# Political Ideology and Racial Preferences in Online Dating

Ashton Anderson[1], Sharad Goel[2], Gregory Huber[3], Neil Malhotra[4], and Duncan J. Watts[2]

[1]Department of Computer Science, Stanford University
[2]Microsoft Research
[3]Department of Political Science, Yale University
[4]Graduate School of Business, Stanford University

**Abstract**

What explains the relative persistence of same-race romantic relationships? One possible explanation is structural—this phenomenon could reflect the fact that social interactions are already stratified along racial lines—while another attributes these patterns to individual-level preferences. We present novel evidence from an online dating community involving more than 250,000 people in the United States about the frequency with which individuals both express a preference for same-race romantic partners and act to choose same-race partners. Prior work suggests that political ideology is an important correlate of conservative attitudes about race in the United States, and we find that conservatives, including both men and women and Blacks and Whites, are much more likely than liberals to state a preference for same-race partners. Further, conservatives are not simply more selective in general; they are specifically selective with regard to race. Do these stated preferences predict real behaviors? In general, we find that stated preferences are a strong predictor of a behavioral preference for same-race partners, and that this pattern persists across ideological groups. At the same time, both men and women of all political persuasions act as if they prefer same-race relationships even when they claim not to. As a result, the gap between conservatives and liberals in revealed same-race preferences, while still substantial, is not as pronounced as their stated attitudes would suggest. We conclude by discussing some implications of our findings for the broader issues of racial homogamy and segregation.

# Introduction

Although interracial marriages have been steadily increasing over time (Fu and Heaton 2008), racial homogamy—the disproportionate prevalence of same-race romantic partners (Fu and Heaton 2008; Schoen and Wooldredge 1989; Blackwell and Lichter 2004)—is a persistent phenomenon. Among all newlyweds in 2008, for example, only 9% of whites and 16% of blacks married someone whose race was different than their own (Passel, Wang, and Taylor 2010). Such racial homogamy is consequential both sociologically and economically. To the extent that information, resources, and opportunities are structured by one's social network (Coleman 1988; Portes 1998), the homogeneity of marital and family ties is likely to affect both individual-level outcomes, such as educational achievement, occupation, and income (Campbell, Marsden, and Hurlbert 1986; Grodsky and Pager 2001), as well as collective phenomena, such as racial inequality, segregation, and polarization (Baldassarri and Bearman 2007).

Population-level statistics indicate the extent of racial homogamy in society. They do not, however, reveal its underlying causes. In particular, there are at least two possible—and qualitatively different—contributing factors. First, relationship partners may be selected from a pool of racially similar candidates because of the preexisting homogeneity of an individual's social environment (Feld 1981)—including their educational institution, profession, and friends. Second, individuals may simply prefer same-race relationships for reasons as diverse as religious beliefs, social or cultural expectations, a sense of shared identity, or race-related physical attributes. Although these two mechanisms, one structural and the other preference-based, are theoretically distinct, differentiating between them empirically can be problematic. As has been previously pointed out (McPherson, Smith-Lovin, and Cook 2001), cross-sectional network data are equally consistent with either mechanism; and although recent work utilizing longitudinal network data has found that observed homophily on both race (Wimmer and Lewis 2010) and non-racial attributes (Kossinets and Watts

2009) is likely due to a combination of structural and psychological forces, these studies were not designed to measure individual preferences directly. When used to elicit attitudes about race, moreover, traditional survey tools are thought to be susceptible to social desirability bias (Krosnick 1999; Crowne and Marlowe 1998); that is, respondents seeking to not appear racist to interviewers and researchers may not be honest about their racial preferences and attitudes. Estimates of racial preferences may accordingly be biased downwards. A second bias, potentially compounding the first, is that individuals often have inaccurate beliefs about their own preferences (Gilbert 2006; Bernard et al. 1984; Nisbett and Wilson 1977). Thus even survey tools that are designed to correct for social desirability bias may underestimate preferences for same-race partners for the simple reason that respondents believe themselves to be more race-blind than they actually are.

We take a novel approach to measuring same-race preferences for romantic relationships, leveraging a unique dataset compiled from an online dating website. Although limited in some respects, online data are increasingly being used to shed light on social scientific questions in general (Lazer et al. 2009), and offer several advantages for addressing this topic in particular. First, our dataset is considerably larger and more diverse than previous, related studies, comprising over 250,000 individuals of widely varying demographic and socio-economic status, from hundreds of U.S. cities and all regions of the country. Second, in contrast to traditional surveys, the data were collected in a natural setting where individuals are less susceptible to social pressures to appeal to an interviewer; hence stated preferences are more likely to reflect actual attitudes. Third, we can account for the entire pool of available online romantic partners in a geographic area, and thereby control for the possibility that homogamy arises due to differences in available dating pools. Fourth, because individuals on the site provide a substantial amount of information about themselves, we can investigate how same-race preferences vary with other factors such as income and education and thereby account for many possible confounding variables.

Finally, because we observe which other personal profiles individuals select to view, we can augment stated attitudes with a behavioral measure of same-race preference, thus allowing us to mitigate biases in self-reported preferences. Importantly, our data allow us to assess these preferences at one of the earliest stages of selection: when a user decides whether to view a candidate's full profile after seeing his or her photo and brief biographical information. We can therefore understand how race affects initial screening decisions in the dating environment, the point at which individuals rule out many potential dating partners from further consideration. Prior work, by contrast, has focused on later-stage selection effects—examining who individuals choose to contact from among those whose full profiles they view—and therefore potentially misses the effect of race and other factors during the initial winnowing of the dating pool. Hitsch, Hortacsu, and Ariely (2010), for example, find that at this later stage, men's observed behavior is in line with their stated preferences, in sharp contrast to our own finding that even those who do not state a racial preference display a strong tendency to prefer same race candidates early in the selection process.

We focus our attention on three particular demographic attributes: sex, race, and political ideology. Given that the outcome variable of interest is a preference for same-race romantic partners of the opposite sex, our focus on sex and race is self-explanatory. Our focus on political ideology, meanwhile, is motivated by a significant body of research that shows political conservatism is correlated with a host of attitudes that may reflect low desire to form personal relationships with people of different races: explicitly stated traditional and symbolic racism, implicit prejudice, affect, and xenophobia (Sidanius, Pratto, and Bobo 1996; Federico and Sidanius 2002; Feldman and Huddy 2005; Nail, Harton, and Decker 2003; Whitley Jr. 1999). There has, however, been relatively little work that directly assesses how preferences for same-race relationships vary by political orientation and whether those differences in expressed preferences predict real behavior.

