either than for Catholic men and women? Prominent scholars of mate choice (e.g., Kalmijn 1998) have decried the gulf between theoretical and empirical work that has long characterized the study of marriage patterns. But mating scholarship itself is also largely sequestered from other fields that are directly relevant to its basic topic of inquiry—in particular, work on symbolic boundaries (Lamont and Molnár 2002), social networks (Rivera, Soderstrom, and Uzzi 2010), and gender as a basic dimension of social evaluation (Ridgeway 2009). The most promising next step would be to combine the theoretical and empirical insights from such work with the emphasis on preferences pursued in this paper. Such an approach would complement the one taken here by privileging explanatory depth over descriptive breadth; focusing narrowly on one or another type of attribute; and addressing the question of contextual variation that has been bracketed for the purposes of this study.

For the moment, I have here provided a broad but unusually nuanced overview of "subjective social distance" in mate selection: how complex social structures (operationalized as a network of romantic ties) are shaped "from the inside" by the preferences and prejudices of individuals and how interpersonal divisions are erected and eroded (Laumann and Senter 1976). As I have noted, traditional data sets tend to focus on formal categories of relationships (e.g., marriage) rather than on the underlying interactions that generated them. Such data obscure the complex, dynamic, and potentially asymmetric processes that take place in romantic pairing—processes that tell us as much about stratification as the outcomes they produce. Collecting longitudinal data, both quantitative and qualitative, and focusing as much on process as on outcomes (Blossfeld 2009; see also Emirbayer 1997; Lin and Lundquist 2013) is a prerequisite to building on the important body of literature that already exists and advancing it to the next level.

Notes

1. Prior to acquisition, the data set was stripped of all identifiers, including photographs, user names, and open-ended text.
2. One could argue that users looking only for "short-term dating" may be particularly open to experimentation. In practice, however, 93 percent of users looking for short-term dating are also looking for long-term dating; and short-term dating is still compatible with the type of general romantic consideration with which this study is concerned. It is important to acknowledge, however, that some users may view "new friends" or "activity partners" as a more appropriate route to finding a partner than explicit dating.
3. This raises two concerns: What if new users to the site display different preferences than veteran users, and what if users' preferences during October and November are not generalizable to other times of the year? To assess these possibilities to the extent I was able, I replicated models on subsets of users defined by (1) when they joined the site and (2) how long they were site members. Results from these robustness checks were generally consistent with those reported below and are available from the author by request.
4. Naturally, this issue will be most problematic for neighboring regions: For instance, users in New York City are 4.5 times more likely to message other users in New York City than they are to message users who are not in New York City but still live in a zip code beginning with "1"; yet users in New York City are 62 times more likely to message other users in New York City than they are to message users whose zip code begins with another number.
5. Comparisons are based on the October 2010 Internet Use Supplement of the Current Population Survey (CPS), and are limited by (1) the mismatch between "single" people in my sample and "unmarried" people in the CPS and (2) the fact that missing data on dating site profiles are unlikely to be randomly distributed.
6. While studies of other sites (Hitsch et al. 2010; Skopek et al. 2011) look only at first messages *conditional* on profile views, this approach underestimates the strength of preferences to an unknown degree: If the aversion between two users is strong

enough, they will never search for or view each other's profiles in the first place (cf. Lin and Lundquist 2013).

