Research Article

Psychological Science
2022, Vol. 33(4) 483–496

DOI: 10.1177/09567976211051272
www.psychologicalscience.org/PS

Racial Bias in Police Traffic Stops: White Residents' County-Level Prejudice and Stereotypes Are Related to Disproportionate Stopping of Black Drivers

Marleen Stelter[1,2], Iniobong Essien[2,3], Carsten Sander[1], and Juliane Degner[1]

[1]Social Psychology Department, Universität Hamburg; [2]Institute of Psychology, FernUniversität in Hagen; and [3]Department of Social and Organisational Psychology of Social Work, Leuphana Universität Lüneburg

Abstract

Racial disparities in policing are well documented, but the reasons for such disparities are often debated. In the current research, we weighed in on this debate using a regional-level bias framework: We investigated the link between racial disparities in police traffic stops and regional-level racial bias, employing data from more than 130 million police traffic stops in 1,413 U.S. counties and county-level measures of racial bias from more than 2 million online respondents. Compared with their population share in county demographics, Black drivers were stopped at disproportionate rates in the majority of counties. Crucially, disproportionate stopping of Black drivers was higher in counties with higher levels of racial prejudice by White residents (rs = .07–.36). Furthermore, county-level aggregates of White people's threat-related stereotypes were less consistent in predicting disproportionate stopping (rs = .00–.19). These observed relationships between regional-level bias and racial disparities in policing highlight the importance of the context in which police operate.

Keywords

regional-level bias, police traffic stops, systemic bias, prejudice, stereotypes, open data, open materials

Received 8/18/20; Revision accepted 8/11/21

Racial disparities in policing, ranging from higher rates of Black Americans being stopped, questioned, or frisked (e.g., Eberhardt, 2016) to incidents of police brutality and shootings (e.g., Hester & Gray, 2018; Voigt et al., 2017), are well documented. But there is an ongoing debate about the reasons for such disparities. Some writers have argued that racial disparities in policing are rooted in widespread racial bias (e.g., Lowery, 2020), whereas others have argued that "a few bad apples" are responsible for police misconduct (e.g., Benner, 2020). Such disparate views indicate that merely documenting racial disparities in policing offers neither sufficient explanation nor ways to overcome them (Hetey & Eberhardt, 2018). In the present research, we contributed to this debate by investigating whether the contexts in which police officers operate relate to racial disparities in policing outcomes.

Our analyses focused on police traffic stops, which are the most common situation in which Americans encounter police (Davis & Whyde, 2018). Broadly, two approaches have been used to investigate racial disparities in policing: field and lab studies. Field research demonstrates that Black drivers are stopped at higher rates than White drivers (e.g., Gelman et al., 2007; Langton & Durose, 2013; Pierson et al., 2020; Warren et al., 2006). Furthermore, traffic stops are often starting points for further racial disparities in policing: When stopped, Black Americans are treated less respectfully (Voigt et al., 2017) and are more likely to be searched (e.g., Higgins et al., 2011; Pierson et al., 2020), arrested

Corresponding Author:
Marleen Stelter, FernUniversität in Hagen, Institute of Psychology
Email: marleen.stelter@fernuni-hagen.de

(e.g., Kochel et al., 2011), and exposed to excessive force than White Americans (e.g., Edwards et al., 2019; Fryer, 2019). Thus, police traffic stops can set in motion a cascade of events that may threaten people's lives and livelihoods, ultimately undermining trust in law enforcement (Camp et al., 2021).

Although such field studies document the extent of racial disparities, they often allow only indirect inferences of why these disparities occur. For example, Pierson and colleagues (2020) conducted a "veil-of-darkness test" to examine racial profiling in police traffic stops. The rationale behind this test is that if disproportionate stopping of Black drivers were indeed caused by biased decision-making, racial disparities should be larger during daylight, when drivers' racial-group membership is more easily categorized, compared with during darkness (Grogger & Ridgeway, 2006). Consistent with this reasoning, results showed that racial disparities in police traffic stops were indeed smaller in darkness than during daylight, suggesting that racial-group membership might play a role in officers' stop decisions (but see Worden et al., 2012). Yet such indirect tests of racial bias in field research are limited because they cannot investigate police officers' attitudes, motives, and behavioral decisions while on duty. Alternatively, lab studies allow more fine-grained analyses of psychological processes in police officers' behavior. For example, some studies have linked racial disparities in use-of-force decisions to threat-related stereotypes (Correll et al., 2015) or prejudice (Eberhardt et al., 2004). This research provides insights into cognitive and affective processes in split-second decisions. However, it is debatable to what extent findings in the lab generalize to real-world police encounters because lab studies potentially underestimate (Jasperse et al., 2022) or overestimate (Cesario, 2021) effects of bias in real-world decision-making. In the current research, we combined both approaches to examine racial disparities in policing by connecting psychological studies of intergroup biases and field data of policing outcomes within a regional-level bias framework (see Hehman et al., 2019).

Applying this approach, a growing number of studies used aggregated measures of prejudice and stereotypes at the regional level (i.e., counties, states) to investigate relationships between racial bias and racial disparities in societal outcomes. Such studies have demonstrated that racial disparities in policing, health care, and education are more likely in regions where residents demonstrate higher levels of racial bias than in regions with lower levels of racial bias (e.g., Hehman et al., 2018; Orchard & Price, 2017; Riddle & Sinclair, 2019). Recently, Payne et al. (2017) proposed a theoretical framework for such findings, conceptualizing racial stereotypes and prejudice as relatively stable characteristics of contexts (rather than people), making the activation of biased thoughts and feelings in some contexts more likely than in others. Building on this idea, we propose that racial disparities in policing may thus not entirely depend on individual attitudes of police officers but may also depend on the contexts in which they act. There is evidence that policing outcomes vary by region (e.g., Police Scorecard, n.d.), and these differences may relate to characteristics of the local environment that affect both local residents' racial biases and policing behavior. For example, region-specific crime levels, racial segregation, quality and quantity of intergroup contact, local politics, and media content may affect whether and how strongly Black people are linked to stereotypes of criminality and threat (e.g., Sim et al., 2013). Consequently, in regions with higher levels of crime- or threat-related stereotyping, police officers might be cued to interpret Black drivers' behavior as suspicious. Another possibility is that regional levels of anti-Black and/or pro-White prejudice might affect racial disparities in police traffic stops. Higher regional levels of hostility toward Black people might increase police officers' inclination to stop Black drivers as a means of nuisance or harassment. Alternatively, higher regional levels of liking of White people might increase police officers' inclination to spare White drivers the nuisance of a traffic stop. Importantly, given that stereotypes and prejudice can be related or distinct

Statement of Relevance

In the United States, Black drivers are stopped with disproportionate frequency, and these traffic stops are often starting points for multiple forms of racial disparities in policing. Current societal debates center on whether this is the result of widespread racial bias or whether it should be attributed to the actions of a few rogue police officers. Analyzing massive data sets of more than 130 million traffic stops from 1,413 U.S. counties, we found that White residents' racial bias aggregated at the county-level relates to racial disparities in police traffic stops. These findings suggest that research on racial disparities in policing should not only focus on individual officers but also examine the social contexts in which police operate.

For additional thoughts on some of the psychological issues of societal importance considered in this research, see the invited Further Reflections piece authored by Payne and Rucker (2022), available online at https://doi.org/10.1177/09567976211056641 and on pages 666–668 of this issue.

components of intergroup attitudes (e.g., Amodio & Devine, 2006; Phills et al., 2020), regional-level stereotypes and prejudice may independently or interactively relate to racial disparities in police traffic stops.