# Data

Our data were assembled from user activity logs for a popular online dating website in which users could view personal profiles and send messages to other members of the site. To protect the privacy of individuals, all data were anonymized prior to analysis. We collected a complete snapshot of activity on the site during a two-month period (October–November 2009). Member profiles consisted of a picture, a short piece of freeform text in which they could describe themselves, and answers to various multiple-choice questions about both the user's characteristics and his or her preferences for a potential partner. For example, for the question, "What is your ethnicity?", users could respond with "White," "Black," "Asian," "Hispanic," or "Other." For each such multiple-choice question, users could also indicate a subset of answers they would prefer from a potential mate, and the strength of that preference. For example, they could state that they would prefer potential partners to have answered the ethnicity question with either "White" or "Asian," and could list this as either a "nice-to-have" preference or a "must-have" preference. Users could also specify that any answer to the question is acceptable. Finally, users were free to answer as few or as many questions as they wished. Political ideology was asked on a five-point response scale: very liberal, liberal, middle-of-the-road, conservative, and very conservative. 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. We also restrict our attention to Whites and Blacks since Hispanics and Asians are sufficiently heterogeneous categories that "same-race" preference may have little meaning. Finally, we limit our sample to heterosexuals. After these restrictions, our dataset consists of 251,701 users for whom we have both profile data and a record of which profiles they chose to view in full.

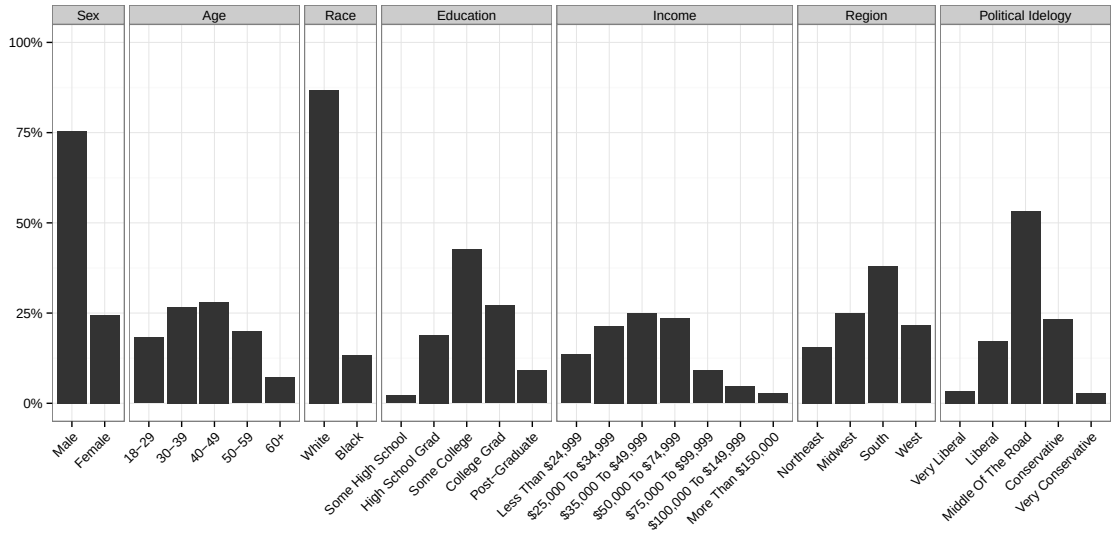


Figure 1: Demographic composition of individuals in study sample.

As shown in Figure 1, the sample of users we study comprises a diverse set of individuals in terms of age, education, income, geography, and political ideology. Although we make no claim that our sample is representative of the general U.S. dating population (which itself differs systematically from the overall U.S. population), it does exhibit significant mass over a broad range of relevant demographics including, for example, both younger (18-29) and older (60+) users, education levels ranging from "some high school" to "post graduate," annual income ranging from less than $25,000 to more than $150,000, substantial populations from all regions of the country, and a variety of political affiliations, where most users describe themselves as "middle of the road." One respect in which our sample is clearly not representative of the general dating population, however, is that males are highly overrepresented[1] (75%)—a disparity that has been noted in other, smaller samples of online dating communities from the same era (Hitsch, Hortacsu, and Ariely 2010).

[1] As we discuss later, this disparity very likely contributes to greater overall selectivity by women relative to men; however, it should not affect our other results, which control for gender.

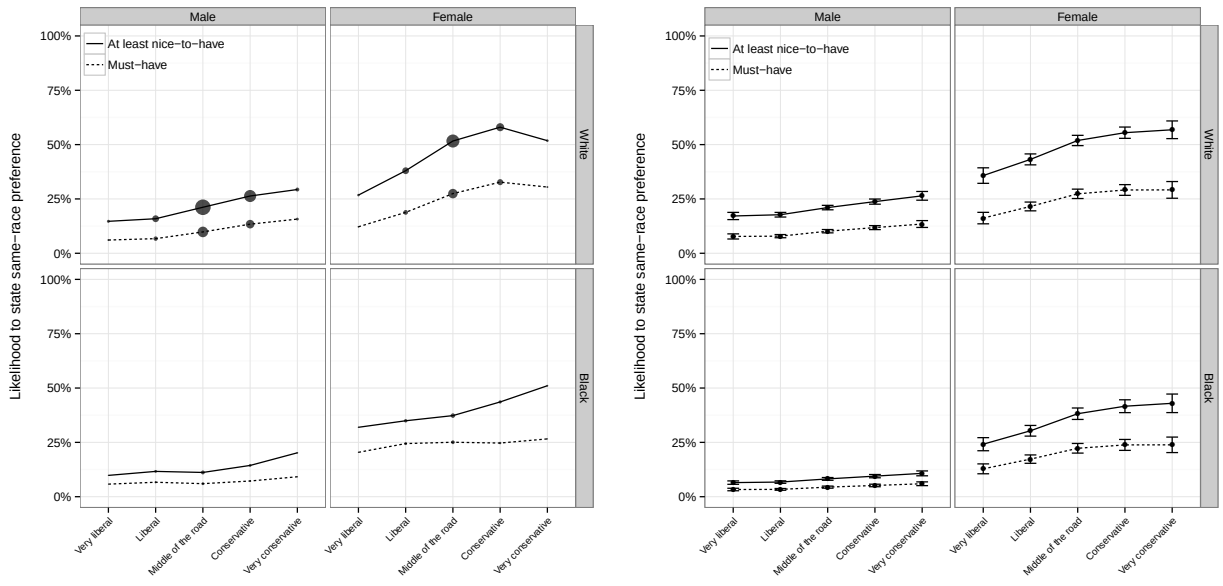


Figure 2: Estimated probability of stating a same-race preference by sex, race and political ideology. The left-hand panel shows unadjusted sample proportions, while estimates in the right-hand panel are derived from a model that controls for all other available demographic attributes. The size of the dots in the left panel corresponds to the number of individuals for each datapoint, while in the right panel the bars are 95% confidence intervals.

