

Historical Trends in Predicting Interracial Behaviors: A Meta-Analysis

Mark H. White II¹

¹ University of Kansas

Author Note

Thank you to Dr. Elizabeth Tipton and Dr. Wolfgang Viechtbauer for their advice in conducting and communicating the results of the meta-analysis, as well as for their essential R packages, `robumeta` and `metafor`, respectively. The Appendix can be found at <https://github.com/markhwhiteii/prej-beh-meta>.

Correspondence concerning this article should be addressed to Mark H. White II.
E-mail: markhwhiteii@gmail.com

Abstract

Researchers often note how self-reports of prejudice have decreased over time, yet evidence of discrimination still persists. Have these explicit measures of interracial attitudes gotten worse at predicting behaviors over time? Contrary to predictions, weak evidence was found for this to be the case. However, this lack of relationship was largely attributable to increasing subtlety of behavioral measures in psychological studies. Meta-regression analyses found that explicit measures predict subtle behaviors worse than more obvious and controllable ones. The current analysis improved upon statistical flaws in past meta-analyses on the topic, contradicting some previous findings. Most notably, affective measures were not stronger predictors of behavior than cognitive measures—they both had the same predictive power. A relatively small overall meta-analytic correlation between interracial attitudes and behavior was found, $r = .21$; however, I argue that psychologists should focus on including more predictors of both person and environment as well as more flexible statistical learning techniques in the focus on predicting prejudiced behavior.

Historical Trends in Predicting Interracial Behaviors: A Meta-Analysis

One of the most common introductions in articles on contemporary forms of prejudice is to note the paradox that self-reports of prejudice have decreased in the recent decades, yet inequality and discrimination continue (e.g., Crosby, Bromley, & Saxe, 1980; Dovidio & Gaertner, 2000, 2004; Frey & Gaertner, 1986; Henry & Sears, 2002; Pearson, Dovidio, & Gaertner, 2009). For example: “Over the past 40 years, public opinion polls have revealed substantial declines in Whites’ endorsement of prejudiced views toward minority groups, and Blacks in particular, in the United States. . . . Despite these declines in overt prejudice, however, evidence of substantial racial disparities and discrimination remain. . . .” (Pearson et al., 2009, p. 315). Many social psychological theories have tried to explain this contradiction with “two-factor theories” of prejudice (Crandall & Eshleman, 2003), including modern (McConahay, Hardee, & Batts, 1981), ambivalent (Katz & Hass, 1988), and aversive racism (Dovidio & Gaertner, 2004). These theories all propose that (a) people continue to feel and want to express and act on their prejudices, but (b) are motivated to keep themselves from doing so for either internal (e.g., personal standards) or external (e.g., social pressure) reasons (Crandall & Eshleman, 2003). Thus, decreases in self-reported prejudice over time are not necessarily decreases in genuine attitudes, so discriminatory behaviors continue.

Researchers across disciplines and settings have examined this contradiction of contemporary racial attitudes, but psychological research provides a unique way to track interracial attitudes and behaviors over time. Academic journals and institutions have kept roughly the same structure for publishing research in the last few decades, providing a consistent record of these relationships over time. Despite the limitations of academic research—such as non-representative samples or contrived laboratory situations—it is a rather valuable source of data, as many researchers have measured interracial attitudes and behaviors within the *same person*, something that data from other institutions and fields cannot speak to.

If the endorsement of negative interracial attitudes is decreasing at a faster pace than

are the corresponding behaviors, an individual's explicit attitudes should be *less* predictive of *that same person's* interracial behaviors today than they were decades ago. Using psychological research from the past 70 years, I will take a meta-analytic approach to test this hypothesis that self-reported racial attitudes have worsened in predictive validity of interracial behaviors over the history of the field.

Trends in Interracial Attitudes

It is essentially axiomatic (and taken for granted, even among scholars) that the tenor of interracial attitudes has grown more positive over time. But is this really the case? The American National Election Studies (ANES; see American National Election Studies, n.d.) and General Social Survey (GSS; see Bobo, Charles, Krysan, & Simmons, 2012) provide nationally representative polling data from the last 60 years that illustrate the complexity of racial attitude trends. Given the history of race relations in America, these data focus on the attitudes of White Americans toward Black Americans.

Support for discriminatory policies, negative responses to social distance and emotional measures, and endorsement of negative stereotypes have all decreased since the 1940s. Favoring segregation in general (polling at 5% in 1978) and keeping Black people out of White neighborhoods (9% in 1976) were items eventually dropped by the ANES; the GSS dropped an item in the 1980s about keeping Black children in separate schools because so few people endorsed it, as well. Social distance items (e.g., having dinner with a Black person) have shown promising improvements. For example, the GSS data show a decrease in opposition to racial intermarriage from 1990 (about 65%) to 2008 (about 25%), and there was an increase in how many Whites reported feeling equally close to Whites and Blacks from 1996 (about 40%) to 2008 (about 50%). Racial stereotypes were not even included at the beginning of the GSS due to a rapid drop in endorsement of anti-Black stereotypes from the 1940s to the 1970s. However, questions asking whether or not Whites are more intelligent and hardworking than Blacks have been included since 1990; there has been a

drop from about 60% to 35% in endorsement of this belief from 1990 to 2008. White Americans are less likely today than decades ago to openly endorse many of these negative attitudes traditionally referred to as “prejudice.”

Not all racial attitudes have improved, however, particularly those about actively promoting Black prosperity—as opposed to just rejecting discriminatory policies—through government programs. According to the ANES, believing the government should (a) *not* help minorities improve their socioeconomic position in society and (b) stay out of ensuring fair treatment for minorities in the job market have *increased* somewhat since the 1960s. The GSS data also show unchanging trends for similar items: The government should *not* give Blacks special treatment to help make up for past discrimination (50-60% range from 1973 to 2007), opposing preferential hiring and promotion of Blacks due to past discrimination (about 90% from 1994 to 2008), and believing that Blacks should work their way up in society without special favors from the government (about 80% from 1994 to 2008). Bobo et al. (2012) also note from the GSS data that an increased number of people over time have been responding to race-related questions with “unsure” or “don’t know,” which is consistent with people suppressing prejudice more over time.

White Americans clearly do not hold the same negative attitudes about racial minorities today as 60 years ago; however, they do not seem to be moving toward supporting concrete policies that would actively address the legacy of racism and discrimination (McConahay et al., 1981). Despite this critical exception, the majority of polling research shows a decline over time in most forms of explicit negative racial measures used by psychologists. Why have these changes occurred?

Norms and Interracial Attitudes

Psychologists and sociologists agree that the decrease in explicitly expressed racial attitudes is due to the change in normative acceptability of openly expressing prejudice since World War II and the Civil Rights Movement (Quillian, 2006). Social psychologists have

argued for more than 70 years that social norms are one of the strongest causes of expressing—or not expressing—prejudice (Crandall & Stangor, 2005). Crandall, Eshleman, and O'Brien (2002) demonstrated that the appropriateness and expression of prejudice are essentially the same construct. Crandall and colleagues asked one group of college students to indicate how much they disliked 105 different targets of prejudice (e.g., Black people, alcoholics, lawyers, police officers) and a separate group of students to indicate how “OK” it is to “feel negative feelings” about each of the groups. Expression and acceptability of prejudice were nearly perfectly correlated, $r = .96$; Cary and Page-Gould (2014) replicated this finding, observing the exact same point estimate. A number of other studies have corroborated this experimentally in the lab and field; making prejudice normative increases its expression, while making it counter-normative decreases expression (Blanchard, Crandall, Brigham, & Vaughn, 1994; Ford, Boxer, Armstrong, & Edel, 2008; Monteith, Deneen, & Tooman, 1996; Paluck, 2011; Sechrist & Stangor, 2001; Stangor, Sechrist, & Jost, 2001; Zitek & Hebl, 2007).

But norms are not the panacea for all of prejudice's ills, and the slope of increasingly positive interracial attitudes is not as linear—or as steep—as we would like it to be. Each step toward racial equality in the United States has provoked a White backlash, from the Black Codes that followed the 13th Amendment to the voter suppression efforts targeted at people of color after Barack Obama's election (Anderson, 2016). Normative pressure can also backfire, causing people to express or re-articulate their prejudices (Lalonde, Doan, & Patterson, 2000; Legault, Gutsell, & Inzlicht, 2011; Omi & Winant, 1994; Plant & Devine, 2001; M. H. White II & Crandall, 2017).

Nevertheless, changing norms about interracial attitudes have limited the expression of traditional forms of prejudice. As these norms have grown stricter over time, I am using the year in which a study was published as a proxy for the normative climate. The primary hypothesis of this paper can thus be rephrased: As norms have gotten more strict against expressing prejudice (i.e., as time has passed), the predictive validity (i.e., correlation

between attitude and behavior) of interracial attitude measures has decreased.

Trends in Interracial Behaviors

There is evidence for the first part of the paradox of contemporary prejudice: Interracial attitudes are improving (however limited) over time. Is there evidence for the second part—that prejudiced behaviors have *not* improved as much over that same time period? Researchers cannot tell if people really “mean it” when respondents express non-prejudiced attitudes, as it is impossible to measure prejudice *directly*—in a way that is immune to social desirability, unconscious cultural knowledge, and so on (Crandall & Eshleman, 2003). When contemporary studies of prejudice do not find much of a link between attitudes and behaviors, researchers infer that the changes in interracial attitudes do not necessarily represent changes in one’s “true” attitudes. Authors often use this contradiction of contemporary racial attitudes to reel readers in, but are these introductions more catchy than they are veridical? Bobo et al. (2012) specifically and unequivocally address the issue: “...the all-to-common sociological assertion that the attitudinal record paints a purely and unduly optimistic picture of race relations at odds with actual behavioral data on segregation, inequality, and discrimination is simply wrong” (p. 73). What does the evidence say?

The Federal Bureau of Investigation (FBI) yearly hate crime reports show that the number of race- and ethnicity-related hate crimes have decreased since 1996, the first year that official data were collected (Middlebrook, 2017), while the Southern Poverty Law Center (SPLC) has counted an increase in the number of hate group chapters since 1999 (Potok, 2016). These sources each have their own biases: Law enforcement agencies voluntarily report hate crimes to the FBI, while the SPLC relies on news, media, and voluntary reports from individuals about both hate crimes and non-criminal incidents (Majumder, 2017). It is unclear if extreme discriminatory behaviors are decreasing, increasing, or being concentrated at fringes of society (i.e., hate groups).

Considering less extreme behaviors, increases in endorsement of interracial marriage have corresponded with actual rates of interracial marriage increasing (Pew Research Center, 2012); similarly, as self-reported support for residential segregation has dropped, field experiments on racial discrimination in housing markets and real-estate agent behavior have found less discrimination in 2000 than in 1989 (Quillian, 2006). Toosi, Babbitt, Ambady, and Sommers (2012) performed a meta-analysis of dyadic interracial interactions, finding that people interacting with someone of their own race expressed more positive attitudes, more friendly nonverbal behavior, feeling less negative affect, and scored higher on performance tasks than those interacting with someone of a different race. They found that these effect sizes have been shrinking over time, although the trend was only significant with attitudes toward the partner and nonverbal behavior. On the other hand, Saucier, Miller, and Doucet (2005) conducted a meta-analysis on interracial helping behavior experiments; looking at 48 effect sizes from 31 articles published since 1967, the authors found *no* relationship between publication year and discrimination in helping behavior.

