

The Biased Enforcement of Rarely Followed Rules

Jordan Wylie¹, Katlyn Lee Milless², John Sciarappo³, and Ana Gantman^{4,5}

Personality and Social
Psychology Bulletin
1–14

© 2024 by the Society for Personality
and Social Psychology, Inc
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/01461672241252853
journals.sagepub.com/home/pspb



Abstract

We examined whether the enforcement of phantom rules—frequently broken and rarely enforced codified rules—varies by the race of the rule breaker. First, we analyzed whether race affects when 311 calls, a nonemergency service, end in arrest in New York City. Across 10 years, we found that calls from census blocks of neighborhoods consisting of mostly White individuals were 65% less likely to escalate to arrest than those where White people were the numerical minority. Next, we experimentally manipulated transgressor race and found that participants ($N = 393$) who were high in social dominance orientation were more likely to route 311 calls to 911 when the transgressor was Black (vs. White). We also explored the subjective experience of phantom rule enforcement; People of color report they are more likely to be punished for violating phantom rules compared to White people. Overall, we find evidence of racism in the enforcement of phantom rules.

Keywords

legal rules, phantom rules, rule enforcement, racial bias

Received May 16, 2023; revision accepted April 6, 2024

Imagine you are walking in New York City on a sunny day during the lunch hustle. You would likely see several people breaking the law—people jaywalking (New York Vehicle & Traffic Law §1152, 2015), loitering, selling souvenirs on the street—and you might think little of it. These behaviors, and many similar ones, are technically illegal and frequently done without consequence. People jaywalk all the time, but they rarely get in trouble for it. Many of these infractions are not reported at all. If people do want to report on these kinds of infractions, they can use New York City’s 311 service—a municipal, nonemergency service—which is explicitly for issues that do not reach the level of an emergency. That is, for many of these technically illegal, but frequent infractions like loitering or jaywalking, it is not clear that the police need to be involved at all.

However, some people do get penalized for breaking these rules, and they experience very real consequences. And, like many patterns of enforcement in the legal system in the United States, the enforcement of these laws is not randomly distributed in the population. For example, despite making up only 23% (30% if you include those who identify as multiracial) of the population in New York City (U.S. Census Bureau, 2022), nearly 70% (47 of 68 where race is specified) of summons for jaywalking were for Black people in 2022 (NYPD Criminal Court Summons, 2022; see also Kunztman, 2020). Breaking a frequently broken and rarely enforced rule, particularly within the context of the

legal system in the United States, is especially consequential for people of color.

Previous research suggests that jaywalking and loitering, behaviors proscribed by official rules in New York City that are rarely followed and rarely enforced, belong to a subclass of rules called phantom rules (Wylie & Gantman, 2023). Because people frequently break these rules without consequence, when and for whom they should be enforced is often ambiguous, leaving their enforcement particularly susceptible to the influence of personal punishment motivations (Wylie & Gantman, 2023).

The existence of these rules and their nonuniform enforcement stands in contrast to how many people expect rules and the law to be applied. People largely expect the law to have generality: it should apply equally to all (Hannikainen et al., 2021). Yet, in practice, that is not always the case. There is often leniency in how rules are enforced (Schafer &

¹Boston College, Chestnut Hill, MA, USA

²Mathematica, Princeton, NJ, USA

³New York University, New York City, USA

⁴The Graduate Center, The City University of New York, New York City, USA

⁵Brooklyn College, NY, USA

Corresponding Author:

Jordan Wylie, Department of Psychology and Neuroscience, Boston College, Chestnut Hill, MA 02467, USA.

Email: wyliej@bc.edu

Mastrofski, 2005; unless leniency is expected; see Moore & Pierce, 2016). Some people get away with breaking the rules while others do not—and the discretion to use leniency or not can be affected by characteristics of the rule-breaker rather than the rule itself (West, 2018). In addition, the existence of “working” rules, or informal rules of thumb that police officers often develop on the job (Stroshine et al., 2008), means that criteria outside of the law is often used to make decisions about who police officers decide to engage with. These and other informal rules or heuristics lead to bias in who sees their rule-breaking enforced. This is especially true in the U.S. legal system where it is often people of color who are disproportionately subject to law enforcement (Samoylov & Kuntzman, 2020; Southern Poverty Law Center, 2018; Thompson, 2017). For example, people with relatively more stereotypically Afrocentric faces tend to be seen as more criminal (Eberhardt et al., 2004), people with more stereotypically Black features are also more likely to receive death sentences when the victim was White (Eberhardt et al., 2006), and all-White jurors are more likely to punish Black defendants compared to White ones (Anwar et al., 2012).

There are also documented racial disparities in the context of more minor offenses or discretionary rules, like routine traffic stops. For example, in two of the largest cities in the United States, disparities are stark: In New York City in 2021 and 2022, around 60% of all traffic stops were for Black drivers (around 30% were Latinx; New York Civil Liberties Union [NYCLU], n.d.). In Chicago, the proportion of traffic stops that involve a Black driver has consistently risen, reaching 63% in 2021 (Jones & Thompson, 2023). These trends also emerge at the state level—in California, where there has been a marked effort to reduce disparities in policing, Black people are still more likely to experience an intrusive traffic stop than White people (e.g., be handcuffed; Lofstrom et al., 2019). These encounters for minor violations are not only unjust and in contrast to expectations about how the law should be enforced (Hannikainen et al., 2021), but they also provide a justification for the introduction of law enforcement into the situation, and thereby opportunity for further interaction with police (Jones & Thompson, 2023), and for escalation.

One explanation (of many) for these patterns of unequal law enforcement is that stereotypes and other sources of bias like racism influence person perception and lead to shifting standards. When determining whether the rules should apply or whether someone meets some criteria or qualification, stereotypes associated with an individual’s group membership may subtly shift which standards come to mind. And indeed, seminal work on shifting standards suggests that stereotypes associated with gender, race, and more all influence how people evaluate others (e.g., Biernat, 2012; Biernat et al., 2009; Biernat & Manis, 1994; Biernat & Sesko, 2013; Biernat & Vescio, 2002; Hodson et al., 2002). Furthermore, given associations between stereotypically Black features and criminality (e.g., Eberhardt et al., 2004), the very same

rule violation may be construed as more serious or worthy of punishment when done by a Black person compared to a White one.

Another explanation is that the police are initiating contact with people of color more frequently than with White people for supposed minor infractions like traffic violations, and so there are more opportunities for interactions to escalate (Barber, 2023). Indeed, there is evidence that the racial composition of a neighborhood affects police interaction rates and subsequent arrest rates (Smith, 1986). When looking at stop and search patterns in New York City, the racial bias is clear: Around 90% of those searched and arrested were Black and Latinx (Barber, 2023; NYCLU, n.d.). Other research shows that Black and Hispanic drivers face a lower threshold for search initiation compared to White drivers. Notably, the rate of stopping Black drivers decreases at night when visible racial information about drivers is more obscured (Pierson et al., 2020). And these increases in interactions are not merely the product of different base rates of criminal activity. The likelihood that Black and Hispanic drivers are searched is 2.5 to four times higher than White drivers, but they are 30% to 50% less likely to actually have contraband (Seguino et al., 2018), and unarmed Black Americans are 3.49 times more likely to be shot by the police compared to unarmed White Americans (Ross, 2015).

We suspect that phantom rules are no exception to the documented racial bias in rule enforcement. Because they are so frequently broken, it is ambiguous whether any given violation should be punished or not. This creates room for racial bias in the enforcement of these rules. Furthermore, because breaking a phantom rule (e.g., one of the many reasons why people call 311) amounts to literally breaking the law, these rules can provide a veneer of legitimacy for the initiation of police involvement. One can easily point to the phantom rule violation as the cause for a police interaction, even when another individual who commits the very same rule violation, may not be subject to the same levels of enforcement (Wylie & Gantman, 2023). Moreover, phantom rules are enforced according to the individual punishment motivations of the enforcer (Wylie & Gantman, 2023), and those with a motive to maintain and enhance current racial hierarchy in America (i.e., are high in social dominance orientation [SDO]; Ho et al., 2012; Pratto et al., 1994; Sidanius et al., 2017), may also be particularly likely to enforce phantom rules for members of minoritized racial groups compared to White individuals as means to satisfy those motives.