# Results

*Stated Preferences.* We begin by examining explicitly stated same-race preferences, where we classify a user as expressing such a preference only if their declared partner race set matches their own self-declared ethnicity (i.e., the only race they prefer is their own). For the reasons outlined in the Introduction, we are mainly interested in three key demographic attributes associated with differences in same-race preferences and behaviors: sex, race, and political ideology. Figure 2 shows the stated same-race preference distribution jointly over these attributes. Specifically, the left-hand panel shows the observed fraction of individuals of different gender and race who express at least a "nice-to-have" preference (solid lines), and separately a "must-have" preference (dotted lines).

Although these raw figures have the benefit of being easy to interpret, they are potentially confounded by other variables such as income and education that are correlated with race and

ideology. To correct for these potential confounds, we estimate the likelihood a user $q_i$ states a "nice-to-have" or "must-have" same-race preference via two separate logistic regression models. Specifically, we fit models of the form

$$\begin{aligned} \Pr\left[q_i \text{ states "nice-to-have" preference}\right] &= \text{logit}^{-1}(\boldsymbol{\beta}_{\text{nice}} \cdot \mathbf{X}_i) \\ \Pr\left[q_i \text{ states "must-have" preference}\right] &= \text{logit}^{-1}(\boldsymbol{\beta}_{\text{must}} \cdot \mathbf{X}_i) \end{aligned}$$

where $\mathbf{X}_i$ is a vector of user $q_i$'s attributes, $\boldsymbol{\beta}_{\text{nice}}$ and $\boldsymbol{\beta}_{\text{must}}$ are vectors of corresponding regression coefficients, and $\text{logit}^{-1}(x) = e^x/(1+e^x)$. These models adjust for every demographic attribute users specify: age, sex, height, ethnicity, education, income, geography, political affiliation, marital status, religion, body type, drinking habits, smoking habits, presence of children, and desire for more children.

The majority of these attributes are categorical, in which case we use indicator variables for each category to allow for the greatest amount of model flexibility. Two exceptions are age and height, which are modeled by including age and age squared and height and height squared in the attribute vector $\mathbf{X}$. Age and height are also normalized to have mean 0 and standard deviation 1. To adjust for geography, we include the population density of the user's declared zip code, an indicator variable specifying whether they live in an urban area (defined as having at least 1,000 people per square mile), a categorical variable for geographic region (Northwest, West, South, and Northeast), and the fraction of people in the user's zip code that are the same-race as the user. We also include two separate continuous variables specifying the number of (non-race) nice-to-have and must-have preferences. These latter two variables capture the user's general selectivity, aside from any race preferences. Finally, given the substantial differences in the number of men and women active on these sites and that heterosexual dating sites are two-sided markets stratified by gender, all of these attributes are interacted with sex.

In sum, the structural form of the "nice-to-have" stated preference model (omitting the individual subscript $q_i$ for clarity) is:

$$\begin{aligned}\Pr\left[q_i \text{ states "nice-to-have" race preference}\right] = \text{logit}^{-1}\Big(&\beta_{\text{race}\times\text{sex}} + \beta_{\text{political}\times\text{sex}} + \\ &\beta_{\text{education}\times\text{sex}} + \beta_{\text{age}\times\text{sex}} + \beta_{\text{age}^2\times\text{sex}} + \beta_{\text{height}\times\text{sex}} + \beta_{\text{height}^2\times\text{sex}} + \dots\Big).\end{aligned} \tag{1}$$

An analogous model is used for "must-have" preferences.

Table A1 in the Appendix lists fitted coefficient values for key variables of interest. Given the complexity of these models and the large number of interactions, the regression coefficients can be difficult to interpret on their own. We therefore use our fitted models to estimate the likelihood of stating nice-to-have and must-have preferences across various demographic groups, holding other factors constant. In particular, after constructing a "typical individual"—based on the median or modal value of the empirical distribution for each attribute—we then vary sex, race, and political affiliation, allowing us to isolate the effects of each of these factors.[2] The right-hand panel of Figure 2 shows these model-adjusted estimates. The similarity between the raw and model-adjusted estimates indicates that the patterns we observe are indeed reflective of race, gender, and political ideology, and not simply driven by the correlation between racial preferences and other demographic characteristics.

Perhaps most strikingly, Figure 2 illustrates that women are substantially more likely than men to express both weak and strong same-race preferences. Specifically, more than half (52%) of White, politically moderate women express at least a "nice-to-have" same-race preference, with 27% explicitly stating a same-race partner is a "must-have"; by comparison, 21% of White, moderate men state having a "nice-to-have" same-race preference and 10%

[2]We separately construct "typical" males and females: height, number of profile views, and number of stated preferences are set to the gender-specific medians; all other attributes are set to the median values over the entire sample.

report having a "must-have" preference.[3] Similar differences between women and men are apparent among Blacks.

Figure 2 further indicates a strong association between political ideology and stated same-race preferences. While the effect is apparent across both sexes, it is particularly salient for women: conservative White women are about 30% more likely to express a preference for same-race partners than their liberal counterparts (56% vs. 43%). Likewise, Conservative Black women are substantially more likely to state a same-race preference than liberal Black women (42% vs. 30%). The percentage of White men with a stated same-race preference is 24% and 18% for conservatives and liberals, respectively. We even find this pattern for Black men, the group with the lowest propensity to state a same-race preference: 9% of conservative Black men state a same-race preference compared to 7% of liberal Black men.[4]

Although the tendency of political conservatives to state same-race preferences at higher rates than political liberals is striking, the underlying cause remains unclear. One possibility is that conservatives are more selective in general—on a variety of traits—and that their same-race preferences are simply a manifestation of this tendency. Indeed, as we have already noted, women in our population are heavily outnumbered by men, and men also tend to be far more active in contacting or approaching women than the reverse. For both these reasons it is plausible that women, seeking to exploit their "market power" or simply to reduce their cognitive load, may elect to state more preferences, including a same-race preference. Possibly, therefore, the observed effect of political ideology can also be explained in terms of overall selectivity, not selectivity on race specifically.

[3]Due to the extremely large sample size, most differences between percentages are highly statistically significant at conventional levels. Accordingly, we focus on the substantive effects and only note when differences are not significant.

