Not Just a Preference: Reducing Biased Decision-making on Dating Websites

Zilin Ma
zilinma@g.harvard.edu
Intelligent Interactive Systems Group
Harvard School of Engineering and Applied Sciences
Allston, MA, USA

Krzysztof Z. Gajos
kgajos@eecs.harvard.edu
Intelligent Interactive Systems Group
Harvard School of Engineering and Applied Sciences
Allston, MA, USA

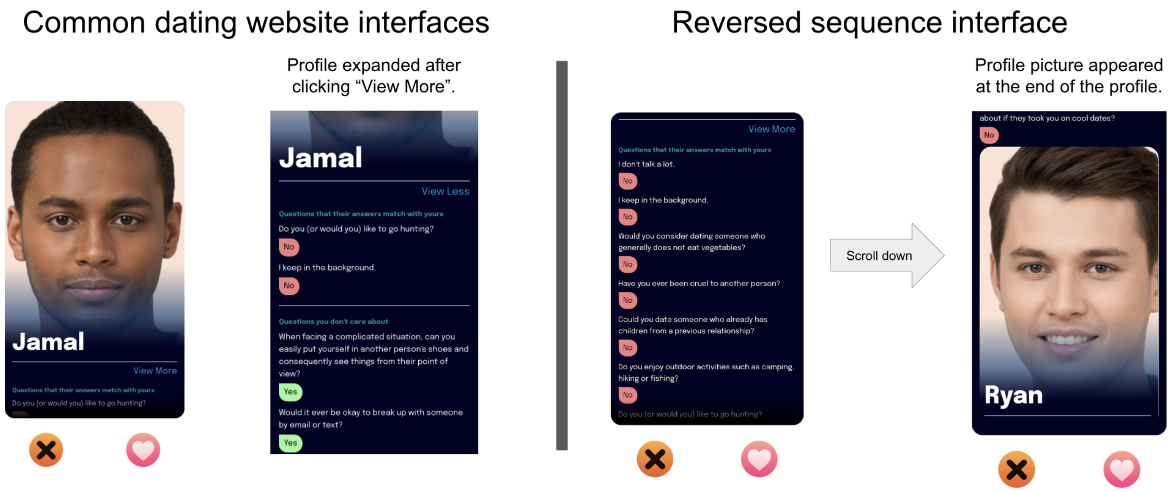

Figure 1: The swipe interface (left), popular with many dating web sites, presents users with a picture of a profiled person along with only some brief information. Users indicate their likes or dislikes by swiping the profile right or left. This design encourages quick decision making based on superficial attributes. In our study with a simulated dating web site, participants who indicated that they had not intended to use race as a criterion in their decision-making, demonstrated a statistically significant anti-Black bias when interacting with the swipe interface. However, when the information sequence was reversed (right) such that participants first saw answers the profiled person gave to substantive questions of importance to the participant, the racial bias was significantly reduced for those who expressed no explicit racial preference.

ABSTRACT

As dating websites are becoming an essential part of how people meet intimate and romantic partners, it is vital to design these systems to be resistant to, or at least do not amplify, bias and discrimination. Instead, the results of our online experiment with a simulated dating website, demonstrate that popular dating website design choices, such as the user of the swipe interface (swiping in one direction to indicate a like and in the other direction to express a dislike) and match scores, resulted in people racially biases choices even when they explicitly claimed not to have considered race in their decision-making. This bias was significantly reduced when the order of information presentation was reversed such that people first saw substantive profile information related to their explicitly-stated preferences before seeing the profile name and photo. These results indicate that currently-popular design choices amplify people's implicit biases in their choices of potential romantic partners, but the effects of the implicit biases can be reduced by carefully redesign the dating website interfaces.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in interaction design.**

KEYWORDS

dating websites, implicit bias

ACM Reference Format:
Zilin Ma and Krzysztof Z. Gajos. 2022. Not Just a Preference: Reducing Biased Decision-making on Dating Websites. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3491102.3517587

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04...$15.00
https://doi.org/10.1145/3491102.3517587

1 INTRODUCTION

Online dating websites have become vital for many individuals to meet and connect with intimate partners. As of today, 30% of Americans have used dating websites [3]. Over 55% of lesbian, gay and bisexual (LGB) individuals have used dating websites and have met at least one person from a dating website [3]. In 2019, Tinder, a popular dating website, was estimated to have 7.86 million active users [95].

In addition to helping people form relationships in general, dating websites create opportunities for people with distant social backgrounds to connect [74]. As the number of couples who met online increased dramatically after 1990, fewer and fewer couples reported having met through family, grade school connections or friends [49, 85]. The ability to find romantic partners outside of one's immediate social circles, which online dating sites enable, has been linked to an increase in interracial marriages [49]. Scholars have argued that marriage is a form of power exchange and facilitates social mobility [53, 84]. Thus, the fact that dating websites bring more interracial marriages could potentially bring more social equity [49].

Despite the potential of online dating sites to bridge distant social groups, racial bias and discrimination can limit opportunities for individuals to form connections [13]. Prior studies conducted by dating websites and academics have shown that race is a significant factor for many individuals when making dating decisions online [2, 60, 64]. Inferred from marriage and socialization data, scientists have found that the preference of people in the US for the same race has exceeded their preference for similarity with respect to any other traits [10, 24, 70, 86].

While instances of discrimination and bias in dating happen offline [87], online spaces offer easy-to-use features for people to act on their biases. For example, dating platforms, such as Grindr, have long been criticized for reifying existing discriminatory social practices [47] and biases [23] through features such as racialized filters [92]. Such features were suspected to make users believe that simplified labels of race can justify the priorities of others in sexual or romantic relationships [19]. The founder of Grindr stated that he did not assume the responsibility for existing social bias manifested on the website [42]. Yet, attempting to take an apolitical stance in design decisions *is*a political stance that upholds the status-quo [35, 101].

In addition to making it easier for people to act on their explicit biases, dating websites may have amplified the effects of implicit biases as well. Implicit bias is a type of automatic evaluation of a person or a group according to the prior belief about them [31]. When making decisions based on implicit bias, the decision-makers are not aware of the bias, even when the bias is inconsistent with their explicit belief [8]. When individuals make decisions on dating websites, they can inadvertently exhibit negative attitudes toward certain social groups, even when they do not want to discriminate. Dating website matching algorithms may pick up such bias and amplify the users' attitudes. Prior research in recommendation systems has discovered that recommendation systems homogenize users' taste [21]. For example, dating websites such as Coffee Meets Bagel still offer racially monotonous profiles even when the users indicate that they have no preference for ethnicity [73]. Interface features such as match scores are intended to evaluate profiles' potential compatibility better. On the contrary, match scores may exert influence on users differently depending on what race of the profile that they examine because users interpret information based on their existing bias [36, 37].

Because dating websites wield a significant influence on users' romantic and sexual relationships, designers have the responsibility to reduce, or at least not amplify, the racial biases that are harmful to minorities. We anticipate objections to this claim. If it is unjust to discriminate based on race in dating (which we refer to as sexual racism [9]), then it should also be unjust to discriminate based on physical features, such as attractiveness [99]. We argue that we should not conflate race with attractiveness and should not perpetuate extant racial hierarchy in romantic and sexual relationships. Conflating race with attractiveness ignores the political and structural forces that have shaped sexual racism [12], affirming the discredited notion that race is biologically fixed [65]. For example, one study of gay men showed that 97% of Asian men, 90% of Latino men, and 88% of Black men stated a preference for White men [79]. This pattern of preference for White people, and the exclusion of people of color, is unlikely due to a taste for specific "types" of people. Instead, it is the already existing racial hierarchy that shaped this notion of "hotness" where Whiteness is prioritized [44]. After all, laws and social institutions have not discriminated based on attractiveness but have on race, which is evident in Jim Crow laws, segregation, and slavery [9, 76].

Sexual racism reinforces racial hierarchy in romantic and sexual relationships, and as designers, we should not uphold it. For example, sexual racism can risk the physical and mental health of excluded people, reinforcing the social disadvantage that minorities already have. Minority men are stereotyped as submissive, and often assume the receptive role of anal sex because of the assumed sexual roles, elevating their risk of HIV and other sexually transmitted infections [7, 9, 43, 100]). Social spaces of people of color and White people are further segregated because sexual racism posts limits on the availability of intimate partners [45]. Thus, sexual racism, when reinforcing racial hierarchy and stereotypes, is not just a personal preference similar to preference over specific physical attraction, but an issue of justice [9]. If one agrees that designers should evaluate whether their designs contribute to racism in general, they should also consider whether their designs contribute to sexual racism.

However, some forms of racial preference that do not contribute to extant social hierarchy can be considered less problematic. Minority people who prefer to date minorities generally do not have racist motives and thus do not reinforce social stigmas. For example, a Black person who prefers Black people in dating challenges the social norm that Whiteness is considered more attractive [9]. Some Black people also prefer to date Black people because they want to avoid racism from the people of other races [61]. On the contrary, racial preference for White people usually implies the racist exclusion of minorities, which is evident in the blatantly racist language used to voice their preferences. For example, many of these open declarations on the dating profiles include racially charged comments such as "I block more Asians than the great wall of China," and "How many times do I have to tell Black guys that I don't like chocolate?" [45]. Additionally, prior research has found an association between the tolerance of sexual racism and racist attitudes in general [19]. For some, racial "preference" is a veil for racist views.

We examined how the design of a simulated dating web site impacted participants' implicit anti-Black bias. We treated the anti-Black bias as an example of anti-minority biases on dating websites. We do not wish to flatten the harmful biases that all minorities face on dating websites because biases that the minorities experience can differ drastically in nature. For example, Asian men are often emasculated on dating websites and considered asexual [58, 60, 78], while Asian women are subjected to fetishization [90]. However, our study's contribution on anti-Black bias can generalize to other biased decision-making against minority races influenced by implicit racial bias on dating websites. These biases share the same unconscious processes that we tackle.