Present Research

We examined how racial identity influences the perceptions and actual enforcement of rarely followed rules across two studies¹ and two supplemental studies (Pilot Study 1 and 2) reported in the Supplemental Materials. First, we examined whether differences in rule enforcement by race are observable in the resolution of 21,400,433 311 municipal calls in

New York City over 10 years. People call 311 for issues including graffiti, loud music, and illegal vending and parking—nonemergency violations which are seen as frequently broken but not grounds for arrest (i.e., phantom rules; validated in a separate study). Previous work has found that discretionary policing, like that used in stop-and-frisk, is affected by the racial composition of the neighborhood in which the stop takes place (Carroll & Gonzalez, 2014). We extend this work to examine evidence of differential law enforcement of phantom rules using 6,200,314 calls (30,073 census blocks) to New York City's 311 call center that are routed to the NYPD. Critically, we can observe in the data that the 311 calls are sometimes diverted to the NYPD sometimes end in arrest, in some neighborhoods more frequently than others.

Next, we sought to experimentally test whether the race of the person violating the rule influences when a 311 call escalates to 911. We recruited a balanced sample of White people ($N \sim 200$) and people of color ($N \sim 200$) and asked participants to put themselves in the shoes of a 311 operator. We manipulated the race of the transgressor and tested whether race influences the decision to route a call for a phantom rule violation (e.g., jaywalking, vending) to 911 (i.e., the police). At the end of the experiment, we also asked participants how likely they themselves would be to be punished by the police for breaking phantom rules and tested for differences by participant race. We hypothesized that participants of color would report being more likely to be punished for breaking the rule.

Study 1

We investigated the nexus between the racial composition of neighborhoods, determined through census block data, and the outcomes of nonemergency municipal service requests. Specifically, we examine calls to 311, a nonemergency municipal phone service for reporting issues that do not rise to the level of an emergency. Some calls to 311 are diverted to the police for resolution. And though the issues themselves seem *prima facie* not to warrant it, the calls sometimes end in arrest. We focused on these 311 calls whose resolution is listed as “arrest” as these calls represent the escalation of the enforcement of rules that are otherwise enforced in more minor ways or not at all. The violations that these 311 calls pertain to include infractions such as public drinking or drug use—rule violations that have been highlighted in previous work on phantom rules (see Wylie & Gantman, 2023). We hypothesized and found that the rule violations that end in arrest in the NYC 311 dataset meet the criteria for phantom rules, rules that are frequently broken and infrequently enforced (Wylie & Gantman, 2023). Specifically, in a separate study reported below, we validated that these rule violations are seen as less worthy of arrest, less illegal, and more common than prototypical legal violations. Importantly, the

Table 1. Full List of 311 Calls That End in Arrest From 2010–2020.

Animal abuse	Homeless encampment	Noise—residential
Bike/roller/skate chronic	Homeless street condition	Noise—street/sidewalk
Blocked driveway	Illegal fireworks	Noise—Vehicle
Derelict vehicle	Illegal parking	Nonemergency police matter
Drinking	Noise—commercial	Panhandling
Drug activity	Noise—house of worship	Urinating in public
Graffiti	Noise—park	Vending
Traffic	Traffic/illegal parking	

Note. In bold are behaviors we include in Study 2.

rule violations from these calls *prima facie* do not rise to the level of a 911 emergency. Notably, however, police are assigned to resolve nonemergency requests at a higher rate than any other government agency, even though other agencies might be better equipped to handle many of these requests (New York City Council, n.d.). To explore the relationship between race and arrests for phantom rules in the real world, we used 311 call data made available by NYC in combination with 2010 demographic data at the census block level.

Validation of 311 Violations as Phantom Rule Violations

We sought to verify that the rule violations that lead to arrest in the 311 dataset meet the criteria for a phantom rule (i.e., are more frequently broken and less frequently enforced than more prototypical rules; Wylie & Gantman, 2023). To do this, we first looked through our dataset to identify the rule codes that corresponded to arrest between 2010 and 2020. Those rules are listed below in Table 1. Next, we investigated judgments of these rule violations compared against serious criminal offenses in NYC (i.e., prototypical rules). The research aimed to examine the frequency, perceived illegality, and appropriateness of arrest for these rule violations.

Design and Procedure. Around 334 participants (Mean age = 19.8, $SD = 3.7$, range: [17, 62]; Gender: 54.8% women, 42.8% men, 2.40% non-binary) were recruited from multiple colleges in New York City (from the same population who might call NYC's 311 or have it called on them) in exchange for course credit. We aimed to recruit as many participants as possible before the end of one semester.

Participants evaluated two sets of rules: the 311 rule violations that have ended in arrest (e.g., blocked driveway, noise complaints, see Table 1) and serious criminal offenses (e.g., felony assault, murder; taken from NYC's Compstat website detailing incidents of crime in the city on a weekly

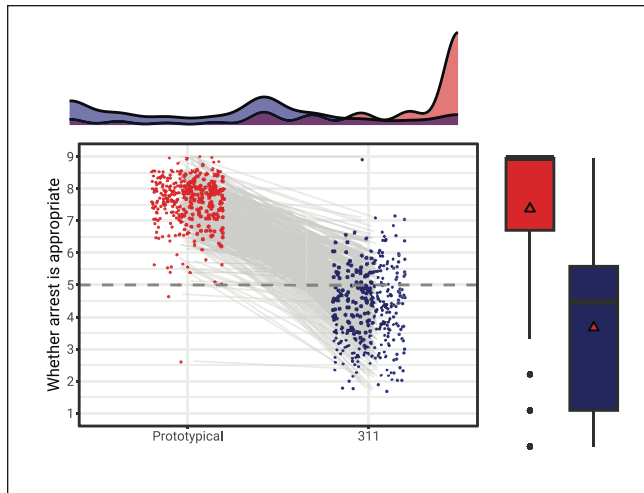


Figure 1. Ratings of How Appropriate Arrest Is for Breaking a Prototypical Rule (e.g., Robbery) Compared to Rules That Ended in Arrest in the 311 Dataset (e.g., Illegal Vending).

Note. Each data point represents the average score for a participant for each of the two categories of rules. Red triangles in the boxplots represent the means for the two groups. Density plots are depicted on top of the plot area.

basis²). Participants made three target ratings: the frequency of the violation (0–100 scale; higher numbers indicating more frequent violations), the illegality of it (1–9 scale; 1 being *not illegal*, 9 being *extremely illegal*), and the appropriateness of arrest (1–9 scale; 1 being *strong disagreement*, 9 being *strong agreement for arrest necessity*). Consenting participants rated all 26 possible rules (which included both 311 and prototypical rules) in completely random order. After completing the rule ratings, participants provided demographic information, were debriefed, and were given course credit for participation.

Results and Discussion. We used linear mixed-effects models that included by-participant and by-stimuli random slopes and intercepts to assess the influence of rule type (311 vs. prototypical rules) on the three target variables. Cohen's *d* effect sizes were calculated using the 'lme.dscore' function in the EMATools package (Kleiman, 2017). First, we compared how common the violations appeared to participants. We found that 311 rules were rated as more frequently broken, $b = 7.54$, $SE = 3.15$, $t = 2.39$, $p = .03$, 95% confidence interval [CI: 1.33, 13.75], $d = 1.08$, than prototypical rules. The 311 rules were also seen as less illegal compared to prototypical rules, $b = -2.01$, $SE = 0.38$, $t = -5.36$, $p < .001$, 95% CI [-2.75, -1.28], $d = 2.23$, and less appropriate for arrest, $b = -3.29$, $SE = 0.71$, $t = -4.60$, $p < .001$, 95% CI [-4.71, -1.87], $d = 2.59$ (see Figure 1).

The findings suggest that—similar to prior research on phantom rule violations (Wylie & Gantman, 2023)—the infractions that lead to arrest in our 311 dataset and differentiable from more serious criminal offenses (i.e., more prototypical rules). In sum, the rule violations that sometimes

have led to arrest in a decade of calls to 311 in New York City meet the *a priori* criteria for phantom rules.

3-1-1 Requests

NYC's 311 Customer Service Center is a telephone and online service that provides nonemergency municipal services and information to users. It is intended to divert non-emergency requests and concerns from 911. All 311 calls going back to 2010 are made available for analysis at the second-by-second level through NYC's open data portal. When a request is made, it is categorized into one of more than 500 request types and gets assigned to a government agency, which then attempts to resolve the request. The 311 dataset contains this information as well as the location of the request and how the request was ultimately resolved.

