Racial Threat and Racial Disparities in Jails: The Mediating Role of Implicit Bias

Christopher J. Marier and Lorie A. Fridell

Department of Criminology, University of South Florida, Tampa, FL

Corresponding author: Christopher J. Marier, cmarier@mail.usf.edu, 4202 E. Fowler Ave. #SOC310, Tampa, FL 33620 https://orcid.org/0000-0002-2445-6315

Christopher J. Marier is a PhD candidate at the University of South Florida. His areas of interest include race and justice, policing, and cross-national research. He is a recipient of the University of South Florida Graduate Fellowship Award.

Lorie A. Fridell is Professor of Criminology at the University of South Florida, former Director of Research at the Police Executive Research Forum, and CEO of *Fair and Impartial Policing*, a national law enforcement training program.

NOTE: Draft version 1.0, 6/10/2019. This paper has not been peer reviewed. This paper has not yet been published and is therefore not the authoritative document of record. Please do not copy or cite without authors' permission.

Racial Threat and Racial Disparities in Jails: The Mediating Role of Implicit Bias

The mixed findings on tests of Racial Threat Theory highlight the need to understand the mechanism(s) that might explain how percent black might impact disparate formal social control. The present study examines the mediating impact of anti-black implicit bias. The study advances the theory and research by measuring implicit bias using groundbreaking data on levels of anti-black implicit bias for U.S. counties and by testing—as part of the mediation examination—competing sociological and psychological predictions regarding whether higher proportions of blacks in a population produce more or less anti-black prejudice. Results indicate that (1) higher proportions of blacks in a population produce less bias, (2) racial disparities in county jails are lower in areas with a larger share of black residents and (3) this latter finding is attributable almost entirely to a reduction in implicit bias, which fully mediates the relationship between percent black and jail confinement disparities.

Keywords: racial threat; intergroup contact; implicit bias; percent black

Introduction

Racial Threat Theory predicts that large or increasing proportions of blacks in a population will result in increased formal social control, especially disparate formal social control. Blalock (1967) originally identified three types of threat—economic, political and symbolic—that might explain why large or increasing numbers of blacks could lead to increases in police force size, arrests, sentences to prison, and so forth. In part, because the empirical tests of this theory have produced mixed results, researchers and theorists continue to explore *why it is* that population makeup might lead to formal social control.

Bias towards minority groups has been identified as a possible mediator between population makeup and formal social control. This proposition reflects the conceptual contributions to Racial Threat Theory of Blumer who explained "threat" in terms of group-based prejudice. Whereas social psychologists have suggested that racial bias emanates from psychological dispositions and personal experiences, Blumer suggested that racial bias is a product of group position. To date, just two studies have formally assessed the mediating impact of bias on the relationship between population makeup and formal social control (or *attitudes* toward formal social control)—producing mixed results. Ousey and Unnever (2012) in their multi-country research, found that prejudice partially mediated the relationship between population diversity and punitive attitudes. Stults and Baumer (2007), in contrast, did not find a mediating impact of prejudice on the relationship between racial composition and police force size in counties.

The explanation for these inconsistent results may be theoretical or methodological, or both. Both of these studies, because they were couched theoretically in Racial Threat assumed that a high or growing percent of blacks in a population would *increase* negative attitudes toward them. The Intergroup Contact Theory from social psychology, however, predicts the opposite. This theory predicts that a high or growing percent of blacks in a population would *decrease* negative attitudes toward them because it would provide more opportunities for positive exposure/interaction between racial groups. This theory, therefore, would predict a negative relationship between population makeup and disparate social control.

In terms of methodology, a deficiency of prior studies is the fact that the researchers measured explicit (conscious), instead of implicit biases. The two key issues associated with measuring explicit, versus implicit, bias are that (1) explicit bias is highly subject to socially desirable responses, and (2) implicit bias is thought to be the "modern" form of how bias manifests in people. It is only in recent decades that science has recognized the existence of implicit bias and only in recent years that researchers have had access to aggregate data on implicit biases for geographical areas.

The purpose of the current study is to assess the impact of black composition in counties on racial disparities in county jail populations and assess the mediating impact of racial prejudice on the relationship between population makeup and disparate social control. In so doing, the study tests competing predictions of how a large or growing minority population impacts on bias. Importantly, the current study further advances the exploration of bias as a potential mediator in the Racial Threat equation by measuring *implicit*, not explicit, bias. A series of multilevel structural equation models are used to assess the relationships based on data for 437 counties. The independent variable is the percent of blacks in the county population based on the 2010 Census; the dependent variable, reflecting disparate formal social control, is racial disparity in detention in county jails. Implicit anti-black bias, the hypothesized mediator, is measured using data produced by the Harvard-based Project Implicit.

Racial threat theory

Racial/ethnic disparities in the criminal justice system are well documented (for an overview, see e.g., Walker, Spohn, & DeLone, 2018). Research, for instance, has found disproportionate representation of racial/ethnic minorities as subjects of various police activities such as arrests or tickets (e.g., Durose & Langton, 2013; Kochel, Wilson, & Mastrofski, 2011), use of force (e.g., Eith & Durose, 2011; Gabrielson, Sagara, & Jones, 2014; Lautenschlager & Omori, 2018), searches (e.g., Baumgartner, Epp, & Love, 2014; Durose & Langton, 2013; Eith & Durose, 2011; Higgins, Vito, & Grossi, 2012), and pedestrian or vehicle stops (e.g., Durose & Langton, 2013; Gelman, Fagan, & Kiss, 2007). Racial/ethnic disparities have also been documented in the workings of prosecutors and defense attorneys (Kutateladze, 2018; Kutateladze, Andiloro, Johnson, & Spohn, 2014; Shermer & Johnson, 2010), judges (U.S. Sentencing Commission, 2018; Salman, Le Coz, & Johnson, 2016), juries (Eberhardt, Davies, Purdie-Vaughns, & Johnson, 2006; Levinson, Cai, & Young, 2010), and others.

While differential offending is one possible contributor to racial/ethnic disparities in the criminal justice system (Hipp, 2011; Loeber et al., 2015; Mears, Cochran, & Lindsey, 2016; Sampson, Morenoff, & Raudenbush, 2005), various theories have been put forth to explain disparities that cannot be wholly explained by differential offending. Key among these theories is Blalock’s Racial Threat Theory and related group-threat propositions, which are situated within the conflict perspective.

Blalock (1967) developed Racial Threat Theory to explain macro-level discrimination. The key proposition of this theory is that large or growing minority populations produce discrimination. Specifically, a large or growing minority population produces perceptions of threat on the part of the majority and this perception leads to various increases in formal social control, particularly those associated with the criminal justice system, such as police force size or expenditures, arrests, convictions, sentences, and so forth.

There has been considerable research testing Racial Threat Theory. Studies have examined the association between minority population size (or growth) and various forms of formal social control linked to criminal justice, such as police use of force (Hehman, Flake, & Calanchini, 2018; Smith & Holmes, 2003), police force size (Kent & Jacobs, 2005), police expenditures (Kent & Jacobs, 2004), arrests (Parker, Stults, & Rice, 2005), sentencing (Johnson, Ulmer, & Kramer, 2008), imprisonment (Greenberg & West, 2001; Jacobs & Carmichael, 2001; Ulmer & Johnson, 2004), and imposition of the death penalty (Baumer, Messner, & Rosenfeld, 2003; Jacobs & Carmichael, 2002). Racial Threat Theory has also been tested looking at forms of social control outside of criminal justice including lynchings, interracial killings, and hate crimes (see e.g., Green, Strolovitch, & Wong 1998; Jacobs & Wood, 1999). Some of the Racial Threat research examines *attitudes linked to formal social control* rather than social control per se. This includes studies using survey methodology to examine the impact of population makeup on punitive attitudes (King & Wheelock, 2007; Ousey & Unnever, 2012; Stewart, Martinez, Baumer, & Gertz, 2015).

Many studies have provided empirical support for Racial Threat Theory (e.g., D'Alessio, Eitle, & Stolzenberg, 2005; Greenberg & West, 2001; Jacobs & Carmichael, 2001, 2004;

Johnson et al., 2008; Kent & Jacobs, 2005; Mosher, 2001; Sharp, 2006; Smith & Holmes, 2003). The research literature is not consistent, however, in showing that higher proportions of racial/ethnic minorities in a population are associated with greater levels of formal social control. Some studies produced findings that are inconsistent with Racial Threat Theory (e.g., Crawford, Chiricos, & Kleck, 1998; Feldmeyer & Ulmer, 2011; Holmes, Painter, & Smith, 2018; Leiber, Peck, & Rodriguez, 2016; Parker et al., 2005; Stolzenberg, D'alessio, & Eitle, 2004; Stults & Baumer, 2007; Thomas, Moak, & Walker, 2013).

Feldmeyer and Cochran (2018) argue that the "mixed results" highlight the need to understand what phenomenon might *explain* the relationship between population make up and formal social control. They report that, "racial threat theory and our understanding of the precise interworking of population composition, threat, and social control is far from settled" (p. 284). The remaining "theoretical ambiguities" include a less than full articulation of *why* population makeup might impact formal social control. Consistent with this important gap in understanding, researchers and theorists have attempted to identify mediating mechanisms.

The search for mediating mechanisms

Blalock originally identified economic, political, and symbolic threats as the key "intervening variables" between large or increasing minority populations and formal social control. Regarding the former, racial/ethnic minorities might compete for jobs and other resources, threatening the privileged economic status of the majority population. A large or increasing minority population might additionally or alternatively produce a threat to political power as the number of minority voters and candidates increase. Symbolic threat reflects the perceived link on the part of the majority population between racial/ethnic minorities and criminal activity.

A handful of studies have explored the potential mediating impacts of political, economic and/or symbolic threat on the relationship between population makeup and formal social control or punitive attitudes. King and Wheelock (2007) found that perceived economic threat (defined as a "strain on public resources") and perceptions of minorities as a threat to public safety partially mediate the relationship between the growth of minority population and punitive attitudes. Stewart, Martinez, Baumer and Gertz (2015) found similar results. This team examined the mediating impact of perceived criminal, economic and political threat on the relationship between minority population size and growth on punitive attitudes—finding that economic and criminal threats, but not political threats, were partial mediators. (See also Stults & Baumer, 2007; in contrast, see Ousey & Lee, 2008.)

As above, there is some partial support for Blalock's originally-identified threats. That those threats do not fully mediate the relationship between population makeup and formal social control highlights the importance of Blalock's comment that his propositions were just a starting point. He suggested that other phenomena might alternatively or additionally be at work and, indeed, current theorists and researchers have continued the quest to identify potential mediators. This exploration includes consideration of *non-threat* possibilities; indeed, in their review and conceptual analysis of Racial Threat Theory, Feldmeyer and Cochran (2018) suggest that future research might determine that it is not actually "threat," but rather "some other phenomenon" that explains the relationship between population makeup and formal social control (p. 299–300).

(See also Stults & Swagar, 2018) The section below discusses one such ”other phenomenon”—anti-black bias.

Racial bias as the mediator

Anti-black bias/prejudice has been identified as a potential mediator between population makeup and formal social control.[1] In her article that summarizes the Racial Threat literature and identifies direction for future research, Dollar (2014) suggests that “prejudicial attitudes mediate the link between minority population size and support for racialized social control practices” (p. 4). This proposition that anti-black prejudice or bias is the link between population composition and mechanisms of formal social control can be traced to the very roots of Racial Threat Theory and reflects the conceptual parallel between Blumer and Blalock. Blalock, as above, claimed that discrimination occurs when groups feel threatened. It was Blumer, writing a decade before Blalock, who explained “threat” in terms of group-based prejudice. Contrary to social psychologists who characterize bias is an individual attitude produced from psychological dispositions and personal experiences, sociologists emphasize group-based functions and structural dynamics (John F. Dovidio, Hewstone, Glick, & Esses, 2010). Blumer, a sociologist, suggested that racial prejudice is a product of group position. He argued that (1958, p. 3), “…prejudice is fundamentally a matter of relationship between racial groups.” According to Blumer, members of the majority group have hostile feelings toward other groups, partly because they think other groups are different and inferior, but also because members of the dominant group have a sense of entitlement and a corresponding fear that their privileged position may be

[1]This article uses “prejudice” and “bias” interchangeably to refer to valenced evaluations of groups.

threatened by a minority group, particularly as that group increases in size. Blalock wrote (1967, p. 5): "The source of race prejudice lies in a felt challenge to this sense of group position." As Chiricos, Pickett, and Lehmann (Chiricos, n.d.) explain: Blumer postulated that an expanding racial group might produce a "felt challenge" and then Blalock built on this formulation, focusing on how this challenge might produce discrimination, including discrimination in the form of formal social control to contain the threat.

