Seeing the state in action: Public preferences about and judgments of common police-civilian interactions

Paige E. Vaughn[1] | Gregory A. Huber[2]

[1]Department of Sociology and Criminology, Spring Hill College

[2]Department of Political Science, Yale University

KEYWORDS

policing; public attitudes; punishment; survey experiments

Abstract

New technologies allow unprecedented public visibility of routine police-civilian interactions, but we know little about how the public wants the police to behave during them. We examine public evaluations about preferred punishment and fair treatment using vignette experiments that randomize multiple features of police-civilian interactions. These causal estimates reveal that for the mass public, officer race generally does not affect public attitudes, while participant demeanor, markers of threat, and civilian race do. Police-civilian interactions are evaluated through a lens of reciprocity: hostile officers are judged as less fair, while hostile and armed civilians are viewed as deserving of harsher punishment. When civilians remain polite and threat is low, there is little support for punitive outcomes, but poor civilian behavior warrants more punitive state action. Moreover, people prefer more punishment for White compared to Black civilians, and in interactions with White officers and civilians compared to those in which both

parties are Black. Interactions with a White officer and a Black civilian are judged as less fair, as are the fairness of assigned punishments in them. Finally, views about fairness are not equivalent to views about appropriate sanctions. These results provide critical evidence about public attitudes regarding police punishment and fairness in order maintenance.

Correspondence

Paige E. Vaughn, Department of Sociology and Criminology, Spring Hill College

4000 Dauphin St. 316 Quinlan Hall, Mobile, AL 36608

Email: pvaughn@shc.edu

Acknowledgements

We thank the Center for the Study of American Politics and the Institution for Social and Policy Studies for research support. Thanks to Kyle Peyton, Ashley Rubin, and Justin Nix for helpful feedback.

Protests against police abuse and widespread calls for police reform have taken place across America in the last several years. These protests have been spurred, in part, by technological innovations such as police body cameras and mobile phone cameras that allow the broader community to observe police-civilian interactions that, in the past, would likely have remained private to all but those involved (e.g., Weitzer, 2015). Consequently, people who do not typically have close contact with the state now regularly find themselves in the position of being able to observe and evaluate firsthand its most frontline actors in their interactions with civilians.

While there is clear evidence of opposition to racist and violent policing, little attention has been given to the mass public's views about what constitutes fair and appropriate officer behavior in typical, low-level, and civilian-initiated interactions. Studying the general public's perceptions of these situations, which constitute the bulk of civilian experiences with the police (Lum, Koper, & Wu, 2021) and involve a great deal of officer discretion (Lum & Nagin, 2017), is important because public perceptions likely affect the political feasibility of any reform efforts (e.g., Lum et al., 2021). Additionally, understanding what the public thinks constitutes fair treatment and appropriate punishment in common police-civilian interactions may also be important for ensuring compliance with the law (e.g., Gau & Brunson, 2015; Tyler, 2006; Xie et al., 2006). Finally, our work can help guide those responsible for police training and reform by providing evidence about how the public will perceive and react to officer choices in different situations.

A large body of literature provides a robust theoretical and empirical foundation for understanding how the police (e.g., Rojek, Alpert, & Smith, 2012) and civilians who encounter the police (e.g., Skogan, 2005; Tyler & Wakslak, 2004) evaluate those interactions. But this

important prior work does not address ordinary people's views of the police or their preferences about criminal justice outcomes. Additionally, a focus solely on police behavior misses the fact that police-civilian interactions are neither unilateral nor homogenous. In addition to officer behavior, people's views are likely affected by civilian behavior and the nature of the situation in which an officer intervenes (e.g., Hickman & Simpson, 2003). Moreover, the recent focus on views about the fairness or demeanor of police during their interactions with civilians obscures the fact that what is at stake in criminal justice reform in most cases is the outcome that the state should assign (e.g., Jenness & Calavita, 2018). We know very little about public preferences about appropriate outcomes in common, low-level police-civilian interactions, and how preferences change with features of those interactions. Additionally, we do not know if views about appropriate punishment are distinct from beliefs about fairness—that is, whether features of police and civilian behavior that affect the public's views of procedural or outcome fairness also affect the sanctions that they want meted out.

In this paper we address these limitations by presenting novel survey-experimental evidence about public evaluations of common, civilian-initiated interactions between civilians and the police. These evaluations provide evidence about how the public would evaluate these interactions if they were able to observe them. While our primary focus is on interactions involving Black civilians because they are disproportionately likely to come into contact with the police (e.g., Epp, Maynard-Moody, & Haider-Markel, 2014) and experience high levels of unfair police treatment and punishment (e.g., Lum & Nagin, 2017), we also examine results from a replication experiment in which civilians are either Black or White to determine whether the public reacts differently to situations involving civilians of different races. Proxying the environment facilitated by new technologies, we examine what the mass public believes is

appropriate police behavior, distinguishing preferences over preferred civilian punishment from evaluations of the fairness of those interactions, using vignette experiments in which we randomize different features of those interactions. We consider four features of interactions that may shape beliefs about these outcomes: Demeanor, operationalized as officer and civilian hostility or politeness; Severity, operationalized as the degree of threat present in an interaction; Race, operationalized as whether the officer and civilian are White or Black; and Punishment outcome, which is the sanction, if any, an officer assigns at the end of an interaction.[1]

Measuring third-party evaluations of police-civilian interactions with randomized features allows us to estimate the causal effects of those features while sidestepping standard concerns about omitted variables bias and measurement error that arise when using observational and survey data. For example, in extant research on the views of those who interact with the police, there are key difficulties in distinguishing the effects of differences in actual police behavior from differences in *reported* behavior, the latter of which may be correlated with baseline evaluations of the criminal justice system (e.g., Nix, Pickett, Wolfe, & Campbell, 2017; but see Reisig, Mays, & Telep, 2018). By directly manipulating features of police civilian interactions we can connect objective features of police-civilian interactions to subjective perceptions of them (e.g., Nagin & Telep, 2017). Additionally, our empirical approach diverges from prior work in two other important ways. First, we separately measure perceptions of preferred outcomes, including punishments, and fair treatment. Second, we randomize multiple features of officer and civilian behavior during routine interactions, including the punishment (if

[1] We use the term punishment broadly to include actions such as the decision to issue a criminal citation or arrest because such actions are perceived as punitive, have downstream consequences on later criminal justice stages, and ultimately shape views of the police and civilian outcomes (e.g., Natapoff, 2018).

any) that an officer assigns. Holding outcomes fixed (e.g., citation or arrest) allows us to estimate the direct effects of demeanor and threat on respondent evaluations without being concerned that differences in either operate by changing expectations about assigned sanctions (Dafoe, Zhang, & Caughey, 2018).

We find that when both officers and Black civilians are courteous and Black civilians are unarmed, the public prefers that interactions with the police end in a less punitive way that nonetheless involves the police taking action rather than withdrawing, suggesting little appetite for not having the police intervene when present or being uniformly punitive. Across scenarios, we find little evidence that officer race affects preferences about punishment or fairness evaluations of police-civilian interactions involving Black civilians. By contrast, the public appears to expect police and civilians to demonstrate courteous treatment toward one another, and deviations from such behavior alter preferences about punishment and assessments of fairness. Specifically, when civilians are hostile to a polite officer, the public believes greater punishment is warranted than if both actors are polite, and interactions are viewed as less fair if the police are hostile while a civilian is polite.

Markers of greater civilian threat to order, including being armed and being rude to a polite officer, are associated with preferences for more severe punishment, but absent such threats, the public does not appear uniformly punitive. We also find that respondents prefer more punitive sanctioning for White compared to Black civilians, and in interactions involving White officers and White civilians compared to interactions in which both the officer and civilian are Black. Moreover, interactions involving a White officer and a Black civilian are judged as less fair, as are the fairness of assigned punishments when this race combination is present. Finally, views about fairness are not equivalent to beliefs about appropriate sanctions. Officer hostility

reduces perceptions of the overall fairness of an interaction while it generally does not affect preferred civilian sanctions.

I. PUBLIC EVALUATIONS OF POLICE-CIVILIAN INTERACTIONS

Our core interest is in understanding how the mass public evaluates interactions between the police and civilians in civilian-initiated, low-level criminal justice interactions. As we explain above, we focus on two outcomes: What sanction, if any, the public prefers to be meted out in a particular interaction and how fair individuals view police behavior.

A key motivation for prior work is understanding how the state can secure civilian compliance with the law. This matters not just for those who experience the state in its coercive capacity, but also for the general public, whose decisions about compliance and support for the state are likely consequential more generally. Two key theoretical concepts undergird most existing explanations of support for the criminal justice system: beliefs about outcome efficacy and perceptions of procedural justice and police legitimacy.

Instrumentally, individuals may choose to obey the law and support the criminal justice system if they believe it is effective at producing desirable outcomes. This includes threatening those who (might) criminally offend with punishment (e.g., Nagin, 1998). But in addition to crime control, the public views the police as important for addressing various threats to order that generate fear and uncertainty, often without assigning severe sanctions (e.g., Wilson, 1968). Notably, while recent evidence shows the public does not support the removal of the police from low-level criminal justice interactions (e.g., Vaughn, Peyton, & Huber, 2022), scholars have not yet examined the public's views of what constitutes appropriate outcomes in day-to-day police-civilian interactions involving threats to order. This is an important omission given the police are increasingly likely to become involved in less serious and more ambiguous order maintenance

matters (e.g., Bittner, 1970; Walker, 1997) and the great discretion they have in such interactions (e.g., Goldstein, 1963; LaFave, 1962).[2]

In addition to instrumental concerns, civilians care about being treated fairly. The extensive literature on procedural justice suggests that police legitimacy is increased when the police engage in respectful treatment, practice neutral decision-making, are trustworthy, and allow civilians a voice (e.g., Tyler & Huo, 2002). People who view the police as legitimate, in turn, tend to report being more cooperative with the police and likely to obey the law (e.g., Tyler, 2004). While beliefs about fair treatment are thought to operate above and beyond beliefs about outcome fairness (e.g., Sunshine & Tyler 2003), there is research suggesting that both are important for police legitimacy. For example, victims who report that the police sanctioned offenders in line with their preferences report being more likely to seek police help and cooperate with law enforcement in the future (e.g., Hickman & Simpson, 2003). Overall, then, it seems to matter whether the public thinks the police behave and sanction fairly.[3]

In light of theoretical concerns about efficacy and fairness, in our experiments, we investigate how four salient features of police civilian interactions affect evaluations of the fairness of those interactions and preferences about officer sanctioning of civilians:

1. *Demeanor:* First, we examine the effects of civilian and police demeanor on our

[2] Extant research about punishment preferences typically focuses on preferences for those found guilty of various crimes (e.g., Warr, Meier, & Erickson, 1983) or examines public support for serious punishments such as the death penalty (e.g., Warr & Stafford, 1984).

[3] Features of police behavior that affect legitimacy perceptions may also affect beliefs about police effectiveness (e.g., an officer who appears biased might be viewed as likely to arrest the wrong people, influencing fairness and efficacy beliefs; see e.g., Tankebe, 2009; Dickson et al., 2022). Moreover, it is likely an oversimplification to fully separate concerns about the instrumental benefits of policing from legitimacy, as legitimacy can concern efficacy (e.g., Easton, 1965). Notions of fairness and outcome preferences may also be directly intertwined. To avoid ambiguity, we directly and separately measure fairness perceptions and punishment preferences.

outcomes of interest. Prior work provides ample evidence that the public prefers that the police treat civilians with respect. Disrespect, for example, undergirds many complaints about over-policing of minority communities (e.g., Soss & Weaver, 2017) and may be viewed as a "moral transgression" (Alpert & Dunham, 2004; Van Maanen, 1978). This leads us to expect that the public will view interactions in which the police are impolite or hostile as less fair. If police hostility undercuts beliefs about the police being fair, it may also lead the public to be less supportive of sanctioning.