We explored requests made between January 1, 2010 at 12:00AM and December 31, 2020 at 11:59PM. On average, 1,945,494 requests were made each year (typically increasing year to year) and, in total, 21,400,433 were made. The New York Police Department (NYPD) is assigned the most requests of any government agency—about 29% of all calls. The most frequent categories of requests include complaints about building conditions and construction, illegal parking, blocked driveways, dirty/unsanitary conditions, noise, and the condition(s) of streets, traffic signals, and sidewalks (NYC Council, n.d.). Requests that are assigned to the NYPD are categorized with one of 13 resolution types (full list available in the Supplemental Material), which range from determining that police action was not necessary to making an arrest. The most common violation resolutions (60% of the resolutions) were nonengagement (the absence of a violation or no party present). Arrests made up <1% of the resolutions made. Our analyses are restricted to requests assigned to the NYPD, examining whether racial information from the census predicts whether or not a given call ended in arrest.

Census Blocks and Demographics

Census blocks are the smallest geographical unit used by the Census Bureau. One census block tends to encompass a single city block. In NYC, there are 30,073 census blocks with demographic and 311 data both available. On average, there were 271 residents in each census block ($SE = 1.8$, $Median = 164$).

The following analyses are limited to populated census blocks. In analyses with categorical variables, we group census blocks with greater than or equal to 70% White residents as "majority-White"; census blocks with fewer than or equal to 30% White residents as "minority-White"; and between 30% and 70% White residents as "mixed." We did this *a priori* to create balanced groups, and we did not try out multiple cut-offs before selection of this one to limit multiple testing. Of the 30,073 populated census blocks used in our analysis, 34.95% were majority-White, 37.16% were minority-White, and 27.89% were mixed.

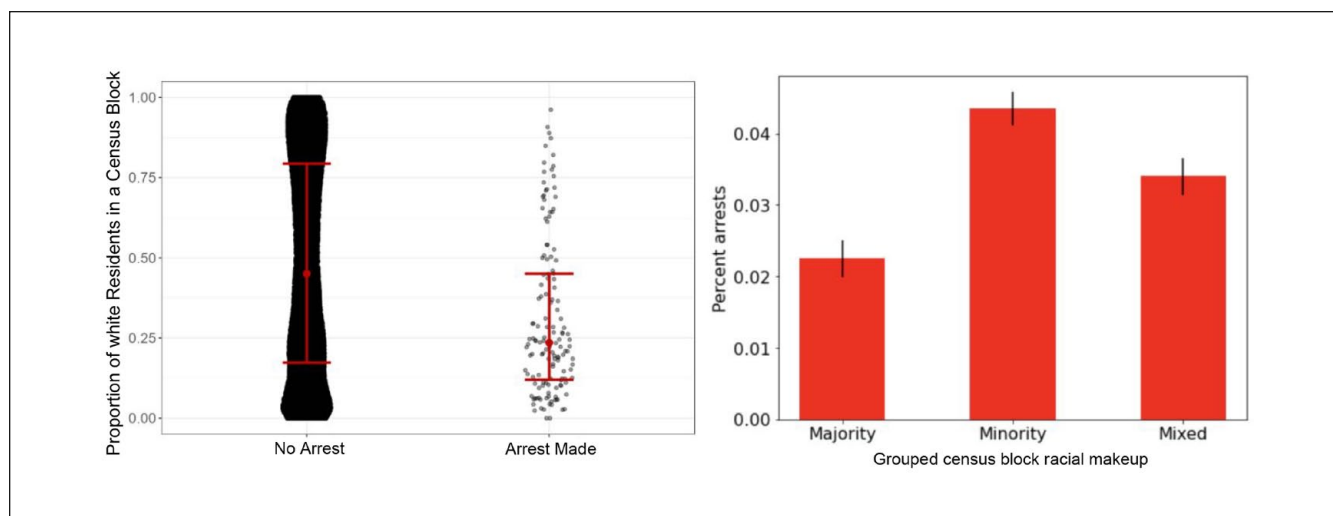


Figure 2. Left Panel: Differences in Calls That End in Arrest as a Function of Census Block Racial Composition. Right panel: Differences in Proportion of Arrests for Majority White, Minority White, and Mixed Racial Blocks.

Results and Discussion

The total number of calls to 311 in the 10-year period is 21,400,433, the total number of calls that were routed to the NYPD is 6,200,314, and the number that ended in arrest is 2,237. The proportion of all calls that get routed to the NYPD is 29% with less than 1% of those ending in arrest. Overall, there was a higher incidence of 311 requests ending in arrests in minority-White census blocks (see Figure 2). To further test this hypothesis, we followed up the initial data exploration with a weighted logistic regression. The total population of each census block, the total volume of 311 requests of each census block (to NYPD and other agencies like the Department of Environmental Protection), the total number of calls that the NYPD received from each census block, and each year (dummy-coded) were included in the model as covariates. Our dependent variable was coded arrest (1), if any arrests occurred in a census block and no arrest (0), if they did not. Importantly, these covariates allow us to estimate the effect of the racial composition of a census block on arrests over and above differences in call volume in those census blocks and are robust to fluctuations by year.

Given the imbalanced nature of the data, we calculated 95% CIs via bootstrapping. That is, we created 1,000 sample datasets using stratified random sampling with replacement from the original dataset. We stratified by year and whether a census block had an arrest or not. This ensures that the sample data distribution across these variables was the same as the original dataset.

Next, we fit each sample dataset with the weighted logistic regression. From the resulting model fits, we determined the lower and upper bounds of the 95% CI for each coefficient by computing the 2.5th and 97.5th percentiles, respectively. We consider any predictor significant if the CI does not contain zero. Results from this analysis supported our hypothesis: the frequency of arrests resulting from 311

Table 2. Predicting Arrests in Census Blocks From 11 Years of Census and 311 Data: Weighted Logistic Regression Coefficients and Confidence Intervals.

Predictor	B	CI ₉₅ Lower	CI ₉₅ Lower
Intercept	-0.7934	-0.8988	-.6874
Proportion White	-1.0646	-1.2452	-0.8934
Total population	0.0004	0.0003	0.0006
Total 311 requests	0.0029	0.0021	0.0036
NYPD 311 requests	0.0205	0.0181	0.0228
2010	-0.0274	-0.1293	0.0717
2011	0.0606	-0.0320	0.1472
2012	-0.1744	-0.0966	-0.0665
2013	0.3708	0.2882	0.4478
2014	0.3929	0.3049	0.4735
2015	0.1942	0.0938	0.2932
2016	0.1519	0.0463	0.2514
2017	-0.0440	-0.1603	0.0566
2018	-0.0296	-0.1281	0.0717
2019	-0.2324	-0.3739	-0.1175
2020	-0.8195	-1.013	-0.6477

Note. Bolded standardized coefficients are statistically significant predictors.

nonemergency requests in a census block was positively associated with the proportion of White residents in that census block³. The log odds ratio of the proportion White was -1.06 with a 95% CI of $[-1.23, -0.89]$. This suggests that between 2010 and 2020, if a 311 request was made in a census block that was 100% White, that request would be 65.5% less likely to end in an arrest than a 311-request made in a census block that was 0% White. Furthermore, if a census block were to experience an increase in the proportion of their White residents from 40% to 60%, we would expect a 19% decrease in the odds of that census block having a 311

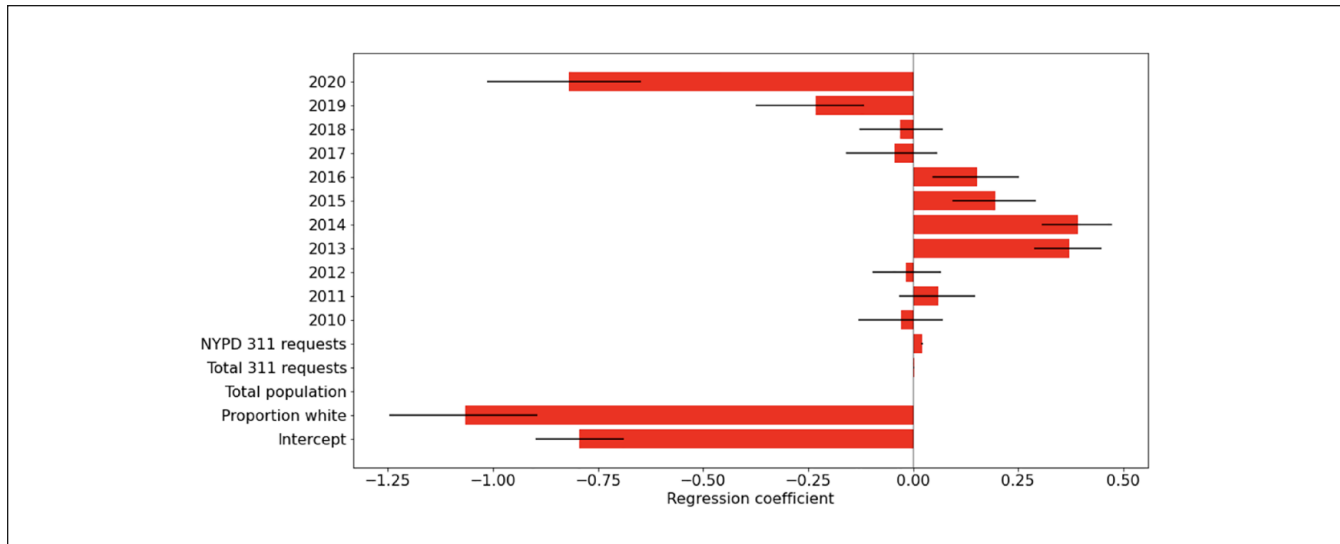


Figure 3. Model Results When Including All Covariates.

