

Studying Online Behavior: Comment on Anderson et al. 2014

Kevin Lewis

University of California, San Diego

Abstract: As social scientists increasingly employ data from online sources, it is important that we acknowledge both the advantages and limitations of this research. The latter have received comparatively little public attention. In this comment, I argue that a recent article by Anderson and colleagues: 1) inadequately describes the study sample; 2) inadequately describes how the website operates; and 3) inadequately develops the paper's central measures – such that it is difficult to evaluate the generalizability, veracity, and importance of their claims and impossible to replicate their findings. These limitations are not unique to the Anderson et al. article; rather, they point to a set of concerns that all researchers in this growing and important line of study need to address if our work is to have enduring impact.

Keywords: political ideology; racial preferences; homogamy; homophily; online dating; computational social science

THE internet is transforming social interaction—but it is also creating exciting new opportunities for social science (Lazer et al. 2009). A recent example of this is provided by Ashton Anderson, Sharad Goel, Gregory Huber, Neil Malhotra, and Duncan Watts in their article, “[Political Ideology and Racial Preferences in Online Dating](#)” (*Sociological Science* 1: 28–40). The authors’ aim is to measure racial preferences in mate selection. Their approach is to examine patterns of who views the profile of whom on an online dating website. They appropriately frame their work in terms of the “initial screening decisions” where “individuals rule out many potential dating partners from further consideration” (P. 29); their ability to measure both stated and revealed preferences represents a powerful potential contribution; and their emphasis on variation by political ideology is interesting and important. Finally, their conclusions are noteworthy and broadly relevant: while conservatives are less racially open than liberals, individuals of *all* political persuasions prefer same-race partners—even when they claim not to.

This article is not the first to use data from an online dating site to study preferences (see, e.g., Feliciano, Lee, and Robnett 2011; Fiore 2004; Hitsch, Hortaçsu, and Ariely 2010; Lewis 2013; Lin and Lundquist 2013; Skopek, Schulz, and Blossfeld 2011; Taylor et al. 2011; Yancey 2009). Research on online dating, in turn, represents but one small corner of a new universe of scholarship using “digital footprints” to better understand human behavior and interaction (see review in Golder and Macy 2014). While there is much reason for enthusiasm about this growing body of work, its limitations are too seldom acknowledged. In this comment, I argue that the strength of the Anderson et al. article is undermined by: 1) inadequate description of the sample (so that we don’t know whom the findings are about); 2) inadequate

Citation: Lewis, Kevin. 2015. “Studying Online Behavior: Comment on Anderson et al. 2014” *Sociological Science* 2: 20–31.

Received: September 19, 2014

Accepted: September 29, 2014

Published: January 21, 2015

Editor(s): Jesper Sørensen, Stephen L. Morgan

DOI: 10.15195/v2.a2

Copyright: © 2015 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited. 

description of the dating site (so that we don't know whether the findings are artifacts of the site's architecture); and 3) inadequate theoretical development and interpretation (such that even if we knew who is in the sample and how the website works, the meaning and importance of the findings are still ambiguous).

These limitations are not unique to Anderson et al.'s work. Rather, I also highlight the relevance of each concern for contemporary research using digital data—research that is often difficult to evaluate and impossible to meaningfully replicate. It is equally important to note that I do not consider my own work exempt from these criticisms. Rather, I am indebted to a number of friends, colleagues, and anonymous reviewers who have helped me identify and appreciate these concerns—concerns that will be most useful to all of us if they are brought into a forum for public discussion.

Who is in their sample?

A common problem with electronic data is that they are “at once too revealing in terms of privacy protection, yet also not revealing enough in terms of providing the demographic background information needed by social scientists” (Golder and Macy 2014:141). However, even when such information is available—such as on many contemporary dating sites—it may mask important distinctions that would drastically alter our interpretation of results.

Who *isn't* in their sample?