We conducted a dating website simulation experiment (total n = 1907) to investigate the biased decision-making in two specific dating website designs — the swipe interface and swipe + match scores interface. The swipe interface is internationally used in multiple dating websites, including Tinder, Tantan, and Bumble [15, 96, 98]. It allows users to quickly judge user profiles by swiping left to indicate a dislike and right to like. Match scores are prediction metrics generated from user profiles and were used by dating apps such as OkCupid [59] to help users choose partners based on calculated compatibility. These interfaces overemphasize the effects of images and encourages quick decision making based on superficial attributes. As a result, implicit racial bias can influence decision-making. In our experiment, we found that users still made racially-biased dating decisions, even when they explicitly indicated that they did not care about race. We reduced the biased decision-making by reversing the order of information on the dating profile so that the profile picture was shown last. We call this interface the *reversed sequence* interface. We also experimented with the *update interface*, where we asked users to remake their decision in light of new information. In prior work, this intervention resulted in improved decision-making [14, 37, 63] presumably by making people less vulnerable to cognitive biases [14]. However, neither the update interface nor the interface that combined the update and reversed sequence designs reduced biased decision-making in our experiment.

In summary, the contributions of our work are:

- We found that people made racially-biased dating decisions using dating interfaces based on the swipe interaction even when they said that they did not have an explicit racial preference.
- We found that reversing the sequence of information to show race-revealing information (such as profile picture and name) at the end resulted in less biased decisions for individuals who said that they had no explicit racial preferences.

2 RELATED WORK

2.1 Dating website that shapes desires

Some argue that designers should keep a "neutral" position when designing dating platforms, as these platforms concern private and intimate matters. However, it is almost impossible for these platforms to maintain a neutral stance. Dating websites wield immense power in shaping individuals' desires. In his "Do artifacts have politics?" [101] essay, Winner argued that specific properties of technologies are intractably connected to power and authority. For example, harnessing nuclear power requires centralized planning, while utilizing solar power decentralizes power production and consumption to individuals. Like other types of technology, dating websites underlie specific power dynamics and shape how users interact with their environment. When users from less privileged positions backgrounds interact with such a system, they conform to the power dynamics that dating websites orchestrate [47]. This section provides evidence for how dating apps have shaped our desires both to promote social justice and to reinforce stigma and discrimination.

Hutson et al. [52] identified several intimate platform design features that shape our desires based on race and other protected attributes. Depending on one's criterion, dating websites offer filters and search tools that select users based on their sexual orientation, gender, age, and other attributes to provide access to potential partners. Users voluntarily fill in their information based on these selectors. Some selectors seem to be innocuous, such as age or location. Other selectors such as race, ethnicity, and HIV status can legitimize discrimination [52, 68]. These filtering designs grant users agencies to include or exclude profiles based on features that historically were used as the basis of discrimination [91], and therefore jeopardizing the benefit of bridging socially distant groups [74].

Dating websites also shape users' desires with matching algorithms, which pair users with "ideal" partners. However, there is no strong evidence showing that matching algorithms help find potential partners [29]. It is hard for machine learning algorithms to predict the romantic compatibility between individuals [55]. While the benefits of these algorithms are still unclear, some discriminating side effects have been documented. Dating websites such as Tinder [97] often claim that their matching algorithms are "racially blind", and therefore fair in terms of distributing dating opportunities across different races. However, algorithms that allocate resources in a racially blind way only reproduce existing social biases [46]. For example, Coffee Meets Bagel, a dating website, recommended users with profiles of the same race, even when users did not explicitly indicate that they had a racial preference [73]. In this case, algorithms create a feedback loop that homogenizes user behaviors because optimizations are conducted on the data of users who have already been influenced by the system [21]. Even in an ideal and unattainable reality where algorithms are perfect and bias-free, human factors mislead these matching systems. Previous research suggests that people filter information to match their bias, even when such information is helpful in decision-making. Green and Chen [36, 37] found that risk assessment scores paired with a recidivism estimation interface lead to systemic discrimination against Black defendants because higher scores of Black defendants exert more substantial influence on users who are evaluating Black defendants. People may interpret these scores differently based on their prior bias, whether it is recommendations on dating websites or risk scores in a recidivism prediction.

Dating websites occasionally make design decisions that can benefit social equality. For example, after months of Black Lives Matter protests, Grindr finally pledged to remove its race filter from its app [51]. Other platforms started to explore alternative ways for categorizations. For example, the Japanese dating app 9Monsters groups users into nine types of “Monsters” [1]. Although these monsters correspond to weights and height, this categorization offered an alternative to attributes such as race and ethnicity. Many dating websites have actively encouraged users to reflect critically on their biased desires. Grindr launched Grindr for Equality (G4E) campaign to educate and empower the community with respect to the equality issues [40]. Other dating websites started to include community guidelines that address racism and anti-social behaviors [52]. Bumble tries to empower women by only allowing women to message first to reduce harassment online [15]. DaddyHunt, a dating app for sexual minority men, addresses the stigma of people living with HIV by allowing user to have a badge that states “open to dating with someone with any [HIV] status.” [5, 68]

This evidence of dating site interfaces shaping people's desires challenges the assumption that dating websites should keep a “neutral” stance in designing for intimacy. Instead, remaining neutral in developing technology only conserves existing social dynamics [35], including social stigma and oppressive structures. Thus, designers should actively consider if their designs create social inequalities and mitigate them.

2.2 Implicit bias

Essential to social psychology is the question of how people are evaluated. Among the different aspects of personal judgment, stereotyping uses the belief about a social group to make judgments about the individuals of the group. Stereotyping reduces the complexity of making personal judgments and serves as an essential social function [39]. It is also crucial to spontaneous discriminatory behaviors such as implicit bias [30, 39, 71]. Prior research on implicit bias has shown that this process is largely unconscious. It can influence one's judgment without the person's awareness of the influence [8, 25, 26, 38, 39].

Implicit bias shows striking effects towards the members of socially stigmatized groups, such as African Americans [62, 81, 82], women and the LGBTQ community [48, 69]. These biased attitudes can be defined partially by the attitude that resides outside of conscious awareness [89]. For example, the attitudes and the behaviors of the health providers have been identified as one of the leading factors of health disparities for people of color [41, 72]. Even when the explicit attitudes are modified, implicit bias towards people of color still remains and causes differences. Thus, when clinicians express an explicit desire to provide equitable care, the unintentional implicit bias can still increase health disparities [11].

Studies have demonstrated the importance of information first available in forming first impressions. Asch [6] showed that the early information shaped an individual's perception of other details. Ratings on other people become more favorable after the evaluators are presented with a list of words that started with high favorability than when they are presented with a list of words that began with low favorability [4, 6]. When new data inconsistent with this impression is presented, people attempt to rationalize inconsistencies by ignoring the new information [83]. Studies showed that people activate implicit bias subconsciously even when they simply imagine an individual in stereotyped groups [26, 27]. The dating website designs, like the swipe interface, that show the picture of a profile at a salient place likely encourage decisions influenced by the profiled person's race rather than other attributes that the user explicitly indicated as relevant. During the brief time that the users interact with a profile, they have already triggered their implicit bias. We argue that when first presented with a profile picture, the users already form biased first impressions based on the traits of the pictures, such as race, skin color, name, or perceived age.

Human decision-making that over-relies on heuristic processing of information also introduces bias [57, 88]. In the case of dating websites, the swipe interface encourages decision-making based on quick heuristics, which is potentially affected by implicit bias. Thus, one can design the interface to disrupt heuristic processing of information and encourage the users to emphasize deliberate judgments rather than implicit bias. Prior work in human-AI collaboration [14, 37] and diagnostic reasoning [63] has explored *cognitive forcing functions* to mitigate overreliance on heuristic thinking. One version of cognitive forcing functions, the *update* design, has been shown to be particularly useful in reducing users' reliance on heuristics. Update intervention asks the users to make another decision in light of new information.

3 EXPERIMENT

3.1 Overview

Because the profile image is the most salient feature in the typical implementations of the swipe interface and because it encourages quick decision-making, we hypothesize that people's choices will be influenced by their implicit racial attitudes even when they explicitly do not care about race in dating.

To mitigate the impact of implicit attitudes and to help people make decisions more closely aligned with their explicit preferences, we hypothesized that first showing people information that they explicitly identified as important for making their partner selections (e.g., answers to questions like “Do you like to go hunting?”) instead of profile pictures and names would reduce the effects of race-related implicit bias. We call this the *reversed sequence* design.

Further, interventions that disrupt heuristic processing of information and encourage more deliberate information processing should also result in decisions more aligned with explicit preferences rather than with the implicit attitudes. We tested one such intervention where people were first shown the the profile and asked to decide. Only then were they shown the system-computed match score and offered a chance to revise their decision. We call this the *update* design. The update design has been previously shown to be successful in reducing over-reliance in AI-assisted decision-making, presumably by encouraging deeper analysis in situations where the person's initial decision and the system-provided information were at odds [14] (though later research suggested that this design might not lead to more analytical processing of information after all [32]).

In summary, our hypotheses include:

- H1: Users make racially-biased decisions when using dating interfaces based on the swipe interaction even when they do not have an explicit racial preference.
- H2: Reversing the sequence of information to show race-revealing information (such as profile picture and name) at the end will result in less biased decisions.

- H3: Asking the participants to make their own initial decision before seeing the match score will result in less biased decisions.
- H4: Reversing the order of information and asking the participants to make their own initial decision before seeing the match score will result in less biased decisions.