Prior experimental work typically considers police behavior in isolation or under the assumption that civilians are compliant and polite, however.[4] For example, Maguire, Lowrey, and Johnson (2017) randomly assigned participants to view video clips of simulated traffic stops that involved procedurally just or unjust officer behavior and measured levels of trust in the police and willingness to cooperate with the police. Respondents who viewed procedurally just interactions had higher ratings of each outcome compared to respondents who viewed procedurally unjust encounters. While such studies provide causal evidence about the effects of police behavior when civilians are polite and compliant, they do not provide evidence about the effect of civilian behavior or any other factor (e.g., officer race and assigned punishments) that may affect these perceptions. Additionally, they do not provide evidence about whether preferred civilian sanctions are affected by officer (or civilian) demeanor.

This limited focus is critical because the public likely cares about civilian behavior by itself and may also view certain civilian behavior as warranting greater punishment or excusing police hostility. The extensive literature on reciprocity shows that people expect good behavior

[4] While a large body of observational policing research has examined how police treat and/or punish respectful versus hostile civilians (e.g., Klinger, 1996; Terrill & Mastrofski, 2002), this work does not provide evidence about how the public perceives such interactions.

by one party to warrant good behavior by another, and likewise poor behavior by one party mitigates expectations of good behavior by someone responding to that poor behavior (e.g., Agnew, 2014). Focusing on police-civilian interactions, it is likely that when the police remain polite but civilians are hostile, people will view officers as behaving more fairly than when the officer becomes hostile. They may also view civilian hostility in this circumstance as a generic proxy for threat and be supportive of greater punishment because of a desire for order maintenance and peacekeeping. Prior work shows that police officers interpret civilian demeanor as a measure of threat, with civilian hostility increasing their support for "get-tough" policing (e.g., Pickett & Nix, 2019), but whether this pattern holds for the *mass public* is unknown (but see e.g., Bayley, 1995).

Moreover, extant work provides little guidance about what happens when either the police or civilians initiate hostile interactions and the other party responds with politeness or hostility. In cases where civilians initiate hostility, police hostility may be excused because that hostility is in response to someone who appears noncompliant and potentially threatening (Nix et al., 2017). A civilian with a bad attitude may be perceived by the public as "a symbolic attack on the law itself" (Van Maanen, 1978, p. 316) or as morally culpable (e.g., Lee et al., 2024) and therefore as deserving of sanctions. Alternatively, people may hold the police to higher standards than civilians, believing they should remain polite even when civilians behave poorly (Mastrofski, Jonathan-Zamir, Moyal, & Willis, 2016; Pickett & Ryon, 2017). Similarly, in cases where officers "start it" by being hostile first, people may excuse civilians who become hostile because they are responding to mistreatment. Overall, it is unclear how much reciprocating to poor treatment will excuse the behavior of the person who responds on the grounds that the other person "started it." Finally, as a global matter, it remains unclear if officer demeanor that alters

perceptions of officer fairness also changes preferred sanctions.

2. *Severity*: People may infer threats to order from demeanor, but an incident's level of severity also likely affects public views about appropriate officer behavior, perhaps by changing perceptions of threat. This may be one reason protests about officer malfeasance often arise in cases where low-level criminal offenses are met with disproportionate officer response. While our first experiment measures reactions to an ambiguous threat to order, our second introduces an incident involving a fight with two direct signals of severity: the presence of a victim and a weapon, the latter of which we manipulate experimentally. Research consistently shows that crimes involving victims are viewed by the public as more serious (e.g., Gottfredson & Hindelang, 1979; Sellin & Wolfgang, 1964; Warr, 1989). We expect this to also affect views of common police-civilian interactions. For instance, it may be that hostile officer behavior is judged as fairer and harsher sanctions are preferred in cases involving greater threats (i.e., armed civilians) because officers are viewed as confronting serious danger or harm (e.g., Bayley, 1995; Lee et al., 2024; Levy, Cohen-Louck, & Herzog, 2021). Importantly, severity and demeanor may interact, as past work shows both victims and police feel more threatened by hostile and/or armed civilians (Holmes & Smith, 2012), in which case each may be individually sufficient to warrant more punitive officer action.

3. *Officer Race*: Many, but not all, instances of police malfeasance are cases in which Black civilians are mistreated by White, rather than Black, officers. Given concerns about intergroup relations generically as well as beliefs about racism and racial stereotyping by police, individuals may interpret actions by White rather than Black officers toward Black civilians differently. Few experimental studies, however, have considered the effect of officer race on perceptions of police-civilian interactions.

Most studies that examine the relationship between officer race and perceptions of police-civilian interactions are retrospective and survey people who have recently experienced traffic stops. For instance, Allen and Monk-Turner (2010) found that civilians whose recent traffic stops involved Black officers had significantly lower ratings of legitimacy and respectful officer treatment than those whose interactions involved White officers. Others have found that civilians' judgements about police performance are not associated with officer race, though such work is limited to explaining retrospective victim accounts (e.g., Chandek, 1999). Of course, because officer and civilian race are both correlated with many other factors that may also affect views of the police, whether this evidence is causal remains uncertain. Overall, what remains unclear is whether in interactions involving Black or White civilians, officer race affects the mass public's preferences about outcomes and perceptions of fairness.

4. *Civilian Race*: People of color have consistently been found to be treated less fairly by the police than Whites and punished more harshly than Whites at all levels of criminal justice (e.g., Omori & Johnson, 2019). It is therefore surprising that few experimental studies have randomized civilian race to assess whether public outcome preferences and perceptions of fair treatment in low-level police-civilian interactions are affected by civilian race. Experimental research that does exist in this realm focuses primarily on race in relation to people's views of the fairness of police treatment. For example, Johnson et al. (2017) studied the effects of race on procedural justice and police ratings in simulated traffic stops that varied procedural justice and race conditions. They found a positive effect of procedural justice on police assessments, and this effect was larger when drivers were White rather than Black. In a related study, Kahn et al. (2017) examined the effects of suspect race and mental illness on support for police use of force and found that mental illness lowered support for police use of force in cases with White, but not

Black, suspects. The interaction of Black suspect race and mental illness increased public support for police use of force. Together, such findings lead us to suspect that hostile treatment toward Black civilians might be tolerated more than hostile treatment against Whites.

Most experimental studies that measure public preferences about punishment for Black versus White people focus on people's views of rare outcomes, such as police use of force or federal sentencing.[5] While people don't explicitly tend to judge use of force on Black suspects as more acceptable than use of force used on Whites (e.g., Girgenti-Malone et al., 2017; Yesberg et al., 2022) or report wanting to assign harsher sentences for Black people than White people (Doherty et al., 2022), there is a need for research that examines how the public perceives less serious police-civilian interactions.[6]

5. *Assigned Punishment*: Finally, our study adds experimental manipulations that randomize outcomes of police-civilian interactions, operationalized as what sanction if any the police impose, so that we can hold constant realized outcomes and assess whether they, in addition to the above factors, affect public perceptions of criminal justice interactions. For one, this manipulation allows us to understand whether people evaluate different outcomes as more or less appropriate. Specifically, people who observe more severe sanctions in incidents involving

[5] General public opinion research consistently finds that Black Americans are more likely than Whites to report concern about racial bias in police treatment and outcomes, including use of force (e.g., Brunson & Weitzer, 2009; Gau & Brunson, 2010; Hagan et al., 2005; Johnson et al., 2017; Weitzer & Tuch, 2002), while White Americans tend to more approving of use of force (Mourtgos & Adams, 2020). However, these observational studies do not provide causally informative evidence about whether different groups would evaluate the same interaction differently.

[6] Respondent demographics and other attitudes appear to be related to punishment preferences. For example, White and male respondents have been found to be more accepting of use of force than others (Girgenti-Malone et al., 2017; Yesberg et al., 2022). Respondent beliefs about the sources of racial inequality are also correlated with preferred punishments for White and Black suspects (Doherty et al., 2022).

polite or unarmed civilians may judge that outcome – and even an overall interaction – as less fair than people who are assigned that outcome in incidents involving hostile or armed civilians.

Additionally, by fixing assigned sanctions we obviate concerns that other features of an interaction affect outcomes by changing expectations about officer behavior that are left unspecified (Dafoe, Zhang, & Caughey, 2018). For example, if civilian or officer demeanor or race affect expectations about what sanction an officer will assign, specifying this outcome directly allows us to rule out expectations about punishment as a potential mechanism explaining any effects of those factors. In our vignettes the interaction's outcome is set by random assignment independently of all other features.

II. EXPERIMENTAL DESIGNS AND DATA COLLECTION

We designed and fielded multiple survey experiments to understand how non-participant evaluators assess routine police-civilian interactions initiated by other civilians. All experiments are between-subject designs using vignettes in which we manipulated multiple features of an interaction and asked respondents what sanction, if any, they think the officer should assign to the civilian, as well as to assess the fairness of the interaction. Each experiment is a detailed textual description of a single police-civilian interaction. The key advantage of the vignette approach is that we can manipulate multiple features of hypothetical interactions while keeping "all else equal" through experimental control, providing exogenous variation in treatment assignment that is unavailable in observational research. Additionally, by fully specifying many details of those interactions, we reduce concerns that respondents form expectations about features left unspecified on the basis of other manipulated features (by, for example, making assessments about demeanor on the basis of officer race if only the latter is specified). As we discuss below, these are simplified textual descriptions of complex interactions, and so a key

question is whether effects we estimate would persist when using other mediums to present the same stimuli (e.g., simulated bodycam footage) or with additional stimuli. At the same time, we view this tradeoff as worth the risk, given our ability to provide unbiased estimates of treatment effects that are unavailable in extant work.

In the first experiment the interaction is an order maintenance call. In a replication of this experiment (R1), we include a civilian race randomization. In the second experiment the interaction is more serious, involving a civilian threat to another civilian. We also vary the level of threat in the second experiment by randomizing whether the apparent assailant is armed. Here, we describe each experiment and our data collection process in detail.

I. Experiment 1

Our first experiment is a vignette describing a 35-year-old Black male, "Samuel Parsons," who is in a public park at night after the park is closed. An officer is called to the scene following a civilian complaint about gunshots. After the officer arrives, but before approaching the civilian, he realizes that the civilian is only lighting off fireworks. The vignette and embedded randomizations, along with outcome measurement questions, are shown in Table 1.

Table 1 about here

In this vignette, we independently randomly assigned the officer's race (2 levels: Black or White) and the demeanor (or tone) of the police and civilian interaction (5 levels: Both always polite; Officer always polite, civilian always hostile; Officer becomes hostile after hostile civilian; Officer hostile before civilian is hostile; Officer always hostile, civilian always polite). We also randomized the outcome of the interaction, which varied in the officer's sanctioning decision. Specifically, respondents were told either that the civilian was escorted from the park

or escorted from the park and issued a citation for disorderly conduct.[7]

Respondents were then asked to answer several outcome questions. First, they were asked how fair the officer's chosen action was (outcome fairness; either escorting from the park or escorting and citing). Second, they were asked which of six punishments (outcomes caused by the officer) they preferred. Finally, they were asked how fair the civilian's overall experience with the officer was (i.e., global fairness). Fairness outcomes were measured using 5-point scales ranging from Extremely Unfair to Extremely Fair (scored 1 to 6). Preferred outcomes in order of increasing severity were: (1) officer left after learning the gunshot call was a false alarm, and did not approach civilian; (2) officer left after speaking with civilian; (3) civilian was escorted from the park; (4) civilian was escorted from the park and given a verbal warning; (5) civilian was escorted from the park and issued a citation for disorderly conduct; and (6) civilian was arrested for disorderly conduct.