Institutional-level disparities have also been thoroughly researched. These inequalities, while not usually considered at the individual level like psychological studies, are still relevant; while the causes of these disparities can exist absent negative behaviors of individuals (Bloome, 2014; Traub & Ruetschlin, 2016), there are nonetheless individuals acting on behalf of institutions making decisions that negatively harm racial outgroups (e.g., teachers, bankers, doctors). For example, the United States Department of Justice found that Wells Fargo bankers led Black and Hispanic borrowers—due to their race—into sub-prime mortgages and paying higher fees in the mid- to late-2000s (U.S. Department of Justice, 2012). Racial inequalities in high school completion, life expectancy, and voter turnout have decreased over time; however, disparities in unemployment, poverty, household income, household wealth, home ownership, and incarceration have remained largely consistent for decades (Economic Policy Institute, 2013; Pew Research Center, 2016; U.S. Census Bureau, 2002; Urban Institute, 2015). The evidence reviewed here is quite mixed; the

objective of the current meta-analysis is to add to this body of research by considering the relationship between attitudes and behaviors within individuals.

How Has Studying Racial Attitudes Changed?

The ways that researchers study prejudice are, of course, subject to the same changing normative environment regarding prejudice expression as participants and respondents in studies and polls. Decisions about what prejudices to study and how to study them reflect the milieu in which the researcher lives; researchers often study what they believe to be wrong, but commonly felt among people (Crandall, Ferguson, & Bahns, 2013). As certain ways of expressing prejudice diminish and other ways emerge, researchers have changed how they study and measure prejudice; Biernat and Crandall (1999) note a large number of scales being published during the 1990s in response to a changing normative climate about racial attitudes and the subsequent social desirability bias of respondents.

Consider Bray (1950), one of the earliest studies in the present meta-analysis. He uses a scale developed by Likert (1932) that contains an item which asks the participant when they think lynching is justifiable, and another item contains the n-word. Many researchers would not consider putting this language in front of participants, and some participants would justifiably react strongly to the language in the items. This language is far different from the items we generally ask participants now, which focus on attitudes toward affirmative action policies and attributions for why Blacks might not be as well-off as Whites. This change of measurement tools is valuable: Researchers are responding to the social climate in real-time, publishing scales that capture how racial attitudes are expressed contemporarily, ensuring that prejudice can still be accurately measured. This poses a problem for the present meta-analysis, however, as the hypothesized negative relationship between publication year and effect size could be absent simply because researchers are skilled in keeping up with the trends of prejudice expression. I return to this point in the Discussion.

Perhaps the greatest change in the study of racial attitudes has been the intense focus

on measuring automatic (Devine, 1989), implicit (Greenwald & Banaji, 1995) racial attitudes, particularly since the creation of the Implicit Association Test (Greenwald, McGhee, & Schwartz, 1998). There has been such a concentration on implicit measures that, in under two decades, there are already various meta-analysis investigating these measurement tools (e.g., Greenwald, Poehlman, Uhlmann, & Banaji, 2009; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005). In just 14 years, Oswald, Mitchell, Blanton, Jaccard, and Tetlock (2013) located over 300 correlations involving race-related IAT scores across 86 samples in 46 reports. Meta-analyses and citation counts do not capture the full movement towards focusing on implicit attitudes over explicit attitudes. “Unconscious biases” have been widely discussed in popular culture—and not just the usual outlets for this type of research (e.g., *The New York Times*, *The Atlantic*, *Vox*). Companies like Google and Microsoft are implementing programs to combat “unconscious bias,” leading the research to be discussed in *Forbes* (Huet, 2015), *Business Insider* (Feloni, 2016), *Fortune* (Olson, 2015), and *The Wall Street Journal* (Van Bavel & West, 2017). Even far-right media outlets have joined in on the discussion, generally lambasting the concept altogether (e.g., French, 2017; Hayward, 2016).

Despite this focus on implicit measures, I will only focus on explicit attitudes in the present meta-analysis. The primary research question—has attitude-behavior correspondence in the interracial domain decreased over time?—is based on the idea that, due to the changing normative climate, people are suppressing their overt expressions quicker than they are changing their genuine attitudes and corresponding behaviors (Crandall & Eshleman, 2003). Implicit attitudes are not as easily controlled and thus not subject to the theoretical phenomenon of interest. Special attention must be paid to this research, nonetheless: Implicit social cognition studies tend to focus on subtler behaviors, while also including measures of explicit attitudes. Various theories predict that implicit attitudes correspond with behaviors that are hard to control (e.g., eye contact in an interracial interaction), while explicit attitudes correspond with behaviors that are controllable (e.g., aggression against someone of another race; for a review, see Frieze, Hofmann, & Schmitt, 2009). Some studies,

therefore, are published to demonstrate that explicit attitudes specifically do *not* predict certain behaviors. Given that these studies are newer, this means more recent studies might have weaker effects simply because the behaviors are more subtle (e.g., time spent smiling during interracial interaction) and have more measurement error (i.e., subtle behaviors are harder to measure); a negative relationship between effect size and time then might be simply due to the changes in research design. I address by coding for subtlety of behavior and type of attitude measurement.

Previous Reviews and Meta-Analyses

Attitude-behavior correspondence (ABC) has long been an interest to social psychologists. There are a number of previous meta-analyses that focus on ABC across fields of study (e.g., intergroup relations, political science, consumer psychology, environmental studies; Armitage & Conner, 2001; Cooke & Sheeran, 2004; Glasman & Albarracín, 2006; Kim & Hunter, 1993; Kraus, 1995; Notani, 1998; Sheppard, Hartwick, & Warshaw, 1988; Wallace, Paulson, Lord, & Bond Jr., 2005). Using various different inclusion criteria and statistical analyses, these reviews have contained meta-analytic correlation coefficients ranging from $r = .21$ (Notani, 1998) to $r = .52$ (Glasman & Albarracín, 2006). Given this heterogeneity, most meta-analyses focus on moderators of ABC; some do not even report the overall meta-analytic correlation (e.g., Armitage & Conner, 2001; Cooke & Sheeran, 2004; Sheppard et al., 1988). Consistent across meta-analyses are findings that ABC is higher when: attitudes are easily retrieved from memory, conscious thought goes into attitudes, attitudes are stable over time, attitudes rely on information relevant to the behavior, attitudinal ambivalence is low and certainty is high, and attitudes are experimentally generated in the laboratory—that is, are *not* preexisting attitudes (e.g., Cooke & Sheeran, 2004; Glasman & Albarracín, 2006).

Authors generally use two theories to frame these relationships: the theory of planned behavior (Ajzen, 1991) and the MODE model (Fazio, 1990), which both argue that one's

willingness and ability to act on their attitudes increase ABC. The MODE adds that if people are unmotivated or unable to cognitively process relevant information about their attitudes and the behavioral situation, then strongly-held, readily-accessible attitudes will automatically guide behavior. Particularly relevant to this meta-analysis is Ajzen's contention that perceived social norms affect intention to behave, which in turn affects actual behavior; the more normative one perceives an attitude to be, the higher their motivation is to perform the behavior, which in turn positively predicts acting in accordance with those attitudes (Armitage & Conner, 2001). My hypothesis that the correlation between interracial attitudes and behaviors should diminish over time is in line with this aspect of the theory of planned behavior, as the publication year is a proxy for subjective norms.

I am predicting that ABC in the *interracial* domain will shrink as time passes. A useful comparison might be historical ABC trends for attitudes in *general*: Kraus (1995) found that ABC was positively correlated with publication year $r = .21$, but was non-significant after controlling for behavior type, $r = .11$; Glasman and Albarracín (2006), in a meta-analysis of only experimentally-formed attitudes, found that articles published since 1996 had a significantly larger ABC ($r = .55$) than those before that year ($r = .40$).

ABC has received meta-analytic attention within intergroup relations specifically, as well: Two meta-analyses on intergroup relations generally (Schütz & Six, 1996; Talaska, Fiske, & Chaiken, 2008), while others have a particular focus on implicit intergroup attitudes, which are outside the scope of this paper (e.g., Oswald et al., 2013). Schütz and Six (1996) found an overall correlation between attitudes and behavior of $r = .29$, but noted that the sizes varied across setting, behavioral measurement type, sample, etc. Their analysis is descriptive in nature; they do not make a firm theoretical position other than stating effect sizes vary and seem to be larger in situations when participants are more in control of their behavior. They urge researchers to not focus on *if* intergroup attitudes predict intergroup behaviors—it is clear they do—but *when* attitudes and behavior are consistent versus inconsistent. Talaska et al. (2008) take up this task by focusing primarily on moderators of

interracial attitude and behavior correspondence; their main argument is that ABC in the interracial domain is higher when attitude measures focus on emotions, not beliefs. They also find that year somewhat negatively predicted effect size in their model, $\beta = -.17, p = .07$. Talaska and colleagues, however, treated this finding as a methodological covariate and do not discuss the possible theoretical implications of the finding. Given Talaska et al. (2008), why is the present meta-analysis necessary to test the relationship?

First, I constructed a data set that was designed to test the year hypothesis specifically. Talaska et al. (2008) include implicit and indirect measures of interracial attitudes; these measures are not as easily-controlled and thus not subject to the normative pressure that is the key phenomenon at play behind the year effect. I do not include effects based on these measures. Second, Talaska and colleagues selected peer-reviewed studies published through 2002; I extended this by including dissertations and theses, as well as including studies published through the end of 2016. I did *not* solicit unpublished or “file-drawer” works from contemporary researchers. Since I am interested in historical effects, I did not want to use a method of searching for a study that I could not apply to the entire historical range (i.e., I cannot go back in time and ask researchers from the 1970s for their unpublished works). Third, I attempt to improve on both methodological and theoretical aspects of the review (see Method section).

The current meta-analysis will test the hypothesis that, due to normative change in recent decades, the ability for explicit measures of interracial attitudes to predict interracial behaviors has decreased; this will be supported if year negatively predicts the effect size of the prediction. I also aim to extend and improve upon previous meta-analyses on ABC in the interracial domain. I account for the dependencies between effect sizes using a robust variance estimation meta-regression technique (Hedges, Tipton, & Johnson, 2010). I will also try to conceptually replicate Talaska et al. (2008) by testing to see if affective measures are better predictors than cognitive ones.

Method

Inclusion Criteria

I sought out studies that measured both attitudes and behaviors toward a racial outgroup. For the purpose of this meta-analysis, *interracial attitudes* are explicitly expressed feelings, thoughts, behavioral intentions, or other evaluations toward a racial outgroup, a specific racial outgroup member, or policies that pertain to a racial outgroup.

I did not include implicit measures due to my interest in the expression and suppression of prejudice over time; implicit measures are not as easily fooled by suppression of prejudice as explicit measures, and these measures have been used only in the last two decades. Race-related beliefs that were about the ingroup (e.g., perception of privilege, racial identification), the self (e.g., intrinsic and extrinsic motivation to suppress prejudice), or society (e.g., multicultural or colorblind ideologies, perceptions of norms) were not included; these beliefs do not focus on the racial outgroup. Attitudes that serve as manipulation checks were not included, because they are likely to be more influenced by the experimental manipulation than an individual's psychological processes.

Interracial behaviors are any actions toward a racial outgroup member or racial outgroup that are—or perceived by the participant to be—consequential (see Talaska et al., 2008). I determined that the behaviors were inconsequential if the authors described the behavior as “mock,” “fantasy,” “fictitious,” or “hypothetical,” or if participants were instructed to act “as if” or “to imagine” they were in a situation. For situations where consequentiality was unclear, I erred on the side of deciding the behavior was not consequential, and I excluded the study; I reasoned if participants were given a story to make them feel that, for example, their judgments of a job applicant actually affected someone's chance of getting the job, then the authors would likely include that key information in the manuscript.

Example behaviors included were: helping an outgroup member, signing a petition for an outgroup, donating money to an outgroup cause, acting aggressively toward an outgroup

member, nonverbal behaviors in an interracial interaction (e.g., smiling, seating distance), and the target's or a confederate's overall perception of the participant in an interracial interaction.