On the basis of these considerations, we investigated regional levels of racial stereotypes and prejudice and their relationships with racial disparities in police traffic stops using large, publicly accessible data sets. We linked county-level aggregates of several measures of racial stereotypes and prejudice to county levels of racial disparities in police traffic stops. Societal racial bias and discrimination is described as relying on White power structures (e.g., Berard, 2008), and classic intergroup theories in social psychology suggest that societal institutions (e.g., law enforcement) are largely shaped by dominant groups (e.g., Sidanius & Pratto, 1999). Consequently, we focused our analyses on White residents' aggregates of stereotyping and prejudice as proxies of regional-level bias.[1]

Method

Data sources

Police traffic stops. We retrieved data on police traffic stops from the Stanford Open Policing Project (Pierson et al., 2020), which collects and standardizes data on vehicle and pedestrian stops from law enforcement departments across the United States. Our analyses included data on 134,016,874 state patrol traffic stops located in 1,413 counties across 24 U.S. states documented between the years 2000 and 2018. We included all traffic stops for which drivers' race and county information was reported. For three states (Illinois, New Jersey, and Vermont), we derived county information on the basis of zip codes and/or municipalities. We analyzed vehicle stops only and excluded data on pedestrian stops. For each county, we calculated the percentage of Black drivers among all stopped drivers. To examine whether Black drivers were stopped at disproportionate rates relative to their population share at the county level, we subtracted the percentage of Black residents in each county (as reported by the 2017 U.S. Census; U.S. Census Bureau, 2017) from the percentage of Black drivers stopped in each county. This resulted in a score of disproportionate stopping of Black drivers; zero indicates no disproportion, and a positive score indicates stopping rates higher than expected on the basis of county population.

Racial stereotypes and prejudice. Measures of racial stereotypes and prejudice were retrieved from Project Implicit (Xu et al., 2022). Threat-related stereotypes were measured with a weapons Implicit Association Test (IAT) as well as with items asking respondents how much more they associate Black people, relative to White people, with weapons and harmless objects. Prejudice was measured with an evaluative race IAT, an item asking respondents for their relative preference for White people relative to Black people, and feeling thermometers. We included data collected between 2002 and 2018 from U.S. respondents who self-identified as White and for whom geographic information was reported. On the basis of respondents' geographic information, we aggregated measures of prejudice and stereotypes at the county level.

Weapons IAT. In the weapons IAT, respondents used two response keys to classify Black and White faces according to race and images of objects as weapons or harmless objects. In trials during which White faces and harmless objects share one response key and Black faces and weapons share another response key (compared with the pairing of White faces and weapons and Black faces and harmless objects), faster responses and fewer errors are interpreted as an indirect indication of threat-related stereotype representations of Black people relative to White people. Responses were converted into IAT D scores (Greenwald et al., 2003). Higher IAT D scores indicate a stronger stereotype effect.

Self-reported racial stereotypes. To measure self-reported endorsement of threat-related racial stereotypes, we analyzed respondents' ratings on two items assessing how much they associate weapons and harmless objects with Black and White people on scales ranging from 1 (*strongly with White people/White Americans*) to 7 (*strongly with Black people/Black Americans*). We converted scales to range from −3 to +3.

Evaluative race IAT. In the race IAT, respondents use two response keys to classify Black and White faces according to race and positive and negative words (e.g., "lovely," "terrible") according to valence. In trials during which White faces and positive words share one response key and Black faces and negative words share another response key (compared with the pairing of White faces and negative words and Black faces and positive words), faster responses and fewer errors are interpreted as an indirect indication of pro-White/anti-Black evaluations (i.e., racial prejudice). Higher IAT D scores in the race IAT indicate a stronger preference for White relative to Black people.

Self-reported racial prejudice. To measure self-reported racial prejudice, we analyzed respondents' relative preference for White people relative to Black people on a scale ranging from "*I strongly prefer African Americans to European Americans*" to "*I strongly prefer European Americans to African Americans.*" The number of scale points varied between different versions of the item (5-point scale used

between 2004 and 2006; 7-point scale used between 2006 and 2017). We converted values from both versions to a combined scale ranging from −3 to +3.

To measure self-reported racial prejudice in another way, we analyzed how much warmth respondents felt toward White people and Black people on feeling-thermometer scales ranging from 0 (*very cold*) to 10 (*very warm*). We converted values to range from 0 to +3 and computed a difference score ranging from −3 to +3; positive values indicate relatively higher warmth felt for White people compared with Black people.

Analysis plan

To increase reliability for measures of stereotypes and prejudice, researchers have applied selection criteria of minimum respondents per county to be included in analyses (e.g., restricting analyses to 100 or 150 respondents per geographical unit; Hehman et al., 2018; Payne et al., 2019). Applying such fixed selection criteria is often arbitrary and without theoretical justification. Also, restricting the number of respondents reduces the number of counties included in analyses and may thereby reduce test power. To circumvent these problems and to increase transparency, we employed a multiverse approach, reporting findings across different selection criteria (i.e., including counties with more than 0, 25, 50, 75, 100, and 150 respondents, respectively).

We first computed zero-order correlations between disproportionate stopping of Black drivers and county-level stereotype and prejudice aggregates. To further examine the incremental value of different prejudice and stereotype measures and to control for potential covariates, we then tested several multiple-regression models. In a first series of models, we compared the contributions of different stereotype measures in predicting disproportionate stopping of Black drivers. Similarly, in a second series of models, we compared the contributions of different prejudice measures in predicting disproportionate stopping. In a third series of models, we compared the contributions of prejudice and stereotypes in predicting disproportionate stopping by adding all stereotype and prejudice measures as predictors. In a fourth and fifth series of models, we controlled for potential effects of county demographics, adding the percentage of Black and White residents as further predictors into the models. In a final set of analyses, we performed a veil-of-darkness test (Grogger & Ridgeway, 2006) and analyzed separate correlations between measures of racial bias and disproportionate stopping of Black drivers during daylight and darkness.

All analyses were conducted in the R programming environment (Version 4.1.0; R Core Team, 2019) using the R packages *broom.mixed* (Version 0.2.7; Bolker & Robinson, 2021), *corx* (Version 1.0.6.1; Conigrave, 2020), *cowplot* (Version 1.1.1; Wilke, 2019), *data.table* (Version 1.14.0; Dowle & Srinivasan, 2019), *gridExtra* (Version 2.3; Auguie, 2017), *here* (Version 1.0.1; Müller, 2017), *jtools* (Version 2.1.3; Long, 2020), *knitr* (Version 1.33; Xie, 2015), *lattice* (Version 0.20.44; Sarkar, 2008), *lme4* (Version 1.1.27.1; Bates et al., 2015), *lmerTest* (Version 3.1.3; Kuznetsova et al., 2017), *lsr* (Version 0.5; Navarro, 2015), *MBESS* (Version 4.8.0; Kelley, 2019), *MuMIn* (Version 1.43.17; Bartoń, 2020), *papaja* (Version 0.1.0.9997; Aust & Barth, 2020), *tidyverse* (Version 1.3.1; Wickham, 2017), and *usmap* (Version 0.5.2; Di Lorenzo, 2021). Our analysis code is publicly accessible via OSF (https://osf.io/7jm53/).