II. Experiment R1.

To assess whether our findings about civilian and officer demeanor, discussed below, are affected by civilian race, we conducted a replication of the first experiment (R1). As in the first experiment, we independently randomly assigned the officer's race (same as experiment 1); the demeanor (or tone) of the police and civilian interaction (here, 4 levels: Both always polite; Officer always polite, civilian always hostile; Officer hostile before civilian is hostile; Officer

[7] After reading the vignette part that specified officer race and demeanor, respondents were asked how they thought the interaction ended. As noted above, in the absence of a specified outcome, people may make inferences about any punishment that will take place from other features of the interaction. We find this is the case for demeanor, meaning any estimated effect of demeanor when officer sanctioning is not specified on any outcome may operate through changes in expected punishment (see Appendix Table A1 and Section A2, Figure 1A). Our analysis therefore focuses on outcomes measured after the randomly assigned interaction outcome was specified.

always hostile, civilian always polite); and the outcome of the interaction (same as experiment 1). In addition, we randomly assigned the civilian's race (2 levels: Black or White).

All outcomes were measured after the randomly assigned interaction outcome was specified. In addition to being asked which of six punishments they preferred; respondents were asked how fair the officer's chosen action was and how fairly the civilian was treated by the officer. Outcomes were measured in the same way as for the first experiment. The vignette and embedded randomizations, along with relevant outcome questions, are shown in Table 2.

Table 2 about here

III. Experiment 2

Our second experiment is a vignette describing a police-civilian interaction involving an individual who has been fighting with another civilian. This experiment builds on the first scenario in two theoretically motivated ways. First, we examine a different type of interaction, one that includes a victim and randomly varies the severity of the incident. Second, we also measure threat perceptions. This is because threat may affect beliefs about acceptable officer demeanor, preferences about punishment, and understandings of fairness. The vignette and embedded randomizations, along with outcome measurement questions, are shown in Table 3.

Table 3 about here

The vignette describes a police-civilian interaction between an officer and two Black men, "Samuel Parsons," who is reported to be fighting with "Austin Storms" on a baseball field. The responding officer is dispatched to the scene by a 911 call reporting a fight in progress. In the vignette, we independently randomly assigned the officer's race (2 levels: Black or White) and the demeanor of the police and civilian (4 levels: Both always polite; Officer always polite, civilian always hostile; Officer always hostile, civilian always polite; Both always hostile

[officer initiates the hostility]). Additionally, we randomized whether the civilian was armed with a baseball bat or was unarmed as a marker for severity.

After reading this first part of their assigned vignette, respondents were asked how much of a threat the civilian was to the victim and how much of a threat the civilian was to the officer. Threat was measured on a 5-point scale: (1) None, (2) A little, (3) Moderate, (4) A lot, and (5) A great deal.[8]

Respondents were then asked to read the second part of the vignette, which as in the first experiment included an additional randomization of the outcome of the interaction. Specifically, respondents were told that the civilian was escorted from the park and either received a ticket or was arrested. Respondents were then asked how fair they believed this outcome was to the civilian (same response options as experiment 1), which of five outcomes they preferred (in order of increasing severity they were: (1) officer released the civilian from handcuffs and left the baseball field; (2) officer released the civilian from handcuffs and escorted him from the baseball field; (3) officer released the civilian from handcuffs, escorted him from the baseball field, and gave him a verbal warning; (4) officer released the civilian from handcuffs, escorted him from the baseball field, and gave him a ticket; and (5) officer formally arrested the civilian and took him to the police station), and finally, how fair the civilian's overall experience was with the officer.

III. Data collection

Data for experiments 1 and 2 were collected from online national samples of the U.S.

[8] We also measured how punitively they thought the interaction ended. As with experiment 1, we show in Appendix Table A5 and Section A2 (Figure A1) that both demeanor and weapon status affect punishment expectations. Theoretically, we do not believe that expected punishment affects threat prior to that punishment being assigned.

adult population provided by the survey vendor Lucid. Lucid provides researchers with a sampling frame of over 11 million U.S. adults and is now one of the most widely used online survey platforms for academic research. To enter the sample, respondents had to provide informed consent, pass a pre-treatment attention check question demonstrating they could recall a salient detail from an unrelated short news article they were asked to read (addressing an important concern about online sample attentiveness), and answer at least one question about an outcome of interest. 2,017 respondents were recruited for experiment 1 in April 2021 and 2,147 respondents were recruited for experiment 2 in June 2021. Section A5 in the Appendix provides summaries of the demographic composition of the samples, although we note that we focus on experimental estimates, which prior research shows are largely similar in samples obtained from Lucid and other online platforms used for academic research (Peyton, Huber, & Coppock, 2021).

Data for the replication experiment, which was fielded in May 2024, come from a different sample vendor, Bovitz, which provides a Census-benchmarked sample (Druckman & Levendusky, 2019). Prior research has compared samples provided by Bovitz to the demographics of the American Community Survey and partisanship data from the American National Election Survey and found that they track closely to these surveys (Druckman & Levendusky, 2019; Gerber, Huber, Tucker, & Cho, 2023). This more representative sample of 1,546 respondents provides assurances about the effect of sample composition on estimated effects and also allows us to assess whether White and Black respondents respond similarly to the factors manipulated in the experiment. Section A5 in the Appendix provides summaries of the demographic composition of this sample.

III. RESULTS

We examine how demeanor, officer race, severity (experiment 2), and punishment

outcome affect preferred punishments, the fairness of assigned punishments, and the fairness of police-civilian interactions. Additionally, for experiment R1, we examine the effect of civilian race and perform subgroup analyses based on respondent race; and for experiment 2, we measure evaluations of threat. Our analysis strategy is to regress each outcome on a vector of treatment assignment variables and report average marginal effect estimates of each manipulation relative to an excluded baseline category. (As we discuss below, we also test for relevant interactions between treatment variables.) As is standard in analyses of survey experiments, we estimate all models using OLS regression with robust (Huber/White) standard errors because coefficients for randomly assigned treatments are equivalent to differences in means for scale score outcomes, allowing for easier interpretation. In the Appendix (Table A9), we demonstrate that our results about preferred punishment are robust to instead running ordered probit models, which have the advantage of allowing the cutpoints between levels of the outcome variable to vary in a flexible fashion, but they are more difficult to directly interpret than OLS estimates. All models displayed in the main text are unadjusted for covariates, with covariate adjusted results in the complete regression tables in the Appendix (Tables A2, A4, and A6). Controlling for covariates has little effect on average estimated treatment effects.

I. Demeanor and threat shape preferred sanctions

Figure 1 plots stated preferred punishment outcomes for each interaction. Panel A shows results for the first experiment, Panel B shows results for the replication experiment, and Panel C shows results for the second experiment. Underlying regression results for all figures, as well as all supplementary analysis, appear in Appendix A1 (Tables A1-A6). In the figure, the horizontal axes measure punishment preferences compared to the baseline condition in which the officer is Black, the civilian is Black (manipulated only in experiment R1), both the civilian and officer

remain polite throughout the interaction, the civilian is given a lesser punishment (experiments 1 and R1, warned and escorted; experiment 2, ticketed), and the civilian is unarmed (experiment 2, does not have a baseball bat). In all panels, points to the left of 0 (indicated by the vertical line) indicate interactions in which respondents preferred less punitive treatment than those assigned the baseline condition.

Figure 1 about here

In experiment 1 (Panel A), the average level of preferred punishment in the baseline condition is 3.2 on the 6-point scale, which corresponds to an average preferred punishment of escorting from the park. In experiment R1 (Panel B), it is 2.6 on the same scale, also corresponding to an average preferred punishment of escorting from the park. In experiment 2 (Panel C), the mean level of preferred punishment in the baseline condition is 3.0 on the 5-point scale, which corresponds to an average punishment preference of releasing from handcuffs, escorting from the baseball field, and giving a verbal warning. This demonstrates that, on average, respondents in these conditions prefer largely non-punitive outcomes that nonetheless had the officers involved in remediating a victimless threat to order (by disturbing the peace at night) or protecting another civilian (following a fist fight).

Our analysis allows us to estimate how punishment preferences change with demeanor, officer race, civilian race (experiment R1), and severity (experiment 2), as well as how preferences differ by respondent race (experiment R1). First, consider the effect of demeanor. Across experiments, compared to the baseline in which both are polite, respondents prefer the most severe punishments when the civilian was hostile but the officer was not ($b=0.71$, $se=0.08$, $p<0.01$ experiment 1; $b= 0.52$, $se=0.07$, $p<0.01$ experiment R1; $b=0.31$, $se=0.07$, $p<0.01$ experiment 2). In the first experiment, this is followed by interactions in which a civilian was

hostile first (b=0.57, se=0.08, p<0.01; this scenario was not included in experiments R1 or 2 for power reasons). The next most severe punishments preferred by respondents are for scenarios involving civilians who became hostile after the officer, although in the second experiment the coefficient is non-significant (b=0.24, se=0.08, p<0.01 experiment 1; b=0.41, se=0.07, p<0.01 experiment R1; b=0.07, se=0.07, p>0.05 experiment 2). In all experiments, if the civilian was polite and the officer was hostile, the preferred punishment is no different than if both were polite.

Overall, these results reveal that on average the mass public prefers more punitive officer behavior when civilians are hostile, indicating that expectations about appropriate behavior apply to civilian behavior even when the remainder of an interaction is specified (i.e., how it ends and what actual sanction was assigned). More directly, when civilians are hostile, the public wants them punished more. This implies that the public has expectations about civilian behavior either because respondents directly prefer respectful treatment of (polite) law enforcement actors or because they view disrespect as signaling potential non-compliance or threats to order. In experiments 1 and R1, this preference persists at a diminished magnitude even when it is an officer who started the hostility, but in experiment 2 the estimate is much smaller and not significant. Importantly, an officer being hostile in response to a hostile civilian (manipulation in experiment 1 only) at most diminished preferences for more severe punishment by a slight amount (difference=0.14, se=0.08, p>0.05). Finally, we note that there is no evidence that the public wants lesser sanctions merely because an officer is hostile while a civilian is polite, meaning that any expectations about fair officer behavior (discussed below) do not translate into a preference for lesser punishment.

Second, we examine the effect of incident severity, which we manipulated in experiment

2. Respondents on average prefer more severe punishment for incidents involving civilians armed with a baseball bat compared to those involving unarmed civilians (b=0.33, se=0.05, p<0.01). This effect exists even though the vignette specifies that the civilian cooperated with the police (by putting down the weapon and stepping away from the victim) and there was no harm to the other civilian. In terms of magnitude, the greater sanction preferred for an armed suspect is about the same as that for a hostile civilian compared to an interaction in which both participants are polite, meaning civilian demeanor has a large effect.

Third, we examine officer race effects. We do not find that officer race significantly affects sanction preferences in experiments 1 or R1. In experiment 2, however, race of officer somewhat matters. Specifically, those assigned vignettes with a Black officer on average preferred harsher sanctions than respondents who read interactions involving a White officer (b=-0.15, se=0.05, p<0.01). This is an interesting pattern given that the vignettes specified other features of the interactions. While there was no victim in the first scenario, in experiment 2 the officer is interceding in an interaction involving two Black individuals, which may explain why respondents are less comfortable with a White officer assigning a more punitive sanction.

Experiment R1 reveals that civilian race has a significant effect on punishment preferences. Holding all else constant, respondents prefer more punitive sanctions for White compared to Black civilians (b=.15, se=0.05, p<0.01). To test whether this effect is explained by differences in officer and civilian race, we estimated a model in which we included all potential pairings between officer and civilian race (Officer and Civilian White; Officer White, Civilian Black; Officer and Civilian Black [omitted baseline]; Officer Black, Civilian White, see Appendix Table A3, Column 2 for regression estimates). This analysis reveals that the only significant difference between pairings in which both the officer and civilian are Black occurs

when both are instead White. In those circumstances, respondents prefer *harsher* sanctions on average (b=0.18, se=0.07, p<0.05; when the civilian is Black and the officer is White, the estimated preferred punishment effect is b=-0.09, se=0.07, p>.05). We also repeated this analysis by the demeanor conditions to see if these effects differed by officer and civilian behavior (See Appendix Table A3, Columns 3-6 for separate regression estimates). There are no statistically significant effects of officer and civilian race pairings when the civilian remains polite. However, when the civilian initiates hostility, preferred punishments are significantly higher when both the officer and civilian are White (b=0.38, se=0.13, p<0.01), and there is a similarly signed but insignificant effect when the officer initiates a jointly hostility interaction (b=0.23, se=0.15, p=0.13). Overall, this means that respondents on average do not favor harsher punishments for interactions between Black civilians and White officers; instead, the pattern is the opposite, and only in the case of civilian hostility. Moreover, when the officer is Black, there is no evidence that civilian race moderates the effect of demeanor.