Self-reports of past behavior were included under certain circumstances. The vast majority of these instances were participants reporting past interracial contact. Many studies look for the causal link from contact to (a decrease in) prejudice; I did not want to include this type of research, because I am interested in the *opposite* causal relationship. Self-reports were thus only included if they were measured at the same time as the attitude. Longitudinal studies were included, but the effect sizes taken from these studies were between self-reported attitudes and self-reported behaviors at the *same* measurement point.

I am only interested, as well, in voluntary contact. I only included self-reports that were explicitly voluntary contact: If the measurement tool asked if participants invited Black people into their home, how much leisure time they spent with Black people, etc., then I included the past reported contact. But contact was not included if it was likely involuntary—such as contact at work, school, or in housing (behaviors which are not as easy to control). I also only included quantity of voluntary contact, not quality; quality is not a behavior, but the participant's subjective assessment of it. Quality of interracial interactions, however, were included if the assessment was from the target or a third party.

Commitments to future action were included as behaviors if the participant made some form of commitment such that it was likely another person would hold them accountable for it in the future (e.g., providing contact information, signing one's name). Lastly, I did not include voting behaviors; while voting for one politician or another consequentially affects racial outgroups, these behaviors are likely to be clouded by other (potentially) more relevant factors, such as political identification.

I established a number of other inclusion criteria that would ensure the collected effect sizes best tested the research question of tracking the correlations over time due to changes in normative acceptability. First, race is socially constructed, and it is sometimes unclear

whether or not a given social group is psychologically represented as a “race” in participants’ minds. Groups that are not “races” per se, such as Muslims or people from a different country (e.g., Turks) were included if they likely signaled a racial or ethnic difference to participants. For example, “Muslim” likely communicates “Arab” to White, Western participants; Turkish people are considered a different ethnic group than Germans, whereas Dutch people are not. Second, the target of the interracial attitudes and behaviors must be the same target group; personality studies that included measures of “generalized prejudice” (i.e., measures of prejudice toward a variety of groups) as predictors of behaviors toward one group or individual were not included. Third, the measurement had to be done at the individual level. Studies that reported correlations at the neighborhood, classroom, or other group level were not included, as these do not necessarily capture the intrapersonal psychological processes of suppression and expression. Lastly, I included only studies where the majority of participants were adults, because (a) children are unlikely to have internalized or fully understand the norms surrounding prejudice and (b) all of these studies were self-reported behavior, which is often closely tied to parental attitudes (e.g., Ata, Bastian, & Lusher, 2009).

Literature Search

I performed backward and forward searches of narrative and meta-analytic reviews of interracial behaviors and attitude-behavior correspondence (e.g., Richeson & Shelton, 2007; Saucier et al., 2005; Talaska et al., 2008; Toosi et al., 2012). I also searched the aggregate database Academic Search Complete and Google Scholar with various combinations of terms, including: prejudice, attitude, stereotype, emotion, racism, behavior, discrimination, etc. These searches included unpublished research, such as theses and dissertations. I read through abstracts and liberally downloaded studies that might meet inclusion criteria. All studies that met criteria upon further reading were then forward and backward searched. I followed this procedure iteratively: If these new searches yielded more studies that met

inclusion criteria, these papers in turn were forward and backward searched, and so on. I performed additional specific searches, including papers introducing new measurements of racism—as behaviors are often included as measures of predictive validity—and for dependent variables often used by prejudice researchers (e.g., lost-letter technique, seating distance, helping behavior, interactions).

Effect Size and Variance Calculation

Effects were reverse-scored when necessary such that a positive correlation means the more negative an interracial attitude was, the more negative was the interracial behavior. I extracted correlation coefficients from studies where they were directly available. A number of studies median split attitudes, behaviors, or both, or the behaviors were dichotomous choices. If both attitudes and behaviors were dichotomized, I converted the 2×2 contingency tables to tetrachoric correlations—and calculated the variances for these effect sizes—using the `escalc` function in the `metafor` R package (Viechtbauer, 2010). In the case of dichotomizing only one of the variables, I converted the (a) means and standard deviations, when they were available, or (b) t - and F - statistics to biserial correlations. I again used the `escalc` function to calculate both the effect size and variance (Jacobs & Viechtbauer, 2016). One study (Raden, 1973) reported a one-way ANOVA with three conditions: high, medium, and low prejudice. These means and standard deviations were used to calculate a polyserial correlation and its variance, using the `polyserial` function in the `polycor` R package (Fox, 2016). Tetrachoric, biserial, and polyserial correlations were all employed because they assume that the underlying latent variables are continuous. Three studies (Fendrich, 1967; L. S. Linn, 1965; Saenger & Gilbert, 1950) provided enough information in the paper (such as detailed frequency tables) to recreate or reasonably approximate the original data, from which correlations were directly calculated. If none of these methods were possible, I e-mailed authors for zero-order correlations between the attitudes and behaviors discussed in their article.

Lastly, all product-moment, tetrachoric, biserial, and polyserial correlations were converted to z using Fisher's transformation, $\text{artanh}(r)$ (R. A. Fisher, 1921). For product-moment correlations ($k = 217$), the variance was calculated using the standard $1/(N - 3)$ formula (Borenstein, Hedges, Higgins, & Rothstein, 2009). The variances for the remaining correlations ($k = 23$) were calculated using the delta method (Severini, 2005, p. 400).

Moderators and Coding Procedures

My primary research question (i.e., has the predictive validity of interracial attitude measures decreased over time?) was tested by using year as a moderator, which is coded in two ways. First, the year a paper was published: I subtracted an article's publication year minus the minimum date of publication for the entire data set (1950). In the case of dissertations that were later published in peer-reviewed journals, I chose the date of publication. The primary benefit of this method is that it is consistent across all studies. In the results section, I will call this "publication year". The problem, however, is that sometimes there was a large discrepancy between when the data were collected and when they were published. Some—but not all—articles reported date of collection, and some—but not all—articles were earlier defended as dissertations. I also coded the earliest possible date from each publication as another measure of year; I refer to this in the results section as "earliest year." Any relationship between effect size and publication year could be due to how the field has changed measuring the phenomenon, so attitude type and behavior subtlety were included as covariates.

Attitude Type

I coded attitudes as (a) cognitive, (b) affective, (c) both, or (d) other or unsure (Table 1). Cognitive measures focused on beliefs or stereotypes about the group. Are they lazy? Would they be better off if only they tried a little harder? This included measures of stereotypes (e.g., rating the outgroup on a number of adjectives, such as lazy or violent) as

well as beliefs, such as the Old-Fashioned and Modern Racism Scales (McConahay et al., 1981) and the Pro-Black/Anti-Black Attitudes Questionnaire (Katz & Hass, 1988).

Table 1

Distribution of Effect Sizes k by Attitude Type

	Cognitive	Affective	Both	Other or Unsure
k	87	70	64	19

Measures were coded as affective if they focused on emotions or overall evaluations of racial outgroups. Does the participant feel positively about the group? Would they be happy to be their neighbor? Do they feel disgusted by them? Overall evaluations (positive, negative) as well as specific emotions (disgust, anger, afraid) were both considered affective measures.

Many popular measures of prejudice include both cognitive aspects and affective aspects of attitudes, such as Attitudes Toward Blacks (Brigham, 1993), Multifactor Racial Attitudes Inventory (Woodmansee & Cook, 1967), and Subtle and Blatant Prejudice Scales (Pettigrew & Meertens, 1995). If a measure included a significant number of items that measured both emotions and beliefs about groups, I coded them as both.

Some studies did not provide enough information to discern whether or not they included affective or cognitive elements, while others could not be categorized as affective or cognitive (e.g., behavioral intentions). There were not enough “other” effect sizes to make their own meaningful categories, so all of these left over effect sizes were coded in an “other or unsure” category.

Attitude type could have been modeled in two ways. The “other and unsure” category does not represent a meaningful theoretical construct. One option is to remove it from any analyses using attitude type as a moderator. This removes any worries that the “other or unsure” category is modeling miscellaneous noise; however, I lose k doing this. Moreover,

reporting standards were lower decades ago, so I lose especially precious old k doing this. I performed meta-regression both including and excluding these “other and unsure” effect sizes in attitude type moderator analyses. Analyzing the data in various ways can be informative, provided that it is transparent (Steege, Tuerlinckx, Gelman, & Vanpaemel, 2016), so I report all permutations of year and attitude coding analyses.

Behavior Subtlety

The behaviors that researchers tend to study may have gotten subtler over time and thus more difficult to predict from explicit behaviors. Indeed, researchers have focused on showing that explicit measures do not predict these behaviors—instead, implicit measures do. To combat this issue, I modeled and controlled for subtlety of behaviors.

I grouped all behavioral outcomes into 39 distinct paradigms, with one more being a combination of two of the others (how many interracial friends and how much time spent with them). I wrote descriptions of these 39 paradigms and asked 171 Amazon Mechanical Turk workers to judge the subtlety of the behavioral measures. After telling participants about the basics of the research process, workers read 10 descriptions (randomly selected from the 39) and answered three questions about each on a 0 (*Not at all*) to 100 (*Very much so*) scale: “Would it be easy for the participant in the study to figure out that the behavior they are doing is about race?” “Would it be easy for the participant to fake it so that they look less prejudiced than they actually are?” and “Is this an obvious way to measure negative behaviors towards people of another race?” Each question was answered by at least 40 people. The mean of each of these questions was taken for each different behavioral measure. Correlations between the three measures were higher than .78, so these three questions were averaged together and then reverse-scored to create a behavior subtlety score. The Appendix includes descriptions of behaviors, means, and standard deviations of subtlety, as well as directions for participants.

Data Analysis

I gathered $k = 240$ effect sizes across $m = 90$ studies. Table 2 describes the distribution across studies: Half of the studies reported more than one effect size, with three studies reporting as many as 15 effect sizes. Traditional meta-analytic methodologies assume that all effect sizes are independent of one another; this assumption is severely violated in the present analysis, as effect sizes drawn from the same participants are dependent on one another.

Table 2
*Number of Studies m with a Given Number of
Effect Sizes*

	1	2	3	4	5	6	7	8	11	12	15
m	44	25	6	4	1	2	1	1	1	2	3

Many meta-analysts ignore these dependencies. For example, Talaska et al. (2008) also find that the majority of their effect sizes are dependent on another effect size; given that their moderator of interest—the type of attitude measure—is at the effect size level, it would have severely limited their k to only choose one effect size per study. The authors thus allow multiple effect sizes per study in their meta-analysis: “We coded effect sizes for every attitude–behavior pair in each study. Although this leads to issues of non-independence in calculating the overall effect size, we were more interested in the effects of moderating variables” (p. 269). Implicit in this statement is that dependencies might bias the overall effect size, but *not* their moderator analyses. Is this the case?

The first issue with this approach is that studies reporting more effect sizes get more weight; all else being equal, a study with three effect sizes will have three times as much weight in the analysis. The primary issue, however, is that effect sizes based on the same participants have correlated errors, and ignoring these dependencies falsely assumes that the covariance between any two errors is zero. This will underestimate the variances of effect

sizes; thus, the standard errors of test statistics in the *moderator* analyses will also be underestimated, leading to an inflation in Type I error (Aarts, Verhage, Veenvliet, Dolan, & Sluis, 2014; Borenstein et al., 2009, Chapter 24; Finch, Bolin, & Kelley, 2014, Chapter 2; James, Witten, Hastie, & Tibshirani, 2013, Chapter 3).

One would ideally use a multivariate meta-analysis to model these dependencies; however, this requires the meta-analyst to have access to the full covariance matrix of all measures in all studies. This is not realistic in many settings (Jackson, Riley, & White, 2011), particularly in the present meta-analysis of a literature where (a) researchers hardly publish this information and (b) the research has been published over the course of 70 years, making acquiring this information from the authors of many of these studies impossible.