On page 30, Anderson et al. describe the size and composition of their sample. What they do not describe is the size or composition of the population they *began* with—such that we have no idea how small or unrepresentative is their slice of the pie. First, the authors explain that “We restrict our analysis to users with relatively complete demographic profiles—those reporting age, sex, location, ethnicity, education, income, political ideology, marital status, religion, height, body type, drinking habits, smoking habits, presence of children, and desire for more children—and who also explicitly express a preference, or lack of a preference, for a potential partner's race.” This is an incredibly demanding set of requirements, and users who are willing to provide information on any one of these dimensions may vary systematically from those who are not.¹ In particular, requiring data on all of these attributes seems to needlessly restrict attention to users who are particularly “open”—and therefore might also feel particularly comfortable divulging discriminatory preferences about race.

Second, Anderson et al. restrict attention to whites and blacks (available options are “white,” “black,” “Asian,” “Hispanic,” and “other”) because “Hispanics and Asians are sufficiently heterogeneous categories that ‘same-race’ preference may have little meaning” (P. 30). In fact, prior work on online dating has documented substantial same-race preferences across all four racial categories (white, black, Asian, Hispanic; see Hitsch et al. 2010; Lewis 2013; Lin and Lundquist 2013); even if users do not identify with these blanket labels, nested dynamics of ethno-racial identification and homophily will still produce racial matching in the aggregate

(Wimmer and Lewis 2010). Historically, a great deal of scholarship on “race” in the United States has been forced for practical reasons to rely on black/white binary measures. Particularly with data on this scale, excluding Hispanic and Asian users bypasses an easy opportunity to expand prior literature and prevents a number of potentially instructive comparisons.²

Third, the authors indicate that they “collected a complete snapshot of activity on the site during a two-month period (October–November 2009)” (P. 30). More detail is needed. On any subscription-based website, membership is constantly evolving as users come and go. Further, most sites show tremendous variation in activity levels—where some users participate a lot and other users do not participate at all. Decisions about how to treat these various individuals and their behaviors—i.e., how to define the “network boundary”—are not at all trivial (Laumann, Marsden, and Prensky 1983). For instance, in their own study of racial preferences in online dating, Lin and Lundquist (2013) identified and excluded “spammer users” whose “preferences” are probably atypical, yet who contribute an unusually large number of data points; they also excluded users who did not send or receive at least one message (i.e., network “isolates”), a decision that can heavily impact the measurement of homophily (Bojanowski and Corten 2014). Strictly speaking, two users whose account periods did not overlap should also not be considered eligible to view one another’s profiles. Finally, on a more substantive level, the authors should consider whether the behavior they have recorded—profile views on an online dating site during the early holiday season—might or might not be representative of other circumstances. One can easily come up with a number of reasons why preferences might be narrower during this time (e.g., because users are particularly concerned about family approval) or more open (e.g., because users feel particularly lonely).

What kind of site are they studying?

Even people who have only the most cursory, secondhand understanding of online dating are probably familiar with the striking variety in online dating sites (for an overview, see Finkel et al. 2012). There are large sites and small sites. There are free sites and sites with fee-based subscriptions. There are sites that are about casual versus serious dating; sites that operate primarily through mobile applications; and sites that incorporate group dates, virtual dates, or even genetic testing. Perhaps most importantly, there are sites that cater to a general audience and sites that cater to a particular market niche—where “niche” has been defined in every conceivable way (from JDate for Jewish singles to Ashley Madison for people seeking extramarital affairs to FarmersOnly—because “city folks just don’t get it”). Users of these various sites are almost certainly very different kinds of people who seek very different characteristics in a partner. And yet when social scientists use these data, the site descriptions they provide are remarkably generic, such as: “a popular online dating website in which users could view personal profiles and send messages to other members of the site” (Anderson et al. 2014:29–30).³