Experiment R1 also allows us to examine differences in reactions between Black and White respondents. We separately analyzed two subgroups: (1) self-identified Black respondents and (2) self-identified White respondents who did not also identify as Black, Asian, or another non-White group (hereafter White) (See Appendix Table A4, Columns 2 and 3). There are almost no differences in preferred punishments at baseline in the Officer and Civilian Polite scenario (2.5 for Black respondents versus 2.6 for White respondents), and the overall pattern of coefficients was broadly similar except that Black respondents react more negatively to poor civilian demeanor, either when the officer remains polite (b=0.60, se=0.18, p<0.01 for Black respondents versus b=0.55, se=0.08, p<0.01 for White respondents; difference not significant) or the officer is hostile (b=0.75, se=0.17, p<0.01 for Black respondents versus b=0.35, se=0.09,

p<0.01 for White respondents; difference significant at p<0.05). As with the overall sample, both subgroups are more punitive when the civilian is White, though this effect is only significant for White respondents (b=0.16, se=0.06, p<0.01 for White respondents, compared to b=0.13, se=0.13, p>0.05 for Black respondents).

Finally, in each experiment, there is clear evidence that merely reading about an interaction in which a more severe punishment was given increases preferences for more punitive outcomes (b=0.37, se=0.05, p<0.01 for experiment 1; b=0.41, se=0.05, p<0.01 for experiment R1; b=0.21, se=0.05, p<0.01 for experiment 2). This likely indicates that some respondents have weak preferences about outcomes or defer to officer judgments.[9]

II. Experiment 2: Threat perceptions may explain preferred sanctions

We hypothesized above that beliefs about threat may explain the greater support for harsher sanctions in cases involving armed or hostile civilians. Our data from experiment 2 allow us to assess whether these factors shape perceptions of threat, either to the officer or victim involved in the interaction. As before, we plot our results graphically in Figure 2 with underlying regressions results in Appendix Tables A5 and A6. In the figure, the horizontal axes measure perceived threat by the civilian to the victim (the person lying on the ground, Panel A) and by the civilian to the officer (Panel B) compared to the baseline condition in which the officer is Black, both the civilian and officer remain polite throughout the interaction, and the civilian is unarmed. In both panels in the figure, points to the left of 0 (indicated by the vertical line) indicate interactions rated less threatening than the baseline condition.

[9] Appendix Section A3 includes a detailed discussion of the effects of interactions between officer race and demeanor (and officer race and weapon status for experiment 2), demeanor and civilian race (for experiment R1), and demeanor and weapon status on preferred punishments. Including these interactions did not improve model fit.

Figure 2 about here

We first examine respondents' perceptions of the civilian's threat to the victim. The average level of perceived threat in the officer Black, both polite, and unarmed civilian condition is 3.2 on the 5-point scale, corresponding to an average threat in this condition of "moderate." The average level of perceived threat to the officer (who has not been attacked and is likely armed) for the same baseline is 2.3 on the 5-point scale, which means that officers are perceived as less endangered and corresponds most closely to an average threat level of "A little."

There is no statistically significant effect of officer race on either threat measure. By contrast to these small effects, there are clear effects of demeanor and weapon status on threat perceptions. Compared to a baseline in which the civilian and officer remain polite, hostile civilians interacting with polite officers are perceived as significantly more threatening to victims ($b=0.16$, $se=0.07$, $p<0.05$) and officers ($b=0.26$, $se=0.07$, $p<0.01$). This may explain the preference for greater punishment in this condition. When both individuals are hostile (and the officer started the hostile interaction), the perceived threat to both the victim and the officer is no greater than if both actors are polite, which is also consistent with the finding from above that preferred punishment in this condition is no different than in the baseline. Finally, when a hostile officer is met by a civilian who remains polite, indicating deference, threat perceptions are diminished relative to the baseline for both the victim ($b=-0.16$, $se=0.06$, $p<0.05$) and officer ($b=-0.14$, $se=0.07$, $p>0.05$), though the coefficient is only insignificant for the latter threat outcome.

Turning to weapon status, armed civilians are perceived as considerably more threatening than unarmed civilians to both the victim ($b=0.58$, $se=0.05$, $p<0.01$) and officers ($b=0.29$, $se=0.05$, $p<0.01$). Once again, this effect is consonant with the greater preference for punishment

in this condition.[10]

Most generally, this analysis highlights the possibility that threat is a mechanism linking both civilian demeanor and weapon status to preferences about punishment. In contrast, we find little evidence that officer race or behavior matter in shaping perceived threat, consistent with the finding that fixing initial civilian behavior (initially polite or hostile) officer behavior has modest effects on preferred sanctions.

III. Interaction characteristics shape fairness of assigned punishments

Up to now, our analysis has focused on what punishment (if any) respondents would like to see assigned and how threating they perceive civilians. The former is a measure of direct preferences about police behavior and criminal justice outcomes, while the latter is a potential mechanism. But what about assessments of fairness, which are the primary outcomes in the extant literature? Figure 3 shows that when we measure beliefs about the fairness of assigned sanctions, we observe somewhat different patterns from when we examine assessments of preferred punishments. Panel A plots the results from experiment 1, Panel B shows results from experiment R1, and Panel C plots results from experiment 2. In all panels, points to the left of 0 (indicated by the vertical line) indicate conditions under which sanctions are perceived as less fair than in the same baseline conditions used above. The most importance difference in terms of effects on preferred sanctions (see Figure 1) versus fairness is that officer hostility does not reduce preferred sentences, but it does reduce assessments of fairness, showing that assessments of fairness are not the same as beliefs about appropriate punishment.

Figure 3 about here

[10] Appendix Section A4 includes a discussion of the effects of interactions between demeanor and weapon status (experiment 2) on perceived threat.

In experiments 1 and R1, average levels of fairness in the officer (and civilian, experiment R1) Black, both polite, and verbally warned and escorted condition is 4.14 in experiment 1 and 3.60 in experiment R1 on the 5-point scale, corresponding to an average fairness rating above the neutral midpoint of "somewhat fair." In experiment 2, the average level of fairness reported in the officer Black, both polite condition, ticketed, and unarmed civilian condition is 3.15 on the 5-point scale, which corresponds most closely to a neutral average fairness rating of "neither fair nor unfair."

In no case does the officer's or civilian's race (experiment R1) affect perceptions of the fairness of the punishment. However, demeanor, severity, and punishment outcome do. In all three experiments, the outcome is perceived as significantly fairer for interactions involving hostile civilians and polite officers (Panel A: $b=0.46$, $se=0.08$, $p<0.01$; Panel B: $b=0.27$, $se=0.07$, $p<0.01$; Panel C: $b=0.24$, $se=0.08$, $p<0.01$), and Panel A also shows that civilian-initiated joint hostility yields higher fairness ratings ($b=0.31$, $se=0.08$, $p<0.01$). In contrast, outcomes for interactions involving hostile officers and polite civilians are viewed as less fair than those involving polite civilians and polite officers in both experiments 1 ($b=-0.42$, $se=0.08$, $p<0.01$) and R1 ($b=-0.36$, $se=0.08$, $p<0.01$), but in experiment 2, this effect is much smaller and not significant ($b=-0.12$, $se=0.08$, $p>0.05$). Finally, joint hostility (initiated by the officer) has insignificant and small effects in all three experiments.

As with the preferred punishment outcome, it appears the public expects civilians to behave politely and officers to respond in turn. An officer's punishment is perceived as fairer when a civilian is either hostile alone or when they initiate hostility in an interaction (experiment 1 only). Officers are also expected to be polite, as the interaction rated least fair is when a hostile officer faces a civilian who remains polite in experiment 1, although in experiment 2 (where the

incident is at baseline more severe), this reduction in fairness is not significant.

In experiments 1 and R1, we find that a civilian becoming hostile eliminates the fairness penalty associated with an officer starting a hostile interaction. By contrast, in experiment 2, while the point estimates are negative for both interactions in which the officer is hostile, neither is statistically significant ($p>0.05$). Overall, these results provide clear support for giving attention to civilian demeanor in addition to officer demeanor in understanding public perceptions of fair officer behavior, and they also emphasize the importance of reciprocity. If anything, civilians are held to a higher standard than officers: sanctions are judged as fairer if a civilian starts a jointly hostile interaction than if an officer does.

Panel C also shows that incident severity also affects the perceived fairness of the outcome. On average, the public views the punishments associated with incidents involving armed civilians as fairer than the punishments associated with incidents involving unarmed civilians ($b=0.31$, $se=0.06$, $p<0.01$). This is the single largest magnitude effect for experiment 2, showing that a proxy for threat has a large effect on how fair any punishment is perceived to be.

Finally, the assigned punishment outcome has a large effect on perceived fairness of the punishment in experiments 1 and R1, but not experiment 2. In the first experiment, vignettes that were assigned a ticket outcome (as opposed to a warning) were rated as less fair ($b=-0.58$; $se=0.05$, $p<0.01$) and in experiment R1 this effect is highly similar ($b=-0.62$, $se=0.05$, $p<.01$). However, the effect of assigned punishment was near zero in experiment 2 ($b=-0.01$, $se=0.06$, $p>0.10$). It therefore appears that respondents employ different standards for the more serious incident involving a victim in experiment 2.

Using the data from experiment R1, we also examined whether there were salient differences between Black and White respondents in their evaluations of the fairness of assigned

punishments (See Appendix Table A4, Columns 5 and 6). We did not find substantial differences between these demographic subgroups. Additionally, we also tested whether there were differences based on the pairing of officer and civilian race (Appendix Table A3, Column 8). As with our measure of preferred punishments, we find that in interactions involving a White officer and a Black civilian, punishments are on average judged as less fair (b=-0.21, se=0.08, $p<0.01$)—all other interactions are evaluated similarly. There is therefore no evidence that respondents are more accepting of punishment of Black civilians by a White officer.

The above analysis assumes that the effects of demeanor and weapon status (experiment 2) are the same regardless of which sanction was assigned, an assumption that may be unwarranted because, as we document above, both factors alter punishment preferences. For this reason, we also estimated models in which we allowed the effects of demeanor and weapon status (experiment 2) to vary with assigned punishment.[11] We did so by estimating models in which we interacted the demeanor (and weapons) manipulations with assigned punishments. In all cases, we can reject the null that these interactions are jointly insignificant (experiment 1 Likelihood ratio $chi^2=37.76$, $p<0.01$; experiment R1 Likelihood ratio $chi^2=18.31$, $p<0.01$; experiment 2 Likelihood ratio $chi^2=9.48$, $p=0.05$). Demeanor and weapon status (experiment 2) therefore affect perceptions of the fairness of the assigned punishment. Table 4 makes the magnitudes of these effects clear by showing the (negative) fairness effect of the officer choosing the more severe sanction in each experiment in three conditions: At baseline (when both actors are polite [and unarmed, experiment 2]), when the civilian is hostile (but the officer is polite),

[11] We also tested whether there were differences in the effects of demeanor, weapon status, and punishment by officer race. Models that allowed for these interactions did not substantially improve model fit (experiment 1 Likelihood ratio $chi^2=3.36$, $p>0.05$; experiment R1 Likelihood ratio $chi^2=5.89$, $p>0.05$; experiment 2 Likelihood ratio $chi^2=7.59$, $p>0.05$).

and when the civilian is armed (experiment 2 only).