7. The following discussion draws heavily on the accessible introduction to ERG models written by Robins et al. (2007).
8. As Goodreau, Kitts, and Morris (2009) have noted, the equation can also be expressed as the conditional log-odds of individual ties, which may help clarify interpretation of the η vector: If forming a tie increases g_A by 1, then *ceteris paribus* the log-odds of that tie forming increase by η_A (p. 109).
9. I estimated all models using the R package ergm (Hunter et al. 2008b) and followed guidelines in the literature (Goodreau et al. 2008; Hunter, Goodreau, and Handcock 2008a) to confirm that all models converged and to assess model fit. Diagnostics, goodness-of-fit plots, and model settings are available from the author by request.
10. In the results that follow, I report categories precisely as they appear on OkCupid, which appears to privilege common usage over logical consistency: "Asian" and "Indian" are treated as distinct categories, as are "Catholic" and "Christian."
11. While the nature of ERG models prevents me from assessing the directionality of these effects, descriptive analyses (see Lewis 2013) suggest that this finding is robust across men and women.
12. Two alternative explanations are possible. One is that fear of rejection prevents users from expressing their "authentic" preferences with initiations but not with replies. Given the low cost of rejection in an online context, however, we might expect that such fear has a minimal impact on behavior—an expectation that was empirically confirmed by Hitsch et al. (2010) in their study of an anonymous dating site (see also Kreager et al. 2014). Second, even if overt interference from OkCupid is relatively low, the site may still implicitly promote a "consumerist" mentality wherein users are encouraged to view each other in terms of objective, tangible attributes (see Frost et al. 2008)—whereas when users reply to each other, they consider each other more "holistically." Unfortunately, I do not have data available to test this possibility.

Appendix A

Table A1. Full Parameter Estimates and Standard Errors for Models 2–5 in Figure 3

	Model 2		Model 3		Model 4		Model 5	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Controls								
Structural effects								
Density	–9.122***	.101	–9.226***	.134	–8.999***	.093	–9.286***	.099
Reciprocity	6.712***	.032	6.784***	.032	6.730***	.032	6.792***	.032
Activity spread	–3.116***	.038	–3.143***	.038	–3.103***	.038	–3.143***	.038
Gender/age effects								
Female-receiver	2.642***	.135	2.914***	.183	2.688***	.127	3.034***	.135
Age-sender	.109***	.002	.106***	.002	.109***	.002	.106***	.002
Age-receiver	.103***	.004	.102***	.004	.105***	.004	.106***	.004
Age × female-sender	–.139***	.005	–.138***	.005	–.140***	.005	–.138***	.005
Age × female-receiver	–.199***	.005	–.199***	.005	–.196***	.005	–.202***	.005
Age-absolute difference	–.154***	.004	–.156***	.004	–.152***	.004	–.155***	.004
Age-(absolute difference)^2	–.002***	3e-4	–.002***	3e-4	–.002***	3e-4	–.002***	3e-4
Propinquity effect								
Same 3-digit zip code	.242***	.012	.249***	.012	.221***	.012	.257***	.012
Site usage effects								
Account duration-sender	–.020***	.001	–.021***	.001	–.020***	.001	–.020***	.001
Account duration-receiver	–.004***	.001	–.004***	.001	–.004***	.001	–.004***	.001
Login time-sender	2e-5***	6e-7	2e-5***	6e-7	2e-5***	6e-7	2e-5***	5e-7
Login time-receiver	4e-6***	7e-7	2e-6***	7e-7	3e-6***	7e-7	2e-6**	6e-7
Has picture(s)-sender	–.039**	.012	–.033**	.012	–.041***	.012	–.042***	.012

Has picture(s)-receiver	.088***	.015	.094***	.015	.075***	.015	.099***	.015
Account overlap	.031***	.001	.031***	.001	.031***	.001	.031***	.001
Race								
Matching								
Male								
Asian	1.022***	.111						
Black	2.253***	.126						
Indian	1.592***	.223						
Hispanic	.780***	.105						
White	.312***	.040						
Other	–.114	.095						
Unknown	–.166	.096						
Female								
Asian	.660**	.227						
Black	1.666***	.218						
Indian	2.224***	.414						
Hispanic	.696**	.218						
White	.577***	.076						
Other	.135	.156						
Unknown	–.154	.184						
Sociality								
Male								
Asian	.009	.047						
Black	.144**	.046						

(*Continued*)

Table A1. *continued*

	Model 2		Model 3		Model 4		Model 5	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Indian	.092	.058						
Hispanic	.487***	.040						
Other	.237***	.038						
Unknown	.046	.036						
Female								
Asian	.346***	.072						
Black	.350***	.095						
Indian	–.149	.146						
Hispanic	.422***	.078						
Other	.440***	.072						
Unknown	.566***	.069						
Competition								
Male								
Asian	–.806***	.114						
Black	–1.060***	.123						
Indian	–1.063***	.170						
Hispanic	–.955***	.102						
Other	–.258***	.076						
Unknown	–.722***	.078						