3.2 Procedure

The study was conducted on LabintheWild.org [80], a crowd-sourcing platform where participants voluntarily access the study in exchange for feedback on how they performed in the study. Participants on LabintheWild do not receive monetary compensation. Results of several validation studies performed on LabintheWild and other similar platforms indicate that results obtained from unpaid online volunteers generally match those obtained in supervised laboratory settings [33, 50, 66, 67, 80].

The study was open to anyone. Because the study could be accessed by minors, we advertised it with a slogan "Who's your best friend?". Upon entering the study, participants were asked immediately to indicate how old they were. Participants under 18 continued with a version of the study that was about identifying traits of a best friend. Results from that version of the study are not included in this manuscript. Participants who were 18 or older transitioned to the slogan "Who's your ideal date?". This version of the experiment is described in this manuscript.

Participants were presented with an informed consent before starting the main part of the study, described below.

3.3 Task description

We first asked the participants to (1) respond to a demographic survey, (2) answer what gender of profiles that they wanted to examine, (3) answer check-in questions, (4) do the dating simulation tasks, (5) answer the check-out questions and (6) view a debrief about what check-in questions they valued the most and the racial composition of their preferences. We designed the task to simulate typical interactions on popular dating websites, with slight variations in interface designs. Below, we explain the tasks in more detail.

3.3.1 Demographic survey. We asked the participants' gender, age, and ethnicity in the demographic survey. We included an attention check question in this survey.

3.3.2 Gender of the profiles. Participants were asked "For the purpose of this study, what kind of profiles do you want to examine?", and were given "men" or "women" as options. In the profile interaction part, the participants would only see the profiles of the gender that they selected here.

3.3.3 Check-in questions. Participants were asked questions adapted from OkCupid's check-in questions [59]. These questions surveyed participants about what they expected their ideal partners would answer. These questions include, for example: "Do you enjoy outdoor activities such as camping, hiking or fishing?" The participants had three options for these questions: yes, no, or not important. For ease of reference in this paper, we denote the questions that were answered "yes" or "no" as **important questions**, and the questions answered "not important" as **unimportant questions**. The participants could proceed only if they answered at least 10 important questions and at least 10 unimportant questions. These questions and answers were later used in the profiles that the participants interacted with. The participants should have interpreted the answers to the important questions as the explicit standard of how much a profile matches them. A sample interface of the check-in questions is shown in Figure 2.

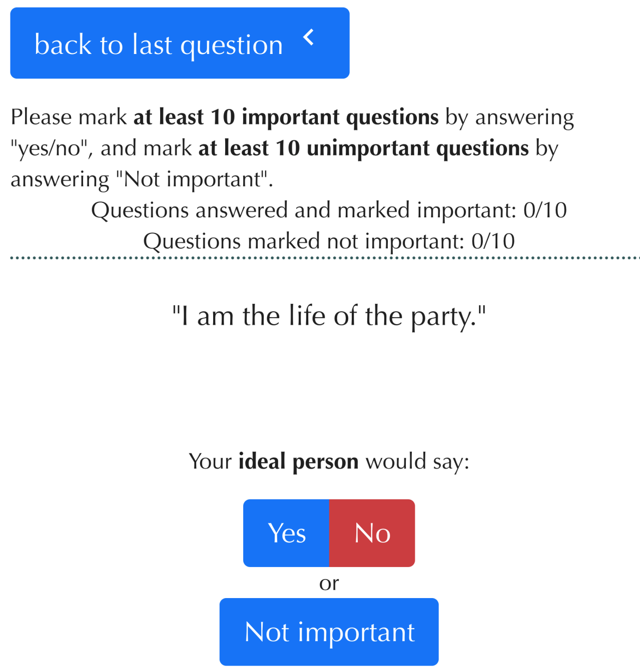

Figure 2: The design of check-in question interface. A participant could answer yes, no or not important to a check-in question.

3.3.4 Dating simulation task. After answering the check-in questions, the participants were asked to interact with fictitious profiles. Participants were shown instructions on how to interact with the profiles according to their assigned interface. Each participant interacted with profiles of 3 races: 5 Asian, 5 Black and 5 White profiles. We focus on anti-Black bias in this study, but we included Asian profiles so that participants would not quickly notice that the study is about race. In addition, participants from all of these three groups would feel welcomed to the study. We would add more races to the dating profile pool if we could, but the images dataset that we utilized only provided three races [75]. Each of the profiles contained a profile picture, name, and profile description.

We generated the profile pools randomly by matching profile pictures, names, and profile descriptions. We selected the names of the profiles to imply the profiles' races. We sampled the profile pictures through GANFaces, a dataset of face images generated by Generative-adversarial Networks [75]. We sampled these faces by hand such that the facial expressions, background color, and camera angles were similar. To mitigate the effect of attractiveness, we asked workers on Amazon Mechanical Turk to evaluate the attractiveness of a pool of images. Workers were given a Likert Scale question asking "Is this person portrayed as attractive?". Workers answered this question by "Very", "Yes", "Somewhat" or "No". For the study, we picked the pictures with the closest attractiveness ratings, regardless of race. Eventually, a total of 30 images, 15 for each of the two genders, 5 per gender for each of the three races, were selected. An ordinal logistic regression conducted separately on each gender did not find a significant difference between the attractiveness ratings of the pictures.

The profile descriptions consisted of 10 check-in questions that the participants already marked. We varied the number of important questions in the descriptions such that for each of the 5 profiles in the 3 races, there were 2, 4, 6, 8 or 10 important questions in the descriptions. The answers to these important questions matched the answers that the participants responded to in the check-in question section so that no *single* important question would make a participant reject immediately because the answers were undesirable. The rest of the profile descriptions consisted of the participants' check-in questions that were marked "unimportant". The answers to these unimportant questions were random. We calculated a match score for each of the profiles as the percentages of the 10 questions in a profile that were important. Therefore, the match scores for the profiles of each race included 20%, 40%, 60%, 80% and 100%. The participants should interpret the match score as how much a specific profile matched the *explicit* standards that the participant expressed.

3.3.5 Checkout form. In this section, the participant first indicated whether they made their choices based on race. The exact question asked was: "Was race an important factor for you to make decisions?" If the participant answered yes, then we considered this participant to have an *explicit* bias. Next, we asked the participants to indicate whether they had technical issues, had cheated or idled in the test. We also provided free-response text input for them to elaborate on when they selected any of these responses. Finally, we presented their ideal date's traits to the participants based on their selections and thanked them for taking the study.

3.4 Interfaces

We designed a dating simulation task with five interfaces (swipe, swipe + match score, update, reversed sequence, and combined). See figure 3 for the visual designs of the interfaces.

swipe: participants liked or disliked by clicking the buttons or swiped accordingly, similar to most of the popular dating websites. By clicking on "View More", participants could examine the whole profile descriptions in an expanded view.

swipe + match score: this interface was identical to the swipe interface except that it also showed the match score next to the profiled person's name.

update: The participants with the update interface first interacted with the profiles without seeing the match scores. The profile presentation in this first interaction was identical to the swipe interface. After participants made their initial decision, a pop up message notified them of the match score of the profile. Then the participants made their decision again. This specific intervention and the text were adapted from Green et al. [37].

reversed sequence: The participants in the reversed sequence interface first viewed the profile descriptions and then the profile pictures instead of vice versa.

combined: This interface is the combined interface of update and reversed sequence, where the users viewed the reversed order of information and had the chance to update their decision after seeing the match scores.

3.5 Participants

A total of 2534 participants finished the study. 91 participants who liked all profiles or disliked all profiles were removed. We removed 102 participants who did not answer the attention check questions correctly. We removed 224 participants who indicated that they had cheated or faced technical problems during the test. Some of the comments for the participants who have cheated or idled included: *"I accidentally answered a bunch of questions without reading them", "I clicked the wrong button by accident a few times".* We removed 28 participants finishing the task outside of the 95% quantile in terms of the time spent on the tasks. We only considered participants who indicated binary genders because we did not have representative gender identities in our dataset to analyze non-binary participants. We were left with 1907 participants. There were 987 men (of which 402 were looking for men and 587 were looking for women), 920 women (of which 633 were looking for men and 287 were looking for women). 1338 participants indicated that race was not a factor in dating. The other 569 participants indicated that race was a factor. Among these participants, most of them (1051) never used dating websites. 446 participants used dating websites several times per month. 97 participants used dating websites once a week. 208 participants users used dating websites a few times a week. 105 participants used dating websites daily.

3.5.1 Participants demographics compared to dating website demographics. We wanted to compare the demographics of dating website users to our participants. We obtained the proportion of US residents of a particular demographic group using dating websites from the the survey by Pew Institute [20]. We obtained the proportion of different demographics (gender, age and race) in the US population from the 2011 US Census data [16]. We then used the Bayes' theorem to compute what fraction of all dating website users belonged to which demographic group. Because different dating websites target different demographics, this estimate is designed to capture the overall demographics of people using dating websites rather than the demographics of any specific dating website. In table 1, a comparison of our user demographics shows that our participant pool generally represents the young, mostly White and slightly more men demographics of dating website users. In out study, participants (the median is 25–34) are generally younger than the dating website demographics (the median is 35–44). We have more women participating in the study, and more Black participants than the users of dating websites.

3.6 Approvals

This research has been reviewed and approved by the Internal Review Board at Harvard University, protocol number IRB20-1308.