[4]Even though we observe similar effects for both Whites and Blacks, the estimated effects for Blacks are harder to generalize from our sample to the population at large because minorities primarily interested in dating within their own race group may seek out racially specific dating sites.

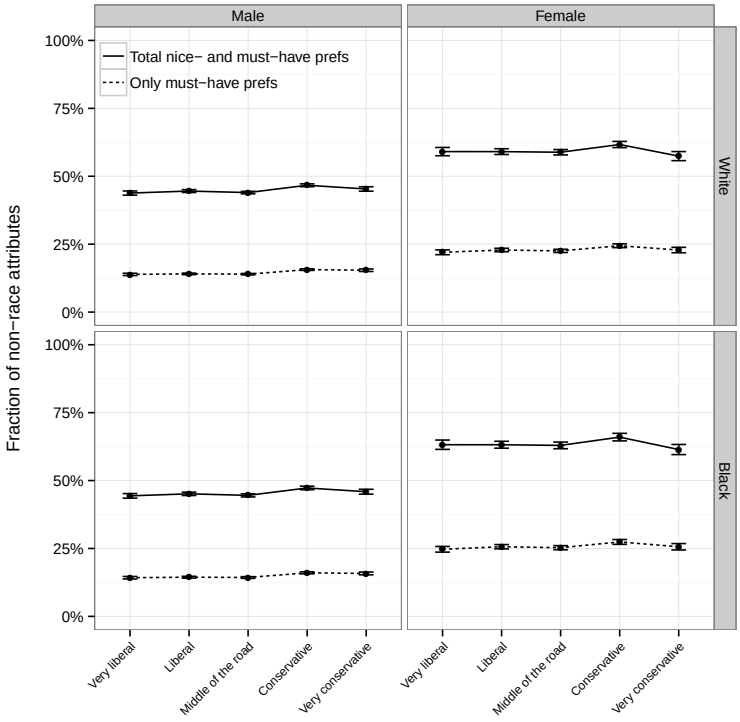


Figure 3: Non-race selectivity: Estimated fraction of non-race attributes for which user states preferences, by sex, race, and political ideology. Bars indicate 95% confidence intervals. See Table A2 for model estimates.

We note that our model includes the number of non-race preferences that each individual states, so that if conservatives were, on average, simply more likely to express any preference, these estimates account for this simple difference in selectivity. Our model also includes a measure of differences in the racial composition of the dating pool in different areas, which mitigates against the possibility that liberals are simply concentrated in areas where expressing a racial preference is less necessary because of greater racial homogeneity in the dating pool. Nevertheless, to further investigate the possibility of differences in overall choosiness by ideology and gender, we measure selectivity by examining the number of attributes other than race (e.g., height, income, education, smoking habits, body type, etc.) for which users express preferences, again broken down by sex, race, and political ideology. Analogous to our analysis framework above, we fit regression models to estimate selectivity as a function of individual attributes. Given that the outcome variable of interest (i.e., the

number of non-race stated preferences) is integer valued, we use Poisson regression. Again omitting the individual subscript $q_i$ for clarity, the form of the models is:

$$\begin{aligned}&\text{Number of non-race stated ``nice-to-have'' and ``must-have'' preferences} = \\ &\text{Poisson}\Big(\exp\big(\beta_{\text{race}\times\text{sex}} + \beta_{\text{political}\times\text{sex}} + \beta_{\text{education}\times\text{sex}} + \beta_{\text{age}\times\text{sex}} + \\ &\beta_{\text{age}^2\times\text{sex}} + \beta_{\text{height}\times\text{sex}} + \beta_{\text{height}^2\times\text{sex}} + \dots\big)\Big)\end{aligned} \tag{2}$$

We also separately fit a model to estimate only the number of "must have" preferences.

Selected model coefficients are listed in Table A2. As before, we also plot model estimates for a prototypical individual, varying sex, race, and political ideology. As expected, Figure 3 shows that women—both White and Black—state "must-have" or "nice-to-have" preferences more than men. Women state such a preference for approximately 60% of these attributes, compared to only about 45% for men. Conservatives, however, state no more preferences on average than liberals (White men: 47% vs. 45%; White women: 62% vs. 59%). The observed propensity of conservatives to state same-race preferences, therefore, is not attributable to some more general selectivity, but rather is specific to race.

*Revealed Preferences.* Our results thus far are based entirely on self-reported racial preferences. But are they accurate proxies for behavior? Prior research suggests that a potential problem with using self-reported data is that they may not reflect people's true preferences. It could be the case, for example, that liberals have exactly the same preferences as conservatives for same-race relationships, but are not inclined to state, or even acknowledge, those views (Sniderman and Carmines 1997). We address this issue by measuring users' *revealed preferences*: 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 him or herself. Specifically, for queriers $q_i$ searching for candidates $c_i$ whose profiles are available on the dating site, we estimate the *relative risk* $R_{RR}$ of selecting racially congruent profiles. Our use of relative risk, motivated by its application in epidemiology, is defined formally as:

$$R_{RR} = \frac{\Pr\left[q_i \text{ views profile of } c_i \mid q_i \text{ is the same-race as } c_i\right]}{\Pr\left[q_i \text{ views profile of } c_i \mid q_i \text{ is a different race than } c_i\right]}.$$

A relative risk $R_{RR}$ greater than 1 means that the querier is disproportionately inclined to view same-race candidates, and hence exhibits a same-race preference, whereas $R_{RR} = 1$ indicates the absence of such a preference. ($R_{RR} < 1$ would indicate a preference for partners of a different race.)

To estimate $R_{RR}$, we must address three complications. First, we need to specify exactly which querier/candidate pairs to consider. One could naively consider all possible pairs of users to be potential matches, but geographic constraints alone suggest that choice is ill-suited for our analysis. In response, we restrict the candidate set to members of the opposite sex living within 25 miles of the querier and who meet the querier's stated age requirements. These constraints—sex, age, and geography—are ostensibly the most important in initial evaluations of candidates, and all queriers are required to specify these. The goal of these constraints was to winnow the full set of dyads to a manageable level while not putting unneeded restrictions on observing potential matches. Accordingly, we did not restrict dyads based on, for example, education, although we do control for such other variables in the analyses. We note, however, that on the actual site, users were only presented with profiles of users who satisfied their age, sex and geography constraints as well as their must-have preferences. Therefore, the "broad pool" of candidates we consider includes some individuals who are, by default, not presented to the querier by the website, although users could still find these candidates by using the site's search functionality. To verify that our results are not being driven by the site's design, we repeat our analysis for a "narrow pool" of candidates that also meets a querier's must-have preferences. As shown in the Appendix, these two pools of candidates lead to similar results, and so we focus on the broad pool for our primary analysis.