Naturally, anonymity is familiar to any consumer of social science—most commonly used to mask the identities of individuals or field sites. However, we generally assume that even when the identity of individual subjects or field sites

is concealed, no *characteristics* of these people or sites are omitted that would *substantially alter our interpretation of the data*. In the case of online dating, the generic description of “popular online dating site” is simply not enough: unless we know the kind of site we are dealing with and the kinds of aims that its users pursue, it is impossible to interpret their behavior.⁴

In sum, while Anderson and colleagues present a limited demographic description of their sample, we still do not know whom exactly they are studying—and therefore it is impossible to compare their results to prior work, replicate their findings using alternative data sources, or assess how broadly their conclusions may be applicable. On one hand, this reflects an increasing gap between contemporary internet-based research and prior work (e.g., on dating or marriage markets) that samples from a well defined frame. On the other hand, this reflects a growing trend in internet-based research where insufficient information is provided about the website to interpret results—even if they are not meant to be statistically generalizable. All sociologists, regardless of methodological orientation, must navigate the dual goals of protecting the privacy of research subjects while providing findings that are meaningful and broadly important. It is unclear why we should alter our standards just because the sample is larger, the information is more fine-grained, or the data were acquired from a private source.

How are profile views generated?

Generalizability and interpretation are not the only concerns that arise from inadequate information about a website. An equally important issue is the extent to which computer-mediated interaction is constrained and/or influenced by the site’s architecture.

Online dating is an attractive tool to sociologists because it represents the possibility of resolving a timeless question about mating patterns, intergroup boundaries, and subjective social distance (Kalmijn 1998; Laumann and Senter 1976): to what extent are these patterns generated by preferences as opposed to the opportunity constraints individuals face when selecting a partner? While online dating sites may seem like relatively “open” markets where preferences reign free of constraints (Skopek et al. 2011:182), these sites are also in the business of *matchmaking*, and the extent to which sites (more or less forcefully) “recommend” potential matches is the extent to which individual preferences are attenuated.

In the case of the site Anderson et al. study, it appears that such influence is substantial. Worse, the precise *way* the site interferes with user behavior is directly derivative of users’ expressed preferences—yet a central goal of their article is to assess the relationship between the two. Specifically, Figure 5 documents the relationship between stated preferences (as listed on users’ profiles) and revealed preferences (as “revealed” by who views the profile of whom). And as the authors conclude on page 37, “Thus stating ‘must-have’ is associated with choosing same-race candidates at higher rates relative to those stating ‘nice-to-have,’ which is associated with choosing same-race candidates at higher rates than those stating no same-race preference.” However, the authors acknowledge that “The effect for those stating ‘must-have’ may be partly due to the mechanics of the site design,

because for those stating a must-have preference, the site automatically displayed only same-race candidates" (P. 37). So of course it is the case that "when a same-race preference is stated, it is highly informative of behavior" (P. 35)—because when a (strong enough) same-race preference is stated, same-race users are the only candidates who are displayed.

Three qualifications are in order. First, if people who express "must-have" preferences regarding race are only shown same-race candidates, why are there any interracial views among such people at all? Anderson et al. go on to clarify, then, that it is still possible for someone who expresses a "must-have" preference to view a cross-race alter—but only if that person conducts a "custom search" (P. 37). However, the authors do not tell us anything about this search function or how it works; they do not tell us whether there are any additional ways that users might "find" one another on the site; and they do not tell us how frequently users employ the constrained approach (the site's default method) versus the apparently unconstrained approach (search)—so we do not know the extent to which the revealed preferences for users who express "must-have" preferences are artificially inflated.

Second, after acknowledging that "must-have" users' revealed preferences are directly constrained by their stated preferences, the authors go on to say that "We also assessed the robustness of these results using a different sampling method that accounts for which profiles were shown in the list presented to the users and found similar results (see the appendix)" (P. 37). This statement is misleading, because it reads as if the authors have conducted a robustness check that corrects for the issue described above. However, if we consult the appendix, we see that when the authors replicate analyses using the "narrow pool" of only those candidates displayed by the site, of course "the narrow pool only allows us to estimate revealed preferences (R_{OR}) for the 'no preference' and 'nice-to-have' groups" (P. S1)—and so there is no robustness check for precisely the subgroup of concern.