Note. Despite variation across years, the proportion of White residents in a census block was the strongest predictor of arrest. Error bars represent 95% confidence intervals. Coefficients are standardized.

nonemergency call end in arrest. Detailed results for covariates are presented in Table 2 and Figure 3. Critically, the proportion of White residents in a census block remains the best predictor of arrest even when including the relevant covariates (see Figure 3).

Overall, we found that rates of arrest for nonemergency 311 calls in New York City overwhelmingly occur within neighborhoods where community members are predominantly people of color compared to when community members are predominantly White. We have validated that these very behaviors that lead to 311 calls escalating to 911 are ones that fit the *a priori* criteria for phantom rules—they are frequently broken and infrequently enforced, especially compared to more prototypical legal violations. Indeed, phantom rules are particularly susceptible to enforcement according to the individual punishment motivations of the enforcer (Wylie & Gantman, 2023). It is likely, given the history of racism in policing in the United States and in New York City in particular (American Civil Liberties Union, 2020) that racial animus is at least partly behind these differences in the punishment of 311 violations. Furthermore, this interpretation is also consistent with previous work on discretionary traffic rules. Prior work has found that Black drivers were 140% more likely to be frisked compared to White drivers (Carroll & Gonzalez, 2014). Furthermore, over a 15-month period, evidence suggests that Hispanic and Black people are subjected to traffic stops at disproportionately higher rates than White people (Gelman et al., 2007). Together, we take this work as suggestive evidence that people of color specifically are more likely to experience arrest following calls to 311, a municipal service for nonemergencies that *prima facie* do not warrant arrest.

An important limitation of this analysis is that there may be overall differences in call volume across census blocks, as

well as differences in the normativity of any of these 311 violations. We adjusted for overall call volume and volume of calls routed to the NYPD in our analyses, and our results hold over and above these differences. That said, an experiment is needed to test the causal role of the race of the rule violator and to keep constant these and other factors that could contribute to the escalation of 311 calls to 911 in some neighborhoods over others.

Study 2

Study 2 was designed to experimentally test whether race influences decision-making about escalation of rarely enforced rule violations. That is, we sought to examine whether the race of a rule violator (White vs. Black) influences how frequently people think 311 calls should be escalated to 911. To do this, we had participants act as 311 operators. We told them that 311 is a municipal service that provides nonemergency information to users and that sometimes the calls get routed to 911, and thereby, the police.

Critically, we also measured right-wing authoritarianism (RWA; Altemeyer, 1981) and SDO (Ho et al., 2012; Pratto et al., 1994; Sidanius et al., 2017)—both of which are correlated with various forms of racism and prejudice (Van Hiel & Mervielde, 2005). Prior work on phantom rules has found that RWA correlated with increases in the desired severity of punishments for phantom rule violations (Wylie & Gantman, 2023). But, as the findings of Study 1 suggest, motivations associated with racism and hierarchy maintenance likely also play a role in who people think should be punished for breaking phantom rules. The biased enforcement of the rule serves as a means to maintain and enhance existing racial hierarchy, subjecting members of minoritized groups to the criminal justice system more frequently than White people. Notably,

police officers tend to be high in SDO (Sidanius et al., 1994), and research suggests that causally manipulating social power increases SDO (Guimond et al., 2003). As such, we expected that these individual difference traits would moderate the relationship between transgressor race and decision to route a call to 911.

We also asked participants whether they themselves would be punished by the police if they broke a rarely followed rule. In a pilot study reported in the Supplemental Materials (Pilot Study 1), we found that men of color report that they are more likely to have rarely enforced rule and social norm violations (which are not even illegal) enforced than women and White men. However, this study did not have enough participants of color for a sufficiently powered comparison. As such, here, we preregistered and powered to detect a main effect of participant race and test whether the race of the participants predicted whether they felt personally vulnerable to phantom rule enforcement by the law. Data collection procedures for all studies were approved by the Brooklyn College Institutional Review Board.

Method

Participants. We recruited 400 participants from Prolific who were located in the United States and had a 99% or above approval rating. We preregistered this sample size based on an *a priori* power analysis conducted using G*Power (Faul et al., 2007), which indicated that a sample of about 199 participants afforded us 80% power to detect an effect as small as $d = .20$. Given the design of the study, and our planned analyses using multilevel models, we also conducted small pilot study to get estimates of our effect of interest. We then conducted an additional power analysis using a summary-statistics-based power analysis for mixed-effects modeling R Shiny App (Murayama et al., 2022). This suggested 208 participants were necessary for 80% power. We opted to recruit 200 participants who identified as White and 200 participants who identified as people of color (using Prolific's simplified racial categories: Black, Asian, Mixed, Other) for a total sample size of 400 participants. We also preregistered excluding people who failed a bot check, attention check, or had trial completion times that exceeded the possible value. This left us with a total sample size of 393 participants ($Mean_{age} = 38.4$, $SD = 13.4$, range: [18, 76], Gender: 61.1% women, 34.7% men, 4.3% nonbinary or other; 50% White, 19% Black, 18% Asian, 1% Hispanic/Latine, 12% Mixed or Other). Participants were compensated \$1.85 for their time.

Design and Procedure. We told all participants to imagine they were a 311 phone call operator and that they would have to decide whether to do nothing about the call or route it up to 911. We told them, "311 is a non-emergency, municipal service that sometimes get routed to the police. 311 is for

making complaints or reporting problems." and participants could not proceed until they correctly answered the question "What is 311 called for?" and chose "non-emergency complaints" with the only other option presented being "emergency complaints." We showed all participants a total of six vignettes that contained a rarely enforced rule violation (e.g., jaywalking, vending) and a social norm violation in completely random order. For each vignette, participants randomly read either a description of a Black transgressor or a White transgressor. There were no repeating vignettes. All participants saw all six of the different rule violations and for each one randomly received either the version where the transgressor is described as White or Black. For each of the six trials, participants read the vignette for a maximum of 14 seconds, and then made a binary decision to route the call to 911 or not (for a maximum of 3 seconds). After that decision, participants then answered questions about the rule violation. These questions were presented in random order.

Materials

Stimuli. Participants were told:

In this study, you will be asked to take on the role of a **311 operator**. Your job is to decide whether the issue reported in the transcript of a call should be **routed to the police or not addressed at all**. This is a critical decision that 311 operators have to make everyday.

Furthermore, we told them that all calls that they will see contain behavior that it is legally possible to route to the police by transferring to 911. Participants read six short transcripts that featured a Black or White transgressor. The vignettes consisted of rarely enforced rule violations (phantom rule violations) from previous work (Wylie & Gantman, 2023) and four of the rule violations from the 311 calls that end in arrest that we analyzed in Study 1. The four rules selected from Study 1 were ones which a majority of participants rated as more frequently broken than prototypical legal rules, and did not think should end in arrest in our validation study. These six rule violations were: jaywalking, loitering, drinking alcohol in public, graffiti, illegal vending, and parking in a no parking zone. The race of the transgressor was always included in the vignette and always followed the description of the rule violation. All calls also contained a description of a social norm violation (like catcalling, aggressive behavior, blocking the sidewalk, offensive language; see Figure 4 for example, see Supplement for full list of stimuli) because previous research found that people do not see these rule violations as worthy of enforcement when broken in isolation (vs. in tandem with a social norm; Wylie & Gantman, 2023). The only thing that differed between the two conditions was the description of the race of the person described in the call. The full list of stimuli is available in the Supplemental Material.