Female				
Asian	.044	.043		
Black	−1.131***	.074		
Indian	.299***	.075		
Hispanic	.150**	.050		
Other	.162***	.041		
Unknown	−.271***	.040		
Income				
Matching				
Male				
0 to 30K			.791***	.185
30 to 60K			.346	.216
60 to 100K			.414	.236
100 to 150K			.681	.379
150K and up			−.014	.342
"Private"			.141*	.056
Unknown			.022	.045
Female				
0 to 30K			−.287	.410
30 to 60K			.589	.352
60 to 100K			−.091	.467
100 to 150K			−.453	.800
150K and up			−.089	.537

(*Continued*)

Table A1. *continued*

	Model 2		Model 3		Model 4		Model 5	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
"Private"			.041	.094				
Unknown			.063	.074				
Sociality								
Male								
0 to 30K			.386***	.052				
60 to 100K			.341***	.050				
100 to 150K			–.098	.059				
150K and up			.162**	.051				
"Private"			.113**	.041				
Unknown			–.085	.049				
Female								
0 to 30K			.032	.094				
60 to 100K			.017	.094				
100 to 150K			–.395**	.133				
150K and up			–.116	.117				
"Private"			–.001	.066				
Unknown			–.050	.074				
Competition								
Male								
0 to 30K			–.092	.136				
60 to 100K			.102	.117				

100 to 150K	.659***	.120		
150K and up	.804***	.106		
“Private”	.350***	.097		
Unknown	.336**	.105		
Female				
0 to 30K	–.062	.079		
60 to 100K	–.061	.087		
100 to 150K	.600***	.102		
150K and up	.512***	.093		
“Private”	.117*	.057		
Unknown	.058	.058		
Education				
Matching				
Male				
High school			1.001**	.329
2-year college			.561*	.263
University			.060	.040
Master’s			.283***	.059
Law/Med/PhD			.280**	.095
Other			–.605	.789
Unknown			.385***	.091

(*Continued*)

Table A1. ***continued***

	Model 2		Model 3		Model 4		Model 5	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Female								
High school					–.522	1.049		
2-year college					.931*	.438		
University					.100	.066		
Master’s					–.013	.093		
Law/Med/PhD					.441**	.137		
Other					.682	.807		
Unknown					.032	.157		
Sociality								
Male								
High school					.283***	.053		
2-year college					.109*	.049		
Master’s					–.037	.038		
Law/Med/PhD					.019	.039		
Other					–.292***	.060		
Unknown					–.281***	.035		
Female								
High school					.318**	.109		
2-year college					–.138	.102		
Master’s					.074	.051		
Law/Med/PhD					–.042	.060		

Other	.068	.106		
Unknown	.216***	.055		
Competition				
Male				
High school	–.662***	.142		
2-year college	–.729***	.125		
Master's	.113	.064		
Law/Med/PhD	.243***	.064		
Other	.100	.111		
Unknown	–.473***	.067		
Female				
High school	–.060	.095		
2-year college	–.254***	.072		
Master's	–.271***	.036		
Law/Med/PhD	–.120**	.042		
Other	.140	.088		
Unknown	–.711***	.043		
Religion				
Matching				
Male				
Agnostic			.228*	.094
Atheist			.569***	.124
Catholic			.286***	.069
Christian			.260***	.069

(*Continued*)

Table A1. ***continued***

	Model 2		Model 3		Model 4		Model 5	
	Coef	SE	Coef	SE	Coef	SE	Coef	SE
Jewish							.378***	.080
Other							.108	.102
Unknown							–.016	.041
Female								
Agnostic							–.030	.146
Atheist							.377	.196
Catholic							.564***	.121
Christian							.435***	.117
Jewish							.865***	.113
Other							.388*	.172
Unknown							–.175*	.069
Sociality								
Male								
Agnostic							–.306***	.038
Atheist							–.331***	.039
Catholic							.095**	.034
Jewish							–.185***	.036
Other							–.077*	.036
Unknown							–.206***	.031
Female								
Agnostic							–.094	.049