Third, even though results for "must-have" users are biased to an unknown degree, the authors reassure us that "'nice-to-have' preferences have no effect on how candidates are displayed to the user" (P. 37). In other words, for users who list "nice-to-have" racial preferences and users who do not list any racial preferences, we should not be concerned (as we should be for the "must-have" users) that the relationship between stated and revealed preferences exists by design. But what other factors influence which candidates are displayed to each user? On page 34, the authors state that "users were only presented with profiles of users who satisfied their age, sex, and geography constraints as well as their must-have preferences." To give an example from my own data, an OkCupid user who lived in New York City in the fall of 2010 and searched only for 30- to 35-year-old women who also lived in New York City would be met with 6,835 matches. Are we to believe that the authors' dating site—assuming it is as large as OkCupid—would present all 6,835 of these people in no particular order? A central concern of most dating sites is identifying "matches"—people who are "presented to the user not as a random selection of potential partners in the local area but rather as potential partners with whom the user will be especially likely to experience positive romantic outcomes" (Finkel et al. 2012:6). Given the vast literature on racial homogamy, it would not be surprising

if the authors' site considered racial similarity to some degree in determining which potential matches to display—even for users who do not explicitly indicate such a preference. However, because most dating sites do not publicize their matchmaking algorithm, this is a possibility even the researchers themselves might not be aware of.

In sum, a central contribution of the Anderson et al. article is that it documents the differences between stated and revealed preferences. This contribution is weakened insofar as the latter are directly constrained by the former. There is also reason to suspect that the site interferes with user behavior in other ways than the authors acknowledge; perhaps this would help explain some of the unusual findings that are otherwise difficult to interpret.⁵ A great deal of the enthusiasm surrounding new digital data stems from their ability to capture actual human behaviors (as opposed to respondents' descriptions of these behaviors) in real time. However, computer-mediated interaction is not merely observable interaction—it is interaction where the computer actively structures *what* options are available and *how* these options are presented to the user. In the large body of literature on digital behavior, this fact is acknowledged remarkably infrequently, and important details about how a website works are commonly omitted.⁶ Particularly when the identity of a given site is masked, we are forced to trust that researchers' descriptions of this site are accurate and complete; the possibility that results were spuriously generated by the site's inner workings is impossible for readers to evaluate.

What does “preference” mean?

The third limitation of the Anderson et al. article is that the authors provide inadequate theoretical justification for and interpretation of the central dependent variables of their analysis. In other words, they are not just studying “stated preferences” and “revealed preferences,” but a certain *type* of stated and revealed preferences that has implications for the meaning of their results. They also do not adequately defend their particular operationalizations of *same-race* preferences — operationalizations that are internally inconsistent and potentially controversial.

Stated preferences

Describing their data on stated preferences, Anderson et al. note that “in contrast to traditional surveys, the data were collected in a natural setting where individuals were less susceptible to social pressures to appeal to an interviewer; hence stated preferences are more likely to reflect actual attitudes” (P. 29). While it is true that the data were collected in a natural (albeit digital) setting, it is unclear why these data should be preferable to traditional survey data. The reason for this is that even though an interviewer is obviously not present online, there are alternative sources of “social pressure” that may be equally strong or stronger—namely, the desire to appear attractive to other members of the dating site. Clearly, from a certain point of view, it is advantageous for site users to be honest about their preferences so that they will not be contacted by potential partners in whom they are not interested. But from another point of view, those potential partners in whom site users *are*

interested may be dissuaded from contacting someone who is so openly racially closed. This problem is exacerbated by the fact that Anderson et al. do not provide data on how many site users provide racial preferences in the first place. Just because someone is unwilling to express racial preferences that will be visible to the very people she is trying to attract does not necessarily mean she would be unwilling to express such preferences in other contexts (e.g., confidentially to a researcher); and it certainly does not require that she is in fact racially unbiased.⁷