Table 4 about here

The results show that the negative fairness penalty is substantial at baseline in both experiment 1 and R1 (-0.92 and -0.74, respectively), but that when the civilian is hostile, this fairness penalty declines substantially, to a non-significant effect of -0.07 (se=0.11, p>0.05) in experiment 1 and -0.27 (se=0.09, p=0.01) in experiment R1. In experiment R2, the baseline fairness penalty associated with the more severe sanction is much smaller (b=-0.16, se=0.13, p>0.05), but it becomes positive when the civilian is hostile (b=0.04, se=0.13, p>.05), although both effects are non-significant. Finally, it is perceived as much fairer to punish more severely when the civilian is armed (b=0.33, se=0.13, p<0.01). In short, assessments of the fairness of punishment are clearly dependent on the civilian's demeanor and whether they are armed. So not only are assessments of fairness distinct from beliefs about appropriate punishment, it is sometimes fairer to punish more harshly on the basis of a civilian's behavior and the threat they pose.

IV. A brief aside on overall fairness

The fairness outcome we analyzed in the previous subsection asked respondents to evaluate how fair the specific punishment the officer chose was. But we also asked respondents a more general question about how fairly the civilian was treated by the officer in their assigned interaction. We repeat our earlier analysis strategy for this outcome in Figure 4, once again with results for each study reported in a separate panel. In all panels, points to the left of 0 (indicated by the vertical line) indicate interactions perceived as less fair than the same baseline conditions used earlier.

Figure 4 about here

Comparing Figure 3 (perceived fairness of punishment) and Figure 4 (overall fairness of interaction), we note that the results are not identical. The most striking difference is that when both civilians and officers are hostile, punishments were on average viewed as either as fair or fairer than in baseline (lines for both hostile are always indistinguishable or to the right of the 0 line in Figure 3), but in the same circumstances, the interaction as a whole is always viewed as less fair to the civilian (lines for both hostile are always to the left of the 0 line in Figure 4).[12] Thus, despite the fact that people believe that poor civilian behavior makes punishment fairer, they do not believe that an interaction as a whole is as fair when an officer is hostile. This also means that evaluations of fairness need to distinguish between the fairness of an assigned sanction and the overall fairness of an interaction, as the two are not the same. Even in experiment 2, when an officer is hostile, the interaction is perceived as less fair but the punishment is perceived as just as fair. More bluntly, views about the fairness of officer behavior are not substitutes for measures of preferences about the outcomes assigned by officers; people view interactions with hostile officers as less fair, but that does not mean they do not want civilians sanctioned. It is therefore a mistake to conclude that officer malfeasance or poor behavior that causes respondents to evaluate interactions as less fair means that the public does not want the police to sanction.[13]

As with evaluations of the fairness of assigned punishments, we find that interactions involving a White officer and a Black civilian are judged as less fair than all other interactions

[12] The other interesting pattern is that in experiment 2, interactions involving a White officer are perceived as less fair ($b=-0.19$, $se=0.06$, $p<0.01$). This is similar to the finding in Figure 1, where respondents also prefer a less severe sanction in experiment 2 when the officer is White.

[13] Another interesting finding, which is not robust across experiments 1 and R1, is that in experiment 1, an interaction involving a hostile civilian is perceived as less fair even when an officer remains polite ($b=-0.27$, $se=0.09$, $p<0.01$). In experiment R1, however, this effect is 0.

(b=-0.23, se=0.09, p<0.01; experiment R1, Table A3, Column 11).We also examined whether there were salient differences between Black and White respondents in their evaluations of the fairness of the overall interaction (experiment R1, Table A4, Columns 8 and 9). White respondents react more negatively to officer hostility than do Black respondents (For example, the coefficient for Polite Civilian, Hostile Officer is -0.15 (se=0.12, p<0.01) for White respondents and -0.03 (se=0.09, p<0.01) for Black respondents, and this difference is significant at p<0.05. We find similar differences for the effect of Officer-initiated joint hostility).

As with the measure of the fairness of the chosen punishment, we also examined whether evaluations of overall fairness were moderated by the interaction between assigned punishment and both demeanor and, for experiment 2, whether the civilian was armed.[14] Once again, we find clear evidence of punishment and demeanor interaction effects, but only for experiments 1 and 1R (experiment 1 Likelihood ratio chi^2=32.05, p<0.01; experiment 1R Likelihood ratio chi^2=10.66, p=0.01; experiment 2 Likelihood ratio chi^2=3.73, p>0.05). These interaction results are summarized in Table 5 in a format identical to that used in Table. 4. It shows that the fairness penalty associated with the more severe sanction is always smaller when the civilian is hostile and vanishes entirely when the civilian is armed.

Table 5 about here

Overall, given the general similarity between the fairness of outcomes and overall fairness results, we prefer the former, but note that the latter is closest to what is used in prior work. At the same time, because it is less specific, we view the general fairness item as

[14] We also tested whether there were differences in the effects of demeanor, weapon status, and punishment by officer race. Models that allowed for these interactions did not substantially improve model fit (experiment 1 Likelihood ratio chi^2=5.23, p>0.05; experiment R1 Likelihood ratio chi^2=0.79, p>0.05; experiment 2 Likelihood ratio chi^2= 2.70, p>0.05).

potentially troublesome because it appears sensitive to features of civilian behavior in ways that depart sharply from the patterns seen for evaluations of punishment fairness. This difference aside, as with the other outcome, we see much larger effects of demeanor for the less serious civilian-police interaction than for the situation involving greater threat to another civilian, highlighting how factors affecting perceptions of fairness are context dependent. Finally, our replication study shows that Black civilians may be viewed as receiving less fair treatment by White officers, even holding all other features of those interactions constant. But race of the officer and civilian do not appear to condition the effect of officer hostility on perceptions of fairness to the civilian.

IV. DISCUSSION AND CONCLUSION

American law enforcement has historically garnered strong public support. But confidence in the police reached a low point in 2020, when a majority of Americans reported not being confident in the police for the first time since Gallup began measuring confidence in institutions in 1993 (Brenan, 2020). At the same time, police behavior is becoming increasingly visible to the public with the rapid development and spread of technologies allowing review of police behavior (Goldsmith, 2010). Together, diminishing confidence in the police and increasing public knowledge of and vicarious experiences with police behavior have put American policing at the forefront of policy concerns and debates, leading to calls for reform and ultimately underscoring the importance of understanding what Americans view as fair and appropriate police behavior.

We are beginning to understand the incidents that Americans want the police to respond to, with research showing that the mass public is reluctant to limit police involvement in the handling of even low-level order maintenance and peacekeeping issues (Peyton, Vaughn, &

Huber, 2020). However, we still lack an understanding of what the mass public believes is appropriate and fair *during* common police-civilian interactions, particularly those involving Black Americans, a group for whom concerns about racist police behavior are most salient. The present study explores support for how the police should respond to civilians' problems when the subject of police action is a White or Black American (and for Black Americans, in an interaction with a Black victim), as well as how the public's evaluations of appropriate officer behavior and fairness depend on the circumstances and features of police-civilian interactions.

Our results underscore the importance of considering the dyadic nature of interactions between the police and civilians. We find that while the mass public's punishment preferences and fairness judgements are generally not affected by officer race, they depend on various civilian *and* officer behaviors. In line with recent work showing that Americans want to involve the police in day-to-day order maintenance interactions, we find that when officers and civilians are courteous and civilians are unarmed, the mass public prefers that an interaction end in less punitive ways that nonetheless involve active police involvement. Police-civilian interactions involving hostility and threat (i.e., armed civilians), however, are judged differently and apparently evaluated through a lens of reciprocity and order maintenance: while hostile officers are evaluated negatively, hostile and threatening (i.e., armed) civilians are perceived to be less deserving of lenient treatment. In other words, and in line with recent work (Lee et al., 2024), people support punishing civilians who pose threats to order, even though they also view hostile officers as less fair. Further, officer hostility is judged more fairly in the presence of civilian-initiated hostility, leading us to believe that while the mass public expects both sides to remain courteous, either party deviating from this expectation reduces the penalty for the other side also doing so.

Additionally, our results show that manipulations of demeanor also affect beliefs about the threat that civilians pose. This is an important mechanism for understanding why civilian demeanor may shape people's beliefs about appropriate punishment and the fairness of interactions. It appears that the mass public equates civilian hostility and weapons with threat, as officers do (e.g., Nix, Pickett, & Mitchell, 2019), and the magnitudes of these effects is roughly equal, showing the powerful role of civilian demeanor.

In judging the fairness of officer behavior, the public also expects civilians to behave politely and officers to respond in turn, underscoring again the importance of reciprocity. Furthermore, what's fair depends on how civilians behave *once* the officer arrives. Experiments 1 and R1 showed that so long as a civilian remains polite, more severe punishment is viewed as unfair. But if a civilian is hostile alone or starts a hostile interaction, more severe punishment is not viewed as unfair. In experiment 2, which involved a more serious incident with a victim, there is no evidence that the effect of demeanor changes with arrest status or vice versa. However, the more severe punishment of arrest is judged as fairer when an objective measure of threat (a weapon) is present, even if it has not been used and the civilian is polite. The mass public therefore appears to have fairness standards that change with incident severity or possibly the presence of a victim.[15]

It is vital that researchers and policymakers understand the mass public's perceptions of what constitutes fair and effective policing and, ultimately, work toward increasing compliance with the law and law enforcement. Our findings suggest that the general public's views about

[15] Methodologically, we show that survey respondents form expectations about the punishments officers will assign based on features of police-civilian interactions. Not accounting for (or specifying) assigned punishment may therefore lead to the mistaken conclusion that characteristics other than expected punishment are operating to influence various outcomes.

fairness and punishment are both dynamic and distinct. Policy responses must be, too. Broadly, our results show that researchers and policymakers should more seriously attend to how the public evaluates everyday features of interactions, such as whether a civilian is respectful and has a weapon, rather than continuing to limit their focus to "demographic" interaction features, such as civilian and officer race. Relatedly, our findings suggest that we must expand our focus beyond the traditional procedural justice framework, which is now arguably the gold standard for police policy and training by researchers and policymakers alike.[16] While procedural justice theory and policy enjoy extensive support (e.g., Vaughn, Feigenberg, & Luben, 2021), procedural justice studies often do not attend to the ways in which other interaction features – and perceptions of them – vary. Moreover, procedural justice theory argues that civilians are more concerned with fair treatment than with outcomes (Sunshine & Tyler 2003). Our work underscores the importance of evaluations of both fairness and punishment, and of *distinguishing* between them, as the mass public has specific views about what constitutes fair and effective treatment that are separate from fair and effective sanctions under different conditions. For example, we find that someone who views an officer's treatment as unfair does not necessarily mean that they will want a civilian to avoid punishment, even in low-level interactions and especially when that civilian is perceived as acting unfairly and/or as threatening. Ultimately, better understanding police-civilian interactions will require us to broaden our consideration of how the public evaluates both the outcome (i.e., punishment) and fairness of common police-

[16] For example, procedural justice was a main theme underlying Obama's 21st Century Policing Task Force, and the federal government gave $4.75 million dollars to the National Initiative for Building Community Trust and Justice to, among other things, develop and implement procedural justice training for police across six cities. But our results show simply understanding whether people believe an officer was fair is not equivalent to understanding what outcome they prefer is appropriate.

civilian interactions and how key features of those interactions shape those evaluations.