A second issue is that even with appropriately defined candidate sets, the number of querier/candidate pairs is large (in the tens of millions). We therefore employ a case control

design (King and Zeng 2001), first constructing a much smaller subset of instances in which a user viewed the profile of another member, and then augmenting this set with randomly selected instances in which a querier did not elect to view a candidate's profile. The size of this latter component is chosen to be approximately three times as large as the former. This selection procedure clearly curbs our ability to estimate either the numerator or denominator in $R_{RR}$. However, as has been observed in the statistics and epidemiology literature, the *odds ratio* can still be estimated from such a sample:

$$R_{OR} = \frac{p_s/(1-p_s)}{p_d/(1-p_d)} \tag{3}$$

where

$$\begin{aligned} p_s &= \Pr\left[q_i \text{ views profile of } c_i \mid q_i \text{ is the same race as } c_i\right] \\ p_d &= \Pr\left[q_i \text{ views profile of } c_i \mid q_i \text{ is a different race than } c_i\right]. \end{aligned}$$

Moreover, given the large number of profiles, $p_s$ and $p_d$ are both relatively small, and so

$$R_{OR} = R_{RR}\left(\frac{1-p_d}{1-p_s}\right) \approx R_{RR}.$$

That is, the odds ratio $R_{OR}$—which we can efficiently estimate—approximates $R_{RR}$.

Finally, as in our analysis of stated preferences, we would like to control for potential confounding variables and isolate the effects of certain key factors of interest, namely sex, race, and political affiliation. To do so, observe that the log odds ratio can be written as:

$$\begin{aligned} &\log(R_{OR}) = \\ &\quad \text{logit}\left(\Pr\left[q_i \text{ views profile of } c_i \mid q_i \text{ is the same-race as } c_i\right]\right) \\ &- \text{logit}\left(\Pr\left[q_i \text{ views profile of } c_i \mid q_i \text{ is a different race than } c_i\right]\right) \end{aligned} \tag{4}$$

where $\text{logit}(x) = \log(x/(1-x))$. We thus estimate $\log(R_{OR})$ by first fitting a logistic regression model that predicts whether any given querier $q_i$ views the profile of a candidate $c_i$, and then examining the difference between model estimates when the candidate $c_i$ is assumed to be the same-race versus a different race than the querier $q_i$, holding all other traits constant. By then varying the demographic attributes of $q_i$ and $c_i$, this approach allows us to investigate how revealed preferences change across subpopulations.

This logistic regression model includes separate terms for the demographic attributes of both the querier and the candidate, as well as joint querier-candidate features (that is, interaction terms between the querier's features and those of the candidate). In particular, analogous to the stated preferences model, for both querier and candidate we include: age, height, education, income, religion, body type, employment status, drinking habits, smoking habits, existence of children, desire to have more children, marital status, population density in zip code, fraction of population in their zip code who are of the same ethnicity, whether or not their zip code is classified as urban, and the number of non-race nice-to-have and must-have preferences, all of which are interacted with the sex of the querier. The joint querier-candidate attributes indicate whether the users have the same political affiliation, level of education, marital status, smoking habits, drinking habits, religion, income, body type, employment status, existence of children, and desire to have more children; we also include the (continuous) distance between the querier and the candidate. Finally, our models include three additional interaction terms: we interact the querier's sex, race and political affiliation with a variable indicating whether the querier-candidate pair is of the same race.

Figure 4 plots model estimates for $R_{OR}$ as a function of political ideology, broken down by gender and race.[5] Consistent with our findings regarding stated preferences, the lines in Figure 4 slope upward for all gender-race groups, indicating that more conservative in-

[5]Since $R_{OR}$ is the difference of model estimates (as described in Eq. 4), it depends only on the coefficients for the three characteristics (race, gender, and ideology) that are interacted with the variable indicating whether the querier is of the same race as the candidate.

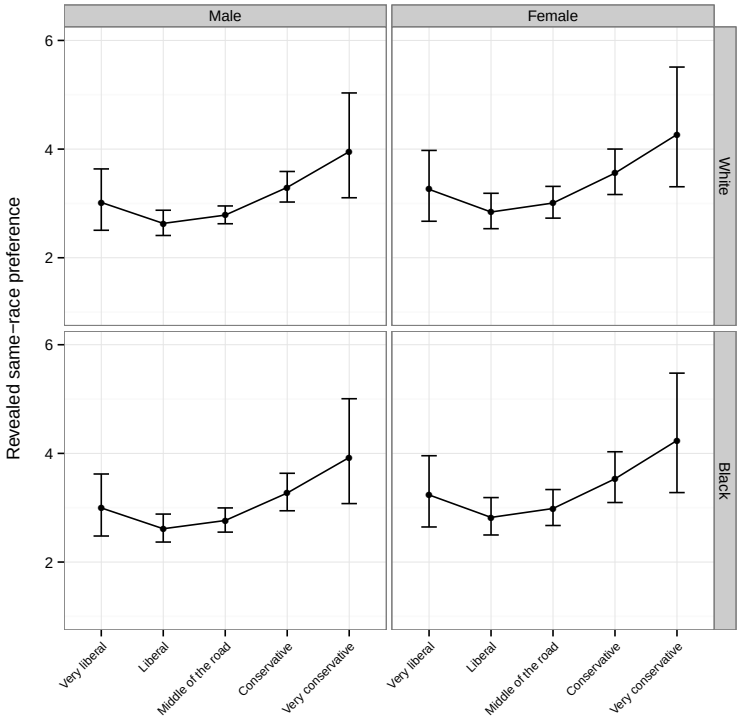


Figure 4: Estimated revealed preferences for same-race partners. Bars indicate 95% confidence intervals.

dividuals exhibit a greater behavioral tendency to select same-race partners. On average, for example, conservative White men have an estimated odds ratio of 3.3 compared to 2.6 for liberal White men. Likewise, conservative White women have an odds ratio of 3.6 compared to 2.9 for liberal White women. Similar patterns exist for all other gender and race groupings. Interestingly, however, our earlier finding that women are more likely than men to state a same-race preference is not present in the behavioral data; both Black and White men are just as likely to reveal a same-race preference as their female counterparts.