Revealed preferences

Just as stated preferences are always expressed in a particular context, so too is it important to keep this context in mind when interpreting results on revealed preferences. Exactly what type of preferences are profile views on an online dating site “revealing,” and why are profile views the relevant marker against which to compare stated preferences in the first place? Simply viewing someone’s profile requires no actual contact or commitment and could reflect a wide variety of possible motivations—from romantic interest to mere curiosity. Profile views are also the stage of user interaction that is most susceptible to site interference; it is perhaps for this reason that Hitsch et al. (2010) focus on first contacts *conditional* on views, while other researchers (e.g., Lewis 2013; Lin and Lundquist 2013; Skopek et al. 2011) compare initial messages and responses.⁸ As network analysts have noted (e.g., Garton, Haythornthwaite, and Wellman 1997; Lewis et al. 2008), one problem with studying online behaviors and relationships is that we often have no idea what they *mean*. This does not mean that the limitations of such data necessarily outweigh the strengths, but it does mean that these concerns deserve our serious consideration—particularly when the focal behavior (profile views) is so equivocal.

Same-race preferences

Finally, independent of concerns regarding the meaning of stated preferences and profile views is the question of how to operationalize “same-race preference.” Anderson et al.’s choice of measure in both cases warrants discussion. First, in the case of stated preferences, Anderson et al. explain that “we classify a user as expressing [a same-race preference] only if the user’s declared partner race set matches the user’s own self-declared ethnicity (i.e., the only race the user prefers is the user’s own)” (P. 30). In other words, “same-race preference” is equivalent to “other-race exclusion.” Methodologically, there are a number of alternative approaches the authors could have employed (such as modeling each category of preference separately; cf. Feliciano et al. 2009; Robnett and Feliciano 2011); and theoretically, the decision to equate homophily (preference for similarity) with heterophobia (aversion to difference) is itself problematic (see Skvoretz 2013). Just because a user *also* prefers partners from background X does not mean that her preference for partners from her own background is necessarily weaker; likewise, expressing a preference for just one racial background that is not one’s own is not the same as being open to anyone. Anderson et al. elide such important distinctions, and in the process discard a great deal of nuance in their data.⁹

Second, in the case of profile views, there are additional decisions that Anderson et al. do not justify. Their use of the *relative risk* R_{RR} —“the relative likelihood that a user views a candidate’s profile given that the candidate is the same race versus a different race as the querier himself or herself” (P. 34)—seems straightforward, although there are a number of alternative measures the authors might have considered (for a recent review, see Bojanowski and Corten 2014). In particular, it is unclear whether one should control for all “main effects” in the estimation of racial homophily (for instance, if black women prefer to message black men, but *all* women prefer to message black men, should this be considered in our estimation of same-race preference?); and, as noted above, controlling for the tendency to reciprocate views might have resulted in substantially weaker estimates of same-race preferences (Wimmer and Lewis 2010). It is also unclear why the authors only consider pairs of users who live within 25 miles of one another. Apparently, in addition to racial preferences, users of this dating site can also list “age, sex, and geography constraints” (P. 34) on their profiles. Where does 25 miles fall on the distribution of listed constraints? What proportion of all profile views was excluded because of this restriction, and how would results vary under different criteria?

Third, Anderson et al. should explain why their analyses of stated and revealed preferences are commensurable in the first place. In short, same-race stated preferences are measured by whether or not a user lists a preference for partners from her own background and no other. Meanwhile, same-race revealed preferences are measured (essentially—page 35) by the odds ratio of viewing a same-race profile compared to viewing a different-race profile. Not only is “same-race preference” operationalized slightly differently in each case (the implications of which are not explored), but in the first case the unit of analysis is the individual while in the second case it is the dyad—such that the preferences of individuals who are unusually active will be disproportionately weighted (see above on “spammer users”).