Still, more research is needed to gain a comprehensive understanding of how the mass public views police-civilian interactions. For one, we analyze only two concrete interactions involving Black civilians, and one involving a White civilian. It would be valuable to extend this analysis to other less serious interactions, traffic stops, and more serious crimes to assess what the public thinks fair and appropriate policing is in these situations. Perhaps there are circumstances where the mass public prefers the police do not get involved at all, and exploring scenarios that vary whether and why the police are involved at all (e.g., a civilian complaint, a call by a social worker, or a traffic issue) is another important line of inquiry. Additionally, while we examine the effect of differences in officer and civilian race, we do not examine the separate question of whether the mass public has different expectations about officer or civilian behavior in interactions involving White rather than Black *victims*. It may be, for example, that members of the public interpret and respond to civilian behavior differently depending on the race of the victim(s) interacting with the police, and such differences may differ systematically across members of the mass public.

Indeed, a more general question beyond the scope of the current paper is to explore systematic heterogeneity in public preferences about police and civilian action to understand observed public divisions on matters of policing policy and criminal justice more generally. While our analysis of experiment R1 provides initial assessments of how Black and White respondents evaluate a single common scenario and shows largely aligned reactions (if with different magnitudes) to the features we manipulate, it would be useful to understand if we find the same patterns in groups that are more likely to interact with the police. A related matter is that while we produce textual descriptions of officer-civilian interactions, many "real world"

treatments are video recordings of these interactions. Thus, randomizing features of interactions captured as movies is also a worthwhile area for future research because the stimuli more fully capture what one would experience if viewing our treatments on YouTube, for example (see e.g., Culhane et al., 2016; Turner et al., 2019; Voigt et al., 2017).

Finally, while survey experiments can help us to better understand how randomized elements of interactions impact people's examinations of police fairness and civilian sanctioning, they are limited in their ability to fully and accurately capture the nuances and dynamics of real-life interactions and people's reactions to them. To better understand civilian and officer behaviors when controlling for the variables we control; researchers should continue to engage in observational analysis of police-civilian interactions, such as body camera and dash cam footage, computer-aided dispatch records, and observational research that examines various types of real-world interactions and civilian preferences of police action in real time (see e.g., Gillooly, 2020; Lum et al., 2022; Worden & McLean, 2014). Traditional policing studies provide key insights about the relationships between civilian demeanor, police treatment, and punishment (e.g., Black & Reiss, 1970; Klinger, 1996; Lundman, 1994; Worden & Shepard, 1996), as well as weapon use and treatment (e.g., Garafolo & Bayley, 1989). Future research should similarly explore how the general public evaluates police-civilian interactions, but in the contemporary context and across less serious interactions.

These areas for future work aside, the key contribution of this work is to provide critical new evidence about public preferences about punishment and fair officer conduct. We find little evidence of either a uniformly punitive or anti-police public. Instead, the public appears to seek and evaluate the police in a sophisticated manner that acknowledges an active role for the police in non-punitive order maintenance alongside respectful treatment of civilians. But the public

becomes more punitive when threats to order are present, either as demonstrated by respondent demeanor or a civilian being armed. The police are similarly held to account for being unnecessarily punitive or hostile in the absence of threat or civilian hostility. Overall, these reactions help understand the likely consequences of expanded public access to observations of police civilian interactions and how observing different sort of interactions will shape attitudes toward the police.

REFERENCES

Agnew, R. (2014). Social concern and crime: Moving beyond the assumption of simple self-interest. *Criminology, 52*(1): 1-32. https://doi.org/10.1111/1745-9125.12031

Allen, J., & Monk-Turner, E. (2010). Civilian perceptions of the legitimacy of traffic stops. *Journal of Criminal Justice 38*, 589-594. https://doi.org/10.1016/j.jcrimjus.2010.04.030

Alpert, G. P., & Dunham, R. G. (2004). *Understanding police use of force: Officers, suspects, and reciprocity*. Cambridge University Press.

Bayley, D. (1995). Getting serious about police brutality. In *Accountability for Criminal Justice: Selected Essays*, edited by P. C. Stenning, pp. 93-109. Toronto: University of Toronto Press.

Black, D. J., & Reiss Jr, A. J. (1970). Police control of juveniles. *American Sociological Review*, 63-77. https://doi.org/10.2307/2093853

Bayley, D. H., & Garofalo, J. (1989). The management of violence by police patrol officers. *Criminology*, *27*(1), 1-26. https://doi.org/10.1111/j.1745-9125.1989.tb00861.x

Bittner, E. (1970). *The functions of the police in modern society: A review of background factors, current practices, and possible role models*.

Brenan, M. (2020). Amid pandemic, confidence in key U.S. institutions surges. *Gallup*, August 12, 2020.

Chandek, M. S. (1999). Race, expectations and evaluations of police performance: An empirical assessment. *Policing: An International Journal of Police Strategies & Management*, *22*(4), 675-695. https://doi.org/10.1108/13639519910299599

Culhane, S. E., Boman IV, J. H., & Schweitzer, K. (2016). Public perceptions of the justifiability of police shootings: The role of body cameras in a pre-and post-Ferguson

experiment. *Police Quarterly*, *19*(3), 251-274. https://doi.org/10.1177/1098611116651403

Dafoe, A., Zhang, B., & Caughey, D. (2018). Information equivalence in survey experiments. *Political Analysis*, 26(4), 399–416. https://doi.org/10.1017/pan.2018.9

Dickson, E. S., Gordon, S. C., & Huber, G. A. (2022). Identifying legitimacy: Experimental evidence on compliance with authority. *Science Advances*, 8, 1-13. DOI: 10.1126/sciadv.abj7377

Druckman, J. N., & Levendusky, M. S. (2019). What do we measure when we measure affective polarization? *Public Opinion Quarterly*, *83*(1), 114-122. https://doi.org/10.1093/poq/nfz003

Easton, D. (1965). *A systems analysis of political life*. New York: John Wiley & Sons.

Epp, C. R., Maynard-Moody, S., & Haider-Markel, D. P. (2014). *Pulled over: How police stops define race and civilianship*. University of Chicago Press.

Gerber, A. S., Huber, G. A., Meredith, M., Biggers, D. R., & Hendry, D. J. (2017). Does incarceration reduce voting? Evidence about the political consequences of spending time in prison. *Journal of Politics*, *79*(4), 1130-46. https://doi.org/10.1086/692670

Gerber, A. S., Huber, G. A., Tucker, P. D., & Cho, J. J. (2023). The Importance of Breaking Even: How Local and Aggregate Returns Make Politically Feasible Policies. *British Journal of Political Science*, 1-18. https://doi.org/10.1017/S0007123423000522

Girgenti-Malone, A. A., Khoder, C., Vega, G., & Castillo, D. (2017). College students' perceptions of police use of force: do suspect race and ethnicity matter?. *Police Practice and Research*, 18(5), 492-506.

Gillooly, J. W. (2020). How 911 callers and call-takers impact police encounters with the public: The case of the Henry Louis Gates Jr. arrest. *Criminology & Public Policy*, *19*(3), 787-804. https://doi.org/10.1111/1745-9133.12508.

Goldsmith, A. J. (2010). Policing's new visibility. *The British Journal of Criminology*, *50*(5), 914-934. https://doi.org/10.1093/bjc/azq033

Goldstein, H. (1963). Police discretion: The ideal versus the real. *Public Administration Review*, 140-148. https://doi.org/10.2307/973838

Gottfredson, M. R., & Hindelang, M. J. (1979). A study of the behavior of law. *American Sociological Review*, 3-18. https://doi.org/10.2307/2094813

Hickman, L. J., & Simpson, S. S. (2003). Fair treatment or preferred outcome? The impact of police behavior on victim reports of domestic violence incidents. *Law & Society Review*, *37*(3), 607-634. https://doi.org/10.1111/1540-5893.3703005

Holmes, M. D., & Smith, B. W. (2012). Intergroup dynamics of extra-legal police aggression: An integrated theory of race and place. *Aggression And Violent Behavior*, *17*(4), 344-353. https://doi.org/10.1016/j.avb.2012.03.006

Jenness, V., & Calavita, K. (2018). "It depends on the outcome': Prisoners, grievances, and perceptions of justice. *Law & Society Review*, *52*(1), 41-72. https://doi.org/10.1111/lasr.12312

Johnson, D., Wilson, D. B., Maguire, E. R., & Lowrey-Kinberg, B. V. (2017). Race and perceptions of police: Experimental results on the impact of procedural (in) justice. *Justice Quarterly*, *34*(7), 1184-1212. https://doi.org/10.1080/07418825.2017.1343862

Kahn, K.B., Thompson, M. & McMahon, J.M. (2017). Privileged protection? Effects of suspect race and mental illness status on public perceptions of police use of force. *J Experimental Criminology,* **13**, 171–191. https://doi.org/10.1007/s11292-016-9280-0

Klinger, D. A. (1996). More on demeanor and arrest in Dade County. *Criminology*, *34*(1), 61-82. https://doi.org/10.1111/j.1745-9125.1996.tb01195.x

LaFave, W. R. (1962). The police and nonenforcement of the law-Part I. *Wisconsin Law Review.*, 104.

Leovy, J. (2015). *Ghettoside: A true story of murder in America*. One World.

Levy, I., Cohen-Louck, K., & Herzog, S. (2021). Predicting support for community corrections: Crime type and severity, and offender, observer, and victim characteristics. *Punishment & Society*, *24*(3), 346-366. https://doi.org/10.1177/1462474521989805

Lum, C., Koper, C. S., & Wu, X. (2021). Can we really defund the police? A nine-agency study of police response to calls for service. *Police Quarterly*, *25*(3), 255-280. DOI: 10.1177/10986111211035002

Lum, C., & Nagin, D. S. (2017). Reinventing American policing. *Crime and Justice*, *46*(1), 339-393. https://doi.org/10.1086/688462

Lundman, R. J. (1994). Demeanor or crime? The Midwest city police-civilian encounters study. *Criminology*, *32*(4), 631-656. https://doi.org/10.1111/j.1745-9125.1994.tb01168.x

Maguire, E. R., Lowrey, B. V., & Johnson, D. (2017). Evaluating the relative impact of positive and negative encounters with police: A randomized experiment. *Journal of Experimental Criminology*, *13*(3), 367-391. https://doi.org/10.1007/s11292-016-9276-9

Mastrofski, S. D., Jonathan-Zamir, T., Moyal, S., & Willis, J. J. (2016). Predicting procedural justice in police–civilian encounters. *Criminal Justice and Behavior*, 43, 119-139. https://doi.org/10.1177/0093854815613540

Nagin, D. S. (1998). Criminal deterrence research at the outset of the twenty-first century. *Crime and Justice*, 23, 1-42. https://doi.org/10.1086/449268

Nagin, D. S., & Telep, C. W. (2017). Procedural justice and legal compliance. *Annual Review of Law and Social Science*, 13, 5-28. https://doi.org/10.1146/annurev-lawsocsci-110316-113310

Natapoff, A. (2018). *Punishment without crime: How our massive misdemeanor system traps the innocent and makes America more unequal*. Hachette UK.