Figure 5 casts additional light on these contrasting findings—that the effect of political ideology is consistent across stated and revealed preferences but that the effect of gender is starkly different—by further breaking down revealed preferences by the stated preference of each gender-race-ideology category.[6] (The vertical axis of Figure 5 is expanded substantially

[6]To generate these estimates, we modify the revealed preferences model to now include three 3-way interaction terms: we interact the querier's sex, race and political affiliation with both the querier's stated

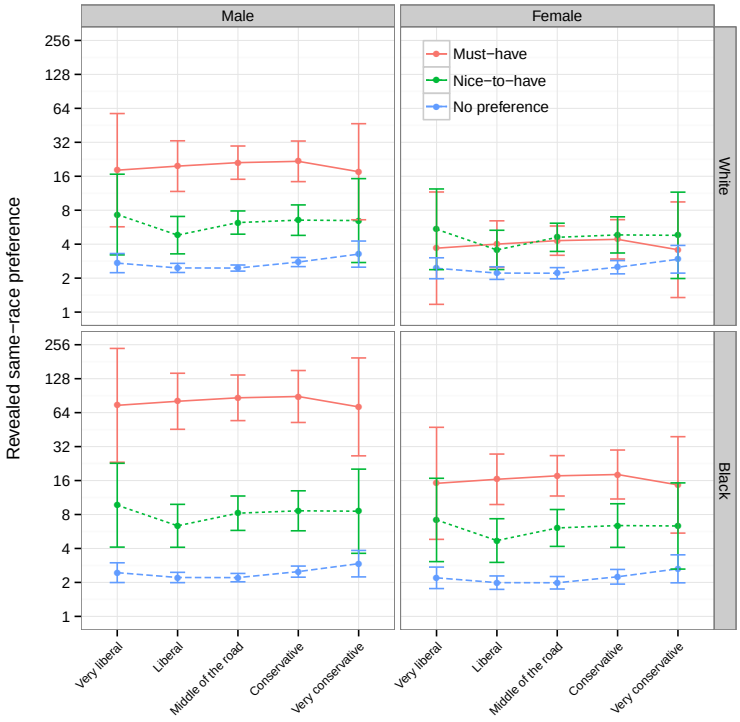


Figure 5: Estimated revealed preferences for same-race partners by stated same-race preference. Bars indicate 95% confidence intervals.

relative to Figure 4; it now ranges up to 256 rather than just 6, and is scaled exponentially.) We note three main results. First, we find that individuals of both genders and races who explicitly state not having a same-race preference do in fact exhibit a substantial tendency to favor same-race partners. In quantitative terms $R_{OR}$ for these "no preference" individuals (dashed line) ranges between between 2 and 3 across demographic groups, meaning that even after controlling for a host of other factors they are two to three times more likely to select a candidate of the same race (all of these differences are statistically distinguishable from 1).

Second, we find that when a same-race preference is stated, it is highly informative of behavior, particularly for men. In Figure 5, that is, the solid line is the estimated behaviorally revealed preference for individuals stating a "must-have" same-race preference, while the dotted line is the same quantity for those expressing "nice-to-have" preferences. Apart

race preference (i.e., no preference, nice-to-have or must-have) and a variable indicating whether the querier-candidate pair is of the same race. Fitted values for these interaction terms are listed in Table A3.

from White women, for whom these two lines cross and point estimates are statistically indistinguishable from one another, the solid line is consistently above the dotted line, which is consistently above the dashed line. 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. In terms of magnitude, for White men, those who say "must-have" are about 20 times more likely to select same-race candidates than different-race candidates, and those saying "nice-to-have" are about 6 times more likely to select same-race candidates. Both effects are much larger than the estimates for the "no preference" category. Similar patterns are apparent for both Black men and Black women. For White women, the effect of "must-have" and "nice-to-have" are indistinguishable, although for most ideological categories each is distinguishable from the odds ratio among those stating no same-race preference.[7]

Finally, Figure 5 provides no evidence of differences across ideological groups in the meaning of distinct statements of racial preference. If liberals were in fact more racially discriminating for a given level of stated racial preference, we would expect the lines for different stated preferences to slope downward. Liberals who declared a "nice-to-have" same-race preference, for example, would show stronger patterns of same-race behaviors than conservatives who had expressed that preference. In fact, across ideological groups, the lines are largely flat for all four race-gender groups. Thus, our findings indicate that liberals and conservatives—unlike men and women—are not using these terms in different ways, which in turn suggests there is little ideological effect on a willingness to express same-race preferences relative to acting on them.

[7]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 unless the user conducted a custom search. However, the effect for nice-to-have preferences is not due to preferential ranking, because "nice-to-have" preferences have no effect on how candidates are displayed to the user. 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).

# Discussion

Returning to our initial motivation, our results suggest that individual preferences are an important explanation for the relative dearth of same-race relationships. In particular, we find not only that a large proportion of our population states a same-race preference and acts on it, but that even individuals who state that they do not have a preference act as if they do. To the extent that we see a discrepancy between stated and revealed preferences, this difference may be due to unconscious bias of which the respondent is unaware. Alternatively, it could be that these individuals have some acknowledged level of same-race preference, but believe it is weaker than "nice-to-have," and so continue to state "no preference."

We also find that although women are substantially more likely than men to state a same-race preference, for any given stated preference level, men display a stronger propensity to act in a same-race preferential manner and that these competing effects largely cancel out; consequently, overall revealed same-race preferences for men and women are very similar. There also appears to be little difference between Blacks and Whites in these patterns. For political ideology, meanwhile, we see a different pattern: conservatives are more likely than liberals to state a same-race preference, but for any given level of stated preferences both conservatives and liberals show a similar propensity to act; thus both revealed and stated preferences for same-race partners increase with political conservatism.

We close by noting that the patterns we have observed have implications for any policy interventions designed to influence homogamy levels. In particular, our findings imply that merely altering the structural environment—say by creating more opportunities for individuals of different races to interact—would not necessarily ameliorate persistent patterns of racial homogamy in romantic relationships. Especially among political conservatives, homogamy appears to derive in part from same-race preferences, where these preferences are shared broadly across racial and gender groups. Finally, a general behavioral preference for racial homogamy is evident across all groups, even among those that state a lack of such

preference, implying that survey data underestimate the proportion of the population that will choose same-race romantic partners. Although our study is silent on the malleability of these preferences, or how they might change with more interracial contact, our results nonetheless imply that opportunity alone will not eliminate the preponderance of same-race romantic relationships.

# References


Baldassarri, Delia, and Peter Bearman. 2007. "Dynamics of Political Polarization." *American Sociological Review* 72 (5): 784–811.