Multilevel meta-analysis has also been proposed as a way to deal with unknown dependencies between effect sizes (Cheung, 2014; Konstantopoulos, 2011; Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013). While some argue that individuals could be modeled at Level 1, effect sizes at Level 2, and study at Level 3 (e.g., Cheung, 2014), three-level meta-analyses still assume that residual errors are orthogonal within clusters (Tanner-Smith, Tipton, & Polanin, 2016). This assumption is violated when multiple effect sizes are drawn from the same participants.

The most recent development (Polanin, Hennessey, & Tanner-Smith, 2017) for modeling these dependencies and avoid underestimating standard errors is to use robust variance estimates (RVE; Hedges et al., 2010; Tanner-Smith et al., 2016). I performed my meta-analysis using RVE for correlated effects in the **robumeta** R package (Z. Fisher & Tipton, 2015).

As mentioned above, I am able to calculate the variances of effect sizes directly from sample size, but I am *not* able to calculate the covariance between effect sizes. RVE solves this problem by using the cross products of the residuals for each study to estimate the variance-covariance matrix of effect sizes within a study. While the estimate of the covariance matrix in each study is not very good, the combined variance estimate converges

to the true variance as the number of studies approaches infinity (Hedges et al., 2010).

Traditional meta-analyses weight effect sizes by using the inverse of the variance. RVE weights each effect size using (a) the inverse of the average variance across all effect sizes in a study (assuming a constant correlation across effect sizes) (b) divided by the number of effect sizes in the study. This ensures that a study does not get “extra” weight simply by having more effect sizes.

This method is used primarily for the purposes of this meta-analysis: interpreting meta-regression coefficients. The variance estimates found in other meta-analyses (e.g., Q, I^2, τ^2) are not precise when using RVE. Given this shortcoming of RVE—and my main focus in interpreting meta-regression coefficients—I will not focus on these estimates. The Appendix includes data, code, and detailed analyses.

Results

I first ran an intercept-only model first to get the overall meta-analytic effect, $r = .21$ [.17, .25]. Figure 1 displays the distribution of these effect sizes, marking the meta-analytic average with the dotted vertical line. This summary effect size is equal to the smallest effect size found in previous ABC meta-analyses (Notani, 1998).

Year

I predicted the effect size from publication year and earliest year in two separate meta-regression equations. In both cases, the results were the same: $b = -.002$ [−.005, .000], $t(26) = -1.91$, $p = .067$. This is debatable evidence that the relationship between interracial attitudes and behaviors decreases over time at a rate of $z = .002$ per year. This is not a standardized regression coefficient; although it may seem small, the predicted r for the first study (i.e., the intercept) was .32, and the model-predicted value in 2016 (the most recent publication) is about .17. The p -value for these analyses warrant skepticism, but there is some evidence that interracial ABC decreases over time. But is this due to the theoretical construct of interest—normative climate—or other factors?

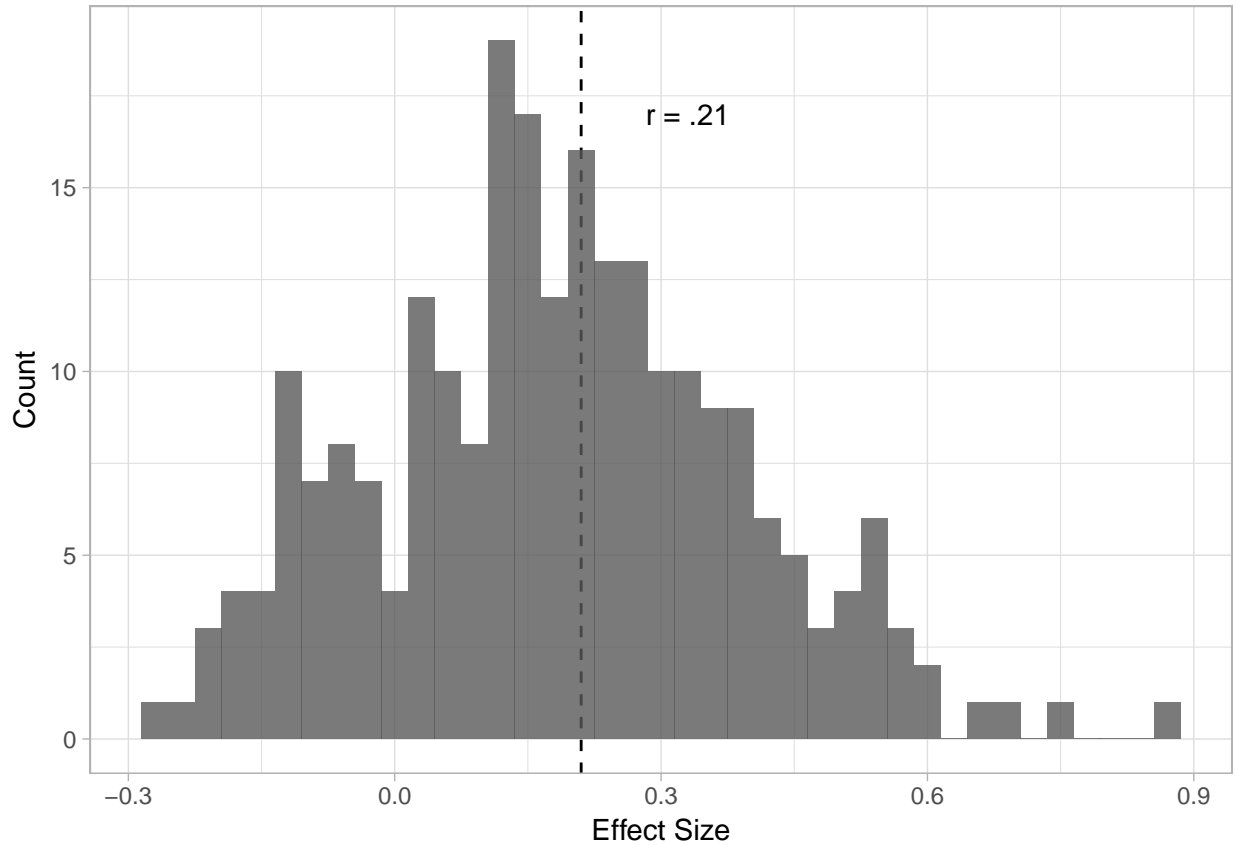


Figure 1. Distribution of effect sizes. Total $k = 240$. Horizontal line represents meta-analytic effect size.

Full Model

I regressed effect size on (a) year, (b) behavior subtlety, (c) attitude type, and (d) the methodological covariate of whether or not the effect size was transformed. Tables 3 and 4 contain four separate meta-regression models, including all combinations of year and attitude type. In each model, effect size does not depend on year, $ps > .802$. Effect size did not depend on attitude type, either: all omnibus test $ps > .313$. The only significant predictor was subtlety, $ps < .009$, which negatively predicted effect size. Note that **robumeta** implements RVE using an adjustment to the degrees of freedom to account for small sample sizes (such as those found in meta-analyses), which causes the degrees of freedom to differ from coefficient to coefficient (Z. Fisher & Tipton, 2015; Tipton, 2015).

The thin evidence for interracial ABC diminishing over time provided by regressing effect size on year alone was nonexistent after controlling for relevant covariates. The only reliable predictor was subtlety: The more subtle a behavior, the less predictive an explicit interracial attitude. This analysis also fails to replicate previous work suggesting emotional measures are better predictors of discrimination than are cognitive ones; the ABC was about $r = .035$ higher for affective than cognitive attitude measures, but not significant $ps > .40$.

Table 3

Meta-Regression Model, Using All Data

Predictor	<i>b</i>	<i>SE</i>	<i>t</i>	<i>df</i>	<i>p</i>	CI UB	CI LB
Intercept	0.407	0.097	4.182	29.621	0.000	0.208	0.606
Publication Year	0.000	0.001	-0.099	13.596	0.922	-0.003	0.003
Subtlety	-0.005	0.002	-2.851	35.523	0.007	-0.009	-0.001
Attitude (Affective)	0.035	0.043	0.813	20.499	0.425	-0.055	0.126
Attitude (Both)	-0.039	0.052	-0.749	36.239	0.459	-0.144	0.066
Attitude (Other)	0.074	0.051	1.459	16.814	0.163	-0.033	0.180
Transformed	0.232	0.129	1.796	16.056	0.091	-0.042	0.506
Intercept	0.404	0.096	4.213	31.061	0.000	0.208	0.600
Earliest Year	0.000	0.001	-0.034	13.567	0.973	-0.003	0.003
Subtlety	-0.005	0.002	-2.879	34.604	0.007	-0.009	-0.002
Attitude (Affective)	0.035	0.044	0.806	20.734	0.429	-0.056	0.126
Attitude (Both)	-0.038	0.052	-0.735	36.308	0.467	-0.143	0.067
Attitude (Other)	0.075	0.050	1.510	16.930	0.150	-0.030	0.181
Transformed	0.234	0.128	1.825	16.048	0.087	-0.038	0.505

Table 4

Meta-Regression Model, Excluding "Other" Attitudes

Predictor	<i>b</i>	<i>SE</i>	<i>t</i>	<i>df</i>	<i>p</i>	CI UB	CI LB
Intercept	0.429	0.112	3.826	21.263	0.001	0.196	0.661
Publication Year	0.000	0.002	-0.255	11.896	0.803	-0.004	0.003
Subtlety	-0.005	0.002	-2.792	31.977	0.009	-0.009	-0.001
Attitude (Affective)	0.036	0.044	0.819	20.790	0.422	-0.055	0.128
Attitude (Both)	-0.035	0.054	-0.654	35.313	0.518	-0.145	0.075
Transformed	0.238	0.138	1.720	13.709	0.108	-0.059	0.536
Intercept	0.428	0.111	3.866	22.410	0.001	0.199	0.657
Earliest Year	0.000	0.002	-0.248	12.225	0.809	-0.004	0.003
Subtlety	-0.005	0.002	-2.791	31.350	0.009	-0.009	-0.001
Attitude (Affective)	0.036	0.044	0.823	21.067	0.420	-0.056	0.128
Attitude (Both)	-0.035	0.054	-0.645	35.388	0.523	-0.144	0.075
Transformed	0.239	0.137	1.737	13.702	0.105	-0.057	0.534

Sensitivity Analysis

The meta-analyst inputs a value ρ as an estimate of the within-study effect size correlation needed to initially calculate τ^2 in the RVE approach to meta-regression with correlated effects. I used the package's default $\rho = 0.80$ for all models in this paper. I conducted sensitivity analyses for each model to test if coefficients and standard errors depended on this ρ ; changing the value of ρ from 0 to 1 altered neither the coefficients nor the standard errors.

Ancillary Analyses

I also ran a model regressing effect size on subtlety alone; as behaviors were more subtle, the effect size was smaller, $b = -0.006$, $SE = 0.002$, $t(35.9) = -3.39$, $p = .002$. I also regressed effect size on attitude type alone. When *including* the “other or unsure” effect sizes in the data set, the omnibus test for attitude type was not significant, $F(32.3) = 2.52$, $p = .076$. There was also no relationship when *excluding* this miscellaneous category from the data set, $F(32) = 1.21$, $p = .313$; the marginal effect in the model including “other or unsure” effect sizes was driven by that category.

Research Trends

These data also allow an examination of research trends over time. Have researchers turned to more subtle behavioral measures? To answer this question, I ran two separate multilevel models, one using publication year and another using earliest year to predict behavior subtlety. Subtlety was measured at Level 1 (effect size) and nested within Level 2 (study), where year was the predictor. A random intercept-only model was compared to models with just publication year or earliest year as predictors, using likelihood ratio tests (Hox, Moerbeek, & Schoot, 2010). Publication year did not predict subtlety, $b = 0.099$, $\chi^2(1) = 2.73$, $p = .099$ and earliest year did somewhat, $b = 0.113$, $\chi^2(1) = 3.57$, $p = .059$.