Atheist	−.285***	.068
Catholic	−.054	.047
Jewish	−.147**	.050
Other	−.117*	.053
Unknown	.083	.044
Competition		
Male		
Agnostic	.647***	.071
Atheist	.672***	.072
Catholic	−.063	.073
Jewish	.408***	.071
Other	.012	.077
Unknown	.163*	.063
Female		
Agnostic	.346***	.038
Atheist	.385***	.047
Catholic	.053	.038
Jewish	.026	.039
Other	.297***	.042
Unknown	.017	.034

Note: For sociality and competition effects, the reference categories are as follows: white (race), 30 to 60K (income), university (education), Christian (religion). Matching effects refer to the tendency to *send* messages to others from the *same* background. Sociality effects refer to the tendency to *send* messages to others from *any* background. Competition effects refer to the tendency to *receive* messages from others from *any* background.
*$p < .05$ **$p < .01$ ***$p < .001$ (two-tailed tests)

Appendix B

Table B1. Full Parameter Estimates and Standard Errors for Models 6–8 in Figure 4

	Model 6		Model 7		Model 8	
	Coef	SE	Coef	SE	Coef	SE
Controls						
Structural effects						
Density	–9.293***	.074	–9.491***	.075	–9.765***	.077
Reciprocity	6.696***	.032	6.643***	.032	7.526***	.067
Activity spread	–3.068***	.038	–2.982***	.038	–2.981***	.038
Gender/age effects						
Female-receiver	2.713***	.102	2.722***	.102	2.836***	.102
Age-sender	.109***	.002	.109***	.002	.109***	.002
Age-receiver	.105***	.004	.104***	.004	.104***	.004
Age × female-sender	–.138***	.005	–.139***	.005	–.137***	.005
Age × female-receiver	–.200***	.005	–.201***	.005	–.202***	.005
Age-absolute difference	–.151***	.004	–.150***	.004	–.150***	.004
Age- (absolute difference)^2	–.002***	3e-4	–.002***	3e-4	–.002***	3e-4
Propinquity effect						
Same 3-digit zip code	.229***	.012	.231***	.012	.234***	.012
Site usage effects						
Account duration-sender	–.020***	.001	–.020***	.001	–.020***	.001
Account duration-receiver	–.004***	.001	–.004***	.001	–.004***	.001
Login time-sender	2e-5***	5e-7	2e-5***	6e-7	2e-5***	6e-7
Login time-receiver	3e-6***	6e-7	3e-6***	7e-7	3e-6***	7e-7
Has picture(s)-sender	–.040**	.012	–.039**	.014	–.038**	.013
Has picture(s)-receiver	.086***	.015	.091***	.016	.091***	.016
Account overlap	.031***	.001	.031***	.001	.031***	.001
Demographic matching						
Race	.318***	.011	.273***	.011	.303***	.017
*Reciprocity					–.122*	.051
Income	.317***	.070	.263***	.070	.344***	.104
*Reciprocity					–.302	.289
Education	.202***	.012	.152***	.012	.171***	.018
*Reciprocity					–.101	.053

(*Continued*)

Table B1. ***continued***

	Model 6		Model 7		Model 8	
	Coef	SE	Coef	SE	Coef	SE
Religion	.329***	.019	.277***	.020	.337***	.029
*Reciprocity					–.221**	.083
Non-demographic matching						
Astrological sign			.047	.027	.066	.038
*Reciprocity					–.084	.112
Occupation			.308***	.022	.349***	.034
*Reciprocity					–.126	.091
Smoking habits			.174***	.011	.298***	.017
*Reciprocity					–.507***	.052
Drinking habits			.011	.011	.022	.017
*Reciprocity					–.066	.050
Drug habits			.012	.011	.093***	.017
*Reciprocity					–.291***	.049
Views on dogs			.095***	.014	.088***	.021
*Reciprocity					.024	.060
Views on cats			.020	.024	.048	.035
*Reciprocity					–.110	.101
Views on children			.070***	.013	.221***	.019
*Reciprocity					–.557***	.056
Body type			.122***	.017	.160***	.025
*Reciprocity					–.149*	.069
Height			.079***	.014	.101***	.020
*Reciprocity					–.101	.057
Number of photos			.015	.016	–.015	.021
*Reciprocity					.120*	.052
Profile length			.190***	.013	.220***	.018
*Reciprocity					–.126*	.053