Bernard, H. Russell, Peter Killworth, David Kronenfeld, and Lee Sailer. 1984. "The Problem of Informant Accuracy: The Validity of Retrospective Data." *Annual Review of Anthropology* 13: 495–517.

Blackwell, Debra L., and Daniel T. Lichter. 2004. "Homogamy among Dating, Cohabiting, and Married Couples." *The Sociological Quarterly* 45 (4): 719–737.

Campbell, Karen E., Peter V. Marsden, and Jeanne S. Hurlbert. 1986. "Social Resources and Socioeconomic Status." *Social Networks* 8 (1): 97–117.

Coleman, James S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 94: 95–120.

Crowne, Doulgas P., and David Marlowe. 1998. *The Approval Motive: Studies in Evaluative Dependence*. New York: Wiley.

Federico, Christopher M., and Jim Sidanius. 2002. "Sophistication and the Antecedents of Whites' Racial Policy Attitudes: Racism, Ideology, and Affirmative Action in America." *Public Opinion Quarterly* 66 (2): 145–176.

Feld, Scott L. 1981. "The Focused Organization of Social Ties." *American Journal of Sociology* 86 (5): 1015–1035.

Feldman, Stanley, and Leonie Huddy. 2005. "Racial Resentment and White Opposition to Race-Conscious Programs: Principles or Prejudice?" *American Journal of Political Science* 49 (1): 168–183.

Fu, Xuanning, and Tim B. Heaton. 2008. "Racial and Educational Homogamy: 1980 to 2000." *Sociological Perspectives* 51 (4): 735–758.

Gilbert, Daniel. 2006. *Stumbling on Happiness*. New York: Vintage.

Grodsky, Eric, and Devah Pager. 2001. "The Structure of Disadvantage: Individual and Occupational Determinants of the Black-White Wage Gap." *American Sociological Review* 66 (4): 542–567.

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

King, Gary, and Langche Zeng. 2001. "Logistic Regression in Rare Events Data." *Political Analysis* 9 (2): 137–163.

Kossinets, Gueorgi, and Duncan J. Watts. 2009. "Origins of Homophily in an Evolving Social Network." *American Journal of Sociology* 115 (2): 405–450.

Krosnick, Jon A. 1999. "Survey Research." *Annual Review of Psychology* 50: 537–67.

Lazer, David, Alex Sandy Pentland, Lada Adamic, Sinan Aral, Albert Laszlo Barabasi, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann et al. 2009. "Life in the Network: The Coming Age of Computational Social Science." *Science* 323 (5915): 721.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27: 415–444.

Nail, Paul R., Helen C. Harton, and Brian P. Decker. 2003. "Political Orientation and Modern Versus Aversive Racism: Tests of Dovidio and Gaertner's (1998) Integrated Model." *Journal of Personality and Social Psychology* 84 (4): 754–770.

Nisbett, Richard E., and Timothy D. Wilson. 1977. "Telling More Than We Can Know: Verbal Reports on Mental Processes." *Psychological Review* 84 (3): 231–259.

Passel, Jeffrey S., Wendy Wang, and Paul Taylor. 2010. "Marrying Out: One-in-Seven New US marriages is Interracial or Interethnic." *Pew Research Center (June 4). http://pewsocialtrends. org/assets/pdf/755-marryingout. pdf* .

Portes, Alejandro. 1998. "Social Capital: Its Origins and Applications in Modern Sociology." *Annual Review of Sociology* 24 (1): 1–24.

Schoen, Robert, and John Wooldredge. 1989. "Marriage Choices in North Carolina and Virginia, 1969-71 and 1979-81." *Journal of Marriage and the Family* 51 (2): 465–481.

Sidanius, Jim, Felicia Pratto, and Lawrence Bobo. 1996. "Racism, Conservatism, Affirmative Action, and Intellectual Sophistication: A Matter of Principled Conservatism or Group Dominance?" *Journal of Personality and Social Psychology* 70: 476–490.

Sniderman, Paul M., and Edward G. Carmines. 1997. *Reaching Beyond Race*. Cambridge, MA: Harvard University Press.

Whitley Jr., Bernard E. 1999. "Right-wing Authoritarianism, Social Dominance Orientation, and Prejudice." *Journal of Personality and Social Psychology* 77 (1): 126–134.

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 (2): 583–642.

# Appendix

Tables A1, A2, A3 below list selected model coefficients for the models discussed in the main text. Specifically, Table A1 lists coefficients for the stated preferences models, Table A2 lists coefficients for the models that estimates the number of stated preferences, and Table A3 lists coefficients for the revealed preferences model.

| | At least nice-to-have | Must-have |
|---|---|---|
| Coefficients for: | Male / Female | Male / Female |
| Very liberal | -2.03 (0.25) / -1.30 (0.12) | -3.02 (0.35) / -2.21 (0.14) |
| Liberal | -1.99 (0.25) / -0.98 (0.10) | -3.00 (0.34) / -1.86 (0.12) |
| Middle of the road | -1.78 (0.25) / -0.63 (0.10) | -2.72 (0.34) / -1.55 (0.12) |
| Conservative | -1.62 (0.25) / -0.49 (0.10) | -2.55 (0.34) / -1.46 (0.12) |
| Very conservative | -1.48 (0.25) / -0.43 (0.12) | -2.40 (0.35) / -1.46 (0.14) |
| Black | -0.54 (0.03) / -0.08 (0.04) | -0.53 (0.04) / -0.05 (0.04) |

Table A1: Coefficients for stated preference models. In each cell, the first coefficient and standard error (in parentheses) are for males, and the second coefficient and standard error (in parentheses) are for females.

| | At least nice-to-have | Must-have |
|---|---|---|
| Coefficients for: | Male / Female | Male / Female |
| Very liberal | 1.78 (0.04) / 1.92 (0.02) | 0.76 (0.07) / 1.17 (0.03) |
| Liberal | 1.80 (0.04) / 1.92 (0.02) | 0.77 (0.07) / 1.21 (0.03) |
| Middle of the road | 1.78 (0.04) / 1.92 (0.02) | 0.76 (0.07) / 1.20 (0.03) |
| Conservative | 1.84 (0.04) / 1.97 (0.02) | 0.88 (0.07) / 1.28 (0.03) |
| Very conservative | 1.81 (0.04) / 1.89 (0.02) | 0.76 (0.07) / 1.21 (0.03) |
| Black | 0.01 (0.00) / 0.06 (0.01) | 0.02 (0.01) / 0.09 (0.01) |

Table A2: Coefficients for number of non-race attributes for which a user expresses a preference, both for at least nice-to-have and must-have preferences. In each cell, the first coefficient and standard error (in parentheses) are for males, and the second coefficient and standard error (in parentheses) are for females.