Nix, J., Pickett, J. T., & Mitchell, R. J. (2019). Compliance, noncompliance, and the in-between: Causal effects of civilian demeanor on police officers' cognitions and emotions. *Journal of Experimental Criminology*, *15*(4), 611-639. https://doi.org/10.1007/s11292-019-09363-4

Nix, J., Pickett, J. T., Wolfe, S. E., & Campbell, B. A. (2017). Demeanor, race, and police perceptions of procedural justice: Evidence from two randomized experiments. *Justice Quarterly*, *34*(7), 1154-1183. https://doi.org/10.1080/07418825.2017.1334808

Peyton, K., Huber, G. A., & Coppock, A. (2020). The generalizability of online experiments conducted during the covid-19 pandemic. *Journal of Experimental Political Science*, 9(3), 379-394. doi:10.1017/XPS.2021.17

Peyton, K., Vaughn, P. E., & Huber, G. A. (2020, August 31). Americans don't support the idea of defunding the police. *The Washington Post*.

https://www.washingtonpost.com/politics/2020/08/31/slogan-defund-police-canturn-americans-away-movement-against-police-brutality/

Pickett, J. T., & Ryon, S. B. (2017). Procedurally just cooperation: Explaining support for due process reforms in policing. *Journal of Criminal Justice*, *48*, 9-20. https://doi.org/10.1016/j.jcrimjus.2016.11.002

Pickett, J., & Nix, J. (2019). Demeanor and police culture: Theorizing how civilian cooperation influences police officers. *Policing: An International Journal, 42*(4), 537-555. https://doi.org/10.1108/PIJPSM-09-2018-0133

Reisig, M. D., Mays, R. D., & Telep, C. W. (2018). The effects of procedural injustice during police–civilian encounters: A factorial vignette study. *Journal of Experimental Criminology*, *14*(1), 49-58. https://doi.org/10.1007/s11292-017-9307-1

Rojek, J., Alpert, G. P., & Smith, H. P. (2012). Examining officer and citizen accounts of police use-of-force incidents. *Crime & Delinquency*, *58*(2), 301-327. https://doi.org/10.1177/0011128710386206

Sellin, T., & Wolfgang, M. E. (1964). *The measurement of delinquency*. Wiley.

Soss, J., & Weaver, V. (2017). Police are our government: Politics, political science, and the policing of race–class subjugated communities. *Annual Review of Political Science*, *20*, 565-591. DOI: 10.1146/annurev-polisci-060415-093825.

Skogan, W. G. (2005). Citizen satisfaction with police encounters. *Police Quarterly* , 8(3), 298–321.

Sunshine, J., & Tyler, T. R. (2003). The role of procedural justice and legitimacy in shaping public support for policing. *Law and Society Review*, *37*, 513-548. https://doi.org/10.1111/1540-5893.3703002.

Tankebe, J. (2009). Public cooperation with the police in ghana: Does procedural fairness matter? *Criminology*, *47*(4), 1265-1293. https://doi.org/10.1111/j.1745-9125.2009.00175.x

Terrill, W., & Mastrofski, S. D. (2002). Situational and officer-based determinants of police coercion. *Justice Quarterly*, *19*(2), 215-248. https://doi.org/10.1080/07418820200095221

Turner, B. L., Caruso, E. M., Dilich, M. A., & Roese, N. J. (2019). Body camera footage leads to lower judgments of intent than dash camera footage. *Proceedings of the National Academy of Sciences*, *116*(4), 1201-1206. https://doi.org/10.1073/pnas.1805928116

Tyler, T. R. (2004). Enhancing police legitimacy. *The Annals of the American Academy of Political and Social Science, 593*(1): 84-99. https://doi.org/10.1177/0002716203262627

Tyler, T. R., & Huo, Y. J. (2002). *Trust in the law: Encouraging public cooperation with the police and courts*. Russell Sage Foundation.

Tyler, T., R., & Wakslak, C. (2004). Profiling and police legitimacy: Procedural justice, attributions of motive, and acceptance of police authority, *Criminology*, 42(2), 253–82.

Vaughn, P. E., Peyton, K., & Huber, G. A. (2022). Mass support for proposals to reshape policing depends on the implications for crime and safety. *Criminology & Public Policy*, 125-146. DOI: 10.1111/1745-9133.12572

Van Maanen, J. (1978). The asshole. *Policing: A view from the street*, 221-238.

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*, *114*(2), 6521-6526. https://doi.org/10.1073/pnas.1702413114

Warr, M. (1989). What is the perceived seriousness of crimes? *Criminology*, *27*(4), 795-822. DOI:10.1111/j.1745-9125.1989.tb01055.x

Warr, M., Meier, R. F., & Erickson, M. L. (1983). Norms, theories of punishment, and publicly preferred penalties for crimes. *Sociological Quarterly*, *24*(1), 75-91. http://www.jstor.org/stable/4106364

Warr, M., & Stafford, M. (1984). Public goals of punishment and support for the death penalty. *Journal of Research in Crime and Delinquency*, *21*(2), 95-111. https://doi.org/10.1177/0022427884021002002

Weitzer, R. (2015). American policing under fire: Misconduct and reform. *Society*, *52*(5), 475-480. https://doi.org/10.1007/s12115-015-9931-1

White, A. (2019). Misdemeanor disenfranchisement? The demobilizing effects of brief jail spells on potential voters. *American Political Science Review*, *113*(2), 311-24. doi:10.1017/S000305541800093X

Wilson, J. Q. (1968). *Varieties of police behavior: The management of law and order in eight communities.* Cambridge, MA: Harvard University Press.

Worden, R. E., & McLean, S. J. (2014). Assessing police performance in civilian encounters: police legitimacy and management accountability. *Report to the National Institute of Justice*.

Worden, R. E., & Shepard, R. L. (1996). Demeanor, crime, and police behavior: A reexamination of the police services study data. *Criminology*, *34*(1), 83-105. https://doi.org/10.1111/j.1745-9125.1996.tb01196.x

Yesberg, J. A., Kyprianides, A., Bradford, B., Milani, J., Quinton, P., & Clark-Darby, O. (2022). Race and support for police use of force:Findings from the UK. *Policing and Society*, 32(7), 878-895.

TABLE 1 Experiment 1 vignette text with associated outcomes

Vignette text (pre-punishment assignment; text in brackets is randomized):

Officer Simmons, a 35-year-old [*Black* OR *White*] male, has worked as a patrol officer in the city for 9 years. On Tuesday June 16th at 10:15PM, Officer Simmons was dispatched to a city park in response to a 911 call about possible gunshots in the area.

Upon arrival, Officer Simmons observed an individual lighting off fireworks from the park's basketball courts. The basketball courts and the park are closed to the public between 8:00PM and 6:00AM.

At 10.25PM, Officer Simmons reported to dispatch that the gunshot call was a false alarm. He then exited his vehicle and activated his body worn camera. The camera footage shows Officer Simmons approaching a young Black male, later identified as Samuel Parsons. Parsons is 19 years old and lives about a mile from the park. According to the body camera footage, the interaction between Officer Simmons' and Parsons went as follows:

> Officer Simmons: "[Polite: *Young man, the park is closed. What are you doing out here?* OR Hostile: *Hey asshole, can't you read the signs – the park is closed. What the hell do you think you're doing?*]"
>
> Parsons: "[Polite: *Hi Officer, sorry about that. I'm just messing around.* OR Hostile: *Woah chill man, I don't have to tell you anything.*]"
>
> Officer Simmons: "[Polite: *It's illegal to light off fireworks in the park. We've gotten complaints from people in the neighborhood, and they think they're hearing gunshots. Can you clean up this mess and go home?* OR Hostile: *It's illegal to light off fireworks in the park you idiot. You've got people calling in reporting gunshots. Clean up your shit and get out of here*.]"
>
> Parsons: "[Polite: *Sorry about that. I wasn't trying to cause any trouble. There aren't any kids around or anything. I'll get out of here.* OR Hostile: *What the fuck man? Do you seriously have nothing better to do than harass me? I'm not hurting anyone. Go fuck yourself.*]"

Outcomes (pre-punishment assignment):

- How fairly was Parsons treated by Officer Simmons?
- How fairly was Officer Simmons treated by Parsons?
- How do you think this interaction ended?

Vignette text (post-punishment assignment):

At 10.48PM, shortly after the exchange on the basketball courts, Officer Simmons [*escorted Parsons out of the park* OR *escorted Parsons out of the park and wrote him a ticket for disorderly conduct*].

Outcomes (post-punishment assignment):

- How fair was it for Officer Simmons to [escort Parsons out of the park OR escort Parsons out of the park and write him a ticket for disorderly conduct]?
- What should Officer Simmons have done, in your view?
- How fair was Parson's overall experience with Officer Simmons?

TABLE 2 Replication of experiment 1 vignette text with associated outcomes

Vignette text (text in brackets is randomized):

Officer Gerald Simmons, a 35-year-old [*Black* OR *White*] male, has been a patrol officer for 9 years. On June 16th at 10:15PM, Officer Simmons was dispatched to a city park in response to a 911 call about possible gunshots.

Upon arrival, Officer Simmons observed an individual lighting off fireworks from the park's basketball courts. The park is closed between 8:00PM and 6:00AM.

At 10:25PM, Officer Simmons reported to dispatch that the gunshot call was a false alarm. He exited his vehicle and activated his body worn camera. The camera footage shows Officer Simmons approaching a young [*Black* OR *White*] male, later identified as Samuel Parsons. Parsons is 19 years old and lives about a mile from the park. According to the body camera footage, the interaction between Officer Simmons' and Parsons went as follows:

> Officer Simmons: "[Polite: *Young man, the park is closed. What are you doing out here?* OR Hostile: *Hey asshole, can't you read the signs – the park is closed. What the hell do you think you're doing?*]"
>
> Parsons: "[Polite: *Hi Officer, sorry about that. I'm just messing around.* OR Hostile: *Woah chill man, I don't have to tell you anything.*]"
>
> Officer Simmons: "[Polite: *It's illegal to light off fireworks in the park. We've gotten complaints from people in the neighborhood, and they think they're hearing gunshots. Can you clean up this mess and go home?* OR Hostile: *It's illegal to light off fireworks in the park you idiot. You've got people calling in reporting gunshots. Clean up your shit and get out of here*.]"
>
> Parsons: "[Polite: *Sorry about that. I wasn't trying to cause any trouble. There aren't any kids around or anything. I'll get out of here.* OR Hostile: *What the fuck man? Do you seriously have nothing better to do than harass me? I'm not hurting anyone. Go fuck yourself.*]"

10:48PM, shortly after the exchange on the basketball courts, Officer Simmons [*escorted Parsons out of the park* OR *escorted Parsons out of the park and wrote him a ticket for disorderly conduct*].

Outcomes:

- What should Officer Simmons have done, in your view?
- How fairly was Parsons treated by Officer Simmons?
- How fairly was Officer Simmons treated by Parsons?
- How fair was it for Officer Simmons to [*escort Parsons out of the park* OR *escort Parsons out of the park and write him a ticket for disorderly conduct*]?

TABLE 3 Experiment 2 vignette text with associated outcomes

Vignette text (pre-punishment assignment; text in brackets is randomized):

Officer Simmons, a 35-year-old [*Black* OR *White*] male, has worked as a patrol officer in the city for 9 years. On Tuesday May 16th at 4:15 PM, Officer Simmons was dispatched to a city park in response to an anonymous 911 call about a fight on a baseball field.

At 4:20 PM, Officer Simmons exited his vehicle and activated his body worn camera. The camera footage shows Officer Simmons approaching two young Black men, later identified as Samuel Parsons and Austin Storms. Parsons is 19 years old, and Storms is 18. Both individuals live about a mile from the field.

Camera footage shows Parsons raising [*his fists* OR *a baseball bat*] above Storms' head, and Storms squatting on the ground, shielding his head with his hands.

According to the body camera footage, the interaction between Officer Simmons' and Parsons went as follows:

> Officer Simmons: "[Polite: *Stop! Put your hands up, and step away from him.* OR Hostile: *Hey asshole! Stop! Drop your weapon, put your fucking hands up, and step away from him now!*]"
>
> Parsons: "[Polite: *Sorry officer, we got into an argument. I wasn't going to hurt him.* OR Hostile: *Fuck off. This is none of your business. I wasn't going to hurt him. Don't you have something better to do?*]"
>
> Officer Simmons: "[Polite: *I'm here because somebody called in reporting an assault. I need to figure out what happened.* OR Hostile: *Look dipshit, somebody called in reporting an assault. I need to figure out what happened.*]"

Camera footage shows Parsons [dropping the baseball bat,] putting his hands up[,] and stepping away from Storms.

Outcomes (pre-punishment assignment):

- How fairly was Parsons treated by Officer Simmons?
- Once Officer Simmons arrived, how much of a threat was Parsons to Storms?
- How much of a threat was Parsons to Officer Simmons?
- How much of a threat was Officer Simmons to Parsons?
- How do you think this interaction ended?