Much of the contemporary literature on mate choice—using both electronic and “traditional” data sources—is built on the foundation of scholarship created by decades of research on marriage patterns. For all of the limitations of this work, it has the advantage of dealing with relatively standardized measures (see Kalmijn 1998) and a relatively institutionalized sense of the meaning and importance of the focal relationship (marriage). As sociologists studying online dating, we cannot continue framing our work in the marriage literature while pretending we are studying the same thing. The behaviors we observe are enacted in an entirely different context with entirely different norms at an entirely different point in the mate selection process—all of which have implications for what the “preferences” revealed through these behaviors actually mean and why (if at all) they are important to study.

Conclusion

Online dating is an important empirical phenomenon that has skyrocketed in popularity in recent years (Rosenfeld and Thomas 2012; Smith and Duggan 2013). It also potentially represents the closest real-world approximation of the “market” metaphor that has so long dominated thinking about mate choice (Heino, Elli-

son, and Gibbs 2010). As such, data from these sites constitute an attractive and potentially powerful tool for better understanding the social world.

The rise of research on online dating also occurs amidst broader, growing enthusiasm about the potential of electronic data to address questions of longstanding interest to social scientists. As more and more researchers—both inside and outside of academia—begin working on this frontier, however, it is important that we give equal attention to the pitfalls as well as the great promise of this approach (Golder and Macy 2014). The article published by Anderson and colleagues embodies both—but it is not the first to do so. Rather, for all of the research advantages of digital data, scholars using these data too often: 1) provide only limited information about their sample and the type of site they are studying; 2) provide limited (if any) information about how the website itself interferes with user behavior; and 3) provide surprisingly little theoretical motivation for their measures. These limitations collectively create a situation where it is impossible to evaluate the authors' findings or meaningfully replicate their results.

It is challenging enough to build a bridge from previous research using offline measures to new research on computer-mediated friendships, digital flirtations, and online behavior more generally. But by applying underdeveloped measures to frameless samples drawn from anonymous sites with unknown features, we are not only distancing ourselves from past work—we are at best speaking past one another and at worst producing scholarship with unknown and unknowable scientific value. Internet data—in all its many forms—will surely play an important role in our discipline's future. As this subfield continues to mature, it is therefore especially important that we hold it to our highest standards.

Notes

- 1 In my own dataset, for instance, the “reports income” criterion alone would reduce the eligible study population by 78.11 percent. (Specifically, of the 1,804,993 users who were active on OkCupid between October 1 and December 17, 2010, 60.58 percent did not provide information about their income, and 17.53 percent explicitly indicated that this information was “private.”)
- 2 For instance, I am still puzzled by the finding (P. 38, Figure 5) that white women are the *only* group for whom users who state “must-have” racial preferences are not actually more racially selective than users who state “nice-to-have” racial preferences. Comparisons with women from other backgrounds would have illustrated whether white women are truly unique in this respect—the authors provide no interpretation.
- 3 Exceptions to this trend are articles by Yancey (2007, 2009) and Feliciano and Robnett (e.g., Feliciano et al. 2011; Feliciano, Robnett, and Komaie 2009; Robnett and Feliciano 2011) in which the authors consistently identify the source of their data (Yahoo! Personals). However, this site was also publicly accessible at the time and contained only information on stated preferences, not behaviors.
- 4 The sample description Anderson et al. provide on page 31 suggests additional reasons for concern: in particular, men are heavily overrepresented on their site (75 percent is much higher than, for example, the 57 percent described by both Hitsch et al. [2010] and Lin and Lundquist [2013]); the highest proportion of site users are from the U.S. south;

and the average age among users appears slightly older compared to the general online dating population (Sautter, Tippett, and Morgan 2010).