INCOMING CALL: 08:01AM, 04.13.2021 Operator: "311, what's your concern?" Caller: "Hi, I'd like to report something I saw on my commute to work today. I noticed a man selling homemade crafts on the street without a vendor's license. And this person was in the middle of the sidewalk, making it difficult to walk by." Operator: "May I have a description of the person?" Caller: "He's a very light-skinned man around 5'11", I'd say, and pretty fit, about 175 lbs. Has really dark, almost black hair, and blue eyes. He's in a green shirt and jeans and I think his shoes are tan or white."	INCOMING CALL: 08:01AM, 04.13.2021 Operator: "311, what's your concern?" Caller: "Hi, I'd like to report something I saw on my commute to work today. I noticed a man selling homemade crafts on the street without a vendor's license. And this person was in the middle of the sidewalk, making it difficult to walk by." Operator: "May I have a description of the person?" Caller: "He's a very dark-skinned man around 5'11", I'd say, and pretty fit, about 175 lbs. Has really dark hair and lighter eyes. He's in a green shirt and jeans and I think his shoes are tan or white."
---	--

Figure 4. Two Examples of a Call Transcript From a Single Trial, Depicting the Violation of the Rule Against Vending in the NYC 311 Dataset.

Note. On the left is a White target trial, on the right, a Black target trial. Participants read the transcript and then made the decision to either route the call to 911 or to do nothing.

311 Binary Decision. On each trial, participants indicated whether they would like to route the 311 call to the police or do nothing. Participants were asked to make this decision as fast as possible, and were only given 3 seconds to make the decision. They were told that there were no right or wrong answers, and to go with their gut. Participants were also warned that if they did not make their selection in time the page would automatically advance and that a timer would be there to remind them to make their decision within the allotted time.

Rule Judgments. After all six 311 call trials, participants saw each of the illegal behaviors again (e.g., vending without a license) and rated how legitimate, harmful, and intentional each of the rarely enforced rule violations were. Each of these items was rated on a scale from 1 = *Not at all* to 7 = *Extremely*.

Judgments of Personal Rates of Enforcement. Finally, we asked participants, if they were to break the rule that they previously rated (e.g., illegal vending), how likely would they be to be punished for it, rated on a scale from 1 = *Not at all likely* to 7 = *Very likely*. Specifically, we asked: "Imagine that you had been the person to [rule violation], how likely is it that a 311 call about it would be routed to the police?"

Demographics. Participants reported their racial identity along with other demographic dimensions. To best match our 311 analyses, we created a binary categorical race variable that consisted of White and people of color (consisting of racial groups identifying as Black, Asian, Latine, or other races) categories.

Results and Discussion

We preregistered using linear mixed models for each analysis. We specified models with by-participant and by-stimuli random intercepts using the R statistical analysis software (R Core Team, 2013), the "lme4" package (Bates et al., 2015), and the "lmerTest" package (Kuznetsova et al., 2017). We had two primary aims: First, to explore whether racial identity of the transgressor affected how frequently calls were routed to 911. Second, whether racial identity of the participant predicted the likelihood of a participant reporting they would be punished for violating a phantom rule. We also preregistered testing whether the racial identity of the transgressor influenced judgments of legitimacy of the rules, and whether individual differences in RWA or SDO moderated these effects. We predicted that RWA would play a larger role than SDO (as by Wylie & Gantman, 2023), but results did not bear that out. Instead, we found that SDO influenced our target analyses. We report the SDO analyses here in the main text. We report the null findings for RWA in the Supplemental Material.

Main Analyses

311 Routing Decision. We tested whether the race of the transgressor in the 311-call vignette affected how often people decided to route calls to 911. No statistically significant effect emerged ($p = .41$). Next, we examined whether individual differences in SDO (scaled and centered) moderated the relationship between race of the transgressor and 911 call routing including by-participant random intercepts ($ICC = .17$) and by-stimulus random intercepts ($ICC = .23$). Results suggested that SDO did indeed moderate this relationship, b

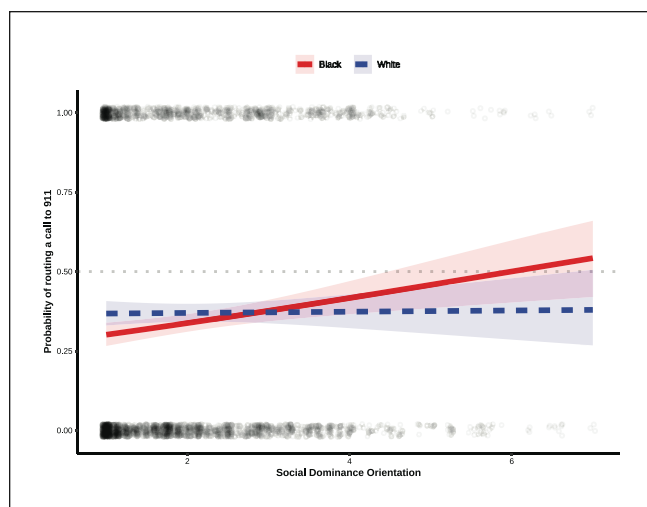


Figure 5. The Probability of Deciding to Route a Call to 911 as a Function of Participant Social Dominance Orientation.

Note. The red line indicates the relationship between routing decision and social dominance orientation for Black targets and the blue line indicates the relationship for White targets. Shading represents SEs.

$= -0.28$, $SE = 0.11$, $t = -2.61$, $p = .009$, 95% CI $[-0.498, -0.066]^4$, $OR = 0.76$, such that people who are high in SDO are less likely to escalate 311 calls to 911 when the transgressor was White compared to when they were Black (see Figure 5). Follow-up simple slopes analyses also suggest that the relationship between SDO and decision is only present for Black transgressor ($b = 0.28$, $SE = 0.09$, $t = 3.23$, $p < .001$). No such relationship emerges when the transgressor was White ($p = .96$). This effect remains robust when we adjust for participant racial identity, age, political orientation, and gender, $b = -0.28$, $SE = 0.11$, $t = -2.56$, $p = .01$. Overall, we found that participants high in SDO were more likely to route calls to 911 that featured Black targets compared to White targets.

Finally, we preregistered testing whether there was an effect of transgressor race on decision on the first trial of the decision task (at which point there is no indication that race is being manipulated within subjects and might create demand). We do not find a significant effect of race ($p = .71$) or an interaction with SDO ($p = .38$).

Legitimacy Judgments. We next tested the role of transgressor race on judgments of the legitimacy of the rule proscribing each of the illegal behaviors described in the 311 transcripts. We used transgressor race as the predictor and included by-participant and stimuli random intercepts. There were no significant effects of transgressor race. As preregistered, we also explored whether SDO moderated the relationship, and it did not. There was a statistically significant effect of SDO on legitimacy, $b = 0.15$, $SE = 0.07$, $t = 2.26$, $p = .02$, 95% CI $[0.020, 0.276]$, $d = 0.39$, such that higher SDO was associated with increasing perceptions of legiti-

macy for these rules; there was no significant interaction with target race.

Judgments of Personal Rates of Enforcement. We also sought to examine whether participants of color are more likely to report that they themselves would see the police involved if they broke the rule than White participants are. We used participant race as the predictor and included by-participant and stimuli random intercepts. There was a statistically significant effect of race, $b = -0.49$, $SE = 0.11$, $t = -4.31$, $p < .001$, 95% CI $[-0.719, -0.269]$, $d = -0.44$, such that White people report that they are less likely to have the call routed to the police than people of color⁵ (see Figure 6). This effect remains robust when adjusting for participant age, gender, and political orientation. We also conducted a follow-up analysis where we included each racial category separately; we observe that Black, Latine/Hispanic, and Asian identifying participants are all more likely to say that the police would be involved if they broke this rule than White participants (see Supplemental Material).

Secondary Analyses. We also preregistered exploring whether the race of the transgressor affected participants ratings of how harmful the rule-breaking behavior is, and how intentionally the rule was broken, and whether SDO moderated those effects. To test this, we specified models that included by-participant and stimuli random intercepts and included race of the transgressor, SDO, and their interaction term. There was no effect of transgressor race on harmfulness judgments ($p = .60$). There was a statistically significant effect of SDO—as SDO increased, so too did judgments of harmfulness of the rule breaking, $b = 0.14$, $SE = 0.06$, $t = 2.49$, $p = .01$, 95% CI $[0.029, 0.246]$, $d = 0.20$. There was no significant interaction ($p = .50$).

There was a statistically significant effect of race of the transgressor on intent judgments. Participants rated White transgressors as more intentional in their rule breaking than Black transgressors, $b = 0.12$, $SE = 0.04$, $t = 2.65$, $p = .01$, 95% CI $[0.025, 0.206]$, $d = 0.11$. There were no other statistically significant effects.