*$p < .05$ **$p < .01$ ***$p < .001$ (two-tailed tests)

About the Author

Kevin Lewis is an assistant professor of Sociology at the University of California–San Diego. His research focuses on the mechanisms underlying social network formation and evolution and has been published in *Social Networks*, the *American Journal of Sociology*, the *Proceedings of the National Academy of Sciences of the United States of America*, and *Sociological Science*.

References

Anderson, Ashton, Sharad Goel, Gregory Huber, Neil Malhotra, and Duncan J. Watts. 2014. "Political Ideology and Racial Preferences in Online Dating." *Sociological Science* 1:28–40.

Barraket, Jo, and Millsom S. Henry-Waring. 2008. "Getting It on(Line): Sociological Perspecetives on E-Dating." *Journal of Sociology* 44:149–65.

Blackwell, Debra L., and Daniel T. Lichter. 2000. "Mate Selection among Married and Cohabiting Couples." *Journal of Family Issues* 21:275–302.

Blau, Peter M., and Joseph E. Schwartz. 1984. *Crosscutting Social Circles: Testing a Macrostructural Theory of Intergroup Relations*. Orlando, FL: Academic Press.

Blossfeld, Hans-Peter. 2009. "Educational Assortative Marriage in Comparative Perspective." *Annual Review of Sociology* 35:513–30.

Bratter, Jenifer L., and Karl Eschbach. 2006. "'What About the Couple?' Interracial Marriage and Psychological Distress." *Social Science Research* 35:1025–47.

Buss, David M., and David P. Schmitt. 1993. "Sexual Strategies Theory: An Evolutionary Perspective on Human Mating." *Psychological Review* 100:204–32.

Clark-Ibáñez, Marisol, and Diane Felmlee. 2004. "Interethnic Relationships: The Role of Social Network Diversity." *Journal of Marriage and Family* 66:293–305.

DiMaggio, Paul, and John Mohr. 1985. "Cultural Capital, Educational Attainment, and Marital Selection." *American Journal of Sociology* 90:1231–61.

Edgell, Penny, Joseph Gerteis, and Douglas Hartmann. 2006. "Atheists as 'Other': Moral Boundaries and Cultural Membership in American Society." *American Sociological Review* 71:211–34.

Edwards, John N. 1969. "Familial Behavior as Social Exchange." *Journal of Marriage and the Family* 31:518–26.

Emirbayer, Mustafa. 1997. "Manifesto for a Relational Sociology." *American Journal of Sociology* 103:281–317.

Erickson, Bonnie H. 1996. "Culture, Class, and Connections." *American Journal of Sociology* 102:217–51.

Feliciano, Cynthia, Belinda Robnett, and Golnaz Komaie. 2009. "Gendered Racial Exclusion among White Internet Daters." *Social Science Research* 38:39–54.

Fiore, Andrew T., and Judith S. Donath. 2005. "Homophily in Online Dating: When Do You Like Someone Like Yourself?" In *ACM Computer-Human Interaction*. Portland, OR.

Frost, Jeana H., Zoë Chance, Michael I. Norton, and Dan Ariely. 2008. "People Are Experience Goods: Improving Online Dating with Virtual Dates." *Journal of Interactive Marketing* 22:51–61.

Goodreau, Steven M., Mark S. Handcock, David R. Hunter, Carter T. Butts, and Martina Morris. 2008. "A Statnet Tutorial." *Journal of Statistical Software 24*.