But many of the effect sizes were not *observed* behaviors, but self-reports of past behaviors. Measurements for these types of behaviors have remained consistent over time (e.g., simply asking people how much they spend time with someone of the target race). Out of the 240 effect sizes collected, 149 (62%) were observed. Running the same analyses of year predicting subtlety on just these 149 data points, both publication year, $b = 0.13$, $\chi^2(1) = 5.50$, $p = .019$, and earliest year, $b = 0.126$, $\chi^2(1) = 5.32$, $p = .021$, positively predict behavioral subtlety. When excluding self-reported behaviors, researchers are choosing to observe subtler discriminatory behaviors in their experiments now than they did in the past.

I also examined what types of attitude measures researchers have chosen over time. I

fit three separate logistic multilevel models predicting each type of attitude from year (Begg & Gray, 1984), using the data without the “other or unsure” effect sizes. Each model had a dichotomous outcome: First, cognitive vs. affective or both; second, affective vs. cognitive or both; third, both vs. cognitive or affective. As with subtlety over time, attitude was a Level 1 (effect size) dependent variable nested within Level 2 (study), where year was the predictor. Neither publication year nor earliest year were significant predictors of any of the three categorical dependent variables, all $\chi^2(1) < 0.76$, $ps > 0.385$; there is no relationship between year and researcher’s choice of attitude measure type.

Diagnostics

The overall meta-analytic effect size between interracial attitudes and behavior is $r = .21$. Despite thoroughly searching for articles, dissertations, and theses to include in this meta-analysis, I certainly missed studies; additionally, publication bias exists and may undermine this meta-analytic result. Meta-analytic tools that handle correlated effect sizes are relatively new, and testing for publication biases are rarely mentioned in the literature (e.g., Cheung, 2014; Z. Fisher & Tipton, 2015; Jackson et al., 2011; Tanner-Smith et al., 2016). Nonetheless, I test and examine the possible influences of bias by calculating fail-safe N , creating a funnel plot, and conducting a p -curve analysis.

I conducted the fail-safe N (Orwin, 1983) to estimate the number of studies with an effect size of zero that I would have to observe for the meta-analytic r to become trivial. Along with other researchers, I consider $r = .10$ to be a trivial effect size here (C. J. Ferguson, 1983; Hyde & Linn, 2006; Lambert, Engh, Hasbun, & Holzer, 2012; Murphy & Myers, 1999; Uttal et al., 2013). I would need to observe 105 studies with $r = .00$ for the correlation to become trivial.

Funnel plots may be asymmetric for a number of reasons beside publication bias, including effect sizes being dependent on one another (Nakagawa, Noble, Senior, & Lagisz, 2017). For this reason, I generated a funnel plot based on a standard random-effects,

intercept-only meta-analysis where effect sizes and variances are averaged together at the study level, an approach used by other researchers using an RVE meta-analytic model (Uttal et al., 2013). The model was fit using the `rma` function and in the `metafor` R package (Viechtbauer, 2010). This funnel plot (Figure 2) is mostly symmetric, save for two outliers.

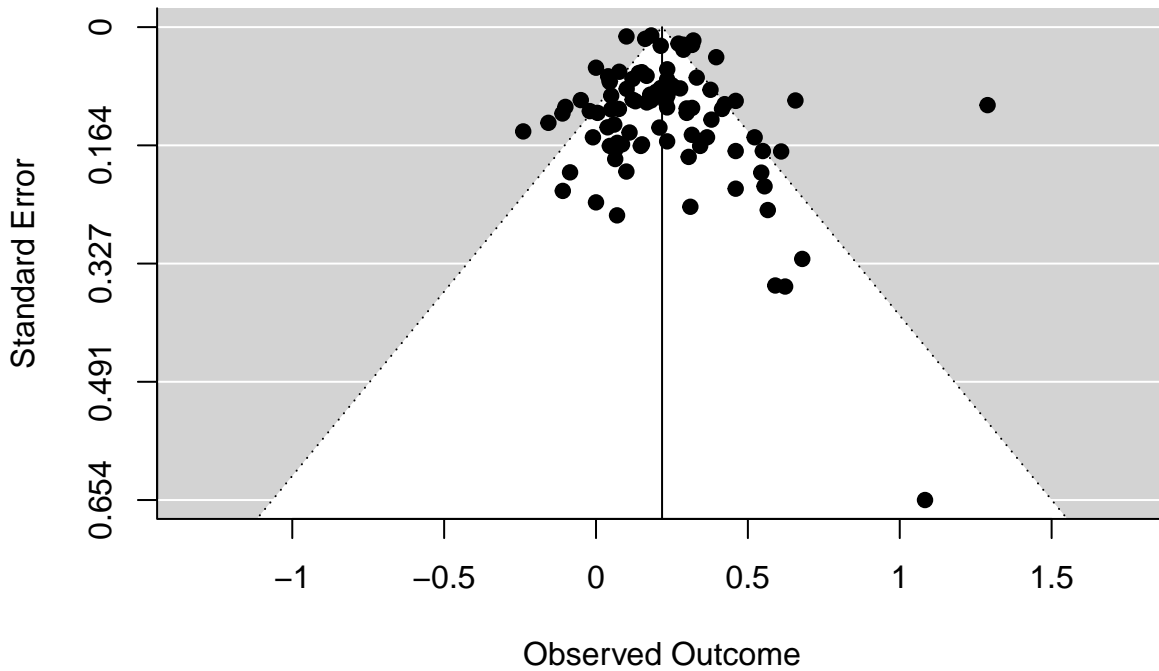
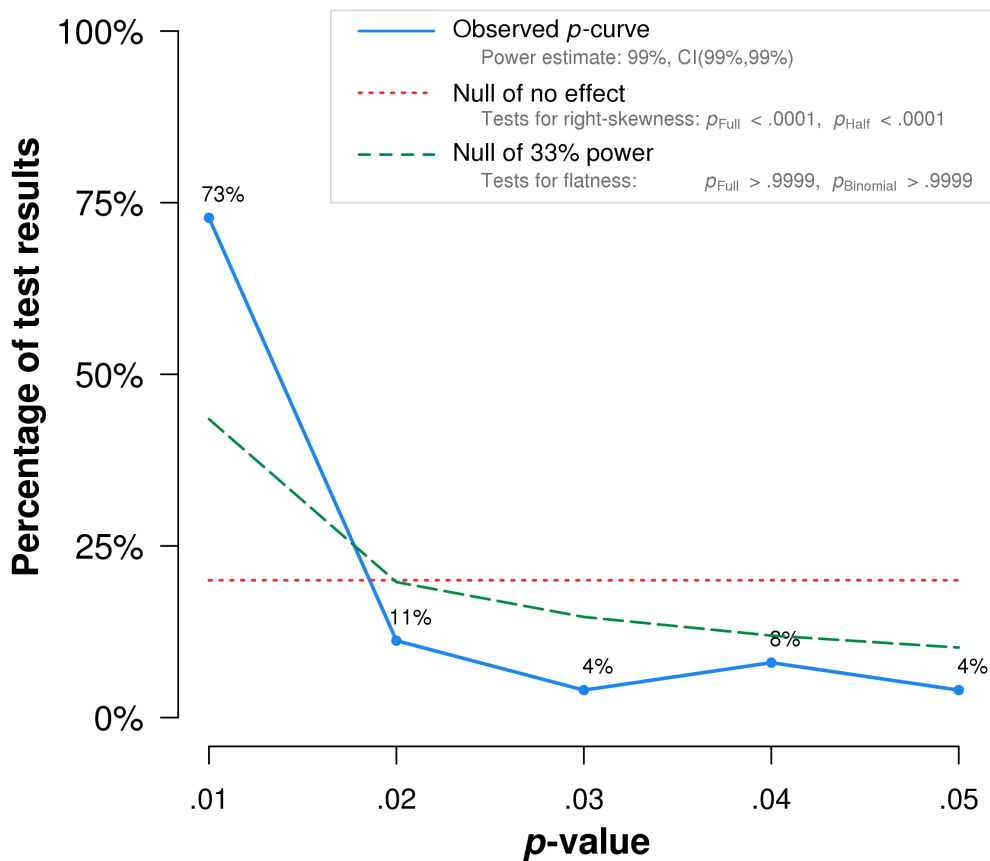


Figure 2. Funnel plot of effect sizes (effect sizes and variances averaged by study).

Lastly, I conducted a *p*-curve analysis to examine evidence of publication bias or *p*-hacking (Simonsohn, Nelson, & Simmons, 2014a, 2014b; Simonsohn, Simmons, & Nelson, 2015). When the null hypothesis is true, *p*-values testing that null hypothesis are uniformly distributed; when the alternative hypothesis is true, *p*-values will be positively skewed, and smaller *p*-values are more likely to be observed. The *p*-curve takes advantage of these probabilities by plotting the percentage of findings at each successive significance level: $p = .05$, $p = .04$, etc. When results are *p*-hacked or lack statistical power, the researcher examines more *p*-values around the $p = .05$ threshold than would be expected. When results contain “evidential value,” the curve is appropriately positively-skewed, with the largest percentage of *p*-values congregating at the $p < .01$ level.



Note: The observed *p*-curve includes 125 statistically significant ($p < .05$) results, of which 107 are $p < .025$. There were 115 additional results entered but excluded from *p*-curve because they were $p > .05$.

Figure 3. *P*-curve of statistically significant results (generated via the *p*-curve app).

Figure 3 displays the *p*-curve. Notably, 115 of the 240 tests were not statistically significant; this is to be expected—even in the presence of a true effect—when a field generally conducts under-powered tests (Lakens & Etz, 2017). It also supports the observation Kraus (1995) made in his meta-analysis that non-significant ABC relationships are more likely to be published than other null results, given their counter-intuitive nature. Of the 125 statistically significant tests, 107 (86%) were $p < .025$. Using a binomial test, this is significantly different from the 50% that would be expected from the uniform distribution if there were no effect, $p < .0001$. I conducted this test using Version 4.052 of the *p*-curve app (www.p-curve.com); recently, the authors have implemented other tests—such as a

continuous test instead of a binomial one—to test for evidentiary value (Simonsohn et al., 2015). Each of these tests demonstrates evidentiary value, $p < .0001$.

Discussion

There is not evidence that the correspondence between interracial attitudes and interracial behaviors in the social psychological literature has decreased over time because of normative changes in the expression of prejudice. Year, when considered alone, predicted smaller effect sizes (although this effect was “marginally,” “nearly,” “almost,” or “somewhat” significant; Hankins, 2013). However, relevant covariates fully accounted for this relationship: When considering year, behavior subtlety, attitude type, and whether or not the effect size was transformed together, only the subtlety of the behavior predicted the size of the effect. The more subtle the observed behavioral measure, the smaller the correlation between explicit interracial attitude and behavior.

Many researchers begin papers by noting the paradox that self-reports of prejudice have decreased in recent decades, yet discrimination continues to persist. If this is true, then psychologists should have gotten worse at predicting discrimination from self-reports over time: Attitude-behavior correspondence (ABC) in the interracial domain should correlate negatively with the year in which the study was published. With this hypothesis, year is treated as a proxy for stricter norms against prejudice. There was debatable evidence for this negative relationship when considering the zero-order relationship, but there was no relationship after considering covariates. There may be indirect evidence for the hypothesis, however. Psychological measurement and methods have improved over time, and meta-analyses of ABC across content domains have found a *positive* relationship with year (Glasman & Albarracín, 2006; Kraus, 1995). ABC in the interracial domain not following the improved correspondence in other domains could be construed as indirect evidence that egalitarian norms have kept the correlation between prejudice and discrimination from similarly increasing. However, researchers have been influenced by the same changing

normative climate; no longer do we ask participants when they think lynching is justified—we ask participants how much they agree with more coded prejudiced statements. Researchers could have preempted finding the hypothesized correlation by switching to better measurements of prejudice as the old measurements exhibited poor psychometric (e.g., floor effect) and predictive properties.