- 5 For instance, if stated “must-have” preferences constrain all users’ viewing behavior equally strongly, why do the revealed preferences of “must-have” users vary so strongly by race and gender—ranging from $R_{OR} = 4$ for white women to $R_{OR} = 64$ for black men (Figure 5)?
- 6 In the case of the Anderson et al. article, many online dating sites allow users to see who has visited their profile. How much racial homogeneity in viewing behavior is generated by reciprocity—the tendency to view other people who have already viewed you (cf. Wimmer and Lewis 2010)? What proportion of views are generated when person B views the profile of person A because A sent B a message? Is it possible to send a message to someone without first viewing her profile at all? These questions are additionally complicated because the site Anderson et al. studied might operate very differently today than it did in 2009 when their data were collected—such that even if the authors had identified the site, details about how it used to work would be virtually impossible to verify.
- 7 For a more nuanced discussion of the strengths and limitations of studying stated preferences on an online personals site, see Feliciano et al. (2009:43–45).
- 8 Conditioning on profile views also risks conditioning away the part of the initial screening process that *is* driven by individual choice—and therefore underestimating same-race preferences to an unknown degree.
- 9 Recall also that Anderson et al. exclude from their analysis anyone who does not explicitly indicate a preference (or lack thereof) for a potential partner’s racial background. On one hand, this implicitly assumes that such omissions are non-informative (as opposed to deliberate or strategic); the authors could easily replicate analyses on this subgroup of users to assess whether their exclusion from the sample influenced results. On the other hand, if this subgroup is sufficiently large and their preferences sufficiently heterophilous, this could undermine Anderson et al.’s central conclusion regarding the inevitability of racial homogamy (P. 39).

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. <http://dx.doi.org/10.15195/v1.a3>.
- Bojanowski, Michał, and Rense Corten. 2014. “Measuring Segregation in Social Networks.” *Social Networks* 39:14–32. <http://dx.doi.org/10.1016/j.socnet.2014.04.001>.
- Feliciano, Cynthia, Rennie Lee, and Belinda Robnett. 2011. “Racial Boundaries among Latinos: Evidence from Internet Daters’ Racial Preferences.” *Social Problems* 58:189–212. <http://dx.doi.org/10.1525/sp.2011.58.2.189>.
- Feliciano, Cynthia, Belinda Robnett, and Golnaz Komaie. 2009. “Gendered Racial Exclusion among White Internet Daters.” *Social Science Research* 38:39–54. <http://dx.doi.org/10.1016/j.ssresearch.2008.09.004>.
- Finkel, Eli J., Paul W. Eastwick, Benjamin R. Karney, Harry T. Reis, and Susan Sprecher. 2012. “Online Dating: A Critical Analysis from the Perspective of Psychological Science.” *Psychological Science in the Public Interest* 13:3–66. <http://dx.doi.org/10.1177/1529100612436522>.