Goodreau, Steven M., James A. Kitts, and Martina Morris. 2009. "Birds of a Feather, or Friend of a Friend? Using Exponential Random Graph Models to Investigate Adolescent Social Networks." *Demography* 46:103–25.

Harris, David R., and Hiromi Ono. 2005. "How Many Interracial Marriages Would There Be If All Groups Were of Equal Size in All Places? A New Look at National Estimates of Interracial Marriage." *Social Science Research* 34:236–51.

Hitsch, Günter J., Ali Hortaçsu, and Dan Ariely. 2010. "What Makes You Click? Mate Preferences in Online Dating." *Quantitative Marketing and Economics* 8:393–427.

Hunter, David R., Steven M. Goodreau, and Mark S. Handcock. 2008a. "Goodness of Fit of Social Network Models." *Journal of the American Statistical Association* 103:248–58.

Hunter, David R., Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris. 2008b. "Ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks." *Journal of Statistical Software 24*.

Kalmijn, Matthijs. 1991. "Status Homogamy in the United States." *American Journal of Sociology* 97:496–523.

_______. 1994. "Assortative Mating by Cultural and Economic Occupational Status." *American Journal of Sociology* 100:422–52.

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

Kalmijn, Matthijs, and Henk Flap. 2001. "Assortative Meeting and Mating: Unintended Consequences of Organized Settings for Partner Choices." *Social Forces* 79:1289–1312.

Kalmijn, Matthijs, and Jeroen K. Vermunt. 2007. "Homogeneity of Social Networks by Age and Marital Status: A Multilevel Analysis of Ego-Centered Networks." *Social Networks* 29:25–43.

Kossinets, Gueorgi, and Duncan J. Watts. 2009. "Origins of Homophily in an Evolving Social Network." *American Journal of Sociology* 115:405–50.

Kreager, Derek A., Shannon E. Cavanagh, John Yen, and Mo Yu. 2014. "'Where Have All the Good Men Gone?' Gendered Interactions in Online Dating." *Journal of Marriage and Family* 76:387–410.

Lamont, Michèle, and Annette Lareau. 1988. "Cultural Capital: Allusions, Gaps and Glissandos in Recent Theoretical Developments." *Sociological Theory* 6:153–68.

Lamont, Michèle, and Virág Molnár. 2002. "The Study of Boundaries in the Social Sciences." *Annual Review of Sociology* 28:167–95.

Laumann, Edward O., and Richard Senter. 1976. "Subjective Social Distance, Occupational Stratification, and Forms of Status and Class Consciousness: A Cross-National Replication and Extension." *American Journal of Sociology* 81:1304–38.

Lewis, Kevin. 2013. "The Limits of Racial Prejudice." *Proceedings of the National Academy of Sciences of the United States of America* 110:18814–19.

_______. 2015a. "Studying Online Behavior: Comment on Anderson et al. 2014." *Sociological Science* 2:20–31.

_______. 2015b. "Three Fallacies of Digital Footprints." *Big Data & Society 2*.

Lichter, Daniel T., Felicia B. LeClere, and Diane K. McLaughlin. 1991. "Local Marriage Markets and the Marital Behavior of Black and White Women." *American Journal of Sociology* 96:843–67.

Lin, Ken-Hou, and Jennifer Lundquist. 2013. "Mate Selection in Cyberspace: The Intersection of Race, Gender, and Education." *American Journal of Sociology* 119:183–215.

Logan, John Allen, Peter D. Hoff, and Michael A. Newton. 2008. "Two-Sided Estimation of Mate Preferences for Similarities in Age, Education, and Religion." *Journal of the American Statistical Association* 103:559–69.

Lundquist, Jennifer H., and Ken-Hou Lin. 2015. "Is Love (Color) Blind? The Economy of Race among Gay and Straight Daters." *Social Forces* 93:1423–49.

Mare, Robert D. 1991. "Five Decades of Educational Assortative Mating." *American Sociological Review* 56:15–32.

Marsden, Peter V. 1990. "Network Data and Measurement." *Annual Review of Sociology* 16:435–63.