The two studies that have assessed the mediating impact of bias on the relationship between minority group representation and punitiveness or social control produced mixed and thus inconclusive results. Ousey and Unnever (2012), in their multi-country research, assessed the mediating impact of attitudes reflecting minority group intolerance on the relationship between ethnic minority group representation and punitiveness. The independent variable, minority group representation, was measured in terms of the "relative presence of ethnic minority groups in the population of a country" (p. 579). The dependent measure reflected attitudes about criminal punishment based on the respondent's level of agreement with the statement: "Nowadays there is too much tolerance. Criminals should be punished more severely." Prejudice or rather "outgroup intolerance," as the mediator, was measured using respondent's level of agreement with the statement, "Immigrants contribute a lot to our country." Consistent with their expectations couched in Racial Threat Theory, the researchers found a positive relationship between population diversity and their measure of punitive attitudes and that their measure of prejudice partially mediated the relationship between the two. They report, "…our results indicate that a greater prevalence of out-group populations increases out-group animus, which in turn increases the likelihood that the public desires the harsher treatment of criminals" (p. 589).

Stults and Baumer (2007) measured the mediating impact of prejudice on the relationship between "racial context" (p. 509) and police force size. They found a significant curvilinear relationship between percent black and police force size; in contrast to Ousey and Unnever (2012), they did not find a mediating impact of prejudice. Using counties and county groups as their unit of analysis, the researchers measured "racial context" in terms of both racial composition and degree of segregation using Census data. The formal social control dependent variable was the number of sworn police officers per 100K residents. "Antiblack racial prejudice," one of several mediator variables tested, was measured using six survey items from the General Social Survey (GSS). County (or county group) measures of bias included the percent of whites who opposed having a family member or close relative marry a Black person and who agreed that they should be able to keep Blacks out of their neighborhoods if they wanted to do so. In Stults and Baumer's analysis, prejudice did not mediate, even partially, the relationship between racial context and formal social control.

The explanation for the inconsistent results may be theoretical or methodological, or both. Regarding the former, the mediation equation in the context of Racial Threat Theory implies that there are positive relationships between (a) level of racial/ethnic minorities in a population and bias against them and (b) bias and formal social control. While the latter has consistent empirical and theoretical support, the former does not. While as above, Racial Threat Theory assumes that a large or increasing minority population will increase negative attitudes toward that population, there is an alternative theory that predicts the opposite. According to the

Intergroup Contact Theory, which emanates from social psychology, higher levels of racial/ethnic minorities could *reduce* negative attitudes towards them. According to this theory, an individual's biases about a group can be reduced through positive interactions with members of that group (Allport, 1954). In describing the Contact Theory, Tausch and Hewstone (2010, p. 544) write, "The notion that contact between members of different groups can, under certain conditions, reduce prejudice is one of the most prominent ideas underlying approaches to improve intergroup relations." There is considerable empirical support for the Contact Theory (see e.g., Thomas F. Pettigrew & Tropp, 2006), who conducted a meta-analysis of 515 studies on the Contact Theory).

Thus, Racial Threat Theory and Intergroup Contact Theory make competing predictions about the effect of racial composition on bias (which, acting as a mediator, predicts subsequent social control). That is, according to Racial Threat Theory larger black populations will be associated with negative attitudes, such as prejudice, which will increase discriminatory social control. Alternatively, Intergroup Contact Theory provides reason to believe that bias would drop as black populations increase, therefore reducing social control disparities.

Another potential explanation for the inconclusive results from studies assessing the mediating impact of minority group bias on the relationship between population makeup and formal social control is the manner in which bias is measured. Indeed, Stults and Baumer report that a key limitation of their mediation assessment pertains to their measure of prejudice. They point out that, in recent years, researchers have identified not one, but two forms of prejudice—the "'old-fashioned' or 'blatant' prejudice" and the "'modern' or 'symbolic' prejudice." Not only is the "modern" form of prejudice thought to be more pervasive (Devine, 1989; Hardin & Banaji, 2013), but it is also less subject to social desirability in its measurement (Kim, 2003). The authors report that (p. 535), “Our null finding for the effect of antiblack prejudice on police size therefore could be due to the fact that our antiblack prejudice scale is composed of items that better tap into old-fashioned rather than symbolic prejudice.” Dollar (2014, p. 4), too, expressed concern about how prejudice has been measured in the Racial Threat literature; she comments: “(s)ociologists have previously argued that prejudice is difficult to measure, especially given the covert and subtle ways that race-based notions are currently expressed.”

Implicit bias

As argued by both Stults and Baumer (2007) and Dollar (2014), the measure of prejudice used in previous mediation studies has reflected “old fashioned” bias rather than “modern” bias. The two issues with this mode of measuring “prejudice” are that (a) measuring old fashioned bias is highly subject to socially desirable responses; and (2) it does not measure what many social psychologists believe is the more predominant form of bias today.

The references to ”old fashioned” and “modern bias” correspond to what social psychologists now refer to as explicit and implicit bias, respectively. This section characterizes the difference between explicit and implicit bias, reviews how racial threat researchers have measured explicit prejudice, reports how problems with explicit-bias measures (such as those used by Racial Threat researchers) led to the discovery of implicit bias, and describes how implicit bias is measured.

Explicit v. implicit bias

Explicit bias is generally what one envisions when thinking about prejudice and bias. With explicit biases, a person associates various groups with negative characteristics. These attitudes are based on animus or hostility toward the groups (Amodio & Mendoza, 2010). As an example, a racist has explicit biases. Explicit biases can impact a person's perceptions and behavior, producing discriminatory behavior. It is conscious and deliberate. That is, the person with explicit bias is aware of his/her hostility toward groups and will justify that hostility; s/he is generally unconcerned about the resulting discriminatory behavior (A. G. Greenwald & Banaji, 1995; Hardin & Banaji, 2013; Staats, 2013).

Implicit biases are similar to explicit biases in that people link individuals to stereotypes or generalizations associated with their group or groups; these are called "implicit associations" (for reviews see Hardin & Banaji, 2013; Staats, 2013; Staats, Capatosto, Wright, & Jackson, 2016). As with explicit biases, these associations can impact perceptions and behavior producing discriminatory behavior (Dasgupta, 2004; John F. Dovidio et al., 2010; Kang, Bennett, Carbado, & Casey, 2011; McConnell & Leibold, 2001). Unlike explicit biases, implicit biases are *not based* on animus or hostility and these implicit associations can impact perceptions and behavior *outside of conscious awareness* (Devine, 1989; Richard E. Petty, Fazio, & Brinol, 2012). Even individuals who, at the conscious level, reject prejudice and stereotyping, can and do manifest implicit biases (Graham & Lowery, 2004; Kang et al., 2011).

Challenges to measuring explicit bias and the discovery of implicit bias

As above, Racial Threat researchers who have incorporated prejudice into their work have measured only explicit biases.[2] Quillian (1996) and Stults and Baumer (2007) used the General Social Survey (GSS) for their studies. Quillian's measures of prejudice included items soliciting degree of agreement with statements, such as "Blacks shouldn't push themselves where they're not wanted" and "White people have the right to keep Blacks out of their neighborhood if they want to, and Blacks should respect that right." Quillian also assessed support or opposition to strategies designed to redress racial/ethnic inequality. For instance, subjects were asked "Are we spending too little, about the right amount, or too much on improving the condition of Blacks?" Stults and Baumer's version of the GSS asked respondents whether they would oppose a "close relative or family member marrying a black person." Giles and Evans (1986) used a common measure of attitude toward blacks; on a "thermometer scale" that ranges from 0 to 100, respondents indicate how "warm" they feel towards blacks.

To create a racial prejudice index for his European study, Quillian's (1996) respondents were read statements, such as "they exploit social security benefits" or "marrying into one of these groups always ends badly," and then were asked to what groups each statement applied, such as "people of another race." There were also two other questions, including "Do you personally, in your daily life, find disturbing the presence of people of another race?" As above, in their multi-country European research, Ousey and Unnever (2012) relied on a single Likert-agreement question to measure prejudice: "Immigrants contribute a lot to our country."

[2]Hehman, et al. (2017) assessed the relationship between bias and disproportionate use of force by police using a measure of implicit bias, but they did not write in the context of Racial Threat Theory.

Hardin and Banaji (2013) describe how various developments and circumstances—including challenges associated with measuring explicit bias—led to the discovery of implicit bias. “Most politically salient,” they suggest (p. 14), was the fact that racial discrimination in its various forms (e.g., social, economic) persisted through the late 20th century, despite the fact that individuals were less and less inclined to endorse or express racist attitudes. The recognition of this reticence to express bias led social scientists to attempt to measure prejudice “unobtrusively.” They wanted to bypass the social-desirability effects of common prejudice measures and/or try to understand why discrimination persisted even when society seemed less and less inclined to endorse it. These aspirations corresponded in time with the growing recognition within psychology of the extent to which information processing occurs outside of conscious awareness and with the scientific abilities to measure implicit cognition. Hardin and Banaji (2013), credit Patricia Devine (in her 1989 article) with documenting and describing the culmination of these forces; she joined implicit cognition and stereotyping, “(marking) the beginning of a paradigm shift in the social-psychological understanding of stereotyping and prejudice more generally” (p. 14–15). Implicit bias had been discovered.

Measuring implicit bias

There are various methods that have been used to measure implicit associations in subjects (for a review, see Staats, 2013). For instance, one method involves ”priming.” In such studies, the subject is exposed to an initial stimulus (the “prime”) that is hypothesized to influence a subsequent response and thereby confirm (or disconfirm) an implicit bias. (Eberhardt, Goff, Purdie, & Davies, 2004), for instance, assessed whether “priming” subjects subliminally with black male faces or white male faces impacted the subjects’ subsequent ability to identify degraded images of crime-related objects. Subjects were shown blurry pictures of objects that became more and more clear with each advancing frame; the researchers measured how long it took the subjects (in terms of frames) to identify the objects. The researchers argued that, if the "black-crime" association exists, exposure to black male faces during the priming phase of the study would make crime images more accessible. The results supported a strong black-crime association.

Another methodology involves response latency measures (see e.g., J. F. Dovidio, Kawakami K Smoak, & Gaertner, n.d.; A. G. Greenwald, McGhee, & Schwartz, 1998; Kang & Lane, 2010; Rudman, 2004). Staats explains (2013, p. 24): "These measures rely on reaction times to specific tasks in order to uncover individual's biases … The underlying premise of these reaction-time studies is that individuals are able to complete cognitively simple tasks relatively more quickly than those that are mentally challenging." An example of a latency measure is the popular Implicit Association Test (IAT) from Project Implicit that is available on line (@ https://implicit.harvard.edu/implicit/). (For an overview, see Lane, Banaji, Nosek, & Greenwald, 2007.) In taking the IAT, respondents are timed as they sort concepts. Since we are faster at completing simple tasks than challenging ones, we will be faster "sorting" concepts that are *linked* in our heads than sorting concepts that are *not linked*. As such, if a person is faster at linking "women" and "childcare" than "men" and "childcare," the implication is that the person has an "association" in his/her head between women and childcare (but not men and childcare). One example is the Race IAT, in which the respondents are directed to categorize white and black faces with positive and negative words. As Staats explains (2013, p. 27), "(f)aster reaction times when pairing White faces with positive words and Black faces with negative terms suggests the presence of implicit pro-White/anti-Black bias." Latency measures have been used in various implicit bias studies, including those examining whether, in laboratory settings, police and non-police subjects are more likely to shoot an unarmed Black subject than an unarmed White subject (e.g., Correll, Park, Judd, Wittenbrink, et al., 2007; Sadler, Correll, Park, & Judd, 2012).

The current study

As summarized above, Racial Threat Theory predicts that the percent of blacks in a population produces discriminatory social control, and yet studies are not consistent in their support for this proposition. The mixed findings highlight the need to understand *what it is* that might explain the relationship between population make up and formal social control. Is it the economic, political and/or symbolic threats posed by a large or increasing minority population, or is it something else? Various theorists and researchers have explored the factors that might mediate the relationship between population makeup and discriminatory social control, and minority group bias is a potential mediator that has been proposed and tested.