Results

Preliminary analyses

Police traffic stops. Compared with their population share in the county demographics, Black drivers were stopped 2.75% (*SD* = 5.38) more often. This score is significantly greater than zero, as determined by a two-sided one-sample *t* test, $t(1412) = 19.25$, $p < .001$, $d = 0.51$, 95% confidence interval (CI) = [0.46, 0.57], indicating disproportionate stopping of Black drivers. Percentiles of county-level disproportionate stopping of Black drivers are illustrated in Figure 1a.

Racial stereotypes and prejudice. Table 1 reports descriptive statistics for stereotype and prejudice measures for different selection criteria, restricted to counties for which we obtained data on police traffic stops. Results from the weapons IAT and from self-reported stereotypes showed that, on average, White respondents associate Black people with weapons, whereas they associate White people with harmless objects. Furthermore, White respondents' race IAT D scores indicate an overall preference for White people relative to Black people. Similarly, White respondents' self-reported measures of prejudice indicate an overall preference for White people and higher perceptions of warmth toward White people compared with Black people. Self-reported measures of prejudice were highly correlated at the county level (*r*s = .71–.85). We therefore converted both measures into *z* scores and averaged them to a combined score of self-reported racial prejudice for the regression analyses. Figures 1b and 1c illustrate percentiles of White residents' county-level race IAT D scores and the combined score for self-reported racial prejudice, respectively.

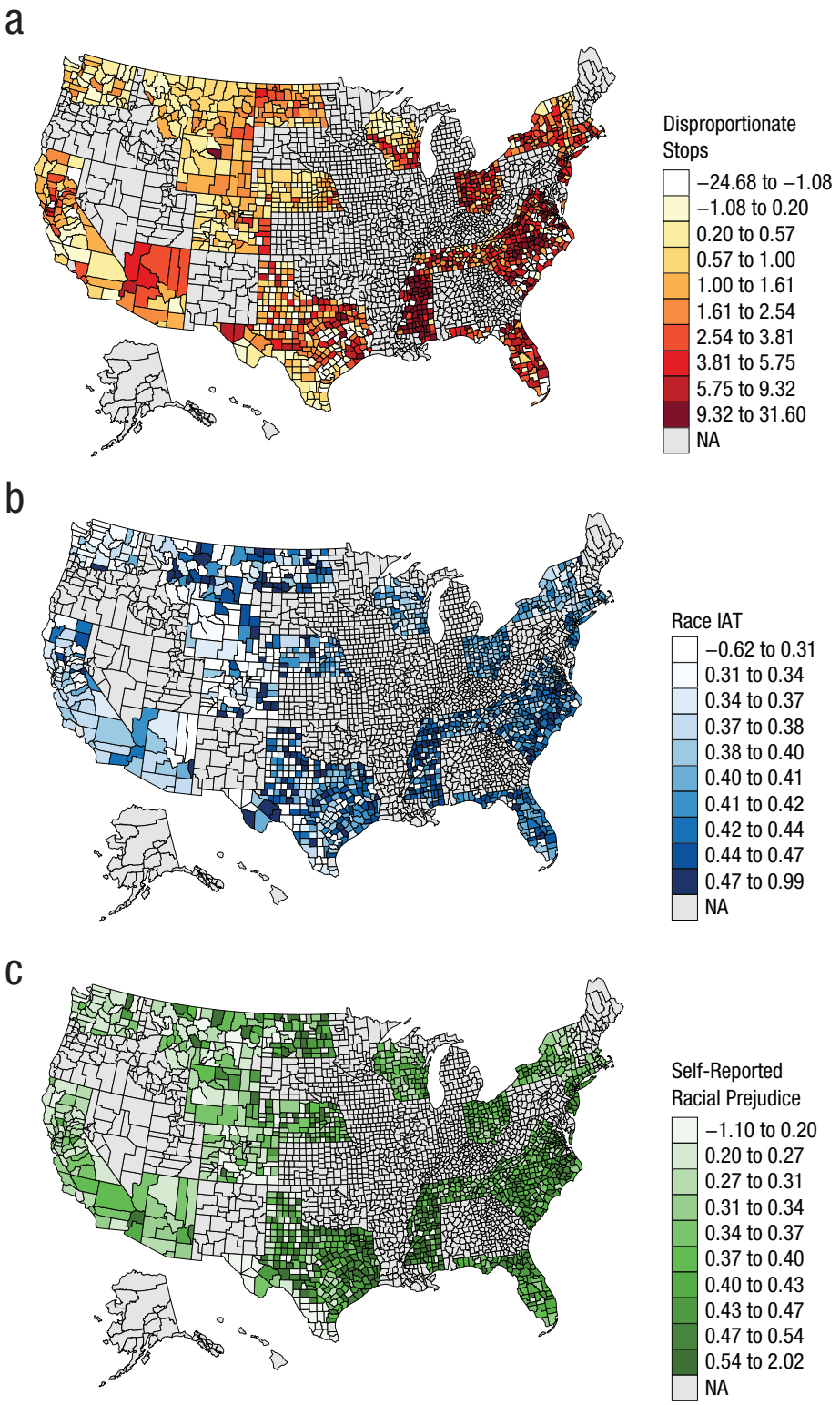

Fig. 1. Maps indicating percentiles of county-level disproportionate stopping of Black drivers (a), county-level race Implicit Association Test (IAT) D scores (b), and county-level self-reported racial prejudice (c).

Table 1. County-Level Descriptive Statistics for White Residents' Racial Stereotype and Prejudice Measures

	Weapons IAT		Stereotype harmless		Stereotype weapons		Race IAT		Preference score		Feeling thermometer	
Sample	*M* (*SD*)	*N*	*M* (*SD*)	*N*	*M* (*SD*)	*N*	*M* (*SD*)	*N*	*M* (*SD*)	*N*	*M* (*SD*)	*N*
> 0 respondents	0.35 (0.15)	1,413	–0.15 (0.27)	3,050	0.16 (0.42)	3,050	0.39 (0.09)	1,413	0.53 (0.22)	3,126	0.23 (0.14)	3,126
> 25 respondents	0.36 (0.05)	749	–0.16 (0.10)	1,410	0.20 (0.18)	1,413	0.39 (0.05)	1,108	0.53 (0.14)	2,398	0.23 (0.08)	2,429
> 50 respondents	0.36 (0.04)	585	–0.16 (0.08)	995	0.21 (0.16)	1,002	0.39 (0.04)	915	0.52 (0.12)	1,930	0.23 (0.07)	1,960
> 75 respondents	0.36 (0.04)	491	–0.16 (0.07)	800	0.22 (0.15)	803	0.39 (0.04)	806	0.52 (0.12)	1,653	0.23 (0.07)	1,683
> 100 respondents	0.36 (0.03)	438	–0.16 (0.06)	673	0.22 (0.14)	676	0.39 (0.04)	740	0.52 (0.11)	1,484	0.23 (0.07)	1,504
> 150 respondents	0.36 (0.03)	368	–0.16 (0.06)	530	0.23 (0.14)	533	0.39 (0.04)	635	0.52 (0.11)	1,234	0.23 (0.07)	1,264

Note: Means for the weapons and race Implicit Association Tests (IATs) are D scores. Stereotype scores are self-reported ratings of the extent to which harmless objects are associated with Black and White people (stereotype harmless) and the extent to which weapons are associated with Black and White people (stereotype weapons). Preference scores indicate preference for White people relative to Black people on a one-item measure, and values for the feeling thermometer are difference scores for White people relative to Black people. Positive scores on all measures indicate less favorable evaluations and more negative stereotypes of Black people relative to White people, whereas negative scores indicate more favorable evaluations and less negative stereotypes of Black people relative to White people.