- Fiore, Andrew T. 2004. "Romantic Regressions: An Analysis of Behavior in Online Dating Systems." Masters' Thesis, Program in Media Arts and Sciences, Massachusetts Institute of Technology.
- Garton, Laura, Caroline Haythornthwaite, and Barry Wellman. 1997. "Studying Online Social Networks." *Journal of Computer-Mediated Communication* 3.
- Golder, Scott A. and Michael W. Macy. 2014. "Digital Footprints: Opportunities and Challenges for Online Social Research." *Annual Review of Sociology* 40:129–152. <http://dx.doi.org/10.1146/annurev-soc-071913-043145>.
- Heino, Rebecca D., Nicole B. Ellison, and Jennifer L. Gibbs. 2010. "Relationshopping: Investigating the Market Metaphor in Online Dating." *Journal of Social and Personal Relationships* 27:427–447. <http://dx.doi.org/10.1177/0265407510361614>.
- 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. <http://dx.doi.org/10.1007/s11129-010-9088-6>.
- Kalmijn, Matthijs. 1998. "Intermarriage and Homogamy: Causes, Patterns, Trends." *Annual Review of Sociology* 24:395–421. <http://dx.doi.org/10.1146/annurev.soc.24.1.395>.
- Laumann, Edward O., Peter V. Marsden, and David Prensky. 1983. "The Boundary Specification Problem in Network Analysis." Pp. 18–34 in *Applied Network Analysis: A Methodological Introduction*, edited by R. S. Burt and M. J. Minor. Beverly Hills, CA: Sage.
- 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* 81:1304–1338. <http://dx.doi.org/10.1086/226225>.
- Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. 2009. "Computational Social Science." *Science* 323:721–723. <http://dx.doi.org/10.1126/science.1167742>.
- Lewis, Kevin. 2013. "The Limits of Racial Prejudice." *Proceedings of the National Academy of Sciences of the United States of America* 110:18814–18819. <http://dx.doi.org/10.1073/pnas.1308501110>.
- Lewis, Kevin, Jason Kaufman, Marco Gonzalez, Andreas Wimmer, and Nicholas Christakis. 2008. "Tastes, Ties, and Time: A New Social Network Dataset Using Facebook.com." *Social Networks* 30:330–342. <http://dx.doi.org/10.1016/j.socnet.2008.07.002>.
- 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. <http://dx.doi.org/10.1086/673129>.
- Robnett, Belinda and Cynthia Feliciano. 2011. "Patterns of Racial-Ethnic Exclusion by Internet Daters." *Social Forces* 89:807–828. <http://dx.doi.org/10.1093/sf/89.3.807>.
- Rosenfeld, Michael J. and Reuben J. Thomas. 2012. "Searching for a Mate: The Rise of the Internet as a Social Intermediary." *American Sociological Review* 77:523–547. <http://dx.doi.org/10.1177/0003122412448050>.
- Sautter, Jessica M., Rebecca M. Tippett, and S. Philip Morgan. 2010. "The Social Demography of Internet Dating in the United States." *Social Science Quarterly* 91:554–575. <http://dx.doi.org/10.1111/j.1540-6237.2010.00707.x>.
- Skopek, Jan, Florian Schulz, and Hans-Peter Blossfeld. 2011. "Who Contacts Whom? Educational Homophily in Online Mate Selection." *European Sociological Review* 27:180–195. <http://dx.doi.org/10.1093/esr/jcp068>.

- Skvoretz, John. 2013. "Diversity, Integration, and Social Ties: Attraction Versus Repulsion as Drivers of Intra- and Intergroup Relations." *American Journal of Sociology* 119:486–517. <http://dx.doi.org/10.1086/674050>.
- Smith, Aaron and Maeve Duggan. 2013. "Online Dating & Relationships." Pew Research Center, Washington, D.C.
- 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* 37:942–954. <http://dx.doi.org/10.1177/0146167211409947>.
- 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. <http://dx.doi.org/10.1086/653658>.
- Yancey, George. 2007. "Homogamy over the Net: Using Internet Advertisements to Discover Who Interracially Dates." *Journal of Social and Personal Relationships* 24:913–930. <http://dx.doi.org/10.1177/0265407507084190>.
- . 2009. "Crossracial Differences in the Racial Preferences of Potential Dating Partners: A Test of the Alienation of African Americans and Social Dominance Orientation." *The Sociological Quarterly* 50:121–143. <http://dx.doi.org/10.1111/j.1533-8525.2008.01135.x>.

Acknowledgements: I am grateful to Brayden King and Chana Teeger for their helpful feedback on an earlier draft of this comment and to the editors of *Sociological Science* for encouraging me to broaden its framing.

Kevin Lewis: Department of Sociology, University of California, San Diego. E-mail: lewis@ucsd.edu.