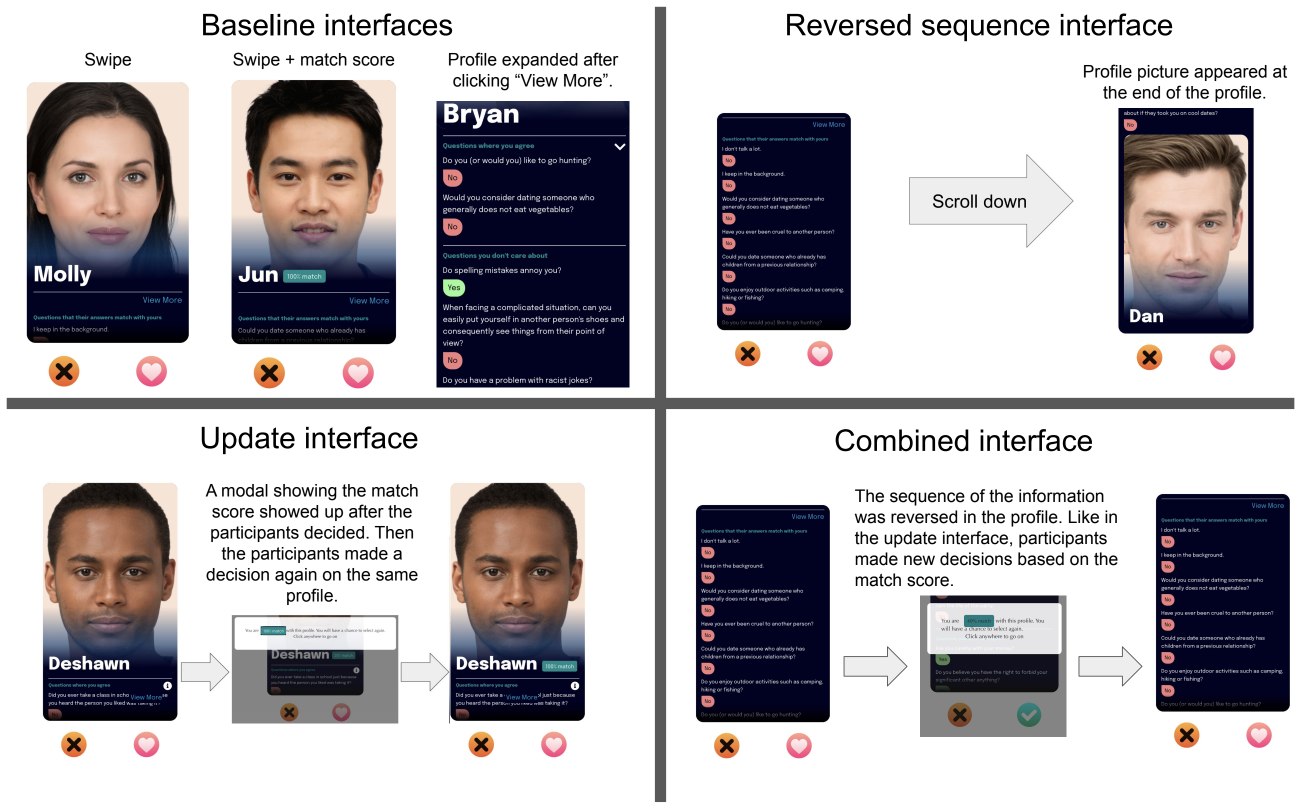

Figure 3: Baseline interfaces: Baseline interface contained the *swipe* interface the *swipe + match score* interface. The swipe interface contains the name, the profile picture, the like button, and the dislike button. The swipe + match score interface also displays the match score. The match score is the ratio of the 10 questions that are important questions on which the participants agreed. The users can swipe left or right to indicate like or dislike. After participants clicked on "view more", the important questions were shown first, while the other questions were shown later. In our preliminary analysis, we did not find a significant difference in bias exhibited by participants who interacted with the swipe interface and swipe + match score interface. We treat the results from these two interfaces as a single interface, which we refer to as the Baseline interface. Update interface: participants first made decisions without seeing the match scores. Afterward, participants saw a prompt with the match score of the profile and made their decisions again. The second decision will be their final decision on the profile. Reversed sequence interface: the profile descriptions were shown first in the profile rather than the profile image and names. After participants opened a profile and scrolled down, they saw the profile image and the name of the profiles after the profile descriptions. Combined interface: participants see a reversed order of information, made their initial decision and then update their decisions after they see match scores.

3.7 Design and Analysis

The study was a between-subjects study with one factor, the interface, with up to 5 levels: swipe, swipe + match score, update, reversed sequence, and combined. Because our preliminary analyses (Section 4.1) showed no statistically significant differences between swipe and swipe + match score interfaces, in subsequent analyses we modeled them together as a single factor level: baseline interface.

We used the following measures as dependent variables in our analyses:

- **Bias.** This variable captured the bias in favor of White profiles compared to Black profiles. We computed bias by pairing participants' responses to Black and White profiles with identical scores (i.e., profiles that should be equally attractive given their explicitly stated preferences). For each such pairing of profiles, the Bias was 1 if the participant chose the White profile but not the Black profile. The Bias was -1 if the participant chose the Black profile but not the White one. If the participant made the same choice for the two profiles, the Bias was 0.
- **Time spent.** The time that participants spent on a profile.
- **Details viewed.** Whether the participants have toggled open each profile.

We also captured the following information for use as covariates in some analyses:

	Study participants	Dating website users
	Gender	
Men	0.517	0.583
Women	0.483	0.415
	Race	
White	0.730	0.839
Black	0.270	0.161
	Age	
18-24	0.417	0.160
25-34	0.348	0.332
35-44	0.141	0.256
45-54	0.065	0.132
55-64	0.021	0.080
65+	0.008	0.044

Table 1: A comparison of participant demographics and the estimates of actual dating website users' demographics. Dating website demographics data on ethnicities other than White or Black is missing from the dating website demographics. Therefore, the probabilities in the race rows are normalized by Black + White. This table indicates that there are more women, more Black participants in our study than among the users of actual dating websites. Our participants are also younger on average.

- **Score** (20%, 40%, 60%, 80% or 100%) captured the percentage of the 10 questions in a profile that the participant earlier indicated were important to them (and on which the participant and the profile agreed). Score was modeled as a continuous variable.
- **Participant gender** (man or woman).
- **Profile gender** (man or woman) captured the gender of the profiles a participant chose to view.
- **Participant sexuality** (men looking for men, men looking for women, women looking for men, women looking for women). Participant sexuality was modeled as an interaction between participant gender and the profile gender.

Because our data were not normally distributed, we used ordinal logistic regression for most analyses. We modeled participant as a random effect because we had multiple observations per participant (one for each score level).

We reported odds ratios (OR) [93] to help in interpretation of the model parameter estimates. Notice that $OR = exp(\beta)$, where β is the parameter estimate of the model. OR greater than 1 indicate a positive effect, while OR smaller than 1 indicate a negative effect. OR can be interpreted as effect sizes similar to Cohen's d [77]. Values between 1.5 and 3 (or between 0.33 and 0.66) are interpreted as a small effect, between 3 and 5 (or between 0.2 and 0.33) as medium, and above 5 (or below 0.2) as large [22, 77, 93].

Because there is evidence that people of different genders and sexuality exhibit different racial bias in selecting intimate partners [28, 34, 54], we included these attributes in our models.

A participant's race is also likely to be associated with their choices. However, because a large fraction (24%) did not report their race and because some races were sparsely represented in our data (e.g., only 0.5% identified as Pacific Islanders compared to 44.4% who identified as White; see Table 2), we decided not to include race directly in our analysis.

4 RESULTS

Except where explicitly stated otherwise, all results pertain to participants who indicated that race was not important in their decisions.

4.1 Preliminary Analyses

First, we checked if there were significant differences in bias exhibited by participants who interacted with the swipe interface and those who interacted with the swipe + match score interface. We computed two models to analyze the impact of these interfaces (Table 4). An analysis on the goodness of fit found a significant difference between the two models ($\chi^2(4, 2460) = 11.72, p = 0.02$). In the first model, we included participant gender (both as a main effect and as an interaction term with the interface design). In the second model, we included both participant gender and sexuality, which we modeled as an interaction between participant gender and profile gender. We observed no significant association between interface and bias in either model ($OR = 1.03, Z = 0.21, n.s.$ in the first model, and $OR = 0.96, Z = -0.183, n.s.$ in the second). Consequently, in subsequent analyses, we combined the swipe and swipe + match score interfaces into one baseline interface.

4.2 People using the swipe interfaces made biased choices

On average, people using the baseline interfaces (i.e., the swipe or the swipe + match score interfaces) demonstrated a statistically significant anti-Black bias ($M = 0.043, SD = 0.31$, 1-sample Wilcoxon signed-rank test $Z = -2.65, p = 0.0081$). Specifically, given a pair of profiles with identical match scores, participants on average chose the White profile but not the Black one 25.27% of the time, the Black profile but not the White 15.13% of the time, and treated both profiles equally (i.e., both chosen or both rejected) 59.59% of the time. This result supports Hypothesis 1.