The Racial Threat prediction for this mediating variable is that the large or increasing presence of minority group members in a population will increase the negative attitudes toward them and result in increased discriminatory social control. Figure 1 provides an illustration of the key variables and relationships explored in this study.[3] Racial Threat Theory would predict a positive relationship between percent black and racial disparity in jail populations (path c'). With regard to potential mediation, Racial Threat Theory would predict positive relationships between

[3] For clarity, error terms and random intercepts are depicted, but control variables are not. Several paths are identified according to conventional mediation analysis (Baron and Kenny 1986).

(1) percent black and bias (see path a), and (2) bias and disparity in jail populations (see path b). Theoretically, Contact Theory challenges those predictions. As with Racial Threat Theory, Contact Theory would predict a positive relationship between bias and disparate formal social control (path b). Contact Theory, however, would predict a negative relationship between percent black and racial disparity in jail populations (path c'), because Contact Theory would predict a negative relationship between percent black and bias (path a).

[FIGURE 1 ABOUT HERE]

To conduct a meaningful assessment of the role of prejudice in explaining the relationship between population, a valid measure of bias must be used. Whereas previous research examining the role of bias in the Racial Threat context has used information on explicit bias, this study will take advantage of data on implicit bias that only relatively recently became available to researchers.

The present study uses a series of multilevel structural equation models to evaluate (1) the impact of black composition of a county on racial disparities in detention in county jails, and (2) the mediating impact of prejudice on the relationship between population makeup and formal social control. It advances the research in this area in two ways. First, as part of the mediation examination, the study explores whether higher proportions of blacks in the population produce more anti-black prejudice as Racial Threat Theory posits, or less prejudice as predicted by the Intergroup Contact Theory. Second, the study uses a measure of implicit bias, instead of explicit bias.

Methods

Data from 437 counties were used to evaluate the impact of the size of the black population on racial disparity in jail populations and to assess the mediating impact of anti-black implicit bias. As detailed below, the data for the independent variable, the dependent variable, the mediating variable and control variables come from the U.S. census, the jail census, and the Harvard-based Project Implicit.

Sample

The sampling frame for this study consists of all U.S. counties with populations of 100,000 or more residents; the exclusion of small counties ensures large enough populations to calculate stable estimates of both implicit bias and per-capita rates of formal social control.[4] As of the 2010 census, 574 such counties existed. These 574 counties encompass nearly 78% of all U.S. residents. After listwise deletion, the final sample consists of 437 counties (clustered within 42 states). The exclusion of counties is generally due to missing arrest or incarceration data; the jail census used as our dependent variable does not include data from Alaska, Connecticut, Delaware, Hawaii, Rhode Island, and Vermont, as these states use combined jail-prison systems. Missing data analysis indicated that counties in these states are not statistically significantly

[4] There are over 3,000 counties (or equivalents) in the U.S., but the vast majority of these are rural and sparsely populated. These small counties demonstrated very high rates of missingness on a variety of measures, but especially on implicit bias scores, jail population data, and arrest data. Furthermore, among those small counties providing prisoner data, the large number of counties with zero black prisoners or zero white prisoners made the calculation of black to white ratios problematic.

different than other states on our key measures, except that they demonstrate slightly higher income inequality and slightly larger county populations, on average.

Measures

Dependent variable: jail confinement disparities

Racial disparities in jail populations are used to measure discriminatory social control. Disparity is measured as the black to white ratio of race-specific confinement rates, log-transformed to account for substantial right-skew.[5] The 2013 jail census is used to measure the total number of black and white prisoners confined in all city and county jails, aggregated to the county level (United States Department of Justice, 2018). The census includes inmates held beyond arraignment, whether pre-trial or post-conviction, and therefore captures a range of potentially racially-influenced criminal justice processes, including prohibitive bail, conviction, the in/out decision, and sentence length. Race-specific confinement rates are determined by taking the number of countywide black prisoners divided by the number of countywide black citizens, and similarly the number of white prisoners divided by the number of white citizens. Our data do not include state prisoners, and therefore reflect disparities for lower-level crimes where discretion is greater and bias is more likely to manifest (Hester & Hartman, 2017).[6] A few studies have considered incarceration rates in the context of Racial Threat Theory, making this

[5] A small constant (0.1, slightly less than half of the lowest non-zero raw value) was added to each raw value before log transformation, because of the presence of several zero values, which have no natural log. Transformation resulted in a dependent variable that was much more normally distributed.

[6] Furthermore, prisoners are frequently held in state prisons outside their original county of residence or offense, which obfuscates county-level social control responses.

an important extension of this body of research (Bridges & Crutchfield, 1988; Campbell, Vogel, & Williams, 2015; Jacobs & Carmichael, 2001). However, no racial threat research has focused on *local* jails, despite the fact that over a third of all prisoners are in jails rather than prisons (Kaeble & Cowhig, 2018). There are reasons to suspect that disparities would be even greater in jails than in prisons, where lower-level offenses, race & poverty, and prohibitive pre-trial bail may disproportionately affect African Americans even more than in state prisons, where discretion (and confinement) is more narrowly constrained by seriousness of offense (Hester & Hartman, 2017).

Mediating variable: implicit bias

The measure of implicit bias at the county level comes from the aggregated Implicit Association Test (IAT) scores produced by the Harvard-based Project Implicit (PI). The IAT, based on response latency, is the most widely used method for measuring implicit bias (Xu et al., 2019). The scores are geo-located such that they can be aggregated to different geographic units such as counties. The present study uses data from the "Race IAT" available through the Project Implicit Demo Website Dataset (Xu et al., 2019). The race IAT measures the extent to which respondents associate black and white faces with either "good" (e.g., joy, wonderful, pleasure) and "bad" (e.g., evil, terrible, awful) words. Higher scores reflect the extent to which "good" is associated with whites and "bad" is associated with blacks—suggesting increased implicit bias toward blacks. The data consist of over a quarter million IAT scores from individuals within the 437 counties (based on zip code data provided by respondents). The number of responses within each county ranges from 28 to 8,072. We limit responses to the years 2011 and 2012, which falls between our predictors and our outcome variable, and therefore makes a more compelling case for temporal order and causality.[7]

The great value of the PI data to researchers is the availability of aggregated measures of implicit bias based on huge samples. That individuals choose to take an IAT on the Harvard site, however, raises significant issues of selection bias (and thus the generalizability of the scores), and yet there are several reasons why these data are, in fact, viable for research. First of all, Pinkston (2015) has shown that PI respondents show patterns of bias that are similar to that of nationally representative samples. Second, research has demonstrated that geographically-aggregated implicit bias measures from PI have predicted expected outcomes such as participation in demonstrations/rallies (Zerhouni, Rougier, & Muller, 2016), health care availability (Leitner, Hehman, Ayduk, & Mendoza-Denton, 2016a), Medicaid expenditures (Leitner, Hehman, & Snowden, 2018), racial disparities in health outcomes (Leitner et al., 2016a; Leitner, Hehman, Ayduk, & Mendoza-Denton, 2016b; Miller, Varni, Solomon, DeSarno, & Bunn, 2016; Orchard & Price, 2017), gender disparities in math and science achievement (Nosek et al., 2009), and disproportionate shootings of blacks by police (Hehman et al., 2018). Third, although the IAT scores for the counties might not reflect the absolute level of racial implicit bias within them, what is important for this study is the *relative level* of implicit bias for the counties, compared to the others. The selection bias will not threaten the validity of the relationship between population makeup and disparate formal social control so long as it operates in the same way across counties (see Zerhouni et al., 2016).

[7] The model presented in this study was repeated with a much broader range of years of IAT data (2006–2012), which substantially expanded the total number of IATs. The results of that model were nearly identical to the model presented here, suggesting that the choice to limit response years and ensure temporal order did not lead to sampling bias.

Independent variable

In order to examine the influence black population size has on implicit bias and social control, we obtain for each county 5-year estimates of the total population and racial subgroups from the U.S. Census Bureau's American Community Survey (U.S. Census Bureau, 2011). Five-year estimates provide more stable and reliable estimates than alternative one-year measures, and cover a broader range of counties (U.S. Census Bureau, 2008). We use the percentage of the county population that is black as our focal independent variable, which is the most commonly used measure of racial threat in the extant research (Feldmeyer & Cochran, 2018).

Control variables

Given that criminal justice disparities may reflect the influence of other systemic disparities and inequities, several control variables are used to account for potential confounders. Racial disparity in employment is measured using 5-year estimates from the Census Bureau's American Community Survey (U.S. Census Bureau, 2011). The ratio between the black unemployment rate and the white unemployment rate is calculated for each county. Values greater than one indicate that blacks are disproportionately unemployed, values near one indicate parity, and values less than one indicate that whites are disproportionately unemployed. The ratio between black and white median income, as well as the ratio of black and white marriage rates are measured in order to account for disparities in economic and social capital. Each of these measures may indicate black disadvantage that could contribute to criminal justice outcomes.

Other social-structural factors may also contribute to criminal justice policy, and so this study includes a measure of the GINI index of income inequality, the (log transformed) total county population, and percentage Hispanic. Because of the history of both overt and covert racial discrimination in the American South, a dummy variable is included to identify counties in Southern states, where 1=yes. Additionally, total arrest disparities are controlled, including all reported violent and non-violent arrests by race at the county level between the years 2006 and 2010 (Kaplan, 2019). Thus, the model examines the effect of race and bias on jail confinement *beyond* disparities introduced by either differential racial offending or differential enforcement by police.

Analytic strategy

In order to assess the influence of black population size on criminal justice disparities, as well as the mediating role of implicit bias, the analysis proceeds in two stages. First, descriptive statistics and bivariate correlations are produced. In the second stage, a multilevel mediation model is developed using Generalized Structural Equation Modelling in Stata 15. This technique provides at least two distinct advantages. First, it allows for simultaneous estimation of the direct and indirect effects (Preacher, Zyphur, & Zhang, 2010). Second, it accounts for the clustering of counties within states, calculating random intercepts at the state level. Because state law, law enforcement priorities, funding, and criminal sentences are often established by state legislatures, analyses must account for the tendency of counties clustered within states to demonstrate substantial within-state similarities, or they risk underestimating standard errors (McNeish, 2014). Grand mean centering is used for all predictors except percent black, percent Hispanic, and the southern region indicator, all of which have legitimate and meaningful zero values. Because macrosocial data, such as census estimates, oftentimes violate assumptions of normality, this study reports robust standard errors, which are not sensitive to violation of such assumptions and reduce the risk of type-I error, especially in large samples such as the one used here (Dietz, Frey, & Kalof, 1987).

Results

Descriptive statistics are presented in Table 1, before transformations. In terms of black and white jail confinement rates, the data indicate that just 17 U.S. counties (3.9% of the entire sample) report higher white confinement rates than black confinement rates. The black county-level confinement rate is, on average, over nine times higher than the white confinement rate. The range of values is also striking. Four of 437 counties report no black prisoners.[8] The independent city of Norfolk Virginia has the lowest non-zero black to white incarceration rate, where the black incarceration rate is one quarter the white incarceration rate. In stark contrast, three counties report black jail confinement rates *over 100 times higher* than whites'.[9]

Among the 437 counties, the average IAT score is .333, indicating moderate white preference/anti-black bias. These counties have, on average, black populations consisting of 11.5% of the total population and Hispanic populations around 10.3% of the total. In addition to the incarceration disparities and implicit bias measures, our controls revealed several other

[8] Faulkner County, Arkansas; Lake County, Illinois; Kootenai County, Idaho; and Knox County, Tennessee

[9] Bonneville County, Idaho; Waukesha County, Wisconsin; and St. Clair County, Michigan.

indications of black disadvantage. On average, the county-level black unemployment rate is twice as high as whites'. Blacks' median incomes are just two-thirds that of whites', and blacks' marriage rates are, on average, just 61% that of whites'. The black arrest rate is on average nearly four times higher than whites'. The average total population of counties included here was 373,292. The county-level GINI index averaged .437, and nearly 34% of these counties are located in a Southern state.

[TABLE 1 ABOUT HERE]

Table 2 presents bivariate statistics. Several correlations stand out. First, the percentage of the population that is black is strongly and negatively correlated with implicit bias ($r = -.679$). Second, implicit bias is moderately correlated with racial disparities in jail confinement ($r = .379$). Finally, the correlations reveal that the size of the black population is negatively correlated with jail confinement disparities ($r = -.352$). This provides some preliminary evidence that larger percentages of black residents reduce criminal justice disparities, at least in part, it seems, by reducing overall levels of implicit bias.