McClintock, Elizabeth Aura. 2010. "When Does Race Matter? Race, Sex, and Dating at an Elite University." *Journal of Marriage and Family* 72:45–72.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27:415–44.

Oppenheimer, Valerie Kincade. 1977. "The Sociology of Women's Economic Role in the Family." *American Sociological Review* 42:387–406.

_______. 1988. "A Theory of Marriage Timing." *American Journal of Sociology* 94:563–91.

Qian, Zhenchao, and Daniel T. Lichter. 2007. "Social Boundaries and Marital Assimilation: Interpreting Trends in Racial and Ethnic Intermarriage." *American Sociological Review* 72:68–94.

Ridgeway, Cecilia L. 2009. "Framed before We Know It: How Gender Shapes Social Relations." *Gender & Society* 23:145–60.

Rivera, Lauren A. 2012. "Hiring as Cultural Matching: The Case of Elite Professional Service Firms." *American Sociological Review* 77:999–1022.

Rivera, Mark T., Sara B. Soderstrom, and Brian Uzzi. 2010. "Dynamics of Dyads in Social Networks: Assortative, Relational, and Proximity Mechanisms." *Annual Review of Sociology* 36:91–115.

Robins, Garry, Pip Pattison, Yuval Kalish, and Dean Lusher. 2007. "An Introduction to Exponential Random Graph (P*) Models for Social Networks." *Social Networks* 29:173–91.

Robins, Garry, Pip Pattison, and Peng Wang. 2009. "Closure, Connectivity and Degree Distributions: Exponential Random Graph (P*) Models for Directed Social Networks." *Social Networks* 31:105–17.

Rosenfeld, Michael J., and Reuben J. Thomas. 2012. "Searching for a Mate: The Rise of the Internet as a Social Intermediary." *American Sociological Review* 77:523–47.

Rudder, Christian. 2014. *Dataclysm: Who We Are (When We Think No One's Looking)*. New York: Crown Publishers.

Sautter, Jessica M., Rebecca M. Tippett, and S. Philip Morgan. 2010. "The Social Demography of Internet Dating in the United States." *Social Science Quarterly* 91:554–75.

Schoen, Robert, and Robin M. Weinick. 1993. "Partner Choice in Marriages and Cohabitations." *Journal of Marriage and the Family* 55:408–14.

Schwartz, Christine R. 2010. "Pathways to Educational Homogamy in Marital and Cohabiting Unions." *Demography* 47:735–53.

Schwartz, Christine R. 2013. "Trends and Variation in Assortative Mating: Causes and Consequences." *Annual Review of Sociology 39*.

Skopek, Jan, Florian Schulz, and Hans-Peter Blossfeld. 2011. "Who Contacts Whom? Educational Homophily in Online Mate Selection." *European Sociological Review* 27:180–95.

Smith, Aaron, and Maeve Duggan. 2013. "Online Dating & Relationships." *Pew Research Center, Washington, DC*.

Snijders, Tom A. B., Philippa E. Pattison, Garry L. Robins, and Mark S. Handcock. 2006. "New Specifications for Exponential Random Graph Models." *Sociological Methodology* 36:99–153.

Stevens, Gillian, Dawn Owens, and Eric C. Schaefer. 1990. "Education and Attractiveness in Marriage Choices." *Social Psychology Quarterly* 53:62–70.

Taylor, Lindsay Shaw, Andrew T. Fiore, G. A. Mendelsohn, and Coye Cheshire. 2011. "'Out of My League': A Real-World Test of the Matching Hypothesis." *Personality and Social Psychology Bulletin* 37:942–54.

Verbrugge, Lois M. 1977. "The Structure of Adult Friendship Choices." *Social Forces* 56:576–97.

Waite, Linda J. 1995. "Does Marriage Matter?" *Demography* 32:483–507.

Wimmer, Andreas, and Kevin Lewis. 2010. "Beyond and Below Racial Homophily: ERG Models of a Friendship Network Documented on Facebook." *American Journal of Sociology* 116:583–642.