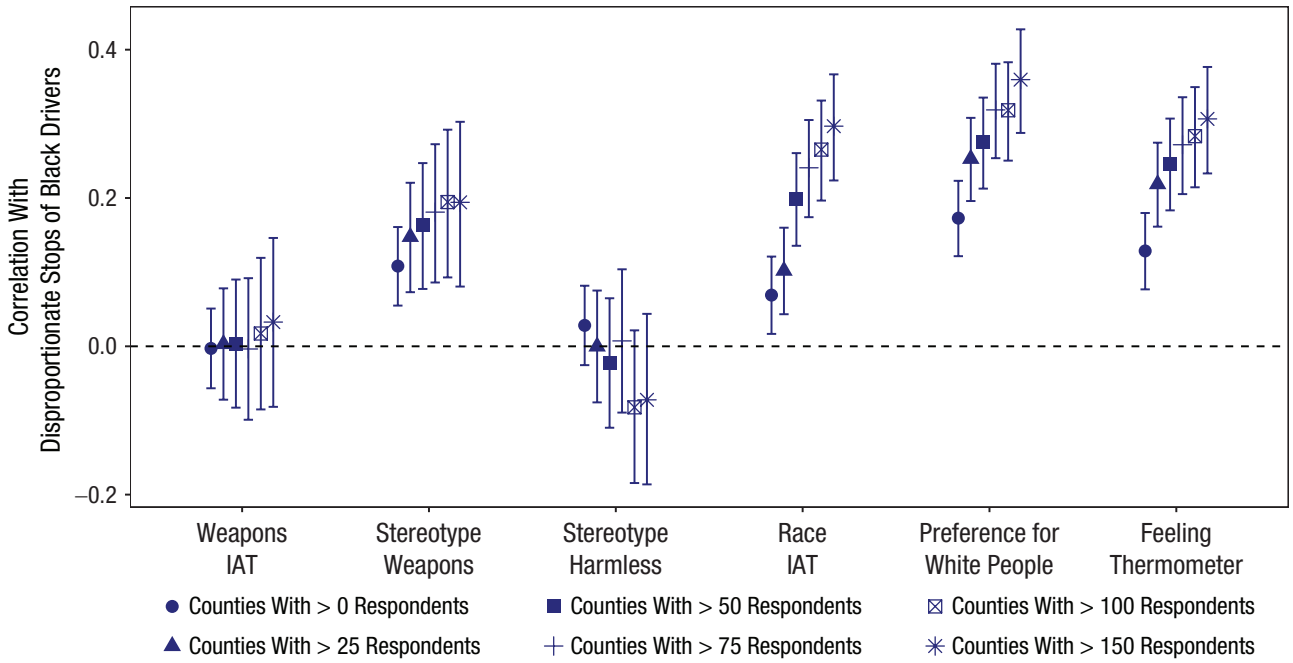

Fig. 2. Zero-order correlations between county-level disproportionate stopping of Black drivers and county-level measures of White residents' racial stereotypes and prejudice. Results are shown separately for analyses with full county-level data and analyses with only subsets of the data. Error bars represent 95% confidence intervals. IAT = Implicit Association Test.

Correlational analyses

Zero-order correlations. Figure 2 illustrates correlations between disproportionate stopping of Black drivers and county-level White residents' weapons IAT D scores, self-reported racial stereotypes, race IAT D scores, and self-reported preference and felt warmth for White people relative to Black people (see also Table S2 in the Supplemental Material available online). There were no correlations between White residents' weapons IAT D scores and disproportionate stopping of Black drivers (rs = .00–.03). However, we observed consistently positive correlations between self-reported associations of Black people with weapons and disproportionate stopping of Black drivers, regardless of selection criteria (rs = .11–.19). Furthermore, we observed no correlations between disproportionate stopping of Black drivers and self-reported associations of Black people with harmless objects (rs = −.08 to .03). Finally, we observed positive correlations between disproportionate stopping of Black drivers and race IAT D scores (rs = .07–.30), self-reported preference for White people (rs = .17–.36), and felt warmth for White people (rs = .13–.31), regardless of selection criteria. Overall, results indicate that disproportionate stopping of Black drivers was more prevalent in counties with higher levels of racial bias on four out of six measures and across different selection criteria.

Multiple-regression models. To test unique contributions of different measures of racial stereotypes and prejudice in predicting disproportionate stopping of Black drivers, we conducted several regression analyses (Fig. 3; see also Tables S3–S8 in the Supplemental Material). To account for the hierarchical data structure with counties nested within states, we fitted linear mixed-effects models. In a first model, we included stereotype measures (i.e., weapons IAT D scores and self-reported racial stereotypes) as predictors. In this model, none of the predictors were consistently related to disproportionate stopping of Black drivers (conditional Rs = .20–.40; marginal Rs = .00–.01). In a second model, we included prejudice measures (i.e., race IAT D scores and self-reported racial prejudice as the combined score of preference for White people and feeling-thermometer bias). In this model, only self-reported racial prejudice was a consistently significant predictor of disproportionate stopping of Black drivers (conditional Rs = .18–.35; marginal Rs = .01–.05). In a third model, we included all stereotype and prejudice measures. In this model, again only self-reported racial prejudice was a consistently significant predictor of disproportionate stopping of Black drivers (conditional Rs = .19–.33; marginal Rs = .02–.06).

Previous research suggests that racial bias is related to regional demographics, such as the percentage of Black or White people living in a region (e.g., Rae et al., 2015). Consequently, we conducted analyses, adding