[TABLE 2 ABOUT HERE]

Table 3 reports the results of the MSEM with mediation for race-specific differences in jail confinement rates. MSEM simultaneously estimates two random-intercept models: the influence of the independent and control variables on the mediating variable (implicit bias); and the influence of the independent/control variables and mediator on the outcome (black to white ratios in local jail confinement). The results are also demonstrated graphically in Figure 2, although control variables are omitted from the figure for brevity. Figure 2 reports both unstandardized coefficients, as is common in SEM path diagrams, as well as standardized coefficients, as is common in mediation path diagrams; each path is identified (a, b, and c') according to the conventions of traditional mediation analysis (Baron & Kenny, 1986).

[TABLE 3 ABOUT HERE]

[FIGURE 2 ABOUT HERE]

The percentage of a county that is black strongly predicts a reduction in overall implicit bias ($b = -.397$, $p < .001$, path a). Southern region ($b = .007$, $p < .01$) and the black-to-white arrest ratio ($b = .003$, $p < .001$) predict a small but significant increase in implicit bias.

Contrary to Racial Threat Theory, these data do not indicate that a county's percentage of black residents directly increases racially disparate criminal justice outcomes. The results among this sample suggests that as the size of the black population increases, the magnitude of racial disparities in confinement rates decreases ($b = -.105$, path c'), although this effect is non-significant.[10] Furthermore, implicit bias appears to significantly and substantially increase the county-level differences in black/white jail confinement rates ($b = 2.744$, $p < .001$, path b).

A few non-theoretical control variables attain statistical significance. Southern counties demonstrated lower incarceration disparities ($b = -.308$, $p < .01$). The size of the Hispanic population was associated with higher levels of disparity ($b = 1.190$, $p < .01$). Curiously, higher

[10] Supplemental analyses, not shown, indicate that black population size exerts a large and statistically significant influence on jail confinement disparities when implicit bias is not included in the model, but drops in both size and significance when controlling for implicit bias, further supporting a mediation effect.

black-to-white marriage ratios predict larger black-to-white confinement disparities ($b = .801$, $p < .01$).

The direct, indirect, and total effects of black population size and jail confinement rate disparities are reported at the bottom of Table 3. Indirect effects are estimated using bootstrap standard errors with 5,000 replications. A bootstrapping technique uses non-parametric sampling with replacement, repeated many times to create an empirically-based sampling distribution, which allows corrections for bias to be made; if the bias-corrected confidence interval of the overall indirect effect does not contain the value of zero, then there is strong evidence of mediation, and this technique does not depend on assumptions of normality that traditional parametric estimation requires (Hayes & Scharkow, 2013). While bias-corrected confidence intervals are commonly reported, they have a tendency to be too liberal; to reduce the risk of type-I error, percentile confidence intervals are reported (Hayes & Scharkow, 2013). The indirect effect here is $b = -1.091$, $p < .001$, with a 95% bootstrap percentile confidence interval of -1.686 to -.328. The total effect is $b = -1.196$, $p < .01$. Thus, over 90% of the relationship between black population size and local incarceration rate differences is mediated by implicit bias, suggesting full mediation.

As illustrated in Figure 2, a one standard deviation increase in the percentage of African Americans is associated with nearly a .721 standard deviation decrease in implicit bias (path a, $p < .001$). Furthermore, a one standard deviation increase in implicit bias is associated with a .209 standard deviation increase in black to white incarceration disparities in local jails (path b, p

<.001). The small and non-significant c' path (B = -.015,) suggests that the relationship between racial composition and incarceration disparities is fully mediated by implicit bias.

Discussion

Blalock's Racial Threat Theory predicts that large or growing black populations lead to discrimination (Blalock, 1967). One such form of discrimination includes formal social control, and a substantial body of research has examined how racial composition predicts various disparate applications of the criminal justice system (Feldmeyer & Cochran, 2018). Much remains unclear about this relationship, however. For instance, some research has found that larger black populations *are not* associated with disproportionate social control (Feldmeyer & Ulmer, 2011; Leiber et al., 2016), or predict *lower* levels of racial disparity (Liska & Chamlin, 1984; Parker et al., 2005; Stolzenberg et al., 2004; Stucky, 2012). The mixed body of findings has led to calls for a closer examination of the mechanisms involved in the relationship between racial composition and associated racial disparities in criminal justice outcomes (Dollar, 2014; Feldmeyer & Cochran, 2018).

The proposition that anti-black bias might mediate this relationship reflects Blumer's claim that racial prejudice is a product of group position. Members of a majority group may develop more hostile attitudes toward minority groups when the minority population is large or growing, and those increased hostile attitudes would lead to increased social control of outgroups (Ousey & Unnever, 2012; Stults & Baumer, 2007). Intergroup Contact Theory, in contrast, suggests that majority group attitudes toward minority groups will be *more positive* if the group is large, because larger minority groups present more opportunities for intergroup interaction. If this is the case, the relationship between percent black and disparate social control will be negative.

Using a sample of 434 U.S. counties, this study used multilevel mediation analysis under a structural equation modelling approach to assess the impact of black population makeup on racial disparities in county jail populations and to assess the mediating impact of prejudice on the relationship between population makeup and disparate social control. The study tests the competing predictions of the relationship between minority group population size and bias.

The key findings, as summarized in Figure 2, pertain to (1) the relationship between percent black in the population and racial disparity in the jail population (Figure 2, path c'), (2) the extent to which anti-black implicit bias mediates that relationship (Figure 2, path a x b), (3) the relationship between percent black and anti-black implicit bias (Figure 2, path a), and (4) the relationship between anti-black bias and disparate social control (Figure 2, path b).

Contrary to the predictions of Racial Threat Theory, the results show that a large black population does not predict greater disparate social control. (See path c' in Figure 2.) In fact, although not significant, the findings are consistent with some previous research showing that larger black populations are associated with *smaller* racial disparities in formal social control. Although blacks are confined at higher rates than whites in nearly all U.S. counties, these differences actually diminish in counties with larger black populations when other variables are controlled.

The finding above, contrary to what Racial Threat Theory would predict, highlights the importance of examining potential mediators, including the potential mediating impact of anti-black implicit bias. The results produced here indicate that the relationship between black population size and disparate social control is mediated by anti-black bias; 90 percent of the relationship is explained by implicit bias. (See path a x b in Figure 2.) The nature of the mediation is consistent with Contact Theory that predicts that positive intergroup contact—such as the positive and routine social interaction that may occur in racially diverse counties—tends to reduce out-group bias (Allport, 1954). Counties with the largest share of black residents demonstrates some of the lowest mean implicit bias scores, and vice versa. (See path a in Figure 2.)

Consistent with expectations from Racial Threat Theory and social psychological research and theory on bias, implicit racial biases produce racially disparate formal social control. (See path b in Figure 2.) A county's mean white preference/anti-black bias, as measured via implicit association tests, predicts significantly larger disparities between blacks' and whites' confinement rates. Thus, counties demonstrating more implicit bias, on average, incarcerate a larger share of their black residents (relative to their white residents) than counties with lower levels of implicit bias. This is true even after accounting for differences between black and white arrest rates, suggesting that it is not a function of differential offending, nor differential enforcement by the police.

This finding is consistent with the research showing that implicit biases can produce discriminatory behavior (Dasgupta, 2004; Kang et al., 2011; Rooth, 2007). In laboratory studies, the behavior that has been impacted by implicit biases ranges from "non-verbal friendliness" (e.g., John F. Dovidio, Kawakami, & Gaertner, 2002) to shooting (Correll, Park, Judd, & Wittenbrink, 2002, 2007; Correll, Park, Judd, Wittenbrink, et al., 2007). Biases have been shown to impact employment decisions (Kushins, 2014; for an overview, see e.g., Fiske &

Krieger, 2013), school discipline (Okonofua & Eberhardt, 2015), and medical treatment (Hirsh, Hollingshead, Ashburn-Nardo, & Kroenke, 2015; Weisse, Sorum, Sanders, & Syat, 2001), to name a few. Most relevant to this study, research has documented the impact of implicit biases on the behavior of various actors within the criminal justice system such as police (Correll, Hudson, Guillermo, & Ma, 2014; Correll & Keesee, 2009; Goff, Jackson, Di Leone, Culotta, & DiTomasso, 2014; Sadler et al., 2012), judges (Rachlinski, Johnson, Wistrich, & Guthrie, 2009), and jurors (Eberhardt et al., 2006; Levinson et al., 2010; Levinson, Smith, & Young, 2014; for a meta-analysis, see T. L. Mitchell, Haw, Pfeifer, & Meissner, 2005).

Theoretical and empirical implications

This study has theoretical and empirical implications. First of all, the finding here of no direct relationship between percent black and racial disparities in local jails (path c') is consistent with other research that has found that racial composition predicts a null or inverse relationship to criminal justice disparities (e.g., Andersen, 2015; Liska & Chamlin, 1984; Parker et al., 2005; Stolzenberg et al., 2004; Stucky, 2012). The theoretical reason for these null or contradictory findings has not been clear. Benign Neglect theorists explain the null findings by suggesting that larger black populations will not lead to more criminal justice involvement when they are concentrated into segregated neighborhoods where black offenders generally prey on black victims (Liska & Chamlin, 1984). This study provides another explanation—with the finding that racial bias is a powerful mediator in the relationship between percent black in a population and disparate social control. Thus, the present study contributes to a growing body of evidence that the basic Racial Threat Theory—asserting that the size of the black population predicts discriminatory criminal justice—is overly simplistic, at best, and it highlights the need for further theoretical attention.

Second, the results provide support for Intergroup Contact Theory with the finding that areas with a larger share of African American residents are associated with lower overall levels of implicit bias (path a). Counties with larger black populations plausibly have more interactions between white and black residents, providing more opportunities for the positive intergroup interaction, which may reduce anti-black bias. While this study does not directly measure the nature or frequency of interactions between majority and minority groups, it provides evidence—via the mediating role of implicit bias—that larger black populations reduce discrimination. Thus, this study reinforces theoretical arguments that whites' interaction with blacks may promote bias reduction (Mancini, Mears, Stewart, Beaver, & Pickett, 2012), rather than promote perceptions of "threat" and discrimination, as Blalock proposed.

Third, this study provides support for the argument that implicit biases impact behavior. When it comes to issues of race and justice, even subconscious, unintentional implicit racial bias contributes to significant increases in racial disparities. This finding challenges claims by some researchers that the association between implicit attitudes and actual behaviors are non-existent or at least too small to be of significance (Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2015). This study indicates that implicit anti-black bias can have substantial theoretical relevance and broad social implications (Anthony G. Greenwald, Banaji, & Nosek, 2015; Payne, Vuletich, & Lundberg, 2017). While traditionally, disparities in criminal justice have been attributed to overt racism, the emerging body of evidence, including the present study, suggests that theories of systemic inequity must increasingly consider implicit social cognition and all of its associated causes and consequences.

Implications for social policy

The key findings from this study with the most important policy implications are: (1) the negative relationship between size of black population and anti-black bias, and (2) the confirmation that negative attitudes toward blacks drives confinement disparities. Two policy implications follow. First, this study suggests that criminal justice disparities may be reduced by promoting racially diverse areas, since the most homogenous (and suburban) areas were also those with the most discriminatory patterns of jail confinement. Therefore, in addition to the various other benefits it provides, there appears to be substantial value in promoting diverse neighborhoods to reduce systemic biases in criminal justice (Turner, 2009). Not only do highly segregated areas tend to concentrate poverty and crime, indirectly contributing to differential justice involvement (Lee & Ousey, 2005; Peterson & Krivo, 1993; Wilson, 1978), segregated areas also tend to encourage disproportionate police activity independent of crime (Kent & Carmichael, 2014). Consistent with the Intergroup Contact Theory, racial diversity may reduce stereotypes and biases that contribute to systemic inequities. Of course, mere contact is insufficient to reduce bias; researchers have long pointed out that intergroup contact must also be positive to reduce bias (Allport, 1954; Aronson & Patnoe, 1997; T. F. Pettigrew, 1998). This underscores the importance of, not just diverse counties and neighborhoods, but diverse schools and workplaces (Novak, Feyes, & Christensen, 2011; Roch & Edwards, 2017). All of this is to say, given that anti-black bias contributes to racial injustice, reductions in anti-black bias are most likely in areas with a greater share of black residents. It also suggests that integrated schools and affirmative action programs may promote intergroup contact and reduced bias, and ultimately reduce racial disparities in the criminal justice system.