Overall, here we explored whether race affects decision-making about the enforcement of rarely enforced rules and how vulnerable people feel to it. The experimental manipulation of transgressor race influenced decisions for some of our participants: People high in SDO were more likely to route 311 calls to 911 when the transgressor was Black compared to when they were White. Critically, with a well-powered sample, we also found that people of color are more likely to report that the police would be involved if they were to break a rarely followed rule than White people are. These data fit well with previous work which finds that Black and minoritized individuals in the United States feel the police are especially hostile toward minority groups (Gau & Brunson, 2015; Peck, 2015; Pryce & Chenane, 2021) and worry more about the police than White people do (Graham et al., 2020).

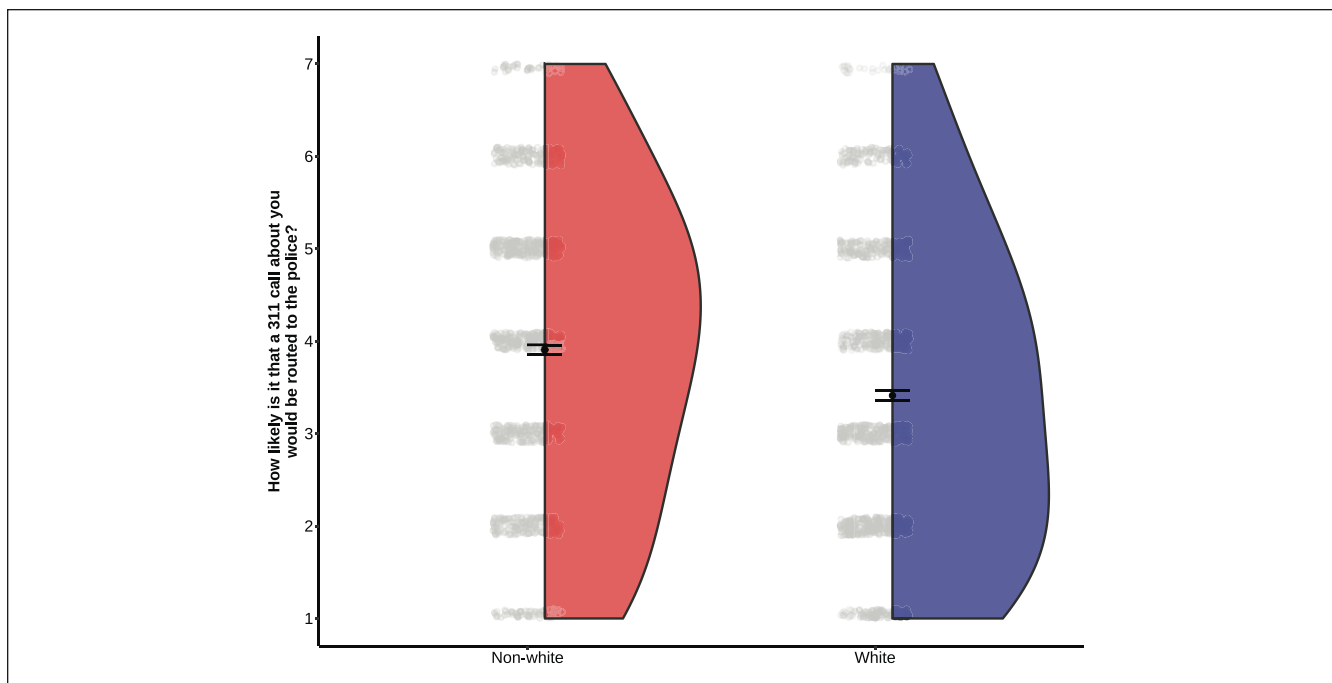


Figure 6. Ratings of How Likely Individual Participants Thought They Would Be to Have the Police Involved If They Were to Break a Rarely Enforced Rule Violation, Broken Down by Race.

Note. The black dot represents the mean, and the error bars represent bootstrapped 95% confidence intervals. Individual data points are represented in gray.

General Discussion

We examined whether there are differences in the rates of escalation of enforcement for phantom rules—rules that are frequently broken and rarely enforced (Wylie & Gantman, 2023)—depending on whether the rule breaker is White or not, using both municipal data from New York City and a tightly controlled experiment. We investigated municipal data from New York City’s 311 nonemergency service, and examined how often 311 calls that are routed to the police end in arrest. We selected calls to 311 for two reasons: *prima facie* the types of complaints that are initially to 311 are not likely to merit arrest, and so we interpret an arrest as an escalation of the situation. Second, some violations that lead to 311 calls likely fit the criteria for phantom rules. Indeed, in a separate sample, we validated that nearly all of the rule violations that ended in arrest in the decade of calls to 311 were rated as both frequently broken and unworthy of arrest (with the exception of animal abuse), which match *a priori* criteria for a phantom rule. Then we analyzed the 311 calls that ended in arrest and merged these data with census block data to determine the racial composition of the neighborhood. We found that in census blocks with a higher number of people of color, 311 calls are more likely to result in arrest. This is particularly notable because 311 is a municipal service explicitly designed to respond to requests for nonemergency issues and minor offenses. The kinds of issues that people call 311 about are *prima facie* not offenses where arrest is proportionate (even within the carceral framework).

Next, we used an online experiment to test whether, keeping all other features of the rule violation constant, the race of the rule breaker would affect whether participants, acting as 311 phone operators, would escalate the call to the police. We found that people who are high in SDO, a trait that is common among police officers (Sidanius et al., 1994), are more likely to escalate 311 calls to 911 for Black but not White rule breakers. In other words, people high in SDO are more likely to let White rule breakers slide. This is in keeping with prior research on the biased enforcement of phantom rules; the appropriateness of enforcement of any given phantom rule violation is ambiguous—the rules are broken without consequence all the time—leaving room for personal motivations to determine whether to enforce the rule (Wylie & Gantman, 2023). People who are high in SDO are motivated to maintain and enhance social hierarchy, which is often expressed in the form of racial hierarchy in the United States. The selective enforcement of phantom rules can provide a means through which actors motivated to keep members of minoritized groups in a position of disadvantage can do so by subjecting them to the full extent of the criminal justice system. And in this case, for behavior that White people can perform with little consequence. It is the testimony of people of color that led us to hypothesize that they experience the enforcement of phantom rules at higher rates—and indeed empirically we see that this is reflected in the reported experience of people of color, and in the rates of arrest after calls to 311 for phantom rule violations.

These data make visible another way in which discrimination in policing occurs in the United States (see also “working rules”; e.g., Carroll & Gonzalez, 2014). This is especially important because research suggests that such links are often underestimated because of bias in reporting in police records (Knox et al., 2020). Alongside flagrant racism in the enforcement of more serious violations of the law and decades of work showing bias in the enforcement of discretionary rules like stop-and-frisk (Carroll & Gonzalez, 2014), these patterns uncover another subtle form of racial bias in legal enforcement. That is, we take these data to be evidence that racial identity plays a role in both who feels more policed (see also Najdowski et al., 2015) and who is actually more policed, which research suggests impacts communities (Garcia & Cao, 2005) and trust in institutions (Pryce & Chenane, 2021; Pryce & Gainey, 2022)—and the escalation of these interactions can cost people their lives. The selective enforcement of rules like loitering, noise complaints, or jaywalking likely builds up frustration, distrust, anger, and fear, and provide justification for police presence. Prior research on phantom rules suggests individuals typically support enforcement of such rules only when they wish to punish a person for a behavior that falls outside the purview of the rules themselves (Wylie & Gantman, 2023). The 311 calls may provide a justifiable avenue for police to enter a situation, and the opportunity for escalation.

We think it is important to identify the selective enforcement of these kinds of rules because we suspect their selective enforcement affects a large number of people (by virtue of them being commonly broken). And the key findings here suggest that although it is not surprising that instances of racism exist within the United States criminal justice system, it is surprising that individuals are being arrested after a call to 311 *at all*. These differences in punishment are especially pernicious because their enforcement is uniquely frustrating (see Wylie & Gantman, 2023). As such, their enforcement may be particularly likely to lead to escalation, especially in neighborhoods where tensions between citizens and police are already high. Phantom rule enforcement reveals one of the mechanisms of over-policing, and evidence suggests that more police do not always decrease crime (Sullivan & O’Keeffe, 2017). In fact, more police may have little effect at all, even on citizen trust in policing (Blair et al., 2021; Nägel & Vera, 2021) and may instead foster the accumulation of frustrating encounters that may potentially escalate to police violence.