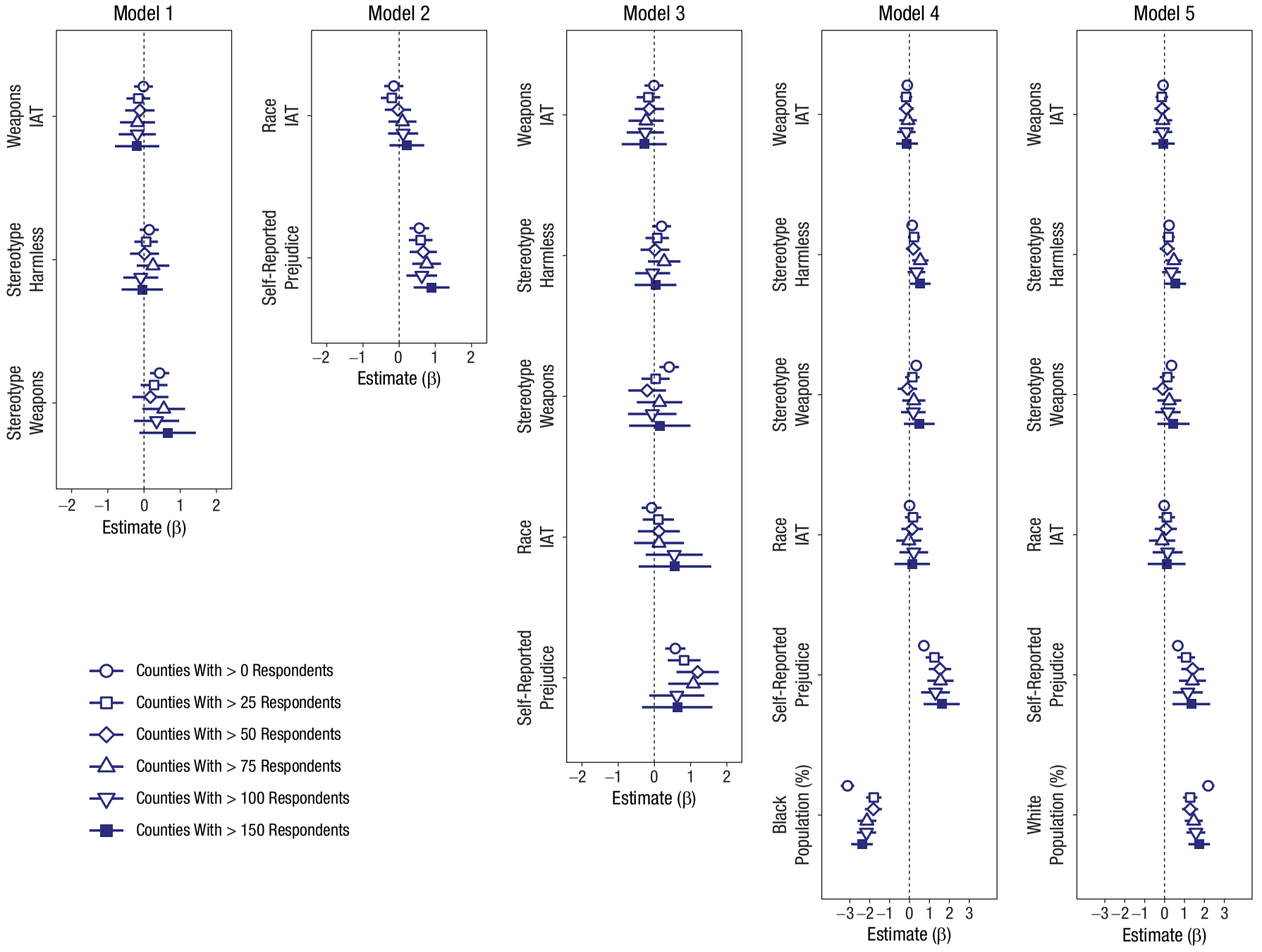

Fig. 3. Estimates from linear mixed-effects models predicting disproportionate stopping of Black drivers from county-level measures of White residents' racial stereotypes and prejudice. Results are shown separately for models with full county-level data and models with only subsets of the data. Error bars represent 95% confidence intervals. IAT = Implicit Association Test.

county proportions of Black and White residents as control variables. Because these variables were highly correlated ($r = -.91$), we fitted two separate models, adding county proportions of Black residents in a fourth model (conditional Rs = .51–.59; marginal Rs = .08–.21) and White residents in a fifth model (conditional Rs = .43–.46; marginal Rs = .07–.14). These models indicated that county demographics were significant predictors; lower proportions of Black residents and higher proportions of White residents were related to higher disproportionate stopping of Black drivers. Importantly, self-reported racial prejudice remained a significant independent predictor of disproportionate stopping of Black drivers in both models.

Veil-of-darkness test. To perform a veil-of-darkness test, we followed recommendations from previous research (e.g., Grogger & Ridgeway, 2006; Pierson et al., 2020; Worden et al., 2012), restricting data for police stops to those times of the day that can be either light or dark, depending on seasonal shifts of sunset. By restricting the data to this intertwilight period, we could rule out the possibility that effects were potentially confounded by differences in driving patterns (i.e., one group driving more often during earlier or later hours of the day). In addition, we removed stops during the ambiguous twilight period of approximately 30 min between sunset and dusk. Note that these restrictions considerably reduced the number of police stops included in our analyses to 9,506,820 stops during daylight and 9,140,738 stops during darkness, measured in 1,022 counties in 15 states. Our analyses demonstrate that proportions of stopped Black drivers in comparison with the Black county population were lower during daylight ($M = 2.23$, $SD = 5.21$) than during darkness ($M = 2.44$, $SD = 4.56$), $t(1009) = -5.21$, $p < .001$, $d_z = -0.16$, 95% CI = [–0.23, –0.10]. Thus, in contrast with the analyses by Pierson et al. (2020), ours did not reveal greater disparities at daylight (during which drivers' race is presumed to be more visible). Note that our analyses differed from those reported by Pierson et al. in two ways. First, ours were less restrictive in terms of data inclusion because Pierson et al. further restricted their analyses to data from the 60-day windows before and after the beginning of daylight saving time. The second difference from Pierson et al.'s analyses is that ours were based on county aggregates, relating proportions of stopped Black drivers to proportions of Black county populations. These differences in analysis strategies may account for the different results. When we applied the same correlational analyses as above to police stops at daylight versus darkness, results showed that relationships between racial bias and disproportionate stopping of Black drivers were almost identical for both lighting conditions (see Fig. 4).

General Discussion

Combining data of more than 130 million traffic stops with regional aggregates of racial prejudice and stereotypes, we observed that racial disparities in police traffic stops were related to White people's local levels of racial bias. These relationships were observed across different model specifications, highlighting their robustness. Relationships between racial bias and police traffic stops were consistent across different measures of racial prejudice but less consistent across measures of threat-related racial stereotypes. Furthermore, these relationships were observed regardless of daytime and after controlling for county demographics. The present findings are consistent with results of previous field studies documenting racial disparities in policing (e.g., Pierson et al., 2020). Connecting policing data to psychological studies of intergroup biases, our approach adds another layer by demonstrating robust relationships between racial disparities in policing outcomes and the contexts in which police officers operate.

Our finding that White county residents' racial bias was associated with county-level racial disparities in traffic stops is consistent with other findings, such as that Black people were disproportionately shot by police in regions with higher levels of racial prejudice and threat-related stereotypes (Hehman et al., 2018). Although consistent overall, the present results depart from this prior finding in two ways. First, we observed stronger relationships for self-report measures than for indirect measures of racial bias. Second, our findings suggest that in stop decisions, prejudice toward Black people might be more relevant than threat-related racial stereotypes. These discrepancies might be explained by situational differences between shoot and stop decisions. Shoot decisions may occur more often in situations involving perceived immediate threats of physical harm, eliciting high arousal and fear and potentially reducing deliberate behaviors while increasing influences of threat-related stereotypes on police officers' behavior (e.g., Correll et al., 2014). Consequently, situations surrounding shoot decisions should relate more closely to measures of threat-related stereotypes (March et al., 2020). Conversely, relationships between stop decisions and measures of threat seem less clear.