Second, given that implicit bias has a substantial (and troublesome) influence on patterns of jail disparities, efforts to reduce the influence on criminal justice personnel of implicit biases may prove worthwhile. Interventions have been suggested for various components of the criminal justice system. There are calls for police to, for instance, enhance opportunities for positive engagement between police and community members (L. A. Fridell, 2017; Hall, Hall, & Perry, 2016) and implement safeguards to reduce the risk of biases impacting high-discretion, crime controlled activities such as stop and frisk (Epp, Maynard-Moody, & Haider-Markel, 2014; L. A. Fridell, 2017; Glaser, Spencer, & Charbonneau, 2014; White & Fradella, 2016). Courts could implement decision-making protocols to reduce the potential for implicit biases to impact decisions, provide attorneys including judges with periodic feedback on disparities in their decision-making, and ensure diversity on juries (Casey, Warren, Cheesman, & Elek, 2013; Kang et al., 2011; Rachlinski et al., 2009; Richardson & Goff, 2012). There are calls for ensuring that the legal analyses within equal protection jurisprudence focus on disparate impact rather than intentions (Clemons, 2014).

A key additional intervention is providing implicit bias awareness training to personnel, the purposes of which are to raise individuals' awareness of implicit biases, identify the negative consequences of biased behavior, and provide individuals with skills for reducing and managing biases.[11] Such trainings have been developed for various criminal justice actors such as law

---

[11] Although controlled evaluations of the effectiveness of these trainings are few, Devine et al., 2012 found reductions in implicit bias as a result of an intervention that involved educating participants on their implicit biases and their deleterious effects, and training them in strategies to reduce and manage their biases. (See also (Casey, Warren, Cheesman, & Elek, 2013).)

enforcement (L. A. Fridell, 2017; L. Fridell & Brown, 2015), prosecutors and defense attorneys (American Bar Association, n.d.), judges (Casey et al., 2013), and jurors (Gayla, 2017). The "skills" taught in these courses reflect the voluminous body of research indicating that individuals can reduce and/or manage their biases (see Bennett, 2010; Hernandez, Haidet, Gill, & Teal, 2013; Kang et al., 2011; for an overview, see Staats, 2013; Staats et al., 2016). The most well-known and most-studied bias-reducing mechanism is positive contact, consistent with the Intergroup Contact Theory (for a meta-analysis, see Thomas F. Pettigrew & Tropp, 2006). As Tausch and Hewstone (2010, p. 544) write, "The notion that contact between members of different groups can, under certain conditions, reduce prejudice is one of the most prominent ideas underlying approaches to improve intergroup relations." Because it is slow and difficult to reduce the biases that took a lifetime to develop, it is important that individuals can manage their biases. If individuals recognize their implicit biases and are motivated to be impartial, they can "self-regulate" (Dasgupta & Rivera, 2006; Devine, Plant, Amodio, Harmon-Jones, & Vance, 2002; Hausmann & Ryan, 2004; Monteith, Mark, & Ashburn-Nardo, 2010).

The Racial Threat Theory implies that it is broader societal bias against minority groups, *not just criminal justice professionals' bias*, that leads to disparate social control through the criminal justice system. Our results demonstrate that county-wide biases correspond to county-level incarceration disparities. This is all the more understandable given that biases of complainants, victims, witnesses, and jurors may contribute to these disparities. The implication is that bias-reducing mechanisms (such as implicit bias awareness training/education) need to be implemented beyond criminal justice professionals in order to impact criminal justice outcomes.

Study limitations and future research

This study makes both theoretical and empirical contributions to the Racial Threat literature. It advances the field by assessing the mediating impact of anti-black bias on the relationship between population make up and disparate formal social control. It does so using advanced statistical techniques applied to data from 434 counties and measures implicit, not explicit, bias. It provides a plausible explanation for previous research that has found null findings or findings that are the opposite of what Racial Threat Theory would predict. Those contributions notwithstanding, there are a number of weaknesses of the current study with implications for future research.

As above, a key contribution of this study is the measurement of implicit bias at the county level using IAT data. The use of these data is a strength or a weakness depending on where one stands with regard to the ongoing academic debate on the validity, reliability, and generalizability of the IAT. (For summaries of the debates, see G. Mitchell & Tetlock, 2017; Nagai, n.d.; Oude Maatman, 2017).[12] There are some studies that report low reliability of the IAT, such as test-re-test reliability (Bar-Anan & Nosek, 2014; Rezaei, 2011). In contrast are the studies that show evidence of both internal consistency and test-retest reliability (Cunningham, Preacher, & Banaji, 2001; Nosek, Greenwald, & Banaji, 2007). (For reviews of test-retest studies, see Gawronski, Morrison, Phills, & Galdi, 2017; Lane et al., 2007.)[13] Jost (2019, p. 2)

[12]Importantly, some of this discourse is not just about the IAT, per se, but rather about the challenges associated with measuring implicit cognitions (see DeSchryver 2018).

[13] Commenting on the findings of some studies that show low test-retest reliability, Payne et al. (2017) speculate that it may not be that the IAT measure is at fault; it may be that implicit bias is not as stable as some have assumed. Instead, implicit biases may be impacted significantly by context.

claims, “The IAT exhibits higher (within-persons) test-retest reliability than other response-latency measures commonly used in psychological research …” and, in fact, he argues, for many versions of the IAT, the test-retest reliability “is as high as (or higher than) that for self-administered blood-pressure readings” (p. 2, citing Brody, Erb, Veit, & Rau, 1999).

In terms of validity, Lane et al reviewed the various assessments of the IAT and found solid evidence for construct validity, mixed results for convergent validity, and evidence of discriminant validity.[14] (See also Nosek et al., 2007.) The discussion of predictive validity has produced the battle of the meta-analyses with one team claiming that the IAT does predict behavior (Anthony G. Greenwald, Andrew Poehlman, Uhlmann, & Banaji, 2009; Anthony G. Greenwald et al., 2015; Jost et al., 2009; Kurdi et al., 2018) another saying that it does not (Oswald, Mitchell, Blanton, Jaccard, & Tetlock, n.d.). (Kurdi et al., 2018, p. 1), suggest that the answer is somewhere in the middle of these two polar extremes; it is not *whether or not* the IAT predicts behavior, but rather identifying the “conceptual and methodological conditions” under which the IAT predicts it.

Some researchers have commented, in particular, on the reliability and validity of IAT scores that are aggregated, as was done in the current study. Jost (2019) observed that the IAT is “highly reliable at the aggregate level” and the Payne et al., (2017) claim that aggregate scores—assessing social climate—may be more valid than individual ones.

A problem with the IAT is the fact that individuals self select to take the exam. The individuals taking the exams are not necessarily representative of the population as a whole and individuals can and do take the IAT test more than once. As previewed above, this issue is

[14]They found that discriminant validity findings varied by the attitude to which the implicit association was compared.

countered by the facts that: (1) PI respondents show patterns of bias that are similar to that of nationally representative samples, (2) research that has demonstrated that geographically-aggregated implicit bias measures have predicted expected outcomes, and (3) although the IAT scores for the counties might not reflect the absolute level of racial implicit bias within them, what is important for this study is the relative level of implicit bias for the counties, compared to the others. Arguably, confidence can be placed in the findings regarding the relationship between population makeup and disparate formal social control if the selection bias operates similarly across counties.

This study used implicit bias scores for all individuals in the county who had taken the Race IAT during our reference period, regardless of race. Practically, using all Race IAT scores produced a greater number of scores for each county. Theoretically, using all scores produced the desired measure of *county-level anti-black bias*. Because criminal justice outcomes are influenced by all residents and social groups—as voters, complainants, victims, witnesses, jurors, and criminal justice employees—we chose not to limit our measure to whites only. That said, there are some complications introduced by using both white and black scores. While the social psychology research shows that members of minority groups can have biases about their own group, black IAT scores are generally lower than white IAT scores (Jost, Banaji, & Nosek, 2004; Pinkston, 2015; Wallace, 2018). The finding that counties with higher proportions of blacks had lower levels of anti-black implicit bias may be produced at least in part by lower scores produced by a greater number of black IAT participants compared to other counties. This study is not able to determine whether the overall reduction in implicit bias was due to (a) reductions in anti-black bias among white residents exposed to larger black populations, vis-à-vis Intergroup Contact

Theory, (b) a purely compositional effect, wherein a larger share of black residents merely weights the average toward less anti-black bias, or (c) both. Future research may help determine to what extent each of these contributes to aggregate implicit bias, and whether racially diverse areas indeed reduce whites' racial biases.

Relatedly, while positive intergroup contact is implicated via the mechanism examined in this study, it is not directly measured. The percentage black measurement assumes that higher levels of minority group populations produce more positive contact between majority and minority individuals. This may not be the case: even counties with very large black populations may not produce intergroup contact if neighborhoods are extremely segregated. Furthermore, even counties with diverse and integrated communities do not necessarily ensure *positive* interactions required by Intergroup Contact Theory (Allport, 1954; Pettigrew & Tropp, 2006). Future research with individuals as the unit of analysis could produce a superior operationalization of this construct with survey questions regarding subjects' own intergroup experiences. The researchers then could explore the relationship between positive contact with outgroup members, anti-outgroup bias, and punitive attitudes toward outgroup members. Macrosocial studies could incorporate a measure of segregation or intergroup exposure, which would provide a more valid measure of interracial interaction, and evaluate its association with aggregate implicit bias.

Conclusion

Racial Threat Theory has produced a rich body of research regarding macrosocial structure and discriminatory social control. Researchers and theorists have suggested over the years that minority group prejudice might play a role in this relationship, and emerging insights into how bias manifests in people as well as enhanced measurement techniques allows for an effective assessment. This study presents compelling evidence that implicit bias plays a substantial role in the relationship between racial composition and formal social control and this finding has significant implications for theory, social policy, and future research. This enhanced understanding comes during a period of re-intensified national discussion of race and justice following a number of controversial and highly publicized deaths of black men during criminal justice encounters. As this discussion continues, it will be important to further consider how minority group size, multiracial communities, implicit bias, and the criminal justice system interact.

References

Allport, G. W. (1954). *The Nature of Prejudice*. Reading, MA: Addison-Wesley.

American Bar Association. (n.d.). Implicit Bias Initiative. Retrieved June 3, 2019, from https://www.americanbar.org/groups/litigation/initiatives/task-force-implicit-bias/

Amodio, D. M., & Mendoza, S. A. (2010). Implicit intergroup bias: Cognitive, affective, and motivational underpinnings. In B. Gawronski (Ed.), *Handbook of implicit social cognition: Measurement, theory, and applications* (353–374). New York: The Guilford Press.

Andersen, T. S. (2015). Race, ethnicity, and structural variations in youth risk of arrest: Evidence from a national longitudinal sample. *Criminal Justice and Behavior*, *42*(9), 900–916.

Aronson, E., & Patnoe, S. (1997). *The jigsaw classroom: Building cooperation in the classroom*. New York: Longman.

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods, 46*(3), pp. 668–688. https://doi.org/10.3758/s13428-013-0410-6

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*(6), 1173.

Baumer, E. P., Messner, S. F., & Rosenfeld, R. (2003). Explaining spatial variation in support for capital punishment: A multilevel analysis. *The American Journal of Sociology*, *108*(4), 844–875.

Baumgartner, F. R., Epp, D. A., & Love, B. (2014). Police searches of Black and White motorists. *UNC-Chapel Hill, Department of Political Science*. Retrieved from https://fbaum.unc.edu/TrafficStops/DrivingWhileBlack-BaumgartnerLoveEpp-August2014.pdf

Bennett, M. W. (2010). *Unraveling the Gordian Knot of Implicit Bias in Jury Selection: The Problem of Judge-Dominated Voir Dire, the Failed Promise of Batson, and Proposed Solutions*. Retrieved from https://papers.ssrn.com/abstract=2505424

Blalock, H. (1967). *Toward a theory of minority-group relations*. New York: Wiley.

Blumer, H. (1958). Race prejudice as a sense of group position. *Pacific Sociological Review*, *1*(1), 3–7.

Bridges, G. S., & Crutchfield, R. D. (1988). Law, social standing and racial disparities in imprisonment. *Social Forces*, *66*(3), 699–724.

Brody, S., Erb, C., Veit, R., & Rau, H. (1999). Intraocular pressure changes: The influence of psychological stress and the Valsalva maneuver. *Biological Psychology, 51*(1), pp. 43–57. https://doi.org/10.1016/s0301-0511(99)00012-5

Campbell, M. C., Vogel, M., & Williams, J. (2015). Historical contingencies and the evolving importance of race, violent crime, and region in explaining mass incarceration in the United States. *Criminology*, *53*(2), 180–203.

Casey, P. M., Warren, R. K., Cheesman, F. L., & Elek, J. K. (2013). Addressing implicit bias in the courts. *Ct. Rev.*, *49*, 64.