4.3 Reversing information order helped people make less biased choices

We computed two models to analyze the impact of interface designs on the bias in participants' choices (Table 5). There was no statistically significant difference in the goodness of fit between the two models ($\chi^2(8, 6690) = 3.2403, p = 0.356$). In the first model, we included participant gender (both as a main effect and as an interaction term with the interface designs). In the second model, we also included participant sexuality, which we modeled as an interaction between participant gender and profile gender. In both models, we observed a very small but significant main effect of reversing information order ($OR = 0.78, Z = -2.36, p = 0.018$ in the first model, and $OR = 0.69, Z = -2.20, p = 0.028$ in the second model). On average, the bias for the baseline interfaces was M = 0.043, SD = 0.31, while the bias in the reversed sequence interface was M = 0.012, SD = 0.28 (Table 3). In both models, we observed a significant main effect of gender, with women making slightly but significantly less biased choices than men ($OR = 0.78, Z = -3.62, p = 0.0003$ in

Interfaces	Asian	Black	Latino	MENA	Native American	Other/ No answer	Pacific Islander	White	Total
swipe + match score	39	56	25	1	1	88	3	148	361
swipe	28	56	19	1	3	74	3	160	344
reversed	48	100	29	4	3	149	2	262	597
update	27	53	21	1	5	75	1	133	316
combined	22	49	13	2	1	57	1	144	289
Total	164	314	107	9	13	443	10	847	1907

Table 2: Participants breakdown by demographics and interfaces. MENA: Middle Eastern and North African. Other/ No answer: this category includes people who self-described as categories other than the categories that we have in the survey and people who declined to answer this question

	Bias (SD)			
	baseline	reversed	update	combined
All	0.04(0.31)	0.01(0.28)	0.02(0.27)	0.04(0.27)
Man	0.09(0.32)	0.03(0.29)	0.05(0.26)	0.06(0.29)
Woman	0.00(0.30)	0.00(0.27)	0.00(0.28)	0.01(0.25)
Men looking for men	0.11(0.34)	0.01(0.30)	0.04(0.30)	0.07(0.28)
Men looking for women	0.08(0.31)	0.03(0.28)	0.06(0.22)	0.06(0.29)
Women looking for men	0.04(0.28)	0.01(0.29)	0.03(0.28)	0.04(0.25)
Women looking for women	-0.07(0.31)	-0.03(0.22)	-0.06(0.26)	-0.05(0.24)

Table 3: Summary statistics: Bias exhibited by participants who said that race was not an important factor in their decisions, aggregated by interface and participant gender and sexuality. SD: standard deviation.

Parameter	Model 1: OR (95% CI)	Model 2: OR (95% CI)
interface: swipe + match score	1.03 (0.78 , 1.37)	0.96 (0.61 , 1.5)
participant gender: woman	**0.75 * (0.56 , 0.99)**	0.83 (0.56 , 1.23)
profile gender: women		0.82 (0.54 , 1.26)
score	0.99 (0.97 , 1.02)	0.99 (0.97 , 1.02)
participant gender: woman * profile gender: women		0.64 (0.36 , 1.14)
interface: match score*participant gender: woman	0.92 (0.62 , 1.36)	0.85 (0.49 , 1.48)
interface: match score * profile gender: women		1.14 (0.64 , 2.02)
interface: match score*participant gender: woman*profile gender: women		1.4 (0.63 , 3.1)
Observations	2460	

Table 4: The model assessing whether showing match score had an effect on the participants who said that race was not an important factor in their decisions. We did not find a significant impact of showing the match scores. The reference level for interface is swipe. $^{*}\ p < .05$, $^{**}\ p < .01$, $^{***}p < .001$.

the first model, and $OR = 0.75, Z = -2.04, p = 0.042$ in the second model.).

In the first model, we observed a marginal interaction effect between reversing information sequence and participant gender ($OR = 1.27, Z = -1.76, p = 0.078$), indicating that the behavior of men and women may be affected differently by this design. Indeed, in the follow up analyses separated by participant gender (Table 6) we observed a significant main effect of reversing information order for men ($OR = 0.78, Z = -2.294, p = 0.022$) but not for women ($OR = 1, Z = 0.043, n.s.$). This difference in impact can be explained in part by the fact that, as shown in Table 3, women in baseline interface condition exhibited no bias on average ($M = 0, SD = 0.3$). These results support Hypothesis 2, but not 3 or 4.

4.4 Additional analysis

Participants who indicated that race *was* important in their decisions demonstrated statistically significant bias when using baseline interfaces ($M = 0.23, SD = 0.45$, 1-sample Wilcoxon signed-rank test $Z = -3.89, p = 0.0001$). We analyzed the effects of the interventions on the people who said that they cared about race in dating. As expected, the interventions do not affect the participants who explicitly considered race in their decisions (Table 7).

4.5 Behavioral data when bias was reduced

We analyzed the behavior data, such as time spent on each profile or whether participants opened a profile, to explore evidence indicating why sequence intervention was effective for people who

Parameter	Model 1: OR (95% CI)	Model 2: OR (95% CI)
interface: reversed	**0.78 * (0.64 , 0.96)**	**0.69 * (0.5 , 0.96)**
interface: update	0.84 (0.66 , 1.07)	0.76 (0.52 , 1.12)
interface: combined	0.89 (0.69 , 1.13)	0.83 (0.55 , 1.27)
profile gender: women		0.88 (0.66 , 1.16)
participant gender: woman	**0.71 *** (0.58 , 0.85)**	**0.75 * (0.58 , 0.99)**
score	1.01 (0.99 , 1.02)	1.01 (0.99 , 1.02)
profile gender: women*participant gender: woman		0.75 (0.51 , 1.1)
interface: combined*profile gender: women		1.11 (0.66 , 1.85)
interface: reversed*profile gender: women		1.22 (0.8 , 1.85)
interface: update*profile gender: women		1.18 (0.72 , 1.93)
interface: combined*participant gender: woman	1.17 (0.83 , 1.65)	1.2 (0.72 , 2.01)
interface: reversed*participant gender: woman	1.28 (0.97 , 1.69)	1.32 (0.88 , 1.97)
interface: update*participant gender: woman	1.17 (0.83 , 1.65)	1.28 (0.78 , 2.09)
interface: combined*profile gender*women*participant gender*woman		0.98 (0.48 , 2.02)
interface: reversed*profile gender: women*participant gender: woman		1.06 (0.6 , 1.89)
interface: update*profile gender: women*participant gender: woman		0.9 (0.45 , 1.82)
Observations	6690	

Table 5: The ordered mixed model for participants said that race was not an important factor in their decisions. We compared two models. In the first model, we considered participant gender. In the second model, we considered participant gender and participant sexuality, which we modeled as an interaction between participant gender and profile gender. There was no significant difference between the two models in terms of model fit. Overall, the reversed sequence interface reduced bias in both models. We also observed a marginal interaction effect between the reversed sequence intervention and the gender of the participant ($p = 0.09$) in the simpler model. The reference level for interface is baseline. * $p < .05$, ** $p < .01$, $^{***}p < .001$.

	Women	Men
Parameter	OR (95% CI)	OR (95% CI)
interface: reversed	1 (0.84 , 1.21)	**0.78 * (0.64 , 0.96)**
interface: update	0.99 (0.78 , 1.25)	0.84 (0.66 , 1.08)
interface: combined	1.04 (0.82 , 1.31)	0.89 (0.69 , 1.14)
score	1.02 (0.99 , 1.04)	0.99 (0.97 , 1.02)
Observations	3515	3175

Table 6: The ordered mixed models for men and women who said that race was not an important factor in their decisions. The reversed sequence intervention was effective for men, but not for women. The reference level for interface is baseline interfaces. * $p < .05$, ** $p < .01$, $^{***}p < .001$.

Parameter	Odds Ratio (95% CI)
interface: combined	0.79 (0.53 , 1.18)
interface: reversed	0.77 (0.56 , 1.06)
interface: update	0.86 (0.6 , 1.24)
participant gender: woman	0.91 (0.65 , 1.28)
profile gender: women	1.24 (0.9 , 1.73)
score	1.01 (0.98 , 1.04)
participant gender: woman * profile gender: women	**0.39 ** (0.21 , 0.7)**
Observations	2845

Table 7: The ordered mixed model for participants who said that race was an important factor in their decisions. Because they showed explicit bias, none of the interventions have an effect on these participants. The reference level for interface is baseline interfaces. There is no effects of * $p < .05$, ** $p < .01$, $^{***}p < .001$. **Odds ratio** $= exp(\beta)$.

said that race was not an important factor in their decisions. We used the Wilcoxon rank-sum test to compare the metrics in the baseline interfaces and the reversed sequence interface. We found that reversed sequence intervention changed several metrics to the desired direction.

Compared to the baseline interfaces ($Mdn = 13.9\ s$), participants in the reversed sequence interface ($Mdn = 15.04\ s$) spent more time on the profiles overall ($Z = -2.65, p = 0.008$). We compared the difference of the differences of time spent on White and Black profiles between the two conditions. When using the reversed sequence interfaces, participants spent more equal time ($Mdn = 0.14\ s$) looking at White and Black profiles than in the baseline interface ($Mdn = 0.46\ s$), but this difference in differences was not statistically significant ($Z = -0.21, n.s.$).

We compared the difference in the percentages of While and Black profiles that the participants toggled open. Over half of the participants opened the profiles equally in both baseline and reversed sequence interfaces ($Mdn = 0$ for both interfaces), so we report means instead of medians. Participants opened the profiles of the Black and White profiles more equally in the reversed sequence ($M = -0.002, SD = 0.12$) interface than in the baseline interfaces ($M = 0.022, SD = 0.14$). This difference is significant ($Z = -2.44, p = 0.015$).

Additionally, participants were more likely to pick Black profiles in the reversed sequence interface ($M = 0.65, SD = 0.12$) than in the baseline interfaces ($M = 0.58, SD = 0.14$). This difference in the fraction of likes that Black profiles get in control versus in reversed sequence condition is significant: $Z = -3.719, p =< 0.0001$.) White profiles received more likes in the reversed sequence interface too ($M = 0.62, SD = 0.14$ in the control interface. $M = 0.66, SD = 0.12$ in the reversed sequence interface. This difference in the fraction of likes that White profiles get in control versus in reversed sequence condition is significant: $Z = -2.57, p = 0.01$.)

5 DISCUSSION

Through a study using a dating website simulation, we showed that even when people indicated no racial preference explicitly, they still made racially-biased decisions in selecting potential dating partners when using swipe interfaces—a popular choice on contemporary dating websites. Our results challenge the assumption that the current dating website interfaces are politically "neutral". The basic swipe interface—where people are presented with a profile picture and name, and are encouraged to express their like or dislike by swiping right/left—did invite individuals who had no explicit racial preference in dating to make racially-biased decisions. Thus, the current dating website deigns is not neutral, as it preserves the racial biases embedded in our society. A seemingly helpful feature like match scores, which are meant to inform the user how well the profiled person matches their explicit preferences, did not reduce the impact of implicit bias on people's choices. We hypothesize that users' implicit preferences will be forwarded to downstream tasks such as profile recommendations based on these results. Then algorithms may recommend racially-biased profiles even when the users explicitly indicate that they do not have a racial preference. This result can potentially explain why matching algorithms still recommended users racially similar profiles on Coffee Meets Bagel, even when users did not explicitly indicated that they had a preference in race.