| | Same-race preference | | |
|---|---|---|---|
| | No preference | Nice-to-have | Must-have |
| Coefficients for: | Different Race / Same Race | Different Race / Same Race | Different Race / Same Race |
| Very liberal | -0.89 (0.11) / NA | -1.03 (0.39) / 0.67 (0.16) | -1.06 (0.55) / 0.23 (0.17) |
| Liberal | -0.80 (0.06) / NA | -1.19 (0.19) / 0.07 (0.08) | -1.27 (0.23) / 0.11 (0.08) |
| Middle of the road | -0.79 (0.06) / NA | -1.30 (0.14) / 0.23 (0.05) | -1.24 (0.15) / 0.22 (0.05) |
| Conservative | -0.91 (0.07) / NA | -1.38 (0.18) / 0.19 (0.07) | -1.27 (0.20) / 0.22 (0.06) |
| Very conservative | -1.08 (0.14) / NA | -1.40 (0.42) / 0.16 (0.19) | -1.09 (0.48) / 0.18 (0.16) |
| Male | -0.11 (0.06) / NA | -0.34 (0.13) / -0.04 (0.05) | -1.59 (0.17) / 0.00 (0.05) |
| Black | 0.11 (0.04) / NA | -0.17 (0.14) / 0.11 (0.09) | -0.77 (0.19) / 0.64 (0.08) |

Table A3: Main model coefficients and standard errors for our revealed preferences model. The first set of entries in each cell corresponds to coefficients and standard errors when the candidate pair are of difference races, and the second corresponds to when the candidate pair are of the same race. All coefficients are relative to White female queriers who state no preference and who match on race with the candidate (as indicated by the NAs).

A key component of our analysis involves selecting which querier/candidate pairs to consider when estimating $R_{OR}$. For the results given in the main text we constructed what we call the "broad pool," which for any given querier comprised of all members of the opposite sex living within 25 miles of the querier, and who meet the querier's stated age requirements. We noted earlier, however, since not all candidates in the broad pool were shown to queriers—namely candidates that did not satisfy a querier's must-have preferences—estimates based on the broad pool could reflect a certain self-fulfilling prophecy, in which users' stated preferences directly constrain their future actions. We thus repeated our analysis for an additional "narrow pool" of querier/candidate pairs, where for each querier we constructed a candidate set of all members who meet the requirements for the broad pool (live within 25 miles of the querier, and who meet the querier's stated age requirements) and also satisfy the querier's must-have preferences. As a consequence, the estimates of the narrow pool are purged of any selection effects arising from the site's recommendation algorithm. By construction, however, the narrow pool only allows us to estimate revealed preferences ($R_{OR}$) for the "no preference" and "nice-to-have" groups, when ideally we would like to estimate them for the "must have" group as well—it is for this reason that we display results for the broad pool in the main text. Table A4 and Figure A1 show selected coefficients and model estimates

from the revealed preferences analysis using the narrow pool. The results are qualitatively the same as the analogous results in the main text, providing reassurance that our findings are not artifacts of the site's design.

| | Same-race preference | |
|---|---|---|
| | No preference | Nice-to-have |
| Coefficients for: | Different Race / Same Race | Different Race / Same Race |
| Very liberal | -0.70 (0.12) / NA | -0.81 (0.47) / 1.01 (0.16) |
| Liberal | -0.67 (0.07) / NA | -1.32 (0.23) / 0.16 (0.08) |
| Middle of the road | -0.64 (0.07) / NA | -1.37 (0.17) / 0.30 (0.06) |
| Conservative | -0.78 (0.08) / NA | -1.59 (0.21) / 0.23 (0.07) |
| Very conservative | -0.84 (0.15) / NA | -1.72 (0.45) / 0.44 (0.19) |
| Male | -0.15 (0.07) / NA | -0.26 (0.16) / -0.10 (0.06) |
| Black | 0.15 (0.05) / NA | -0.12 (0.16) / 0.13 (0.09) |

Table A4: Main model coefficients and standard errors for our revealed preferences model based on the "narrow pool" of candidates. The first set of entries in each cell corresponds to coefficients and standard errors when the candidate pair are of difference races, and the second corresponds to when the candidate pair are of the same race. All coefficients are relative to White female queriers who state no preference and who match on race with the candidate (as indicated by the NAs).

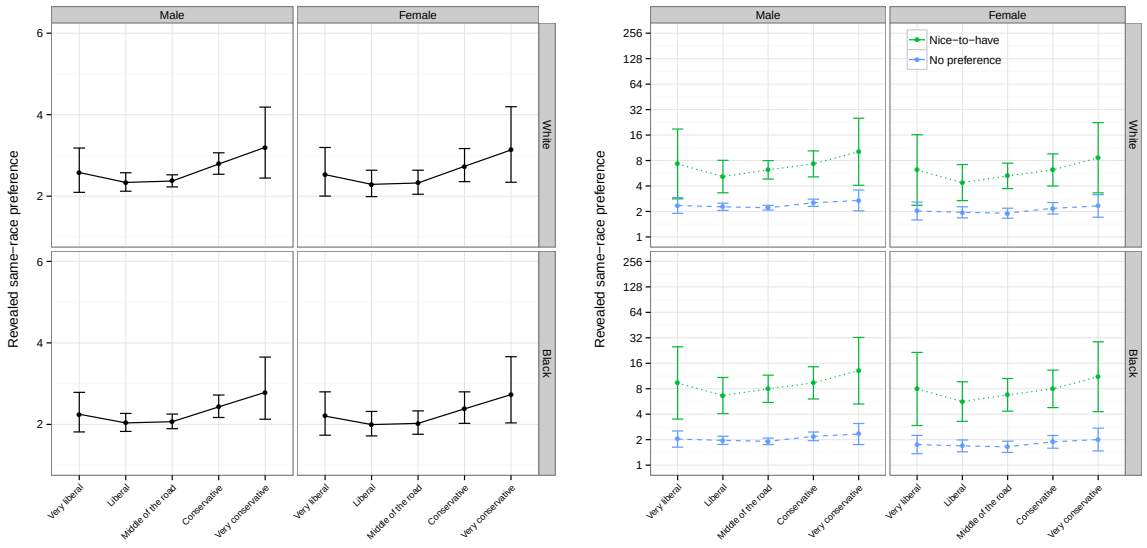


Figure A1: Estimated revealed same race preferences based on the "narrow pool" of candidates. Estimated revealed preferences for same-race partners by stated same-race preference. Bars indicates 95% confidence intervals.