The role of prejudice in stop decisions is worth highlighting. Some scholars have argued that disparities in policing do not necessarily reflect racial bias but instead reflect that Black people are more often involved in crime, which in turn shapes people's racial stereotypes (e.g., Cesario, 2021). The present findings challenge such notions, suggesting that relative liking and preference

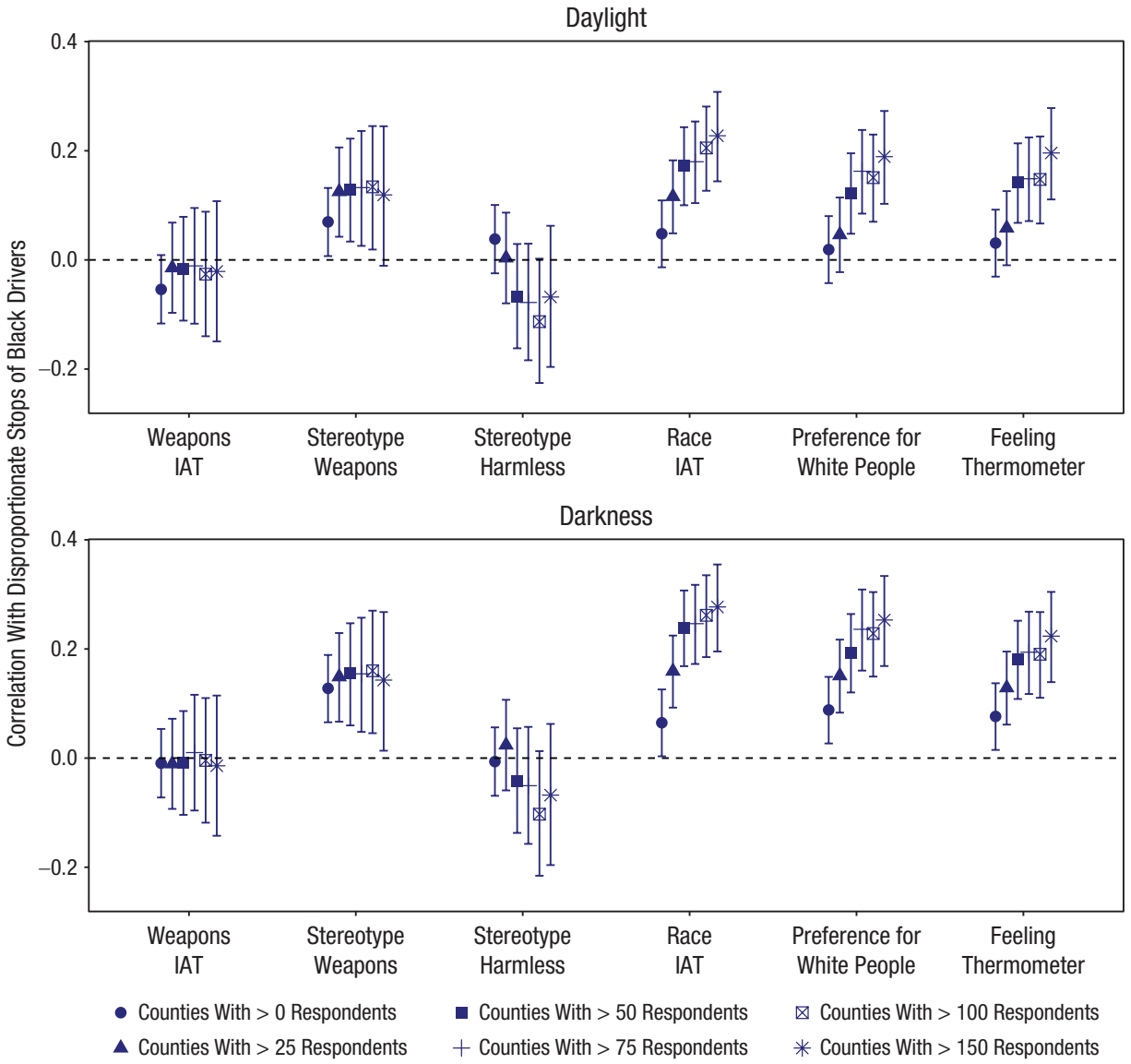

Fig. 4. Zero-order correlations between county-level disproportionate stopping of Black drivers and county-level measures of White residents' racial stereotypes and prejudice during daylight (upper panel) and during darkness (lower panel). Results are shown separately for analyses with full county-level data and analyses with only subsets of the data. Error bars represent 95% confidence intervals. IAT = Implicit Association Test.

for White people over Black people played a more important role in racial disparities in police traffic stops than stereotypes (see also Essien et al., 2022).

The observed relationships between regional-level bias and police traffic stops underscore the role of the context in which police officers operate. Our findings are consistent with theorizing by Payne et al. (2017), who argued that some contexts expose individuals more regularly to stereotypes and/or prejudice, increasing mental accessibility of biased thoughts and feelings, in turn influencing individual behavior. Consequently, behavioral expressions of prejudice and stereotypes often reflect properties of contexts rather than stable dispositions of people (but see Connor & Evers, 2020).

One possible explanation of how regional-level bias might affect policing is that it reflects regional norms that interact with institutionalized practices in law enforcement. For example, police traffic stops in the United States are regularly used as investigative tools that encourage officers to maximize the number of traffic stops based on "suspicion of criminal activity" (Epp et al., 2017, p. 172). Although racially neutral on the surface, such practices may connect regional-level bias with police officers' behavior: Incentives to maximize the number of stops in conjunction with local norms

that favor White people might encourage police officers to disproportionately stop Black drivers. Moreover, when instructed to base initial stop decisions on "gut feelings," officers may rely on contextually available prejudice or stereotypes (De Houwer & Tucker Smith, 2013; Epp et al., 2017; Swencionis et al., 2021). Additionally, racial bias of supervisors may affect decisions to direct officers to regions where they more likely encounter Black drivers (Beckett et al., 2006), consistent with documentations of overpatrolling of minority areas (Vomfell & Stewart, 2021). Thus, racial bias may affect investigatory stops even in the absence of prejudice or stereotyping on the part of the individual police officers involved. Future research needs to identify the psychological mechanisms that mediate between community racial bias and officer behavior, for example, by including direct investigation of institutionalized practices in law enforcement.

The veil-of-darkness test did not demonstrate greater racial disparities in police traffic stops during daylight compared with darkness, thus departing from previous research relying on similar data (Pierson et al., 2020). Although the hypothesis that darkness eliminates social categorization of drivers is intuitively convincing, it may not necessarily be true, especially because racial categorization has been shown to be robust to suboptimal viewing conditions (e.g., Kawakami et al., 2017). Given that the validity of the veil-of-darkness test hinges on a number of assumptions (e.g., that police officers use person features and neglect other information, such as vehicle type) and given other studies in which relationships between lighting conditions and racial disparities in police traffic stops were not observed (e.g., Worden et al., 2012), future research is needed to understand the boundary conditions of the veil-of-darkness test. Importantly, correlations between racial-bias measures and disproportionate stopping rates were unaffected by lighting conditions, underlining the robustness of these relationships.

The present research has limitations. First, the correlational approach does not allow causal interpretations. On the basis of previous theorizing (Payne et al., 2019), we suggest that regional-level bias provides a context that potentially affects police officers' behavior. Alternatively, racial disparities in police traffic stops might affect regional-level bias by triggering negative evaluations linking Black people with crime (e.g., Hetey & Eberhardt, 2018). It is also possible that third variables affect both racial bias and policing behavior. For example, regional differences in socioeconomic disparities between Black and White residents might affect both racial bias and policing. This nonexhaustive list of alternative causal relationships illustrates the need for future research. Another limitation originates from the use of Project Implicit data (Xu et al., 2022), which are not representative of the U.S. population. Participants visit the demonstration website voluntarily and skew liberal (e.g., Essien et al., 2021). Similarly, because of the available data sets, our analyses included only one third of U.S. counties, excluding data for many midwestern and southern states. Consequently, we do not know whether or how including these regions might alter conclusions of this research.

Despite these limitations, it is important to highlight that the observed relationships were consistent across different measures of racial bias and relied on millions of online respondents and tens of millions of traffic stops. Our findings suggest that researchers analyzing racial disparities in policing should not only focus on individual officers but also consider the contexts in which police operate.