Chiricos, T. (n.d.). Group threat and social control: A review of theory and research. In C. Chouhy, J. C. Cochran, & C. L. Jonson (Eds.), *Criminal Justice Theory: Explanation and Effects.*

Clemons, J. T. (2014). Blind Injustice: The Supreme Court, Implicit Racial Bias, and the Racial Disparity in the Criminal Justice System. *Am. Crim. L. Rev.*, *51*, 689.

Correll, J., Hudson, S. M., Guillermo, S., & Ma, D. S. (2014). The police officer's dilemma: A decade of research on racial bias in the decision to shoot. *Social and Personality Psychology Compass*, *8*(5), 201–213.

Correll, J., & Keesee, T. (2009). Racial bias in the decision to shoot. *The Police Chief*, *76*(5), 54–57.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2002). The police officer's dilemma: Using ethnicity to disambiguate potentially threatening individuals. *Journal of Personality and Social Psychology*, *83*(6), 1314–1329.

Correll, J., Park, B., Judd, C. M., & Wittenbrink, B. (2007). The influence of stereotypes on decisions to shoot. *European Journal of Social Psychology*, *37*(6), 1102–1117.

Correll, J., Park, B., Judd, C. M., Wittenbrink, B., Sadler, M. S., & Keesee, T. (2007). Across the thin blue line: Police officers and racial bias in the decision to shoot. *Journal of Personality and Social Psychology*, *92*(6), 1006–1023.

Crawford, C., Chiricos, T., & Kleck, G. (1998). Race, racial threat, and sentencing of habitual offenders. *Criminology*, *36*(3), 481–512.

Cunningham, W. A., Preacher, K. J., & Banaji, M. R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science, 12*(2), pp. 163–170. https://doi.org/10.1111/1467-9280.00328

D'Alessio, S. J., Eitle, D., & Stolzenberg, L. (2005). The impact of serious crime, racial threat, and economic inequality on private police size. *Social Science Research*, *34*(2), 267–282.

Dasgupta, N. (2004). Implicit Ingroup Favoritism, Outgroup Favoritism, and Their Behavioral Manifestations. *Social Justice Research*, *17*(2), 143–169.

Dasgupta, N., & Rivera, L. M. (2006). From automatic antigay prejudice to behavior: The moderating role of conscious beliefs about gender and behavioral control. *Journal of Personality and Social Psychology*, *91*(2), 268–280.

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

Devine, P. G., Plant, E. A., Amodio, D. M., Harmon-Jones, E., & Vance, S. L. (2002). The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice. *Journal of Personality and Social Psychology*, *82*(5), 835–848.

Dietz, T., Frey, R. S., & Kalof, L. (1987). Estimation with cross-national data: robust and nonparametric methods. *American Sociological Review*, 380–390.

Dollar, C. B. (2014). Racial threat theory: Assessing the evidence, requesting redesign. *Journal of Criminology*, *2014*. https://doi.org/10.1155/2014/983026

Dovidio, J. F., Hewstone, M., Glick, P., & Esses, V. M. (2010). Prejudice, stereotyping and discrimination: Theoretical and empirical overview. *The SAGE Handbook of Prejudice, Stereotyping and Discrimination*, 3–29.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*(1), 62–68.

Dovidio, J. F., Kawakami K Smoak, & Gaertner, S. L. (n.d.). The nature of contemporary racial prejudice. In R. E. Petty, R. H. Fazio, & P. Brinol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 165–192). New York: Psychology Press.

Durose, M. R., & Langton, L. (2013). *Police Behavior during Traffic and Street Stops, 2011* (Bureau of Justice Statistics Publication No. 242937). Washington, DC: U.S. Government Printing Office.

Eberhardt, J. L., Davies, P. G., Purdie-Vaughns, V. J., & Johnson, S. L. (2006). Looking deathworthy: Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes. *Psychological Science*, *17*(5), 383–386.

Eberhardt, J. L., Goff, P. A., Purdie, V. J., & Davies, P. G. (2004). Seeing black: Race, crime, and visual processing. *Journal of Personality and Social Psychology*, *87*(6), 876–893.

Eith, C., & Durose, M. R. (2011). *Contacts between police and the public, 2008*. Retrieved from U.S. Department of Justice website: https://www.bjs.gov/content/pub/pdf/cpp08.pdf

Epp, C. R., Maynard-Moody, S., & Haider-Markel, D. P. (2014). *Pulled Over: How Police Stops Define Race and Citizenship*. Chicago: University of Chicago Press.

Feldmeyer, B., & Cochran, J. C. (2018). Racial threat and social control: A review and conceptual framework for advancing racial threat theory. In J. D. Unnever, S. L. Gabbidon, & C. Chouhy (Eds.), *Building a Black Criminology: Race, Theory, and Crime* (Vol. 24, pp. 283–316). New York: Routledge.

Feldmeyer, B., & Ulmer, J. T. (2011). Racial/ethnic threat and federal sentencing. *The Journal of Research in Crime and Delinquency*, *48*(2), 238–270.

Fiske, S. T., & Krieger, L. H. (2013). Policy implications of unexamined discrimination gender bias in employment as a case study. In *The Behavioral Foundations of Public Policy* (pp. 52–74). Princeton, NJ: Princeton University Press.

Fridell, L. A. (2017). *Producing bias-free policing: A science-based approach*. Springer.

Fridell, L., & Brown, S. (2015). Fair and impartial policing: A science-based approach. *The Police Chief, June,* 20–25.

Gabrielson, R., Sagara, E., & Jones, R. G. (2014, October 10). Deadly Force, in Black and White — ProPublica. Retrieved June 3, 2019, from ProPublica website: https://www.propublica.org/article/deadly-force-in-black-and-white

Gawronski, B., Morrison, M., Phills, C. E., & Galdi, S. (2017). Temporal stability of implicit and explicit measures. *Personality and Social Psychology Bulletin, 43*(3), pp. 300–312. https://doi.org/10.1177/0146167216684131

Gayla, M. (2017, June 21). A Federal Court Asks Jurors to Confront Their Hidden Biases. Retrieved June 3, 2019, from The Marshall Project website: https://www.themarshallproject.org/2017/06/21/a-federal-court-asks-jurors-to-confront-their-hidden-biases

Gelman, A., Fagan, J., & Kiss, A. (2007). An analysis of the new york city police department's "stop-and-frisk" policy in the context of claims of racial bias. *Journal of the American Statistical Association*, *102*(479), 813–823.

Giles, M. W., & Evans, A. (1986). The Power Approach to Intergroup Hostility. *Journal of Conflict Resolution, 30*(3), pp. 469–486. https://doi.org/10.1177/0022002786030003004

Glaser, J., Spencer, K., & Charbonneau, A. (2014). Racial bias and public policy. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 88–94.

Goff, P. A., Jackson, M. C., Di Leone, B. A. L., Culotta, C. M., & DiTomasso, N. A. (2014). The essence of innocence: Consequences of dehumanizing Black children. *Journal of Personality and Social Psychology*, *106*(4), 526–545.

Graham, S., & Lowery, B. S. (2004). Priming unconscious racial stereotypes about adolescent offenders. *Law and Human Behavior*, *28*(5), 483–504.

Greenberg, D. F., & West, V. (2001). State prison populations and their growth, 1971-1991. *Criminology*, *39*(3), 615–654.

Green, D. P., Strolovitch, D., & Wong, J. (1998). Defended neighborhoods, integration, and racially motivated crime. *American Journal of Sociology, 104*(2), 372-403.

Greenwald, A. G., Andrew Poehlman, T., 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), pp. 17–41. https://doi.org/10.1037/a0015575

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4–27.

Greenwald, A. G., Banaji, M. R., & Nosek, B. A. (2015). Statistically small effects of the Implicit Association Test can have societally large effects. *Journal of Personality and Social Psychology*, *108*(4), 553–561.

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.

Hall, A. V., Hall, E. V., & Perry, J. L. (2016). Black and blue: Exploring racial bias and law enforcement in the killings of unarmed black male civilians. *The American Psychologist*, *71*(3), 175–186.

Hardin, C. D., & Banaji, M. R. (2013). The nature of implicit prejudice: Implications for personal and public policy. *The Behavioral Foundations of Public Policy*, 13–31.

Hausmann, L. R. M., & Ryan, C. S. (2004). Effects of external and internal motivation to control prejudice on implicit prejudice: The mediating role of efforts to control prejudiced responses. *Basic and Applied Social Psychology*, *26*(2-3), 215–225.

Hayes, A. F., & Scharkow, M. (2013). The relative trustworthiness of inferential tests of the indirect effect in statistical mediation analysis: Does method really matter? *Psychological Science*, *24*(10), 1918–1927.

Hehman, E., Flake, J. K., & Calanchini, J. (2018). Disproportionate use of lethal force in policing is associated with regional racial biases of residents. *Social Psychological and Personality Science*, *9*(4), 393–401.

Hernandez, R. A., Haidet, P., Gill, A. C., & Teal, C. R. (2013). Fostering students' reflection about bias in healthcare: Cognitive dissonance and the role of personal and normative standards. *Medical Teacher*, *35*(4), 1082–1089.

Hester, R., & Hartman, T. K. (2017). Conditional race disparities in criminal sentencing: A test of the Liberation Hypothesis from a non-guidelines state. *Journal of Quantitative Criminology*, *33*(1), 77–100.

Higgins, G. E., Vito, G. F., & Grossi, E. L. (2012). The impact of race on the police decision to search during a traffic stop: A Focal Concerns Theory perspective. *Journal of Contemporary Criminal Justice*, *28*(2), 166–183.

Hipp, J. R. (2011). Spreading the wealth: The effect of the distribution of income and race/ethnicity across households and neighborhoods on city crime trajectories. *Criminology*, *49*(3), 631–665.

Hirsh, A. T., Hollingshead, N. A., Ashburn-Nardo, L., & Kroenke, K. (2015). The interaction of patient race, provider bias, and clinical ambiguity on pain management decisions. *The Journal of Pain*, *16*(6), 558–568.

Holmes, M. D., Painter, M. A., & Smith, B. W. (2018). Race, place, and police-caused homicide in U.S. municipalities. *Justice Quarterly*, 1–36.

Jacobs, D., & Carmichael, J. T. (2001). The politics of punishment across time and space: A pooled time-series analysis of imprisonment rates. *Social Forces*, *80*(1), 61–89.

Jacobs, D., & Carmichael, J. T. (2002). The political sociology of the death penalty: A Pooled Time-Series Analysis. *American Sociological Review*, *67*(1), 109.

Jacobs, D., & Carmichael, J. T. (2004). Ideology, social threat, and the death sentence: Capital sentences across time and space. *Social Forces*, *83*(1), 249–278.

Jacobs, D., & Wood, K. (1999). Interracial conflict and interracial homicide: Do political and economic rivalries explain white killings of blacks or black killings of whites? *The American Journal of Sociology*, *105*(1), 157–190.

Johnson, B. D., Ulmer, J. T., & Kramer, J. H. (2008). The social context of guidelines circumvention: The case of federal district courts. *Criminology*, *46*(3), 737–783.

Jost, J. T. (2019). The IAT is dead, long live the IAT: Context-sensitive measures of implicit attitudes are indispensable to social and political psychology. *Current Directions in Psychological Science, 28*(1), pp. 10–19. https://doi.org/10.1177/0963721418797309

Jost, J. T., Banaji, M. R., & Nosek, B. A. (2004). A decade of System Justification Theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, *25*(6), 881–919.

Jost, J. T., Rudman, L. A., Blair, I. V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C. D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior*, *29*, 39–69.

Kaeble, D., & Cowhig, M. (2018). Correctional populations in the United States, 2016. *Bureau of Justice Statistics Special Report*, 1–13.

Kang, J., Bennett, M., Carbado, D., & Casey, P. (2011). Implicit bias in the courtroom. *UCLA Law Review*, *59*(5), 1124-1186.

Kang, J., & Lane, K. (2010). Seeing through colorblindness: Implicit bias and the law. *UCLA Law Review*, *58*(2), 465–520.

Kaplan, J. (2019). *Uniform Crime Reporting (UCR) Program Data: County-Level Detailed Arrest and Offense Data V3* [Data set]. https://doi.org/10.3886/E108164V3

Kent, S. L., & Carmichael, J. T. (2014). Racial residential segregation and social control: A panel study of the variation in police strength across U.S cities, 1980–2010. *American Journal of Criminal Justice*, *39*(2), 228–249.

Kent, S. L., & Jacobs, D. (2004). Social divisions and coercive control in advanced societies: Law enforcement strength in eleven nations from 1975 to 1994. *Social Problems*, *51*(3), 343–361.