Our results showed that a different interface design significantly reduced anti-Black bias in the decisions made by participants who said that race was not a factor in their decision making. In this alternative *reversed information order*design, people first saw the information related to the attributes they explicitly identified as important and only then saw race-related information (such as profile picture and name). When using this interface, participants spent more time viewing each profile compared to the standard swipe interface. Moreover, participants opened profiles more equally, giving Black and White profiles equal chances of close examination. As expected, this intervention only reduced biased decision-making of people who said that race was not an important factor in their decisions.

We only demonstrated the effects of the reversed sequence intervention on the anti-Black bias rather than other anti-minority biases. The racism that different minority groups face differs depending on the context, and it is impossible to enumerate all kinds of implicit biases against minorities. However, we postulate that this intervention can help reduce other race-related biased decision-making, as long as implicit racial bias influences the decision-making.

We currently do not have evidence that update interface (where people first make a decision using the swipe interface and only then are presented with the match score and given a chance to revise their decision) is useful in reducing biased decision-making, despite its effectiveness in some human-AI collaboration settings. The update interface may tackle cognitive biases different from implicit bias. In addition, tasks in human-AI collaboration generally have a correct answer. Users are highly motivated to answer correctly, so the update intervention encourages people to think analytically. However, there is no "correct" decision on the dating websites, and people are less motivated to reconstruct their biased choices.

We also found that a combination of update interface and reversed sequence interface did not reduce bias. Prior research shows that effortful decision-making only occurs when motivations and cognitive capacities are present [88]. For example, several studies have shown that increased cognitive load contributes to racially biased clinical decision-making [17, 18, 56]. Perhaps the combination of the two interventions made the task too cognitively demanding or so tedious that participants lost their motivation for the task.

Additionally, we propose that the reversed sequence intervention can be generalized into designs other than simply reversing the order of information presented, as long as the interface does not immediately show the profile's race. Dating interfaces can first present images that have no salient racial information, such as images of hobbies or photography portfolios. The interfaces can also be designed such that users have to maintain conversations before seeing each others' profile pictures. For example, Taffy [94] is a dating website that prioritizes chat over appearance. On Taffy, users can only see other users' profile pictures after meaningful conversations. Although Taffy focuses on reducing the effects of the appearance of a user rather than race, their interface showed a different and viable implementation of the reversed sequence intervention.

Although these alternative interfaces, such as the reversed sequence intervention, can decrease biased behaviors, they can also sacrifice the users' engagement because swiping images is more entertaining than reading text. Prior studies that used the update intervention also raised similar concerns that adding friction to the human-AI interaction in real-life systems may interrupt user experience [14]. Buçinca et al, suggested adaptively deploying such interventions by predicting when and where the deployment of the intervention will yield the optimal improvement [14]. Designers may need to accurately model users' implicit biases and deploy the reversed sequence intervention accordingly on dating websites.

One limitation of our work is that our participant demographics do not align completely with the dating website demographics. We have participants more diverse than the dating websites in real life. For example, a larger fraction of women and Black participants participated in our study than use dating websites in real life. Another limitation is that we did not include the participants who indicated non-binary genders.

Some users may not want to interact with other users who have explicit or implicit biases. The reversed sequence interface impacts individuals who only show implicit bias, while the effects on people with explicit biases are limited. Therefore our intervention will not increase the match of minority users with people with explicit bias. We are optimistic that people with implicit biases are willing to interact with minorities respectfully.

Potential future work includes understanding bias and the effectiveness of our intervention among non-binary users. In that case, gender socialization, instead of self reported gender, might have a significant effect on how bias is manifested. Additionally, making decisions to initiate contact on a dating website is only the first step in finding partners in real life. After matching with a minority profile, users may still hold implicit biases against these profiles. While we have demonstrated the effectiveness of reversed sequence intervention on reducing biased-decision making, we cannot achieve equitable interaction on dating websites without considering how holistically different components of dating websites work together.

6 CONCLUSION

As dating websites are becoming a crucial part of meeting potential sexual and romantic partners, it is vital to ensure that people of diverse identities have equal access to love and romance. Our study found that the current popular dating website designs, the swiping interface with or without match scores, do not support this goal. People, even those who said that race was not an important factor in their decisions, exhibited anti-Black bias when using those interface designs. Interface designers should actively examine the equity effects of their system and come up with viable ways to reduce bias. The results of our experiment suggest that the designers consider changing the presented order of information to reduce the harmful effects of the swipe interface and match scores: The interface should provide race-revealing information (such as the person's photo or name) after the information that the searcher explicitly indicated was important to their decision-making. Our findings may only be the first step in anti-bias design on dating websites because different components of dating websites such as algorithms and community building can also harbor bias. Therefore, further research is necessary to assess the effects of implicit bias on other components of dating websites and how these components interact to maintain or amplify racial discrimination that is already present in offline settings.

7 ONLINE ACCESS

Data available at: https://doi.org/10.7910/DVN/NYRUGS.

ACKNOWLEDGMENTS

We extend special thanks to Prof. Mahzarin Banaji for her mentorship throughout this project. We thank Lucy Yip for the visual designs of the dating website interface. We thank Herman Soksano, Zana Buçinca, Brendan O'melia, Vineet Pandey and Sohini Upadhyay for valuable feedback on the paper.

This work was funded in part by the National Science Foundation (grant IIS-2107391).

REFERENCES

[1] 9Monsters. 2021. 9Monsters. https://ninemonsters.com/en

[2] Saleem Alhabash, Kayla Hales, Jong-hwan Baek, and Hyun Jung Oh. 2014. Effects of race, visual anonymity, and social category salience on online dating outcomes. *Computers in Human Behavior* 35 (2014), 22–32. https://doi.org/10.1016/j.chb.2014.02.016

[3] Monica Anderson, Emily A. Vogels, and Erica Turner. 2020. Americans' personal experiences with online dating. https://www.pewresearch.org/internet/2020/02/06/americans-personal-experiences-with-online-dating/

[4] Norman H. Anderson and Alfred A. Barrios. 1961. Primacy effects in personality impression formation. *The Journal of Abnormal and Social Psychology* 63, 2 (1961), 346–350. https://doi.org/10.1037/h0046719

[5] Daddyhunt App. 2020. Faqs. https://www.daddyhuntapp.com/faqs

[6] S. E. Asch. 1946. Forming impressions of personality. *The Journal of Abnormal and Social Psychology* 41, 3 (1946), 258–290. https://doi.org/10.1037/h0055756 Place: US Publisher: American Psychological Association.

[7] George Ayala, Trista Bingham, Junyeop Kim, Darrell P Wheeler, and Gregorio A Millett. 2012. Modeling the impact of social discrimination and financial hardship on the sexual risk of HIV among Latino and Black men who have sex with men. *American journal of public health* 102 Suppl 2, Suppl 2 (May 2012), S242–S249. https://doi.org/10.2105/AJPH.2011.300641 Edition: 2012/03/08 Publisher: American Public Health Association.

[8] Mahzarin R. Banaji, Curtis Hardin, and Alexander J. Rothman. 1993. Implicit stereotyping in person judgment. *Journal of Personality and Social Psychology* 65, 2 (1993), 272–281. https://doi.org/10.1037/0022-3514.65.2.272

[9] Sonu Bedi. 2015. Sexual Racism: Intimacy as a Matter of Justice. *The Journal of Politics* 77, 4 (Oct. 2015), 998–1011. https://doi.org/10.1086/682749

[10] Brent Berry. 2006. Friends for Better or for Worse: Interracial Friendship in the United States as Seen Through Wedding Party Photos. *Demography* 43, 3 (2006), 491–510. https://doi.org/10.1353/dem.2006.0020

[11] Irene V. Blair, John F. Steiner, and Edward P. Havranek. 2011. Unconscious (implicit) bias and health disparities: where do we go from here? *The Permanente Journal* 15, 2 (2011), 71–78.

[12] Donald Braman. 1999. Of Race and Immutability. *UCLA Law Review* 46 (1999), 1375.