Non-self-report measures of prejudiced behavior have gotten more subtle over time, a trend likely influenced by the focus on implicit prejudice. Various studies show that explicit (i.e., self-reported) prejudice predicts controlled, purposeful behavior while implicit prejudice predicts subtle, automatic behaviors (Pearson et al., 2009). I observed meta-analytic evidence for half of this contention: The subtler the behavior, the worse the predictive power explicit measures have. Therefore, research has changed in two ways that mask the ability to observe the relationship between year (as a measurement of norms) and ABC: first, researchers are using measures that keep up with changing normative trends; second, researchers turned their interest toward the newer, implicit theories of prejudice, measuring subtler behaviors that implicit measures are said to predict. Given the evidence of controlled, purposeful prejudiced behaviors in the recent political climate (Cuevas, 2016; Holley, Bever, Lowery, & Mettler, 2017; Lombroso & Appelbaum, 2017; Stolberg & Rosenthal, 2017; Thrush & Haberman, 2017; Wang, 2017), psychologists might consider a refocus on theories of explicit prejudice and behavior.

Small Effects, and A Focus on Prediction

Many researchers have lamented how social and personality psychology lacks a focus on measuring *actual* behaviors, calling for more of a focus on them, with a particular interest in *predicting* behaviors (e.g., Baumeister, Vohs, & Funder, 2007; Cialdini, 2009; Furr, 2009). In the area of stereotyping, prejudice, and discrimination specifically, Fiske (2000) argued that social and personality psychologists need “to get serious about predicting behavior... The alarm is urgent” (p. 312). How well can we predict interracial behaviors from interracial

attitudes?

These data yield a meta-analytic correlation of $r = .21$, a smaller estimate than found by both Talaska et al. (2008), $r = .26$, and Schütz and Six (1996), $r = .29$. A first reaction to this estimate might be to square it and bemoan how we can only account for about 4% of the variance in interracial behaviors with interracial attitudes. Researchers have felt this sense of disappointment in attitudes ever since LaPiere (1934) found that, while most establishments said over the phone that they would *not* permit Chinese people to eat in their restaurant or stay in their hotel, only *one* of these establishments actually denied service to a Chinese couple.

Researchers have debated how to interpret the size of ABC effects for decades (see Ajzen & Fishbein, 2005). In response to the earliest studies showing a lack of consistency between attitudes and behaviors, DeFleur and Westie (1963) called it a “fallacy” to expect strong ABC (p. 27). Wicker (1969) similarly remarked that “it is considerably more likely that attitudes will be unrelated or only slightly related to overt behaviors than that attitudes will be closely related to actions” (p. 65); recent meta-analysts came to similar conclusions that interracial attitudes do not predict behaviors as much as we would like them to (Oswald et al., 2013; Schütz & Six, 1996).

Others are not as pessimistic: Kraus (1995) argues that his observed $R^2 = .14$ is not really that small and that R^2 is often misinterpreted; Ajzen and Fishbein (2005) argue correlations will be higher if researchers choose “representative” behavioral measures (or average across many behaviors). Many have made suggestions on how to improve ABC (e.g., Oswald et al., 2013 suggests using multiple, heterogeneous attitude measures in latent variable models to account for measurement error), but I wish to offer two related suggestions on increasing our ability to *predict* discrimination: first, adapting statistical methods aimed at *prediction* more than at *inference*; and second, relaxing the focus on just interracial attitudes.

James et al. (2013) present two motivations for conducting statistical analyses:

inference and *prediction*. The goal of inference is to understand the relationship between variables. This is the approach social and personality psychologists take: We generally perform null hypothesis significance tests to tell us whether or not some relationship between variables is present, and we make theoretical inferences from these data. We know that there is a positive relationship between interracial attitudes and behavior. What we should move toward now is prediction, where the goal is to make accurate guesses at some outcome, given a collection of input variables. This is not what social and personality psychologists do. However, as Fiske (2000) noted, we should be moving toward using our various theories to *predict* discrimination.

Lewin's famous equation says that behavior is a function of a person and their environment, $B = f(P, E)$ (Lewin, 1951). When prejudiced behavior is our dependent variable, our job is to find some function f that explains B . Personality and social psychologists rely on general linear models as our estimation of f , which serves the purpose of inference well. These models, such as ordinary least squares (OLS) regression, are easy to interpret. However, these models make rigid assumptions: In the present case, we assume that the relationship between attitudes and behavior is strictly linear.

If we are to move toward a focus on *prediction*, we should embrace the vast world of algorithms beyond OLS. Our foremost concern for prediction is to minimize the difference between what our models predict and what we actually observe. Unless the relationship is strictly additive and linear, we can often accomplish this better by using, for example, non-parametric models (e.g., regression and classification trees) that do not assume a rigid form for f (see James et al., 2013 for an overview). However, this comes with the trade-off that the models are not always as interpretable.

But shiny, cutting edge algorithms to estimate f are not a magic solution. If we wish to estimate behavioral prejudice B , we should remember Lewin's equation that it is a function of both the person *and their environment*. If we assume $B = f(P, E)$ is correct, it should be no surprise that we observe modest R^2 values when we consider only *one* aspect of

the wide variety of factors that make up P . The field of social and personality psychology has been tremendously successful as a field at generating many useful theories of prejudice, discrimination, and stereotyping. Each of these theories contribute key insights into which aspects of P and E we should consider when trying to predict discrimination. Instead of competition between theories, prediction requires using important variables from *many theories* as inputs in our model. Predicting prejudiced behaviors accurately will involve a large number of these inputs [e.g., social norms, Crandall and Stangor (2005); moral credentials, Monin and Miller (2001); motivation, Plant and Devine (1998); fatigue, Govorun and Payne (2006); etc.] and, of course, a correspondingly larger n on which to train and test models.

Programmers without the crucial domain expertise in prejudice, discrimination, and stereotyping have produced racist and otherwise biased models (Angwin, Larson, Mattu, & Kirchner, 2016; O’Neil, 2016). One lesson is to enlist those with domain expertise, such as social and personality psychologists, when creating these models. Another lesson is to take Fiske’s suggestion seriously: If predicting discrimination is our goal, we should be leveraging our theories to create models (a) with *many* predictors across *many* theories and (b) that are less rigid than traditional regression and analysis of variance.

There is great potential here. For example, consider the data scientists who are predicting and preventing adverse police incidents, using various inputs—from past allegations to sick days taken—and various methods (especially tree-based algorithms). When compared to the existing systems, the model trained by those working with the Data Science for Social Good organization “would have correctly flagged 10-20% more officers who were later involved in adverse incidents and reduced the number of incorrect flags by 50% or more,” while a model trained for another police department “was able to correctly flag 80% of officers who would go on to have an adverse interaction, while only requiring intervention on 30% of officers to do so” (Carton et al., 2016; Mitchum, Genova, & Taylor, 2017). A linear approximation using one predictor—interracial attitudes—is simply not enough to

accurately predict something as complex as discrimination; we should not be surprised by the modest meta-analytic relationship found here and elsewhere.

Attitude Type

I did not find evidence that affective measures predicted behavior better than cognitive ones; however, Talaska et al. (2008) subtitle their article “Emotions, Not Beliefs, Best Predict Discrimination” from their meta-analysis of the same phenomenon and many of the same studies. Why does their conclusion differ from the present meta-analysis?

First, the authors treated *aggregated* measures of specific emotions (e.g., fear, disgust) to be qualitatively different than general affective measures, a differentiation that is not supported by theories of specific prejudices (Cottrell & Neuberg, 2005). For example, Talaska and colleagues coded measures as “emotion” if they went “beyond overall positive/negative valence” (p. 268). There were two other affective categories: “Overall valence” measures were “*overall* positive versus negative evaluation, without reference to specific beliefs, stereotypes, emotions, or behaviors,” while “valence and emotion” measures included “feelings limited to simple positive-negative connotations” (p. 268). Talaska et al. (2008) do not differentiate between differentiated emotions, however. All measures of specific emotions are aggregated into an “emotion” category; proponents of specific prejudiced emotions argue *against* aggregating across multiple types of emotions, because it masks the very specifics that these theories try to explain (see Cottrell & Neuberg, 2005).

I did not distinguish between differentiated emotions and overall valence, because there are not enough studies using specific emotions to code for each emotion (e.g., contempt, fear, anger, disgust) separately. This was also done for theoretical reasons. Talaska et al. (2008) cite the primacy of affect—that people often evaluate attitude objects, with little effort, and with no thought required—to argue why emotions should be the primary motivators for behavior. Zajonc (1980) did argue that feelings would be better predictors of behaviors: “Quite often ‘I decided in favor of X’ is no more than ‘I liked X.’ Most of the time,

information collected about alternatives serves us less for making a decision than for justifying it afterward” (p. 155). However, the theory is about affect, not specific emotions: Zajonc (1980) refers to “general” feelings that are about “approach-avoidance”, while “other emotions, such as surprise, anger, guilt, or shame... are ignored” (p. 152). Aggregated measures of specific emotions and overall evaluations both reflect overall affect, given that these measures average across specific emotions that fall on a general negative-to-positive continuum (Lindquist, 2013; Russell, 2003).

Second, they do not report any pairwise comparisons between overall evaluations and any other attitude type, so it is unclear how general affective measures performed in their meta-analysis. There were nine attitudinal categories in total. Talaska and colleagues collapsed across all stereotype and belief categories, reducing the total number to seven categories. However, the comparisons between attitude types (shown in Figure 2, p. 282) only display five of these categories. The “emotion and overall valence” category and “stereotype and overall valence” category (e.g., identifying how good vs. bad are outgroups members) are missing. It is unclear if these two measures were included with the “mixed/other” category (e.g., IMS/EMS, Plant & Devine, 1998). Yet, the median effect sizes for differentiated emotions ($r = .35$) and emotion and overall valence ($r = .32$) measures were of about the same strength (Table 7, p. 281).

An omnibus test for type of attitude was reported to be significant, $p < .001$, but only one follow-up test was reported: emotion-focused attitudes had a larger effect size than stereotypes and/or beliefs attitudes, $p < .001$. The direct comparison between differentiated emotions and overall evaluations was not tested, which is the strongest test for the importance of specific emotions. It is unclear if this test was done, why only five of the seven categories were included in the omnibus test, and if the necessary alpha-level corrections were done if multiple comparisons were made.

Third, the majority of the effect sizes in Talaska and colleagues’ meta-analysis are dependent on at least one other effect size, as many effects are gathered from the same

participants. I correct for the underestimation in standard errors that this causes, while Talaska and colleagues do not.

Is their argument that emotions are better predictors of discrimination than beliefs wrong? Not necessarily. Their meta-analysis does not test the predictions derived from theories of specific prejudices and differentiated emotions. To test these hypotheses, one must be able to differentiate between the specific emotions in analysis. There is not enough literature examining the link between *specific* prejudices (e.g., fear, disgust) and correspondingly *specific* behaviors yet. Researchers interested in these theories should examine actual behaviors, not just behavioral tendencies or intentions. For example, *anger* might outperform a feeling thermometer when it comes to predicting *aggressive* behaviors specifically; *pity* might be the best predictor of benevolently prejudiced behaviors. These are interesting empirical questions that need more study before they be examined meta-analytically.

Is *affect* no more important than beliefs in predicting discrimination? This analysis provides evidence that it is not. Yet, it is unclear if affective self-reports are bona fide measurements of affect. “Prejudice” is a latent construct—we do not measure directly and resort to measure indicators of it. Affective self-reports are *expressions* of prejudice, which are often quite different from genuine prejudice itself (Crandall & Eshleman, 2003). Affective self-reports likely measure underlying affect itself, but also: willingness to admit that one is prejudiced, how justified one thinks the prejudice is, perceived normative support, among many other constructs.