These findings are also consistent with biased rule enforcement in other domains—such as those documented in childhood. In the United States, Black students are more than three times more likely to be suspended than their White peers (U.S. Department of Education Office for Civil Rights, 2016). Teachers tend to hold more negative implicit attitudes about and choose harsher punishments for minoritized students compared to majority ones (Glock, 2016; Pit-ten Cate & Glock, 2019). Racial disparities seen within the classroom

can also be explained by the ease with which teachers are able to imagine themselves in a scenario where they suspend a Black student compared to a White student (Okonofua & Eberhardt, 2015). This bias perpetuates a dangerous cycle where minor incidents escalate into major ones (see Okonofua, Paunesku et al., 2016; Okonofua, Walton et al., 2016). And the bias in the enforcement of rules begins early in life; harsher, more frequent punishment is a frequent reality for Black students across the country (for a review on teacher bias, see Turetsky et al., 2021).

Finally, there are likely other intergroup differences that contribute to who and how harshly people decide to punish, including when punishments are doled out by unaffected third parties. For example, while research suggests that third-party punishment is influenced by the group membership of both the victim and the perpetrator (Bernhard et al., 2006), people favor ingroup members with relatively lenient punishment, even when the groups are about sports team loyalty (Yudkin et al., 2016). Even young children favor their ingroup when it comes to third-party punishment (Jordan et al., 2014). General biases in how people reason about punishment in intergroup contexts alongside racist policing patterns create many opportunities for the motivated enforcement of rules, especially when the decision to enforce the rule or not is ambiguous. We have documented one important way that motivated reasoning leads to differential punishment of phantom rule violations. Specifically, we have shown that this phenomenon, originally demonstrated in economic games and with vignettes (Wylie & Gantman, 2023), generalizes to the population of New York City 311 phone operators over a decade and to a balanced sample of White and nonwhite participants imagining themselves making this same decision. Future research should explore other settings where phantom rules may be particularly subject to biased enforcement, like in the workplace and in schools to further explore the generalizability of these effects. Further research could also examine the role of other intergroup processes in motivating when and for whom phantom rules apply; we suspect that race is not the only group identity that people use for selective punishment. Finally future research should investigate the policy interventions that reduce the influence of these selectively enforced rules.

In sum, we find evidence of racism in the enforcement of frequently broken, rarely enforced rules in a 21,400,433 calls to NYC’s 311 over a decade, in a controlled experiment, and in the subjective experience of people of color.

Acknowledgments

The authors thank Daniel Yudkin for his helpful insights in the early stages of this project, Ariel Abramova, our anonymous reviewers, and the members of the Gantman Lab for their valuable feedback. We also thank Michael Kozlov, Ryan Tracy, and Quinn Hartman for their help with data collection, data collating, and editing.

Author Contributions

JW developed the concept, conducted the analyses for Study 1, contributed to the study design, and wrote and edited the manuscript. KLM contributed to editing and providing feedback on the manuscript. JS conducted analyses for Study 2. AG contributed to concept development and refinement, study design, and editing the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: NSF (NSF SMA- 2313957) awarded to the first author, and the Tow Faculty Research and Creativity Grant and SPSSI grants in aid awarded to the last author.

ORCID iDs

Jordan Wylie  <https://orcid.org/0000-0001-5142-8515>

Katlyn Lee Milless  <https://orcid.org/0000-0003-1669-4038>

Supplemental Material

Supplemental material is available online with this article.

Notes

1. All code and materials are available on the project's OSF page (<https://osf.io/9nuxm/>).
2. <https://compstat.nypdonline.org> Accessed on October 10, 2023.
3. This pattern of results remains consistent when we remove the few rule violations that were not phantom rules from the set (i.e., animal abuse).
4. These results remain unchanged when we include by-participant random slopes (which makes fit model slightly worse).
5. When we replace self-reported race of the participant (binarized) with the sample that the data came from (i.e., race information via Prolific), the patterns remain unchanged.

References

- Altemeyer, B. (1981). *Right-wing authoritarianism*. University of Manitoba Press.
- American Civil Liberties Union. (2020). *What 100 years of history tells us about racism in policing*. <https://www.aclu.org/news/criminal-law-reform/what-100-years-of-history-tells-us-about-racism-in-policing>
- Anwar, S., Bayer, P., & Hjalmarsson, R. (2012). The impact of jury race in criminal trials. *The Quarterly Journal of Economics*, 127, 1017–1055. <https://doi.org/10.1093/qje/qjs014>
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Grothendieck, G., Green, P., & Bolker, M. B. (2015). Package 'lme4'. *Convergence*, 12(1), 2.
- Barber, J. (2023, March 24). Black, Latinx people were 90 percent of those arrested in NYPD traffic stops. *New York Civil Liberties Union*. <https://www.nyclu.org/en/news/black-latinx-people-were-90-percent-those-arrested-nypd-traffic-stops>
- Bernhard, H., Fischbacher, U., & Fehr, E. (2006). Parochial altruism in humans. *Nature*, 442(7105), 912–915.
- Biernat, M. (2012). Stereotypes and shifting standards: Forming, communicating, and translating person impressions. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 45, pp. 1–59). Academic Press.
- Biernat, M., Collins, E. C., Katzarska-Miller, I., & Thompson, E. R. (2009). Race-based shifting standards and racial discrimination. *Personality and Social Psychology Bulletin*, 35(1), 16–28.
- Biernat, M., & Manis, M. (1994). Shifting standards and stereotype-based judgments. *Journal of Personality and Social Psychology*, 66(1), 5.
- Biernat, M., & Sesko, A. K. (2013). Communicating about others: Motivations and consequences of race-based impressions. *Journal of Experimental Social Psychology*, 49(1), 138–143.
- Biernat, M., & Vescio, T. K. (2002). She swings, she hits, she's great, she's benched: Implications of gender-based shifting standards for judgment and behavior. *Personality and Social Psychology Bulletin*, 28(1), 66–77.
- Blair, G., Weinstein, J. M., Christia, F., Arias, E., Badran, E., Blair, R. A., Cheema, A., Farooqui, A., Fetzer, T., Grossman, G., Haim, D., Hameed, Z., Hanson, R., Hasanain, A., Kronick, D., Morse, B. S., Muggah, R., Nadeem, F., Tsai, L. L., Nanes, M., . . . Wilke, A. M. (2021). Community policing does not build citizen trust in police or reduce crime in the Global South. *Science*, 374(6571), Article eabd3446.
- Carroll, L., & Gonzalez, M. L. (2014). Out of place: Racial stereotypes and the ecology of frisks and searches following traffic stops. *Journal of Research in Crime and Delinquency*, 51(5), 559–584.
- 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, 383–386. <https://doi.org/10.1111/j.1467-9280.2006.01716.x>
- 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.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191.
- Garcia, V., & Cao, L. (2005). Race and satisfaction with the police in a small city. *Journal of Criminal Justice*, 33(2), 191–199.
- Gau, J. M., & Brunson, R. K. (2015). Procedural injustice, lost legitimacy, and self-help: Young males' adaptations to perceived unfairness in urban policing tactics. *Journal of Contemporary Criminal Justice*, 31(2), 132–150.
- 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.
- Glock, S. (2016). Stop talking out of turn: The influence of students' gender and ethnicity on preservice teachers' intervention strategies for student misbehavior. *Teaching and Teacher Education*, 56, 106–114.
- Graham, A., Haner, M., Sloan, M. M., Cullen, F. T., Kulig, T. C., & Jonson, C. L. (2020). Race and worrying about police brutality: The hidden injuries of minority status in America. *Victims & Offenders*, 15(5), 549–573.