Kent, S. L., & Jacobs, D. (2005). Minority threat and police strength from 1980 to 2000: A fixed-effects analysis of nonlinear and interactive effects in large US cities. *Criminology*, *43*(3), 731–760.

Kim, D.Y. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social*

Psychology Quarterly, 66(1), 83–96. https://doi.org/10.2307/3090143

King, R. D., & Wheelock, D. (2007). Group threat and social control: Race, perceptions of minorities and the desire to punish. *Social Forces*, *85*(3), 1255–1280.

Kochel, T. R., Wilson, D. B., & Mastrofski, S. D. (2011). Effect of suspect race on officers' arrest decisions. *Criminology, 49*(2), pp. 473–512. https://doi.org/10.1111/j.1745-9125.2011.00230.x

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., … Banaji, M. R. (2018). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *The American Psychologist*. https://doi.org/10.1037/amp0000364

Kushins, E. R. (2014). Sounding like your race in the employment process: An experiment on speaker voice, race identification, and stereotyping. *Race and Social Problems*, *6*(3), 237–248.

Kutateladze, B. L. (2018). Tracing charge trajectories: A study of the influence of race in charge changes at case screening, arraignment, and disposition. *Criminology*, *56*(1), 123–153.

Kutateladze, B. L., Andiloro, N. R., Johnson, B. D., & Spohn, C. C. (2014). Cumulative disadvantage: Examining racial and ethnic disparity in prosecution and sentencing. *Criminology*, *52*(3), 514–551.

Lane, K. A., Banaji, M. R., Nosek, B. A., & Greenwald, A. G. (2007). Understanding and using the Implicit Association Test: IV: Procedures and validity. *Implicit Measures of Attitudes: Procedures and Controversies*, 59–102.

Lautenschlager, R., & Omori, M. (2018). Racial threat, social (dis)organization, and the ecology of police: Towards a macro-level understanding of police use-of-force in communities of color. *Justice Quarterly*, pp. 1–22. https://doi.org/10.1080/07418825.2018.1480792

Lee, M. R., & Ousey, G. C. (2005). Institutional access, residential segregation, and urban black homicide. *Sociological Inquiry, 75*(1), 31–54. https://doi.org/10.1111/j.1475-682x.2005.00111.x

Leiber, M. J., Peck, J. H., & Rodriguez, N. (2016). Minority threat and juvenile court outcomes. *Crime & Delinquency*, *62*(1), 54–80.

Leitner, J. B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016a). Blacks' death rate due to circulatory diseases is positively related to whites' explicit racial bias. *Psychological Science, 27*(10), 1299–1311. https://doi.org/10.1177/0956797616658450

Leitner, J. B., Hehman, E., Ayduk, O., & Mendoza-Denton, R. (2016b). Racial bias is associated with ingroup death rate for Blacks and Whites: Insights from Project Implicit. *Social Science & Medicine*, 170, pp. 220–227. https://doi.org/10.1016/j.socscimed.2016.10.007

Leitner, J. B., Hehman, E., & Snowden, L. R. (2018). States higher in racial bias spend less on disabled medicaid enrollees. *Social Science & Medicine*, *208*, 150–157.

Levinson, J. D., Cai, H., & Young, D. (2010). Guilty by implicit racial bias: The guilty/not guilty Implicit Association Test. *Ohio St. J. Crim. L.*, *8*, 187.

Levinson, J. D., Smith, R. J., & Young, D. M. (2014). Devaluing death: An empirical study of implicit racial bias on jury-eligible citizens in six death penalty states. *NYUL Rev.*, *89*, 513.

Liska, A. E., & Chamlin, M. B. (1984). Social structure and crime control among macrosocial units. *American Journal of Sociology, 90*(2), 383–395. https://doi.org/10.1086/228084

Loeber, R., Farrington, D. P., Hipwell, A. E., Stepp, S. D., Pardini, D., & Ahonen, L. (2015). Constancy and change in the prevalence and frequency of offending when based on longitudinal self-reports or official records: Comparisons by gender, race, and crime type. *Journal of Developmental and Life-Course Criminology*, *1*(2), 150–168.

Mancini, C., Mears, D. P., Stewart, E. A., Beaver, K. M., & Pickett, J. T. (2012). Whites' perceptions about black criminality: A closer look at the contact hypothesis. *Crime & Delinquency*. https://doi.org/10.1177/0011128712461900

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, *37*(5), 435–442.

McNeish, D. M. (2014). Analyzing clustered data with OLS regression: The effect of a hierarchical data structure. *Multiple Linear Regression Viewpoints*, *40*(1), 11–16.

Mears, D. P., Cochran, J. C., & Lindsey, A. M. (2016). Offending and racial and ethnic disparities in criminal justice. *Journal of Contemporary Criminal Justice, 32*(1), pp. 78–103. https://doi.org/10.1177/1043986215607252

Miller, C. T., Varni, S. E., Solomon, S. E., DeSarno, M. J., & Bunn, J. Y. (2016). Macro-level implicit HIV prejudice and the health of community residents with HIV. *Health Psychology: Official Journal of the Division of Health Psychology, American Psychological Association*, *35*(8), 807–815.

Mitchell, G., & Tetlock, P. (2017). *Popularity as a Poor Proxy for Utility: The Case of Implicit Prejudice*. Retrieved from https://papers.ssrn.com/abstract=2973929

Mitchell, T. L., Haw, R. M., Pfeifer, J. E., & Meissner, C. A. (2005). Racial bias in mock juror decision-making: A meta-analytic review of defendant treatment. *Law and Human Behavior*, *29*(6), 621–637.

Monteith, M. J., Mark, A. Y., & Ashburn-Nardo, L. (2010). The self-regulation of prejudice: Toward understanding its lived character. *Group Processes & Intergroup Relations: GPIR*, *13*(2), 183–200.

Mosher, C. (2001). Predicting drug arrest rates: Conflict and social disorganization perspectives. *Crime & Delinquency*, *47*(1), 84–104.

Nagai, A. (n.d.). The Implicit Association Test: Flawed Science Tricks Americans into Believing They Are Unconscious Racists. Retrieved June 3, 2019, from The Heritage Foundation website: https://www.heritage.org/science-policy/report/the-implicit-association-test-flawed-science-tricks-americans-believing-they

Nosek, B. A., Greenwald, A. G., & Banaji, M. R. (2007). The Implicit Association Test at age 7: A methodological and conceptual review. In J. A. Bargh (Ed.), *Frontiers of social psychology. Social psychology and the unconscious: The automaticity of higher mental processes* (pp. 265–292). New York: Psychology Press.

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., … Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10593–10597.

Novak, J., Feyes, K. J., & Christensen, K. A. (2011). Application of intergroup contact theory to the integrated workplace: Setting the stage for inclusion. *Journal of Vocational Rehabilitation*, *35*(3), 211–226.

Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science*, *26*(5), 617–624.

Orchard, J., & Price, J. (2017). County-level racial prejudice and the black-white gap in infant health outcomes. *Social Science & Medicine*, *181*, 191–198.

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2015). Using the IAT to predict ethnic and racial discrimination: Small effect sizes of unknown societal significance. *Journal of Personality and Social Psychology*, *108*(4), 562–571.

Oude Maatman, F. (2017). *Implicit assumptions: A case study on the IAT controversy*. (Unpublished doctoral dissertation). Radboud University Nijmegen, Nijmegen, Netherlands.

Ousey, G. C., & Unnever, J. D. (2012). Racial-ethnic threat, out-group intolerance, and support for punishing criminals: A cross-national study. *Criminology*, *50*(3), 565–603.

Parker, K. F., Stults, B. J., & Rice, S. K. (2005). Racial threat, concentrated disadvantage and social control: Considering the macro-level sources of variation in arrests. *Criminology*, *43*(4), 1111–1134.

Payne, B. K., Vuletich, H. A., & Lundberg, K. B. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, *28*(4), 233–248.

Peterson, R. D., & Krivo, L. J. (1993). Racial segregation and black urban homicide. *Social Forces*, *71*(4), 1001–1026.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, *49*, 65–85.

Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751–783.

Petty, R. E., Fazio, R. H., & Brinol, P. (2012). *Attitudes: Insights from the New Implicit Measures*. New York: Psychology Press.

Pinkston, K. (2015). The black-white malleability gap in implicit racial evaluations: A nationally representative study. *The Journal of Social Psychology*, *155*(3), 189–203.

Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling*, *18*(2), 161–182.

Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological Methods*, *15*(3), 209–233.

Quillian, L. (1996). Group threat and regional change in attitudes toward African-Americans. *The American Journal of Sociology*, *102*(3), 816–860.

Rachlinski, J. J., Johnson, S. L., Wistrich, A. J., & Guthrie, C. (2009). Does unconscious bias affect trial judges? *The Notre Dame Law Review*, *84*(3), 1195–1246.

Rezaei, A. R. (2011). Validity and reliability of the IAT: Measuring gender and ethnic stereotypes. *Computers in Human Behavior*, 27, 1937–1941. https://doi.org/10.1016/j.chb.2011.04.018

Richardson, L. S., & Goff, P. A. (2012). Implicit racial bias in public defender triage. *Yale Law Journal*, *122*(8), 2626–2649.

Roch, C. H., & Edwards, J. (2017). Representative bureaucracy and school discipline: The influence of schools' racial contexts. *American Review of Public Administration*, *47*(1), 58–78.

Rooth, D.O. (2007). *Implicit Discrimination in Hiring: Real World Evidence*. Retrieved from https://papers.ssrn.com/abstract=984432

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological*

Science, *13*(2), 79–82.

Sadler, M. S., Correll, J., Park, B., & Judd, C. M. (2012). The world is not black and white: Racial bias in the decision to shoot in a multiethnic context. *The Journal of Social Issues*, *68*(2), 286–313.

Salman, J., Le Coz, E., & Johnson, E. (2016). Florida's broken sentencing system: Designed for fairness, it fails to account for prejudice. *Sarasota Herald-Tribune. Retrieved from http://projects.heraldtribune.com/bias/sentencing*.

Sampson, R. J., Morenoff, J. D., & Raudenbush, S. (2005). Social anatomy of racial and ethnic disparities in violence. *American Journal of Public Health, 95*(2), 224–232. https://doi.org/10.2105/ajph.2004.037705

Sharp, E. B. (2006). Policing urban America: A new look at the politics of agency size. *Social Science Quarterly*, *87*(2), 291–307.

Shermer, L. O., & Johnson, B. D. (2010). Criminal prosecutions: Examining prosecutorial discretion and charge reductions in U.S. Federal District Courts. *Justice Quarterly*, *27*(3), 394–430.

Smith, B. W., & Holmes, M. D. (2003). Community accountability, minority threat, and police brutality: An examination of civil rights criminal complaints. *Criminology 41*(4), 1035–1064.

Staats, C. (2013). *State of the Science: Implicit Bias Review.* Columbus, OH: Kirwan Institute for the Study of Race and Ethnicity.

Staats, C., Capatosto, K., Wright, R. A., & Jackson, V. W. (2016). *State of the Science: Implicit Bias Review, 2016 Edition*. Columbus, OH: Kirwan Institute for the Study of Race and

Ethnicity.

Stewart, E. A., Martinez, R., Baumer, E. P., & Gertz, M. (2015). The social context of Latino threat and punitive Latino sentiment. *Social Problems*, *62*(1), 68–92.

Stolzenberg, L., D'alessio, S. J., & Eitle, D. (2004). A multilevel test of racial threat theory. *Criminology, 42*(3), 673–698.

Stucky, T. D. (2012). The conditional effects of race and politics on social control: Black violent crime arrests in large cities, 1970 to 1990. *The Journal of Research in Crime and Delinquency*, *49*(1), 3–30.

Stults, B. J., & Baumer, E. P. (2007). Racial context and police force size: Evaluating the empirical validity of the minority threat perspective. *The American Journal of Sociology*, *113*(2), 507–546.

Stults, B. J., & Swagar, N. (2018). Racial and ethnic threat: Theory, research, and new directions. *The Handbook of Race, Ethnicity, Crime, and Justice*, 147–171.

Tausch, N., & Hewstone, M. (2010). Intergroup contact and prejudice. *The Sage Handbook of Prejudice, Stereotyping, and Discrimination*, 544–560.

Thomas, S. A., Moak, S. C., & Walker, J. T. (2013). The contingent effect of race in juvenile court detention decisions: The role of racial and symbolic threat. *Race and Justice*, *3*(3), 239–265.