Zajonc (1980), in his argument that affect is primary, reviews evidence that affect is hard to verbalize; even if people *wanted* to be truly honest on affective self-reports, it would be hard for them to do so. He notes that affect may be more tied up in physiology than it is in communication, and modern constructionist theories of emotion have found evidence that supports the idea that overall effect has its foundations in biological reactions, which are then interpreted and communicated to others (Barrett, 2017; Lindquist, Satpute, Wager,

Weber, & Barrett, 2015; Satpute et al., 2015). More research correlating physiological and neurological measures with interracial discrimination should be done to see if these measures (a) represent negative affect felt toward other groups (i.e., prejudice) better than self-report measures and (b) are particularly strong predictors of behavior. Despite these exciting new ways of measuring affect, I share the sentiment of Crandall and Eshleman (2003), who are “pessimistic about psychologists’ ability to directly measure the unfiltered psychological reality of prejudice” (p. 437).

Conclusion

Omi and Winant (1994) note that matters of race in the United States are “maddeningly complex.” The current meta-analysis aimed to support the idea that self-reports of prejudice are decreasing at a quicker rate than prejudiced behaviors. In line with the muddy findings across various disciplines, I found no clear evidence to support my hypothesis, largely due to researchers adapting to the same norms I set out to capture. Contrary to previous findings, I did not find evidence that affective measures are better predictors of behavior than cognitive ones. I also hope that psychologists might consider broadening their focus if their goal is to predict discrimination.

References

- Aarts, E., Verhage, M., Veenvliet, J., Dolan, C., & Sluis, S. van der. (2014). A solution to dependency: Using multilevel analysis to accomodate nested data. *Nature Neuroscience*, 17, 491–496.
- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211.
- Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 173–221). New York, NY: Psychology Press.
- American National Election Studies. (n.d.). *The ANES guide to public opinion and electoral behavior*. Retrieved from <http://www.electionstudies.org/nesguide/text/CHAP4>
- Anderson, C. (2016). *White rage: The unspoken truth of our racial divide*. New York, NY: Bloomsbury Publishing.
- Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). *Machine bias*. Retrieved from <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- Armitage, C. J., & Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British Journal of Social Psychology*, 40(4), 471–499.
- Ata, A., Bastian, B., & Lusher, D. (2009). Intergroup contact in context: The mediating role of social norms and group-based perceptions on the contact-prejudice link. *International Journal of Intercultural Relations*, 33(6), 498–506.
- Barrett, L. F. (2017). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience*, 12(1), 1–23.
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior?

- Perspectives on Psychological Science*, 2(4), 396–403.
- Begg, C. B., & Gray, R. (1984). Calculation of polychotomous logistic regression parameters using individualized regressions. *Biometrika*, 71(1), 11–18.
- Biernat, M., & Crandall, C. S. (1999). Racial attitudes. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of political attitudes* (pp. 297–411). New York: Academic Press.
- Blanchard, F. A., Crandall, C. S., Brigham, J. C., & Vaughn, L. A. (1994). Condemning and condoning racism: A social context approach to interracial settings. *Journal of Applied Psychology*, 79(6), 993–997.
- Bloome, D. (2014). Racial inequality trends and the intergenerational persistence of income and family structure. *American Sociological Review*, 79(6), 1196–1225.
- Bobo, L., Charles, C. Z., Krysan, M., & Simmons, A. D. (2012). The real record on racial attitudes. In P. V. Marsden (Ed.), *Social trends in American life: Findings from the General Social Survey since 1972* (pp. 38–83). Princeton, NJ: Princeton University Press.
- Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley Online Library.
- Bray, D. W. (1950). The prediction of behavior from two attitude scales. *The Journal of Abnormal and Social Psychology*, 45(1), 64–84.
- Brigham, J. C. (1993). College students' racial attitudes. *Journal of Applied Social Psychology*, 23(23), 1933–1967.
- Carton, S., Helsby, J., Joseph, K., Mahmud, A., Park, Y., Walsh, J., . . . Ghani, R. (2016). *Identifying police officers at risk of adverse events*. Retrieved from <https://dssg.uchicago.edu/wp-content/uploads/2016/04/identifying-police-officers-3.pdf>
- Cary, L. A., & Page-Gould, E. (2014). The prevalence and impact of sexism, racism and homophobia in online gaming environments. Poster presented at the annual meeting

- of the Society for Personality and Social Psychology, Austin, TX.
- Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods*, 19(2), 211–229.
- Cialdini, R. B. (2009). We have to break up. *Perspectives on Psychological Science*, 4(1), 5–6.
- Cooke, R., & Sheeran, P. (2004). Moderation of cognition-intention and cognition-behaviour relations: A meta-analysis of properties of variables from the theory of planned behaviour. *British Journal of Social Psychology*, 43(2), 159–186.
- Cottrell, C. A., & Neuberg, S. L. (2005). Different emotional reactions to different groups: A sociofunctional threat-based approach to “prejudice”. *Journal of Personality and Social Psychology*, 88(5), 770–789.
- Crandall, C. S., & Eshleman, A. (2003). A justification-suppression model of the expression and experience of prejudice. *Psychological Bulletin*, 129(3), 414–446.
- Crandall, C. S., & Stangor, C. (2005). Conformity and prejudice. In J. F. Dovidio, P. Glick, & L. A. Rudman (Eds.), *On the nature of prejudice: Fifty years after Allport* (pp. 295–309). Malden, MA: Blackwell.
- Crandall, C. S., Eshleman, A., & O’Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, 82(3), 359–789.
- Crandall, C. S., Ferguson, M. A., & Bahns, A. J. (2013). When we see prejudice: The normative window and social change. In C. Stangor & C. S. Crandall (Eds.), *Stereotyping and prejudice* (pp. 53–70). New York, NY: Psychology Press New York, NY.
- Crosby, F., Bromley, S., & Saxe, L. (1980). Recent unobtrusive studies of black and white discrimination and prejudice: A literature review. *Psychological Bulletin*, 87(3), 546–563.
- Cuevas, M. (2016). *‘Trump’ as anti-Latino epithet: Ugly incidents at high school games.*

Retrieved from

<http://www.cnn.com/2016/03/01/us/midwest-trump-school-chants/index.html>

DeFleur, M. L., & Westie, F. R. (1963). Attitude as a scientific concept. *Social Forces*, 42(1), 17–31.

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56(1), 5–18.

Dovidio, J. F., & Gaertner, S. L. (2000). Aversive racism and selection decisions: 1989 and 1999. *Psychological Science*, 11(4), 315–319.

Dovidio, J. F., & Gaertner, S. L. (2004). Aversive racism. *Advances in Experimental Social Psychology*, 36, 1–52.

Economic Policy Institute. (2013). *The unfinished march: An overview*. Retrieved from <http://www.epi.org/files/2013/EPI-The-Unfinished-March-An-Overview.pdf>

Fazio, R. H. (1990). Multiple processes by which attitudes guide behavior: The MODE model as an integrative framework. *Advances in Experimental Social Psychology*, 23, 75–109.

Feloni, R. (2016). *Here's the presentation Google gives employees on how to spot unconscious bias at work*. Retrieved from <http://www.businessinsider.com/google-unconscious-bias-training-presentation-2015-12>

Fendrich, J. M. (1967). A study of the association among verbal attitudes, commitment and overt behavior in different experimental situations. *Social Forces*, 347–355.

Ferguson, C. J. (1983). In the eye of the beholder: Thin-ideal media affects some, but not most, viewers in a meta-analytic review of body dissatisfaction in women and men. *Journal of Educational Statistics*, 8(2), 157–159.

Finch, W. H., Bolin, J. E., & Kelley, K. (2014). *Multilevel modeling using R*. Boca Raton, FL: CRC Press.

Fisher, R. A. (1921). On the probable error of a coefficient of correlation deduced from a

- small sample. *Metron*, 1, 3–32.
- Fisher, Z., & Tipton, E. (2015). Robumeta: An R-package for robust variance estimation in meta-analysis. *ArXiv Preprint ArXiv:1503.02220*.
- Fiske, S. T. (2000). Stereotyping, prejudice, and discrimination at the seam between the centuries: Evolution, culture, mind, and brain. *European Journal of Social Psychology*, 30(3), 299–322.
- Ford, T. E., Boxer, C. F., Armstrong, J., & Edel, J. R. (2008). More than “just a joke”: The prejudice-releasing function of sexist humor. *Personality and Social Psychology Bulletin*, 34(2), 159–170.
- Fox, J. (2016). *Polycor: Polychoric and polyserial correlations*. Retrieved from <https://CRAN.R-project.org/package=polycor>
- French, D. (2017). *Implicit bias gets an explicit debunking*. Retrieved from <http://www.nationalreview.com/article/443723/implicit-bias-debunked-study-disputes-effects-unconscious-prejudice>
- Frey, D. L., & Gaertner, S. L. (1986). Helping and the avoidance of inappropriate interracial behavior: A strategy that perpetuates a nonprejudiced self-image. *Journal of Personality and Social Psychology*, 50(6), 1083–1090.
- Friese, M., Hofmann, W., & Schmitt, M. (2009). When and why do implicit measures predict behaviour? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology*, 19(1), 285–338.
- Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality*, 23(5), 369–401.
- Glasman, L. R., & Albarracín, D. (2006). Forming attitudes that predict future behavior: A meta-analysis of the attitude-behavior relation. *Psychological Bulletin*, 132(5), 778–822.
- Govorun, O., & Payne, B. K. (2006). Ego-depletion and prejudice: Separating automatic

- and controlled components. *Social Cognition*, 24(2), 111–136.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, 102(1), 4–27.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.
- Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1), 17–41.
- Hankins, M. (2013). *Still*. Retrieved from <https://mchankins.wordpress.com/2013/04/21/still-not-significant-2/>
- Hayward, J. (2016). *Implicit bias: The new 'original sin'*. Retrieved from <http://www.breitbart.com/big-government/2016/10/08/implicit-bias-youre-guilty-even-though-youre-innocent/>
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65.
- Henry, P. J., & Sears, D. O. (2002). The symbolic racism 2000 scale. *Political Psychology*, 23(2), 253–283.
- Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, 31(10), 1369–1385.
- Holley, P., Bever, L., Lowery, W., & Mettler, K. (2017). *Texas officer who killed black teenager leaving party is charged with murder*. Retrieved from <https://www.washingtonpost.com/news/post-nation/wp/2017/05/05/texas-officer-who-killed-black-teenager-leaving-party-is-charged-with-murder/>
- Hox, J. J., Moerbeek, M., & Schoot, R. van de. (2010). *Multilevel analysis: Techniques and*

- applications*. Routledge.
- Huet, E. (2015). *Rise of the bias busters: How unconscious bias became Silicon Valley's newest target*. Retrieved from <https://www.forbes.com/sites/ellenhuet/2015/11/02/rise-of-the-bias-busters-how-unconscious-bias-became-silicon-valleys-newest-target/>
- Hyde, J. S., & Linn, M. C. (2006). Gender similarities in mathematics and science. *Science*, 314(5799), 599–600.
- Jackson, D., Riley, R., & White, I. R. (2011). Multivariate meta-analysis: Potential and promise. *Statistics in Medicine*, 30(20), 2481–2498.
- Jacobs, P., & Viechtbauer, W. (2016). Estimation of the biserial correlation and its sampling variance for use in meta-analysis. *Research Synthesis Methods*.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. New York, NY: Springer Science & Business Media.
- Katz, I., & Hass, R. G. (1988). Racial ambivalence and american value conflict: Correlational and priming studies of dual cognitive structures. *Journal of Personality and Social Psychology*, 55(6), 893.
- Kim, M.-S., & Hunter, J. E. (1993). Attitude-behavior relations: A meta-analysis of attitudinal relevance and topic. *Journal of Communication*, 43(1), 101–142.
- Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods*, 2(1), 61–76.
- Kraus, S. J. (1995). Attitudes and the prediction of behavior: A meta-analysis of the empirical literature. *Personality and Social Psychology Bulletin*, 21(1), 58–75.
- Lakens, D., & Etz, A. J. (2017). Too true to be bad: When sets of studies with significant and nonsignificant findings are probably true. *Social Psychological and Personality Science*.
- Lalonde, R. N., Doan, L., & Patterson, L. A. (2000). Political correctness beliefs, threatened identities, and social attitudes. *Group Processes & Intergroup Relations*, 3(3),

317–336.