[13] Ashley Brown. 2018. 'Least Desirable'? How Racial Discrimination Plays Out In Online Dating. https://www.npr.org/2018/01/09/575352051/least-desirable-how-racial-discrimination-plays-out-in-online-dating

[14] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z. Gajos. 2021. To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-Assisted Decision-Making. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 188 (apr 2021), 21 pages. https://doi.org/10.1145/3449287

[15] Bumble. 2020. Date, meet, network better. https://bumble.com/

[16] U.S. Census Bureau. 2011. U.S. Census Bureau (2010). https://www.census.gov/prod/cen2010/briefs/c2010br-08.pdf

[17] Diana J. Burgess. 2010. Are Providers More Likely to Contribute to Healthcare Disparities Under High Levels of Cognitive Load? How Features of the Healthcare Setting May Lead to Biases in Medical Decision Making. *Medical Decision Making* 30, 2 (March 2010), 246–257. https://doi.org/10.1177/0272989X09341751

[18] Diana J. Burgess, Sean Phelan, Michael Workman, Emily Hagel, David B. Nelson, Steven S. Fu, Rachel Widome, and Michelle van Ryn. 2014. The Effect of Cognitive Load and Patient Race on Physicians' Decisions to Prescribe Opioids for Chronic Low Back Pain: A Randomized Trial: Table 1. *Pain Medicine* 15, 6 (June 2014), 965–974. https://doi.org/10.1111/pme.12378

[19] Denton Callander, Christy E. Newman, and Martin Holt. 2015. Is Sexual Racism Really Racism? Distinguishing Attitudes Toward Sexual Racism and Generic Racism Among Gay and Bisexual Men. *Archives of Sexual Behavior* 44, 7 (Oct. 2015), 1991–2000. https://doi.org/10.1007/s10508-015-0487-3

[20] Pew Research Center. 2013. Part 2: Dating apps and online dating sites. https://www.pewresearch.org/internet/2013/10/21/part-2-dating-apps-and-online-dating-sites/

[21] Allison J. B. Chaney, Brandon M. Stewart, and Barbara E. Engelhardt. 2018. How algorithmic confounding in recommendation systems increases homogeneity and decreases utility. In *Proceedings of the 12th ACM Conference on Recommender Systems*. ACM, Vancouver British Columbia Canada, 224–232. https://doi.org/10.1145/3240323.3240370

[22] Henian Chen, Patricia Cohen, and Sophie Chen. 2010. How Big is a Big Odds Ratio? Interpreting the Magnitudes of Odds Ratios in Epidemiological Studies. *Communications in Statistics - Simulation and Computation* 39, 4 (March 2010), 860–864. https://doi.org/10.1080/03610911003650383

[23] Christopher T. Conner. 2019. The Gay Gayze: Expressions of Inequality on Grindr. *The Sociological Quarterly* 60, 3 (July 2019), 397–419. https://doi.org/10.1080/00380253.2018.1533394

[24] S. Currarini, M. O. Jackson, and P. Pin. 2010. Identifying the roles of race-based choice and chance in high school friendship network formation. *Proceedings of the National Academy of Sciences* 107, 11 (March 2010), 4857–4861. https://doi.org/10.1073/pnas.0911793107

[25] Jan De Houwer, Sarah Teige-Mocigemba, Adriaan Spruyt, and Agnes Moors. 2009. Implicit measures: A normative analysis and review. *Psychological Bulletin* 135, 3 (May 2009), 347–368. https://doi.org/10.1037/a0014211

[26] P. Devine. 1989. Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology* 56 (1989), 5–18.

[27] Russell Fazio, David Sanbonmatsu, Martha Powell, and Frank Kardes. 1986. On the Automatic Activation of Attitudes. *Journal of personality and social psychology* 50 (03 1986), 229–38. https://doi.org/10.1037/0022-3514.50.2.229

[28] Cynthia Feliciano, Belinda Robnett, and Golnaz Komaie. 2009. Gendered racial exclusion among white internet daters. *Social Science Research* 38, 1 (March 2009), 39–54. https://doi.org/10.1016/j.ssresearch.2008.09.004

[29] Eli J. Finkel, 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, 1 (Jan. 2012), 3–66. https://doi.org/10.1177/1529100612436522

[30] Susan T Fiske. 1998. Stereotyping, prejudice, and discrimination. In *The handbook of social psychology, Vols. 1-2, 4th ed.* McGraw-Hill, New York, NY, US, 357–411.

[31] Chloë FitzGerald, Angela Martin, Delphine Berner, and Samia Hurst. 2019. Interventions designed to reduce implicit prejudices and implicit stereotypes in real world contexts: a systematic review. *BMC Psychology* 7, 1 (dec 2019), 29. https://doi.org/10.1186/s40359-019-0299-7

[32] Krzysztof Z. Gajos and Lena Mamykina. 2022. Do People Engage Cognitively with AI? Impact of AI Assistance on Incidental Learning. In *Proceedings of the 27th International Conference on Intelligent User Interfaces (IUI '22)*. ACM Press, New York, NY, 13 pages.

[33] Laura Germine, Ken Nakayama, Bradley C Duchaine, Christopher F Chabris, Garga Chatterjee, and Jeremy B Wilmer. 2012. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic bulletin & review* 19, 5 (2012), 847–857.

[34] Carol L. Glasser, Belinda Robnett, and Cynthia Feliciano. 2009. Internet Daters' Body Type Preferences: Race-Ethnic and Gender Differences. *Sex Roles* 61, 1-2 (July 2009), 14–33. https://doi.org/10.1007/s11199-009-9604-x

[35] Ben Green. 2020. Data Science as Political Action: Grounding Data Science in a Politics of Justice. arXiv:1811.03435 [cs.CY]

[36] Ben Green and Yiling Chen. 2019. Disparate Interactions: An Algorithm-in-the-Loop Analysis of Fairness in Risk Assessments. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*. ACM, Atlanta GA USA, 90–99. https://doi.org/10.1145/3287560.3287563

[37] Ben Green and Yiling Chen. 2019. The Principles and Limits of Algorithm-in-the-Loop Decision Making. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–24. https://doi.org/10.1145/3359152

[38] Anthony G. Greenwald and Mahzarin R. Banaji. 1995. Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review* 102, 1 (1995), 4–27. https://doi.org/10.1037/0033-295X.102.1.4

[39] Anthony G. Greenwald and Linda Hamilton Krieger. 2006. Implicit Bias: Scientific Foundations. *California Law Review* 94, 4 (July 2006), 945. https://doi.org/10.2307/20439056

[40] Grindr. 2021. Grindr for equality. https://www.grindr.com/g4e/

[41] William J. Hall, Mimi V. Chapman, Kent M. Lee, Yesenia M. Merino, Tainayah W. Thomas, B. Keith Payne, Eugenia Eng, Steven H. Day, and Tamera Coyne-Beasley. 2015. Implicit Racial/Ethnic Bias Among Health Care Professionals and Its Influence on Health Care Outcomes: A Systematic Review. *American Journal of Public Health* 105, 12 (Dec. 2015), e60–76. https://doi.org/10.2105/AJPH.2015.302903

[42] Doron Halutz. 2018. Israeli founder of Grindr talks about growing up gay and coming out to his family. https://www.haaretz.com/israel-news/culture/the-israeli-loins-behind-grindr-1.5253351

[43] Chong-suk Han. 2008. No fats, femmes, or Asians: the utility of critical race theory in examining the role of gay stock stories in the marginalization of gay Asian men. *Contemporary Justice Review* 11, 1 (March 2008), 11–22. https://doi.org/10.1080/10282580701850355 Publisher: Routledge.

[44] Chong-suk Han. 2007. They Don't Want To Cruise Your Type: Gay Men of Color and the Racial Politics of Exclusion. *Social Identities* 13, 1 (Jan. 2007), 51–67. https://doi.org/10.1080/13504630601163379

[45] Chong-suk Han and Kyung-Hee Choi. 2018. Very few people say "No Whites": Gay men of color and the racial politics of desire. *Sociological Spectrum* 38, 3 (2018), 145–161. https://doi.org/10.1080/02732173.2018.1469444 Place: United Kingdom Publisher: Taylor & Francis.

[46] Alex Hanna, Emily Denton, Andrew Smart, and Jamila Smith-Loud. 2020. Towards a critical race methodology in algorithmic fairness. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. ACM, Barcelona Spain, 501–512. https://doi.org/10.1145/3351095.3372826

[47] Jean Hardy and Silvia Lindtner. 2017. Constructing a Desiring User: Discourse, Rurality, and Design in Location-Based Social Networks. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, Portland Oregon USA, 13–25. https://doi.org/10.1145/2998181.2998347

[48] Tanja Hentschel, Madeline E. Heilman, and Claudia V. Peus. 2019. The multiple dimensions of gender stereotypes: A current look at men's and women's characterizations of others and themselves. *Frontiers in Psychology* 10, JAN (jan 2019), 11. https://doi.org/10.3389/fpsyg.2019.00011

[49] Philipp Hergovich and Josué Ortega. 2017. The strength of absent ties: Social integration via online dating. https://doi.org/10.2139/ssrn.3044766 arXiv:1709.10478

[50] Bernd Huber and Krzysztof Z Gajos. 2020. Conducting online virtual environment experiments with uncompensated, unsupervised samples. *Plos one* 15, 1 (2020), e0227629.

[51] Ben Hunte. 2020. Grindr fails to remove ethnicity filter after pledge to do so. https://www.bbc.com/news/technology-53192465

[52] Jevan A. Hutson, Jessie G. Taft, Solon Barocas, and Karen Levy. 2018. Debiasing Desire: Addressing Bias & Discrimination on Intimate Platforms. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 1–18. https://doi.org/10.1145/3274342

[53] Eva Illouz. 2007. *Cold intimacies: the making of emotional capitalism.* Polity Press, Cambridge, UK ; Malden, MA. OCLC: ocn123441703.

[54] Jerry A. Jacobs and Teresa G. Labov. 2002. Gender Differentials in Intermarriage Among Sixteen Race and Ethnic Groups. *Sociological Forum* 17, 4 (Dec. 2002), 621–646. https://doi.org/10.1023/A:1021029507937

[55] Samantha Joel, Paul W. Eastwick, and Eli J. Finkel. 2017. Is Romantic Desire Predictable? Machine Learning Applied to Initial Romantic Attraction. *Psychological Science* 28, 10 (Oct. 2017), 1478–1489. https://doi.org/10.1177/0956797617714580

[56] Tiffani J. Johnson, Robert W. Hickey, Galen E. Switzer, Elizabeth Miller, Daniel G. Winger, Margaret Nguyen, Richard A. Saladino, and Leslie R. M. Hausmann. 2016. The Impact of Cognitive Stressors in the Emergency Department on Physician Implicit Racial Bias. *Academic Emergency Medicine* 23, 3 (March 2016), 297–305. https://doi.org/10.1111/acem.12901

[57] Daniel Kahneman. 2011. *Thinking, fast and slow.* Farrar, Straus and Giroux, New York.