Transparency

Action Editor: Patricia J. Bauer
Editor: Patricia J. Bauer
Author Contributions

M. Stelter and I. Essien are joint first authors of this article; they developed the study concept and wrote the manuscript. C. Sander retrieved and prepared the data sets. M. Stelter prepared and analyzed the data and created the tables and figures. All the authors interpreted the results, revised the manuscript, and approved the final manuscript for submission.

Declaration of Conflicting Interests

The author(s) declared that there were no conflicts of interest with respect to the authorship or the publication of this article.

Open Practices

All data and analysis code have been made publicly available via OSF and can be accessed at https://osf.io/7jm53/. The design and analysis plan for the study were not preregistered. This article has received the badges for Open Data and Open Materials. More information about the Open Practices badges can be found at http://www.psychologicalscience.org/publications/badges.

ORCID iD

Marleen Stelter https://orcid.org/0000-0001-6546-0946

Supplemental Material

Additional supporting information can be found at http://journals.sagepub.com/doi/suppl/10.1177/09567976211051272

Note

1. We report additional analyses with Black residents' racial bias in Table S10 and Figure S3 in the Supplemental Material available online. These analyses show no consistent relationships between Black residents' racial bias and racial disparities in police traffic stops.

References

Amodio, D. M., & Devine, P. G. (2006). Stereotyping and evaluation in implicit race bias: Evidence for independent constructs and unique effects on behavior. *Journal of Personality and Social Psychology*, *91*(4), 652–661. https://doi.org/10.1037/0022-3514.91.4.652

Auguie, B. (2017). *gridExtra: Miscellaneous functions for "grid" graphics* (Version 2.3) [Computer software]. https://CRAN.R-project.org/package=gridExtra

Aust, F., & Barth, M. (2020). *papaja: Create APA manuscripts with R Markdown* (Version 0.1.0.9997) [Computer software]. https://github.com/crsh/papaja

Bartoń, K. (2020). *MuMIn: Multi-model inference* (Version 1.43.17) [Computer software]. https://CRAN.R-project.org/package=MuMIn

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beckett, K., Nyrop, K., & Pfingst, L. (2006). Race, drugs, and policing: Understanding disparities in drug delivery arrests. *Criminology*, *44*(1), 105–137. https://doi.org/10.1111/j.1745-9125.2006.00044.x

Benner, K. (2020, June 7). Barr says there is no systemic racism in policing. *The New York Times*. https://www.nytimes.com/2020/06/07/us/politics/justice-department-barr-racism-police.html

Berard, T. J. (2008). The neglected social psychology of institutional racism. *Sociology Compass*, *2*(2), 734–764. https://doi.org/10.1111/j.1751-9020.2007.00089.x

Bolker, B., & Robinson, D. (2021). *broom.mixed: Tidying methods for mixed models* (Version 0.2.7) [Computer software]. https://CRAN.R-project.org/package=broom.mixed

Camp, N. P., Voigt, R., Jurafsky, D., & Eberhardt, J. L. (2021). The thin blue waveform: Racial disparities in officer prosody undermine institutional trust in the police. *Journal of Personality and Social Psychology*, *121*(6), 1157–1171. https://doi.org/10.1037/pspa0000270

Cesario, J. (2021). What can experimental studies of bias tell us about real-world group disparities? *Behavioral & Brain Sciences*. Advance online publication. https://doi.org/10.1017/S0140525X21000017

Conigrave, J. (2020). *corx: Create and format correlation matrices* (Version 1.0.6.1) [Computer software]. https://CRAN.R-project.org/package=corx

Connor, P., & Evers, E. R. K. (2020). The bias of individuals (in crowds): Why implicit bias is probably a noisily measured individual-level construct. *Perspectives on Psychological Science*, *15*(6), 1329–1345. https://doi.org/10.1177/1745691620931492

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. https://doi.org/10.1111/spc3.12099

Correll, J., Wittenbrink, B., Crawford, M. T., & Sadler, M. S. (2015). Stereotypic vision: How stereotypes disambiguate visual stimuli. *Journal of Personality and Social Psychology*, *108*(2), 219–233. https://doi.org/10.1037/pspa0000015

Davis, E., & Whyde, A. (2018). *Contacts between police and the public, 2015*. Bureau of Justice Statistics. https://www.bjs.gov/content/pub/pdf/cpp15.pdf

De Houwer, J., & Tucker Smith, C. (2013). Go with your gut! Effects in the affect misattribution procedure become stronger when participants are encouraged to rely on their gut feelings. *Social Psychology*, *44*(5), 299–302. http://doi.org/10.1027/1864-9335/a000115

Di Lorenzo, P. (2021). *usmap: US maps including Alaska and Hawaii* (Version 0.5.2) [Computer software]. https://CRAN.R-project.org/package=usmap

Dowle, M., & Srinivasan, A. (2019). *data.table: Extension of 'data.frame'* (Version 1.14.0) [Computer software]. https://CRAN.R-project.org/package=data.table

Eberhardt, J. L. (Ed.) (2016). *Strategies for change: Research initiatives and recommendations to improve police-community relations in Oakland, Calif*. Social Psychological Answers to Real-World Questions, Stanford University. https://stanford.app.box.com/v/Strategies-for-Change

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. https://doi.org/10.1037/0022-3514.87.6.876

Edwards, F., Lee, H., & Esposito, M. (2019). Risk of being killed by police use of force in the United States by age, race-ethnicity, and sex. *Proceedings of the National Academy of Sciences, USA*, *116*(34), 16793–16798. https://doi.org/10.1073/pnas.1821204116

Epp, C. R., Maynard-Moody, S., & Haider-Markel, D. (2017). Beyond profiling: The institutional sources of racial disparities in policing. *Public Administration Review*, *77*(2), 168–178. https://doi.org/10.1111/puar.12702

Essien, I., Calanchini, J., & Degner, J. (2021). Moderators of intergroup evaluation in disadvantaged groups: A comprehensive test of predictions from system justification theory. *Journal of Personality and Social Psychology*, *120*(5), 1204–1230. https://doi.org/10.1037/pspi0000302

Essien, I., Stelter, M., Rohmann, A., & Degner, J. (2022). Beyond stereotypes: Prejudice as an important missing force explaining group disparities. *Behavioral & Brain Sciences*.

Fryer, R. G., Jr. (2019). An empirical analysis of racial differences in police use of force. *Journal of Political Economy*, *127*(3), 1210–1261. https://doi.org/10.1086/701423

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. https://doi.org/10.1198/016214506000001040

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Grogger, J., & Ridgeway, G. (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of*

the American Statistical Association, *101*(475), 878–887. https://doi.org/10.1198/016214506000000168

Hehman, E., Calanchini, J., Flake, J. K., & Leitner, J. B. (2019). Establishing construct validity evidence for regional measures of explicit and implicit racial bias. *Journal of Experimental Psychology: General*, *148*(6), 1022–1040. https://doi.org/10.1037/xge0000623

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. https://doi.org/10.1177/1948550617711229

Hester, N., & Gray, K. (2018). For Black men, being tall increases threat stereotyping and police stops. *Proceedings of the National Academy of Sciences, USA*, *115*(11), 2711–2715. https://doi.org/10.1073/pnas.1714454115

Hetey, R. C., & Eberhardt, J. L. (2018). The numbers don't speak for themselves: Racial disparities and the persistence of inequality in the criminal justice system. *Current Directions in Psychological Science*, *27*(3), 183–187. https://doi.org/10.1177/0963721418763931

Higgins, G. E., Jennings, W. G., Jordan, K. L., & Gabbidon, S. L. (2011). Racial profiling in decisions to search: A preliminary analysis using propensity-score matching. *International Journal of Police Science & Management*, *13*(4), 336–347. https://doi.org/10.1350/ijps.2011.13.4.232

Jasperse, L., Stillerman, B. S., & Amodio, D. M. (2022). Missing context from experimental studies amplifies, rather than negates, racial bias in the real world. *Behavioral & Brain Sciences*.