- Guimond, S., Dambrun, M., Michinov, N., & Duarte, S. (2003). Does social dominance generate prejudice? Integrating individual and contextual determinants of intergroup cognitions. *Journal of Personality and Social Psychology*, 84(4), 697–721.
- Hannikainen, I. R., Tobia, K. P., de Almeida, GdFCF, Donelson, R., Dranseika, V., Kneer, M., Strohmaier, N., Bystranowski, P., Dolinina, K., Janik, B., Keo, S., Lauraitytė, E., Liefgreen, A., Próchnicki, M., Rosas, A., & Struchiner, N. (2021). Are there cross-cultural legal principles? Modal reasoning uncovers procedural constraints on law. *Cognitive Science*, 45, Article e13024. <https://doi.org/10.1111/cogs.13024>
- Ho, A. K., Sidanius, J., Pratto, F., Levin, S., Thomsen, L., Kteily, N., & Sheehy-Skeffington, J. (2012). Social dominance orientation: Revisiting the structure and function of a variable predicting social and political attitudes. *Personality and Social Psychology Bulletin*, 38(5), 583–606.
- Hodson, G., Dovidio, J. F., & Gaertner, S. L. (2002). Processes in racial discrimination: Differential weighting of conflicting information. *Personality and Social Psychology Bulletin*, 28(4), 460–471.
- Jones, L., & Thompson, A. (2023). *A new vehicle for “Stop and Frisk.”* Business and Professional People for the Public Interest. <https://www.impactforequity.org/wp-content/uploads/2023/04/A-New-Vehicle-for-Stop-and-Frisk-Report.pdf>
- Jordan, J. J., McAuliffe, K., & Warneken, F. (2014). Development of in-group favoritism in children’s third-party punishment of selfishness. *Proceedings of the National Academy of Sciences*, 111(35), 12710–12715.
- Kleiman, E. M. (2017). EMAtools: Data management tools for real-time monitoring/ecological momentary assessment data. <https://cran.r-project.org/package=EMAtools>
- Knox, D., Lowe, W., & Mummolo, J. (2020). Administrative records mask racially biased policing. *American Political Science Review*, 114(3), 619–637.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, 82(13).
- Kuntzman, G. (2020, May 7). NYPD’s racial bias in “Jaywalking” tickets continues into 2020. *StreetsBlog*. <https://nyc.streetsblog.org/2020/05/07/nypds-racial-bias-in-jaywalking-tickets-continues-into-2020/>
- Lofstrom, M., Hayes, J., Martin, B., Premkumar, D., & Barton, S. (2019, June 12). *Policy Brief: Racial disparities in law enforcement stops*. Public Policy Institute of California. <https://www.ppic.org/publication/policy-brief-racial-disparities-in-law-enforcement-stops/>
- Moore, C., & Pierce, L. (2016). Reactance to transgressors: Why authorities deliver harsher penalties when the social context elicits expectations of leniency. *Frontiers in Psychology*, 7, Article 550.
- Murayama, K., Usami, S., & Sakaki, M. (2022). Summary-statistics-based power analysis: A new and practical method to determine sample size for mixed-effects modeling. *Psychological Methods*, 27, 1014–1038. <https://doi.org/10.1037/met0000330>
- Nägel, C., & Vera, A. (2021). More cops, less trust? Disentangling the relationship between police numbers and trust in the police in the European Union. *Policing: A Journal of Policy and Practice*, 15(2), 939–949.
- Najdowski, C. J., Bottoms, B. L., & Goff, P. A. (2015). Stereotype threat and racial differences in citizens’ experiences of police encounters. *Law and Human Behavior*, 39(5), 463.
- New York City Council. (n.d.). *311 agency service requests by service*. <https://council.nyc.gov/data/311-agency/>
- New York Civil Liberties Union. (n.d.). *Stop-and-frisk data*. <https://www.nyclu.org/en/stop-and-frisk-data>
- New York Vehicle & Traffic Law §1152 (2015). Crossing at other than crosswalks (1960). <https://law.justia.com/codes/new-york/2015/vat/title-7/article-27/1152>
- NYPD Criminal Court Summons. (2022). *Your preschool child’s speech and language development*. <https://data.cityofnewyork.us/w/f7n3-uv8/25te-f2tw?cur=h5IK0bzZcZL&>
- Okonofua, J. A., & Eberhardt, J. L. (2015). Two strikes: Race and the disciplining of young students. *Psychological Science*, 26(5), 617–624.
- Okonofua, J. A., Paunesku, D., & Walton, G. M. (2016). Brief intervention to encourage empathic discipline cuts suspension rates in half among adolescents. *Proceedings of the National Academy of Sciences*, 113(19), 5221–5226.
- Okonofua, J. A., Walton, G. M., & Eberhardt, J. L. (2016). A vicious cycle: A social-psychological account of extreme racial disparities in school discipline. *Perspectives on Psychological Science*, 11(3), 381–398.
- Peck, J. H. (2015). Minority perceptions of the police: A state-of-the-art review. *Policing: An International Journal of Police Strategies & Management*, 38(1), 173–203.
- 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>
- Pit-ten Cate, I. M., & Glock, S. (2019). Teachers’ implicit attitudes toward students from different social groups: A meta-analysis. *Frontiers in Psychology*, 10, Article 2832.
- Pratto, F., Sidanius, J., Stallworth, L. M., & Malle, B. F. (1994). Social dominance orientation: A personality variable predicting social and political attitudes. *Journal of Personality and Social Psychology*, 67(4), 741–763.
- Pryce, D. K., & Chenane, J. L. (2021). Trust and confidence in police officers and the institution of policing: The views of African Americans in the American South. *Crime & Delinquency*, 67(6–7), 808–838.
- Pryce, D. K., & Gainey, R. (2022). Race differences in public satisfaction with and trust in the local police in the context of George Floyd protests: An analysis of residents’ experiences and attitudes. *Criminal Justice Studies*, 35(1), 74–92.
- R Core Team, R. (2013). R: A language and environment for statistical computing.
- Ross, C. T. (2015). A multi-level Bayesian analysis of racial bias in police shootings at the county-level in the United States, 2011–2014. *PLOS ONE*, 10, Article e0141854. <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0141854>
- Samoylov, M., & Kuntzman (2020, January 28). *NYPD targets Blacks and Latinos for “Jaywalking” tickets*. <https://nyc.streetsblog.org/2020/01/08/nypd-targets-blacks-and-latinos-for-jaywalking-tickets/>
- Schafer, J. A., & Mastrofski, S. D. (2005). Police leniency in traffic enforcement encounters: Exploratory findings from observations and interviews. *Journal of Criminal Justice*, 33(3), 225–238.
- Seguino, S., Brooks, N., Morrison, C. P. C. J., & Merkel, V. P. C. G. (2018). *A deeper dive into racial disparities in policing in Vermont*.

- http://mediad.publicbroadcasting.net/p/vpr/files/201803/a_deeper_dive_into_racial_disparities_in_policing_in_vermont_3.26_final.pdf
- Sidanius, J., Cotterill, S., Sheehy-Skeffington, J., Kteily, N., & Carvacho, H. (2017). Social dominance theory: Explorations in the psychology of oppression. In C. G. Sibley & F. K. Barlow (Eds.), *The Cambridge handbook of the psychology of prejudice* (pp. 149–187). Cambridge University Press.
- Sidanius, J., Liu, J. H., Shaw, J. S., & Pratto, F. (1994). Social dominance orientation, hierarchy attenuators and hierarchy enhancers: Social dominance theory and the criminal justice system. *Journal of Applied Social Psychology*, 24(4), 338–366.
- Smith, D. A. (1986). The neighborhood context of police behavior. *Crime and Justice*, 8, 313–341.
- Southern Poverty Law Center. (2018, October 18). *Alabama's war on Marijuana*. <https://www.splcenter.org/20181018/alabamas-war-marijuana?eType=EmailBlastContent&eId=fbfb3108-c152-47d8-8e23-28fac65ab617>
- Stroshine, M., Alpert, G., & Dunham, R. (2008). The influence of working rules on police suspicion and discretionary decision making. *Police Quarterly*, 11(3), 315–337.
- Sullivan, C. M., & O'Keeffe, Z. P. (2017). Evidence that curtailing proactive policing can reduce major crime. *Nature Human Behaviour*, 1(10), 730–737.
- Thompson, B. Y. (2017). “Good moral characters”: How drug felons are impacted understate marijuana legalization laws. *Contemporary Justice Review*, 20(2), 211–226.
- Turetsky, K. M., Sinclair, S., Starck, J. G., & Shelton, J. N. (2021). Beyond students: How teacher psychology shapes educational inequality. *Trends in Cognitive Sciences*, 25(8), 697–709.
- U.S. Census Bureau. (2022). *Population estimates program data*. <https://www.census.gov/quickfacts/newyorkcitynewyork>
- U.S. Department of Education Office for Civil Rights. (2016). *2013-14 civil rights data collection: A first look*. <https://www2.ed.gov/about/offices/list/ocr/docs/2013-14-first-look.pdf>
- Van Hiel, A., & Mervielde, I. (2005). Authoritarianism and social dominance orientation: Relationships with various forms of racism. *Journal of Applied Social Psychology*, 35(11), 2323–2344.
- West, J. (2018). *Racial bias in police investigations*. University of California, Santa Cruz. <https://people.ucsc.edu/jwest1/articles/WestRacialBiasPolice.pdf>
- Wylie, J., & Gantman, A. (2023). Doesn't everybody jaywalk? On codified rules that are seldom followed and selectively punished. *Cognition*, 231, 105323.
- Yudkin, D. A., Rothmund, T., Twardawski, M., Thalla, N., & Van Bavel, J. J. (2016). Reflexive intergroup bias in third-party punishment. *Journal of Experimental Psychology: General*, 145(11), 1448–1459.