[58] Grace Kao, Kelly Stamper Balistreri, and Kara Joyner. 2018. Asian American Men in Romantic Dating Markets. *Contexts* 17, 4 (Nov. 2018), 48–53. https://doi.org/10.1177/1536504218812869

[59] Emil OW Kirkegaard and Julius D Bjerrekær. 2016. The OKCupid dataset: A very large public dataset of dating site users. *Open Differential Psychology* 46 (2016), 1–10.

[60] Alexis Kleinman. 2017. Black People And Asian Men Have A Much Harder Time Dating On OKCupid. https://www.huffingtonpost.ca/entry/okcupid-race_n_5811840?ri18n=true

[61] Helen Kolawole. 2004. Why I'll only date black men. https://www.theguardian.com/world/2004/jul/31/race.uk

[62] Marlene Kollmayer, Barbara Schober, and Christiane Spiel. 2018. Gender stereotypes in education: Development, consequences, and interventions. *European Journal of Developmental Psychology* 15, 4 (jul 2018), 361–377. https://doi.org/10.1080/17405629.2016.1193483

[63] Kathryn Ann Lambe, Gary O'Reilly, Brendan D. Kelly, and Sarah Curristan. 2016. Dual-process cognitive interventions to enhance diagnostic reasoning: A systematic review. https://doi.org/10.1136/bmjqs-2015-004417

[64] K. Lewis. 2013. The limits of racial prejudice. *Proceedings of the National Academy of Sciences* 110, 47 (Nov. 2013), 18814–18819. https://doi.org/10.1073/pnas.1308501110

[65] Richard C. Lewontin, Steven P. R. Rose, and Leon J. Kamin. 1984. *Not in our genes: biology, ideology, and human nature* (1st ed ed.). Pantheon Books, New York.

[66] Qisheng Li, Krzysztof Z. Gajos, and Katharina Reinecke. 2018. Volunteer-Based Online Studies With Older Adults and People with Disabilities. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) *(ASSETS '18)*. ACM, New York, NY, USA, 229–241. https://doi.org/10.1145/3234695.3236360

[67] Qisheng Li, Sung Jun Joo, Jason D Yeatman, and Katharina Reinecke. 2020. Controlling for participants' Viewing Distance in Large-Scale, psychophysical online experiments Using a Virtual chinrest. *Scientific Reports* 10, 1 (2020), 1–11.

[68] Calvin Liang, Jevan Hutson, and Os Keyes. 2020. Surveillance, Stigma & Sociotechnical Design for HIV. arXiv:2006.04882 [cs.HC]

[69] Milena Ribeiro Lopes and Carl Vogel. 2017. Women's perspective on using tinder: A user study of gender dynamics in a mobile device application. In *SIGDOC 2017 - 35th ACM International Conference on the Design of Communication*. Association for Computing Machinery, Inc, New York, New York, USA, 1–10. https://doi.org/10.1145/3121113.3121220

[70] Elizabeth Aura McClintock. 2010. When Does Race Matter? Race, Sex, and Dating at an Elite University. *Journal of Marriage and Family* 72, 1 (Feb. 2010), 45–72. https://doi.org/10.1111/j.1741-3737.2009.00683.x

[71] Allen R. McConnell and Jill M. Leibold. 2001. Relations among the Implicit Association Test, Discriminatory Behavior, and Explicit Measures of Racial Attitudes. *Journal of Experimental Social Psychology* 37, 5 (Sept. 2001), 435–442. https://doi.org/10.1006/jesp.2000.1470

[72] Alan Nelson. 2002. Unequal treatment: confronting racial and ethnic disparities in health care. *Journal of the national medical association* 94, 8 (2002), 666.

[73] Katie Notopoulos. 2016. The dating app that knows you secretly aren't into guys from other races. https://www.buzzfeednews.com/article/katienotopoulos/coffee-meets-bagel-racial-preferences

[74] Josue Ortega and Philipp Hergovich. 2017. The Strength of Absent Ties: Social Integration via Online Dating. https://doi.org/10.2139/ssrn.3044766

[75] Generated Photos. 2021. Unique, worry-free model photos. https://generated.photos/

[76] David Pilgrim. 2000. What was jim crow. *Ferris State University* 16 (2000), 2007.

[77] Joshua R. Polanin and Birte Snilstveit. 2016. Converting between effect sizes. *Campbell Systematic Reviews* 12, 1 (Jan. 2016), 1–13. https://doi.org/10.4073/cmpn.2016.3

[78] Maurice Kwong-Lai Poon and Peter Trung-Thu Ho. 2008. Negotiating Social Stigma Among Gay Asian Men. *Sexualities* 11, 1-2 (Feb. 2008), 245–268. https://doi.org/10.1177/1363460707085472

[79] Matthew H. Rafalow, Cynthia Feliciano, and Belinda Robnett. 2017. Racialized Femininity and Masculinity in the Preferences of Online Same-sex Daters. *Social Currents* 4, 4 (Aug. 2017), 306–321. https://doi.org/10.1177/2329496516686621

[80] Katharina Reinecke and Krzysztof Z. Gajos. 2015. LabintheWild: Conducting Large-Scale Online Experiments With Uncompensated Samples. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Vancouver, BC, Canada) *(CSCW '15)*. Association for Computing Machinery, New York, NY, USA, 1364–1378. https://doi.org/10.1145/2675133.2675246

[81] Ernesto Reuben, Paola Sapienza, and Luigi Zingales. 2014. How stereotypes impair women's careers in science. *Proceedings of the National Academy of Sciences of the United States of America* 111, 12 (mar 2014), 4403–4408. https://doi.org/10.1073/pnas.1314788111

[82] Wendy Reynolds-Dobbs, Kecia M. Thomas, and Matthew S. Harrison. 2008. From mammy to superwoman: Images that Hinder Black women's career development. *Journal of Career Development* 35, 2 (dec 2008), 129–150. https://doi.org/10.1177/0894845308325645

[83] Neal J. Roese and Michael W. Morris. 1999. Impression valence constrains social explanations: The case of discounting versus conjunction effects. *Journal of Personality and Social Psychology* 77, 3 (1999), 437–448. https://doi.org/10.1037/0022-3514.77.3.437

[84] Renee Romano. 2018. Something Old, Something New: Black Women, Interracial Dating, and the Black Marriage Crisis. *differences* 29, 2 (Sept. 2018), 126–153. https://doi.org/10.1215/10407391-6999802

[85] Michael J. Rosenfeld and Reuben J. Thomas. 2012. Searching for a Mate: The Rise of the Internet as a Social Intermediary. *American Sociological Review* 77, 4 (Aug. 2012), 523–547. https://doi.org/10.1177/0003122412448050

[86] Robert Schoen and Nicola Standish. 2001. The Retrenchment of Marriage: Results from Marital Status Life Tables for the United States, 1995. *Population and Development Review* 27, 3 (Sept. 2001), 553–563. https://doi.org/10.1111/j.1728-4457.2001.00553.x

[87] Jiannbin Lee Shiao and Mia H. Tuan. 2008. "SOME ASIAN MEN ARE ATTRACTIVE TO ME, BUT FOR A HUSBAND ...". *Du Bois Review: Social Science Research on Race* 5, 02 (Sept. 2008), 259. https://doi.org/10.1017/S1742058X08080132

[88] Eliot R. Smith and Jamie DeCoster. 2000. Dual-Process Models in Social and Cognitive Psychology: Conceptual Integration and Links to Underlying Memory Systems. *Personality and Social Psychology Review* 4, 2 (2000), 108–131. https://doi.org/10.1207/S15327957PSPR0402_01

[89] D. A. Stanley, P. Sokol-Hessner, M. R. Banaji, and E. A. Phelps. 2011. Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences* 108, 19 (May 2011), 7710–7715. https://doi.org/10.1073/pnas.1014345108

[90] Lindsay Steenberg. 2009. Book Review: The Hypersexuality of Race: Performing Asian/American Women on Screen and Scene. *Feminist Review* 92, 1 (July 2009), 172–173. https://doi.org/10.1057/fr.2009.11

[91] Ann Laura Stoler. 2010. *Carnal knowledge and imperial power: race and the intimate in colonial rule ; with a new preface*. Univ. of California Press, Berkeley, Calif.

[92] Patrick Strudwick. 2016. The Founders Of This Gay Dating App Won't Stop You Searching By Race. https://www.buzzfeednews.com/article/patrickstrudwick/this-is-how-gay-dating-app-bosses-defend-racial-filtering

[93] Magdalena Szumilas. 2010. Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal De l'Academie Canadienne De Psychiatrie De L'enfant Et De L'adolescent* 19, 3 (Aug. 2010), 227–229.

[94] Taffy. 2017. Taffy: The chat-first way to meet new people. http://taffy.chat/.

[95] H. Tankovska. 2021. U.S. dating apps by audience size 2019. https://www.statista.com/statistics/826778/most-popular-dating-apps-by-audience-size-usa/#:~:text=As%20of%20September%202019%2C%20Tinder,5.03%20million%20U.S.%20mobile%20users.

[96] Tantan. 2020. Tantan. https://tantanapp.com/

[97] Tinder Team. 2020. Powering Tinder® - The Method Behind Our Matching. https://blog.gotinder.com/powering-tinder-r-the-method-behind-our-matching/

[98] Tinder. 2021. Dating, Make Friends & Meet New People. https://tinder.com/

[99] Laurence Watts. 2016. Gay men and women are not more racist. https://www.huffpost.com/entry/gay-racism_b_1307819

[100] Chongyi Wei and H. Fisher Raymond. 2011. Preference for and Maintenance of Anal Sex Roles Among Men Who Have Sex with Men: Sociodemographic and Behavioral Correlates. *Archives of Sexual Behavior* 40, 4 (Aug. 2011), 829–834. https://doi.org/10.1007/s10508-010-9623-2

[101] Langdon Winner. 1980. Do Artifacts Have Politics? *Daedalus* 109, 1 (1980), 121–136. http://www.jstor.org/stable/20024652