Turner, M. A. (2009). *Promoting neighborhood diversity: Benefits, barriers, and strategies*. Washington, D.C.: Urban Institute.

Ulmer, J. T., & Johnson, B. (2004). Sentencing in context: A multilevel analysis. *Criminology* *42*(1), 137–178.

United States Department of Justice. (2018). *Census of Jails, 2013* [Data set]. https://doi.org/10.3886/ICPSR36128.v4

U.S. Census Bureau. (2008). *A compass for understanding and using American Community Survey data: What general data users need to know*. Washington, D.C.: US Government Printing Office.

U.S. Census Bureau. (2011). Tables B02001, B03003, B19083, S1201, S1903, and S2301 [Data set]. *American Community Survey 5-Year Estimates*. Retrieved from https://factfinder.census.gov/faces/nav/jsf/pages/searchresults.xhtml?refresh=t#

U.S. Sentencing Commission. (2018). Demographic differences in sentencing. *Federal Sentencing Reporter, 30*(3), 212–229. https://doi.org/10.1525/fsr.2018.30.3.212

Walker, S., Spohn, C., & DeLone, M. (2018). *The Color of Justice: Race, Ethnicity, and Crime in America*. Boston, MA: Cengage Learning.

Wallace, B. G. (2018, August 28). *Factors Associated with Implicit and Explicit Racial Preferences of African Americans*. (Unpublished doctoral dissertation). Emory University, Atlanta, GA.

Weisse, C. S., Sorum, P. C., Sanders, K. N., & Syat, B. L. (2001). Do gender and race affect decisions about pain management? *Journal of General Internal Medicine*, *16*(4), 211–217.

White, M. D., & Fradella, H. F. (2016). *Stop and Frisk: The Use and Abuse of a Controversial Policing Tactic*. New York: NYU Press.

Wilson, W. J. (1978). *The Declining Significance of Race*. Chicago: University of Chicago Press.

Xu, F. K., Nosek, B. A., Greenwald, A. G., Ratliff, K., Bar-Anan, Y., Umansky, E., … Axt, J. (2019). *Project Implicit Demo Website Datasets*. https://doi.org/10.17605/OSF.IO/Y9HIQ

Zerhouni, O., Rougier, M., & Muller, D. (2016). “Who (really) is charlie?” french cities with lower implicit prejudice toward Arabs Demonstrated larger participation rates in Charlie Hebdo rallies [“Qui est (Vraiment) Charlie?” Les Villes Françaises à plus Faible niveau de Préjugés Implicites envers les Maghrébins ont davantage Participé aux rassemblements de Charlie Hebdo]. *International Review of Social Psychology*, *29*(1). Retrieved from https://www.rips-irsp.com/article/10.5334/irsp.50/

Table 1: Descriptive Statistics of Study Variables

Variable	Mean	S.D.	Min	Max
B:W Confinement Rate	9.147	15.159	0	198.643
Mean Implicit Bias	.333	.069	.016	.514
Percent Black	.115	.127	.003	.648
B:W Unemployment Rate	2.021	.807	0	6.867
B:W Income Ratio	.671	.184	.313	1.506
B:W Marriage Rate	.614	.130	.184	1.092
GINI	.437	.033	.347	.541
Total Population	373,292	596,353	100,139	9,758,256
South	.336	.473	0	1
Percent Hispanic	.103	.115	.008	.819
B:W Arrest Rate Ratio	3.896	2.323	.559	17.031

Table 2: Bivariate Correlations of Study Variables

Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
(1) B:W Confinement Rate										
(2) Mean Implicit Bias	0.379									
(3) Percent Black	-0.352	-0.679								
(4) B:W Unemployment Rate	0.101	-0.086	0.185							
(5) B:W Income Ratio	0.058	0.098	-0.216	-0.412						
(6) B:W Marriage Rate	0.067	0.025	-0.065	-0.255	0.469					
(7) GINI	-0.110	-0.215	0.371	0.144	-0.350	-0.219				
(8) Total Population	0.069	-0.136	0.188	0.069	-0.090	0.110	0.324			
(9) South	-0.358	-0.322	0.449	-0.041	-0.019	0.098	0.069	-0.129		
(10) Percent Hispanic	0.096	-0.021	-0.129	-0.183	0.273	0.221	0.176	0.295	0.002	
(11) B:W Arrest Rate Ratio	0.397	0.383	-0.409	0.086	-0.076	-0.163	-0.260	-0.150	-0.359	-0.269

Table 3: Results of MSEM: Direct and Indirect Effects of Black Population Size on Jail Confinement Disparities

	Direct Effect			
	Implicit Bias		B:W Jail Confinement Rate (ln)	
Predictor	b (Robust S.E.)	95% C.I. Lower, Upper	b (Robust S.E.)	95% C.I. Lower, Upper
Implicit Bias	--	--	2.744*** (.653)	1.465, 4.023
Percent Black	-.397*** (.048)	-.492, -.303	-.105 (.553)	-1.190, .980
B:W Unemployment	.002 (.003)	.004, .009	.166 (.105)	-.039, .372
B:W Income	-.000 (.014)	-.027, .028	.120 (.309)	-.484, .725
B:W Marriage	.036 (.022)	-.008, .079	.801** (.289)	.234, 1.367
GINI	.238* (.112)	-.019, .458	-.048 (1.376)	-2.746, 2.649
Total Population (ln)	.000 (.003)	-.006, .005	.059 (.060)	-.058, .176
South	.007** (.009)	-.011, .025	-.308** (.117)	-.536, -.079
Percent Hispanic	-.030 (.031)	-.091, .031	1.190** (.401)	.405, 1.976
B:W Arrest Rate	.003*** (.001)	.001, .005	.113*** (.022)	.069, .157
Fixed-Effects Intercept	.377 (.008)	.361, .393	-.859 (.440)	-.280, 1.439
Variance Component	.000 (.000)	.000, .001	.014 (.010)	.004, .053
	Indirect Effect		Total Effect	
	b (Bootstrap S.E.)	Percentile Bootstrap 95% C.I.	b (Bootstrap S.E.)	Percentile Bootstrap 95% C.I.
Indirect & Total Effect of %Black →B:W Confinement	-1.091*** (.343)	-1.686, -.328	-1.196** (.417)	-2.202, -.555

N=437, AIC=-383.21, BIC=-281.212, df=25

Figure 1: Proposed MSEM with Mediation (controls not depicted)

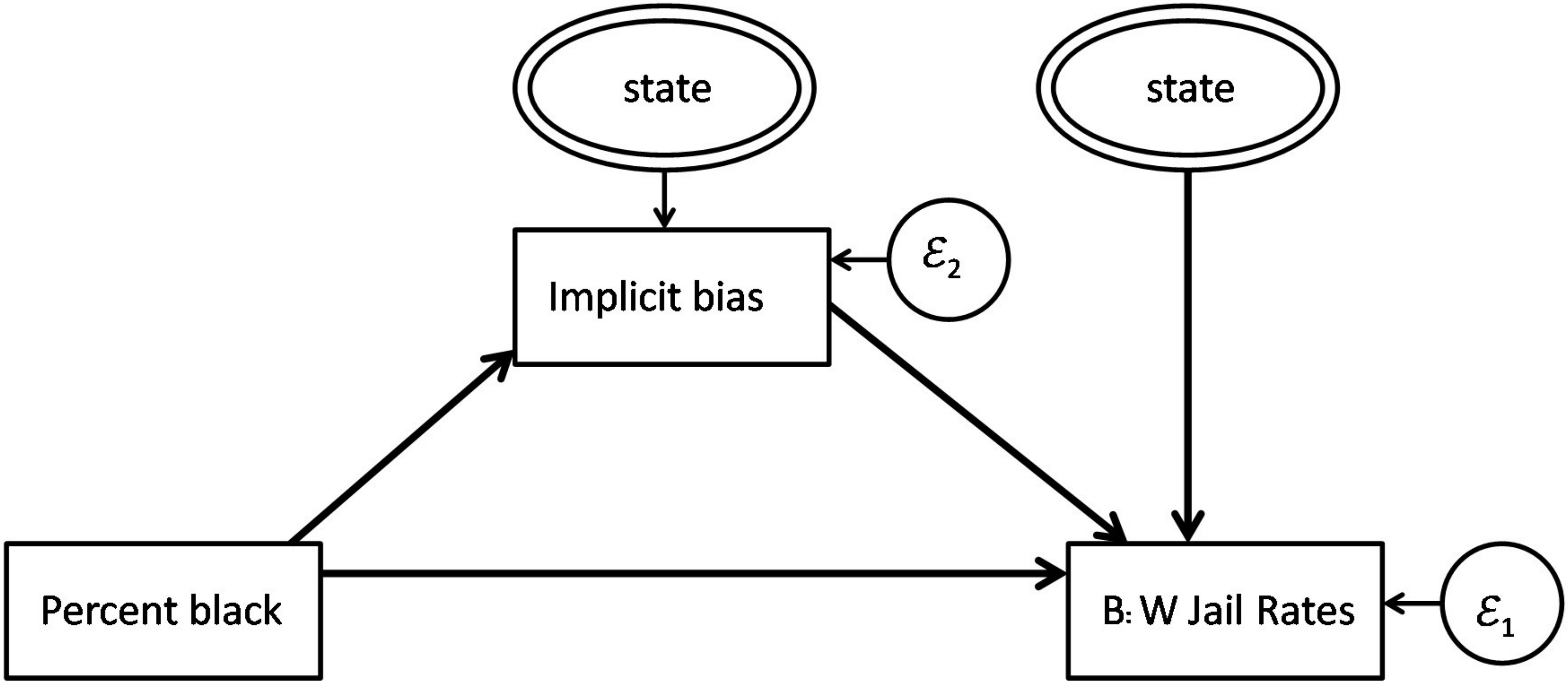
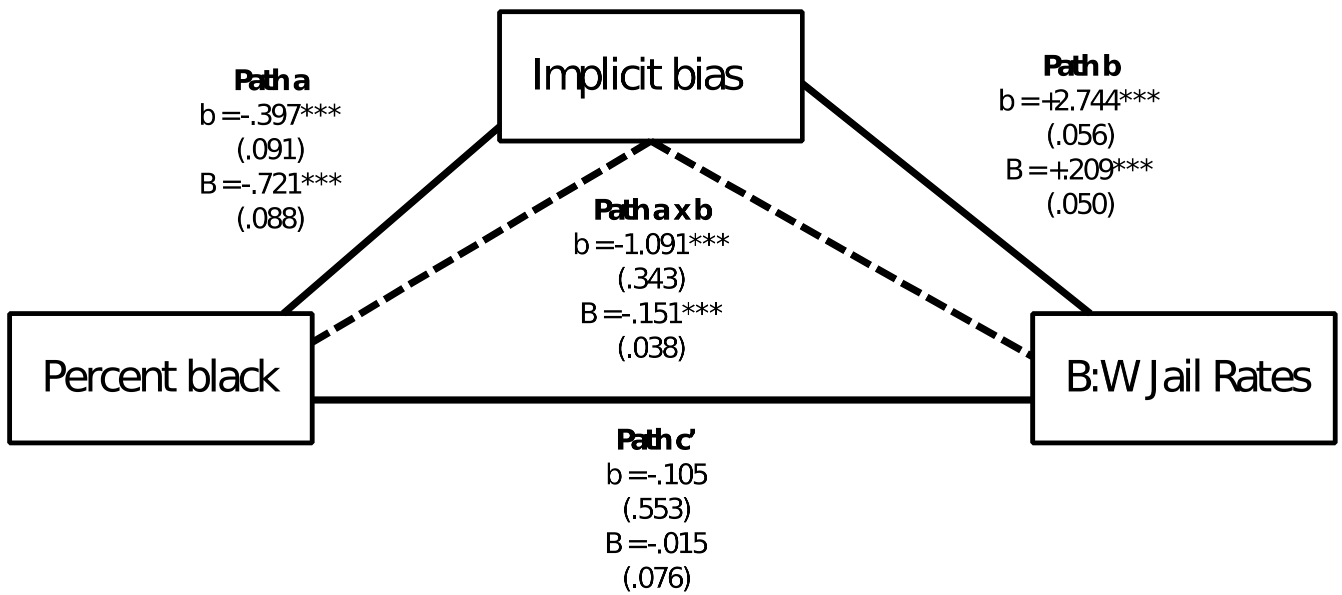


Figure 2: Direct and Indirect Effects of Key Study Variables



b = unstandardized coefficients; B = standardized coefficients. Standard errors in parentheses. Dashed line represents indirect effect. ***p<.001