- Lambert, J. E., Engh, R., Hasbun, A., & Holzer, J. (2012). Impact of posttraumatic stress disorder on the relationship quality and psychological distress of intimate partners: A meta-analytic review. *Journal of Family Psychology, 26*(5), 729–737.
- LaPiere, R. T. (1934). Attitudes vs. actions. *Social Forces, 13*(2), 230–237.
- Legault, L., Gutsell, J. N., & Inzlicht, M. (2011). Ironic effects of antiprejudice messages: How motivational interventions can reduce (but also increase) prejudice. *Psychological Science, 22*(12), 1472–1477.
- Lewin, K. (1951). *Field theory in social science: Selected theoretical papers*. Harper & Brothers.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology, 22*(140), 5–55.
- Lindquist, K. A. (2013). Emotions emerge from more basic psychological ingredients: A modern psychological constructionist model. *Emotion Review, 5*(4), 356–368.
- Lindquist, K. A., Satpute, A. B., Wager, T. D., Weber, J., & Barrett, L. F. (2015). The brain basis of positive and negative affect: Evidence from a meta-analysis of the human neuroimaging literature. *Cerebral Cortex, 26*(5), 1910–1922.
- Linn, L. S. (1965). Verbal attitudes and overt behavior: A study of racial discrimination. *Social Forces, 43*(3), 353–364.
- Lombroso, D., & Appelbaum, Y. (2017). 'Hail Trump!': White nationalists salute the president-elect. Retrieved from <https://www.theatlantic.com/politics/archive/2016/11/richard-spencer-speech-npi/508379/>
- Majumder, M. (2017). *Higher rates of hate crimes are tied to income inequality*. Retrieved from <https://fivethirtyeight.com/features/higher-rates-of-hate-crimes-are-tied-to-income-inequality/>
- McConahay, J. B., Hardee, B. B., & Batts, V. (1981). Has racism declined in america? It depends on who is asking and what is asked. *Journal of Conflict Resolution, 25*(4),

563–579.

Middlebrook, H. (2017). *The fascinating, if unreliable, history of hate crime tracking in the US*. Retrieved from

<http://www.cnn.com/2017/01/05/health/hate-crimes-tracking-history-fbi/>

Mitchum, R., Genova, J., & Taylor, L. (2017). *Police project update: Expanding and implementing the early intervention system*. Retrieved from

<https://dssg.uchicago.edu/2017/01/12/police-project-update-expanding-and-implementing-the-early-intervention-system/>

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81(1), 33–43.

Monteith, M. J., Deneen, N. E., & Tooman, G. D. (1996). The effect of social norm activation on the expression of opinions concerning gay men and blacks. *Basic and Applied Social Psychology*, 18(3), 267–288.

Murphy, K. R., & Myers, B. (1999). Testing the hypothesis that treatments have negligible effects: Minimum-effect tests in the general linear model. *Journal of Applied Psychology*, 84(2), 234–248.

Nakagawa, S., Noble, D. W., Senior, A. M., & Lagisz, M. (2017). Meta-evaluation of meta-analysis: Ten appraisal questions for biologists. *BMC Biology*, 15(18), 1–14.

Notani, A. S. (1998). Moderators of perceived behavioral control's predictiveness in the theory of planned behavior: A meta-analysis. *Journal of Consumer Psychology*, 7(3), 247–271.

Olson, E. G. (2015). *How corporate America is tackling unconscious bias*. Retrieved from [http:](http://fortune.com/2015/01/15/how-corporate-america-is-tackling-unconscious-bias/)

[//fortune.com/2015/01/15/how-corporate-america-is-tackling-unconscious-bias/](http://fortune.com/2015/01/15/how-corporate-america-is-tackling-unconscious-bias/)

Omi, M., & Winant, H. (1994). *Racial formation in the United States: From the 1960s to the 1990s*. New York, NY: Routledge.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational*

- Statistics*, 8(2), 157–159.
- Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, 105(2), 171–192.
- O’Neil, C. (2016). *Weapons of math destruction*. New York, NY: Crown.
- Paluck, E. L. (2011). Peer pressure against prejudice: A high school field experiment examining social network change. *Journal of Experimental Social Psychology*, 47(2), 350–358.
- Pearson, A. R., Dovidio, J. F., & Gaertner, S. L. (2009). The nature of contemporary prejudice: Insights from aversive racism. *Social and Personality Psychology Compass*, 3(3), 314–338.
- Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in Western Europe. *European Journal of Social Psychology*, 25(1), 57–75.
- Pew Research Center. (2012). *The rise of intermarriage: Rates, characteristics vary by race and gender*. Retrieved from <http://assets.pewresearch.org/wp-content/uploads/sites/3/2012/02/SDT-Intermarriage-II.pdf>
- Pew Research Center. (2016). *On views of race and inequality Blacks and Whites are worlds apart*. Retrieved from <http://www.pewsocialtrends.org/2016/06/27/on-views-of-race-and-inequality-blacks-and-whites-are-worlds-apart/>
- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75(3), 811–832.
- Plant, E. A., & Devine, P. G. (2001). Responses to other-imposed pro-Black pressure: Acceptance or backlash? *Journal of Experimental Social Psychology*, 37(6), 486–501.
- Polanin, J. R., Hennessy, E. A., & Tanner-Smith, E. E. (2017). A review of meta-analysis packages in R. *Journal of Educational and Behavioral Statistics*, 42(2), 206–242.
- Potok, M. (2016). *The year in hate and extremism*. Retrieved from <https://www.splcenter>.

org/fighting-hate/intelligence-report/2016/year-hate-and-extremism

Quillian, L. (2006). New approaches to understanding racial prejudice and discrimination.

Annual Review of Sociology, 32, 299–328.

Raden, D. (1973). *The relationship between prejudice and overt hostile behavior* (PhD dissertation). University of Wisconsin.

Richeson, J. A., & Shelton, J. N. (2007). Negotiating interracial interactions: Costs, consequences, and possibilities. *Current Directions in Psychological Science*, 16(6), 316–320.

Russell, J. A. (2003). Core affect and the psychological construction of emotion.

Psychological Review, 110(1), 145–172.

Saenger, G., & Gilbert, E. (1950). Customer reactions to the integration of Negro sales personnel. *International Journal of Opinion and Attitude Research*, 4(1), 57–76.

Satpute, A. B., Kang, J., Bickart, K. C., Yardley, H., Wager, T. D., & Barrett, L. F. (2015).

Involvement of sensory regions in affective experience: A meta-analysis. *Frontiers in Psychology*, 6, 1–14.

Saucier, D. A., Miller, C. T., & Doucet, N. (2005). Differences in helping whites and blacks:

A meta-analysis. *Personality and Social Psychology Review*, 9(1), 2–16.

Schütz, H., & Six, B. (1996). How strong is the relationship between prejudice and

discrimination? A meta-analytic answer. *International Journal of Intercultural Relations*, 20(3), 441–462.

Sechrist, G. B., & Stangor, C. (2001). Perceived consensus influences intergroup behavior

and stereotype accessibility. *Journal of Personality and Social Psychology*, 80(4), 645–654.

Severini, T. A. (2005). *Elements of distribution theory*. New York, NY: Cambridge

University Press.

Sheppard, B. H., Hartwick, J., & Warshaw, P. R. (1988). The theory of reasoned action: A

meta-analysis of past research with recommendations for modifications and future

- research. *Journal of Consumer Research*, 15(3), 325–343.
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014a). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 9(6), 666–681.
- Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014b). P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, 143(2), 534–547.
- Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Better p-curves: Making p-curve analysis more robust to errors, fraud, and ambitious p-hacking, a reply to ulrich and miller (2015). *Journal of Experimental Psychology: General*, 144(6), 1146–1152.
- Stangor, C., Sechrist, G. B., & Jost, J. T. (2001). Changing racial beliefs by providing consensus information. *Personality and Social Psychology Bulletin*, 27(4), 486–496.
- Steege, S., Tuerlinckx, F., Gelman, A., & Vanpaemel, W. (2016). Increasing transparency through a multiverse analysis. *Perspectives on Psychological Science*, 11(5), 702–712.
- Stolberg, S. G., & Rosenthal, B. M. (2017). *Man charged after White Nationalist rally in Charlottesville ends in deadly violence*. Retrieved from <https://www.nytimes.com/2017/08/12/us/charlottesville-protest-white-nationalist.html>
- Talaska, C. A., Fiske, S. T., & Chaiken, S. (2008). Legitimizing racial discrimination: Emotions, not beliefs, best predict discrimination in a meta-analysis. *Social Justice Research*, 21(3), 263–296.
- Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology*, 2(1), 85–112.
- Thrush, G., & Haberman, M. (2017). *Trump gives White supremacists an unequivocal boost*. Retrieved from <https://www.nytimes.com/2017/08/15/us/politics/trump-charlottesville-white-nationalists.html>
- Tipton, E. (2015). Small sample adjustments for robust variance estimation with

- meta-regression. *Psychological Methods*, 20(3), 375–393.
- Toosi, N. R., Babbitt, L. G., Ambady, N., & Sommers, S. R. (2012). Dyadic interracial interactions: A meta-analysis. *Psychological Bulletin*, 138(1), 1–27.
- Traub, A., & Ruetschlin, C. (2016). *The racial wealth gap: Why policy matters*. Retrieved from <http://www.demos.org/publication/racial-wealth-gap-why-policy-matters>
- U.S. Census Bureau. (2002). *Income, poverty, and health insurance coverage in the United States: 2011*. Retrieved from <https://www.census.gov/prod/2012pubs/p60-243.pdf>
- U.S. Department of Justice. (2012). Retrieved from <https://www.justice.gov/opa/pr/justice-department-reaches-settlement-wells-fargo-resulting-more-175-million-relief>
- Urban Institute. (2015). *Wealth inequalities*. Retrieved from <http://www.urban.org/policy-centers/cross-center-initiatives/inequality-and-mobility/projects/wealth-inequalities>
- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, 139(2), 352–402.
- Van Bavel, J. J., & West, T. V. (2017). *Seven steps to reduce bias in hiring*. Retrieved from <https://www.wsj.com/articles/seven-steps-to-reduce-bias-in-hiring-1487646840>
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, 45(2), 576–594.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48.
- Wallace, D. S., Paulson, R. M., Lord, C. G., & Bond Jr., C. F. (2005). Which behaviors do attitudes predict? Meta-analyzing the effects of social pressure and perceived difficulty. *Review of General Psychology*, 9(3), 214–227.
- Wang, A. B. (2017). *“Final act of bravery”: Men who were fatally stabbed trying to stop anti-Muslim rants identified*. Retrieved from

<https://www.washingtonpost.com/news/post-nation/wp/2017/05/27/>

[man-fatally-stabs-2-on-portland-ore-train-after-they-interrupted-his-anti-muslim-rants-police-say/](#)

White, M. H., II, & Crandall, C. S. (2017). Freedom of racist speech: Ego and expressive threats. *Journal of Personality and Social Psychology*.

Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues*, 25(4), 41–78.

Woodmansee, J. J., & Cook, S. W. (1967). Dimensions of verbal racial attitudes: Their identification and measurement. *Journal of Personality and Social Psychology*, 7(3), 240–250.

Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35(2), 151–175.

Zitek, E. M., & Hebl, M. R. (2007). The role of social norm clarity in the influenced expression of prejudice over time. *Journal of Experimental Social Psychology*, 43(6), 867–876.