Kawakami, K., Amodio, D. M., & Hugenberg, K. (2017). Intergroup perception and cognition: An integrative framework for understanding the causes and consequences of social categorization. In J. M. Olson (Ed.), *Advances in experimental social psychology* (Vol. 55, pp. 1–80). Academic Press. https://doi.org/10.1016/bs.aesp.2016.10.001

Kelley, K. (2019). *MBESS: The MBESS R package* (Version 4.8.0) [Computer software]. https://CRAN.R-project.org/package=MBESS

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

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Langton, L., & Durose, M. (2013). *Police behavior during traffic and street stops, 2011*. U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics. https://bjs.ojp.gov/content/pub/pdf/pbtss11.pdf

Long, J. A. (2020). *jtools: Analysis and presentation of social scientific data* (Version 2.1.3) [Computer software]. https://cran.r-project.org/package=jtools

Lowery, W. (2020, June 10). Why Minneapolis was the breaking point. *The Atlantic*. https://www.theatlantic.com/politics/archive/2020/06/wesley-lowery-george-floyd-minneapolis-black-lives/612391/

March, D. S., Olson, M. A., & Gaertner, L. (2020). Lions, and tigers, and implicit measures, oh my! Implicit assessment and the valence vs. threat distinction. *Social Cognition*, *38*, 154–s164. https://doi.org/10.1521/soco.2020.38.supp.s154

Müller, K. (2017). *here: A simpler way to find your files* (Version 1.0.1) [Computer software]. https://CRAN.R-project.org/package=here

Navarro, D. (2015). *Learning statistics with R: A tutorial for psychology students and other beginners* (Version 0.5). The University of Adelaide. https://open.umn.edu/opentextbooks/textbooks/559

Orchard, J., & Price, J. (2017). County-level racial prejudice and the Black-White gap in infant health outcomes. *Social Science & Medicine*, *181*, 191–198. https://doi.org/10.1016/j.socscimed.2017.03.036

Payne, B. K., & Rucker, J. M. (2022). Explaining the spatial patterning of racial disparities in traffic stops requires a structural perspective: Further reflections on Stelter et al. (2022) and Ekstrom et al. (2022). *Psychological Science*, *33*(4), 666–668. https://doi.org/10.1177/09567976211056641

Payne, B. K., Vuletich, H. A., & Brown-Iannuzzi, J. L. (2019). Historical roots of implicit bias in slavery. *Proceedings of the National Academy of Sciences, USA*, *116*(24), 11693–11698. https://doi.org/10.1073/pnas.1818816116

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. https://doi.org/10.1080/1047840X.2017.1335568

Phills, C. E., Hahn, A., & Gawronski, B. (2020). The bidirectional causal relation between implicit stereotypes and implicit prejudice. *Personality and Social Psychology Bulletin*, *46*(9), 1318–1330. https://doi.org/10.1177/0146167219899234

Pierson, E., Simoiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., & Goel, S. (2020). A large-scale analysis of racial disparities in police stops across the United States. *Nature Human Behaviour*, *4*(7), 736–745. https://doi.org/10.1038/s41562-020-0858-1

Police Scorecard. (n.d.). *Policing differs substantially depending where you live*. https://policescorecard.org/findings#policing-differs

Rae, J. R., Newheiser, A.-K., & Olson, K. R. (2015). Exposure to racial out-groups and implicit race bias in the United States. *Social Psychological and Personality Science*, *6*(5), 535–543. https://doi.org/10.1177/1948550614567357

R Core Team. (2019). *R: A language and environment for statistical computing* (Version 4.1.0) [Computer software]. https://www.R-project.org/

Riddle, T., & Sinclair, S. (2019). Racial disparities in school-based disciplinary actions are associated with county-level rates of racial bias. *Proceedings of the National Academy of Sciences, USA*, *116*(17), 8255–8260. https://doi.org/10.1073/pnas.1808307116

Sarkar, D. (2008). *lattice: Multivariate data visualization with R*. Springer. http://lmdvr.r-forge.r-project.org

Sidanius, J., & Pratto, F. (1999). *Social dominance: An intergroup theory of social hierarchy and oppression*. Cambridge University Press.

Sim, J. J., Correll, J., & Sadler, M. S. (2013). Understanding police and expert performance: When training attenuates (vs. exacerbates) stereotypic bias in the decision to shoot. *Personality and Social Psychology Bulletin*, *39*(3), 291–304. https://doi.org/10.1177/0146167212473157

Swencionis, J. K., Pouget, E. R., & Goff, P. A. (2021). Supporting social hierarchy is associated with White police officers' use of force. *Proceedings of the National Academy of Sciences, USA*, *118*(18), Article e2007693118. https://doi.org/10.1073/pnas.2007693118

U.S. Census Bureau. (2017). *Annual county resident population estimates by age, sex, race, and Hispanic origin*. https://www2.census.gov/programs-surveys/popest/datasets/2010-2017/counties/asrh/cc-est2017-alldata.csv

Voigt, R., Camp, N. P., Prabhakaran, V., Hamilton, W. L., Hetey, R. C., Griffiths, C. M., Jurgens, D., Jurafsky, D., & Eberhardt, J. L. (2017). Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences, USA*, *114*(25), 6521–6526. https://doi.org/10.1073/pnas.1702413114

Vomfell, L., & Stewart, N. (2021). Officer bias, over-patrolling and ethnic disparities in stop and search. *Nature Human Behaviour*, *5*, 566–575. https://doi.org/10.1038/s41562-020-01029-w

Warren, P., Tomaskovic-Devey, D., Smith, W., Zingraff, M., & Mason, M. (2006). Driving while Black: Bias processes and racial disparity in police stops. *Criminology*, *44*(3), 709–738. https://doi.org/10.1111/j.1745-9125.2006.00061.x

Wickham, H. (2017). *tidyverse: Easily install and load the 'tidyverse'* (Version 1.3.1) [Computer software]. https://CRAN.R-project.org/package=tidyverse

Wilke, C. O. (2019). *cowplot: Streamlined plot theme and plot annotations for 'ggplot2'* (Version 1.1.1) [Computer software]. https://CRAN.R-project.org/package=cowplot

Worden, R. E., McLean, S. J., & Wheeler, A. P. (2012). Testing for racial profiling with the veil-of-darkness method. *Police Quarterly*, *15*(1), 92–111. https://doi.org/10.1177/1098611111433027

Xie, Y. (2015). *Dynamic documents with R and knitr* (2nd ed.). Chapman and Hall/CRC. https://yihui.org/knitr/

Xu, F. K., Nosek, B. A., Greenwald, A. G., Ratliff, K., Bar-Anan, Y., Umansky, E., Banaji, M. R., Lofaro, N., Smith, C., Axt, J., Lai, C. K., Simon, L., O'Shea, B., & Frost, N. (2022). *Project Implicit demo website datasets*. OSF. https://doi.org/10.17605/OSF.IO/Y9HIQ