Vignette text (post-punishment assignment):

Storms left the baseball field around 4.40PM. Officer Simmons [*formally placed Parsons under arrest and he was taken to the police station about 20 minutes later* OR *issued Parsons a ticket and escorted him out of the park about 20 minutes later*].

Outcomes (post-punishment assignment):

- How fair was it for Officer Simmons to [*formally place Parsons under arrest and take him to the police station about 20 minutes later* OR *issue Parsons a ticket and escort him out of the park about 20 minutes later*]?
- How fair was this outcome to Storms?
- What should Officer Simmons have done, in your view?
- How fair was Parson's overall experience with Officer Simmons?

TABLE 4 Interaction effects of demeanor and weapon with assigned punishment on fairness of assigned punishment

Study	Baseline	Civilian Hostile, Officer Polite	Civilian Armed
Experiment 1	-0.92	-0.07	
	(0.11)***	-(0.11)	
Experiment R1	-0.74	-0.27	
	(0.11)***	(0.09)***	
Experiment 2	-0.16	0.04	0.33
	-(0.13)	-(0.13)	(0.13)**

Note: Entries are estimated marginal effects and standard errors of assigning more punitive sanction on evaluations of fairness of the sanction, by column heading.

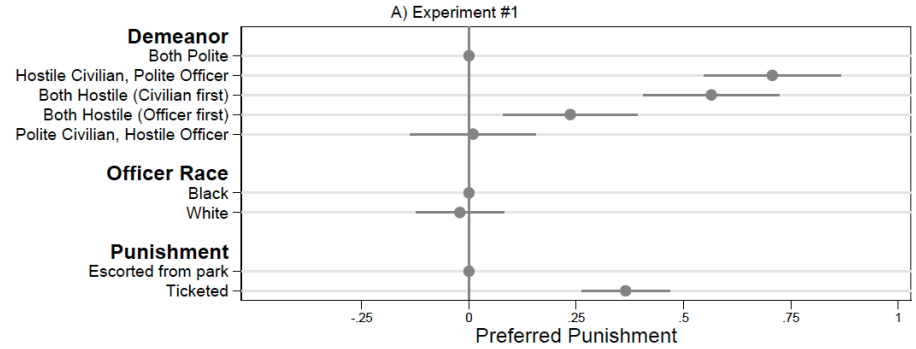
* significant at 10%; ** significant at 5%; *** significant at 1%

TABLE 5 Interaction effects of demeanor and weapon with assigned punishment on evaluations of overall fairness

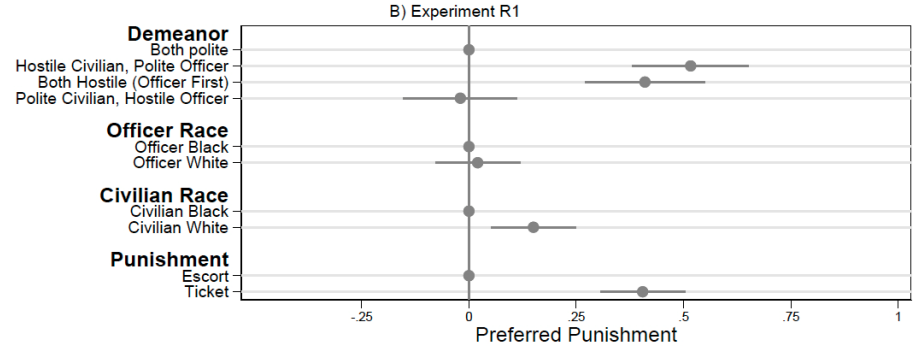
Study	Baseline	Civilian Hostile, Officer Polite	Civilian Armed
Experiment 1	-0.71	-0.57	
	(0.11)***	(0.12)***	
Experiment R1	-0.45	-0.28	
	(0.09)***	(0.08)***	
Experiment 2	-0.19	-0.14	0.20
	-(0.13)	-(0.14)	-(0.14)

Note: Entries are estimated marginal effects and standard errors of assigning more punitive sanction on evaluations of fairness of officer's treatment of civilian, by column heading.
* significant at 10%; ** significant at 5%; *** significant at 1%

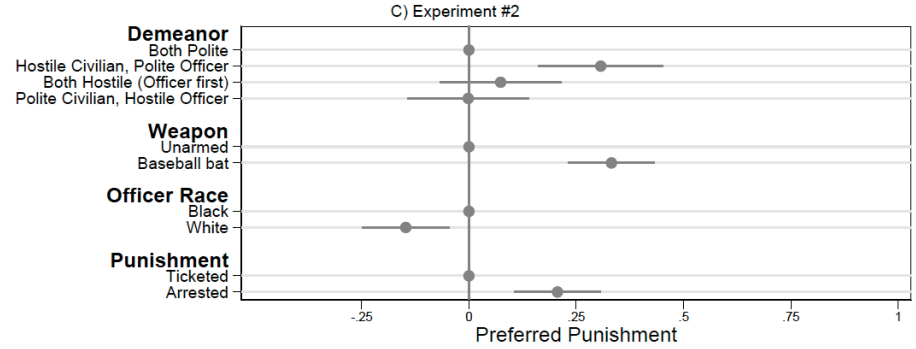
FIGURE 1 The effect of vignette characteristics on preferred punishment



Preferred punishment ranges from 1 to 6.
Predicted preferred punishment for Black officer/Both polite/Escorted from park is 3.22



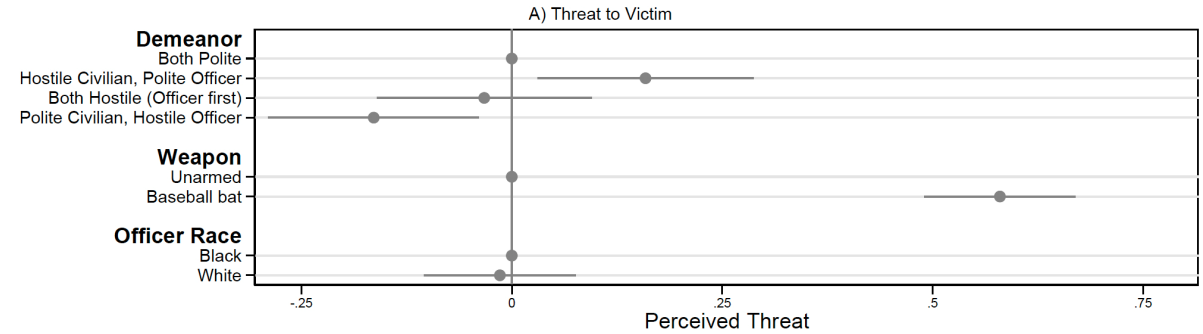
Preferred punishment ranges from 1 to 6.
Predicted preferred punishment for White officer/White civilian/Both polite/Escorted from park is 2.57



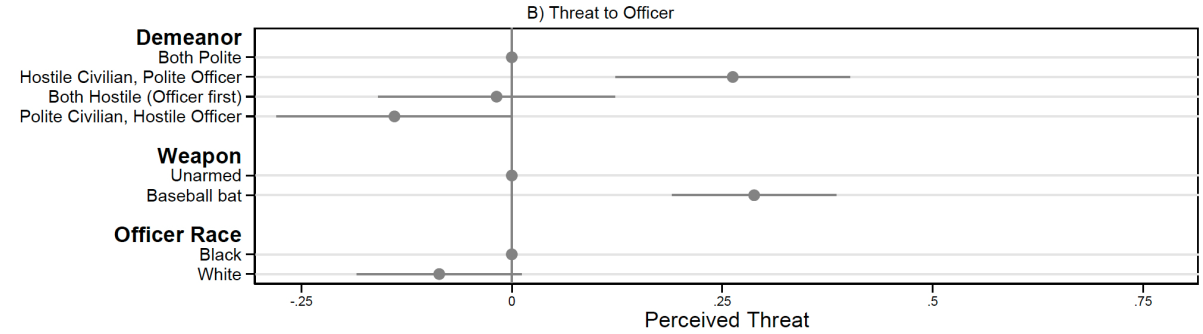
Preferred punishment ranges from 1 to 5.
Predicted baseline preferred punishment for Black officer/Both polite/No weapon/Ticketed is 3.00

Point estimates displayed with 95% confidence intervals. Regression results in Appendix Tables A1-A6.

FIGURE 2 The effect of vignette characteristics on perceived threat (experiment 2)



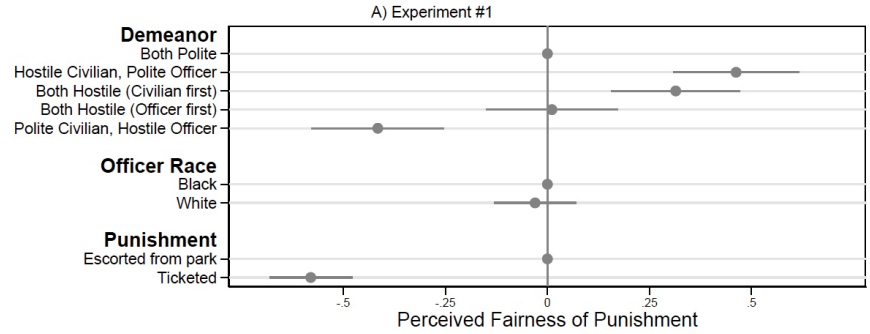
Threat to victim ranges from 1 to 5.
Predicted baseline threat for Black officer/Both polite/No weapon is 3.20



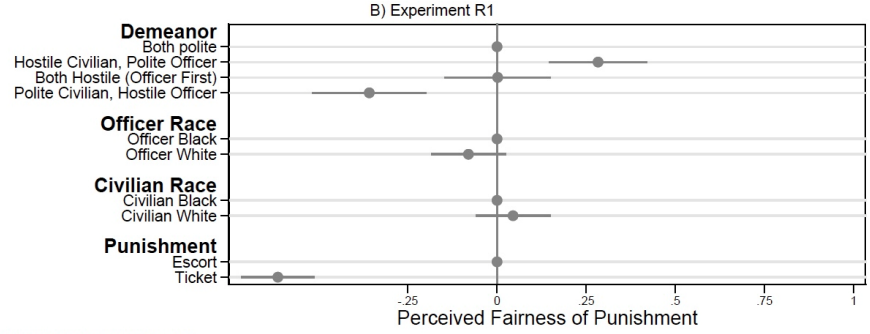
Threat to officer ranges from 1 to 5.
Predicted baseline threat for Black officer/Both polite/No weapon is 2.25

Point estimates displayed with 95% confidence intervals. Prior to punishment being revealed. Regression results appear in Appendix Tables A5 and A6.

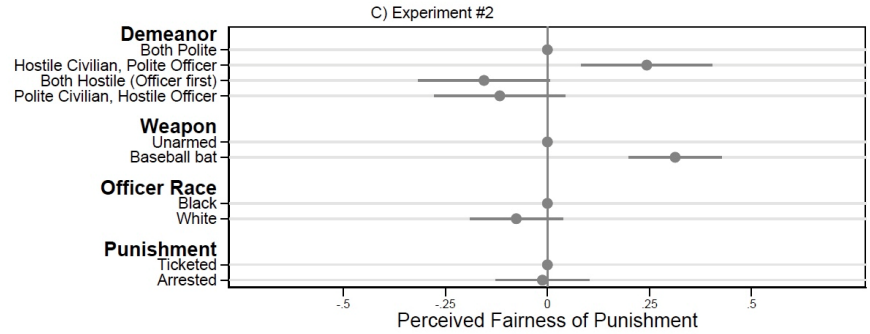
FIGURE 3 The effect of vignette characteristics on perceived fairness of punishment



Perceived fairness ranges from 1 to 5.
Predicted perceived fairness for Black officer/Both polite/Escorted from park is 4.14



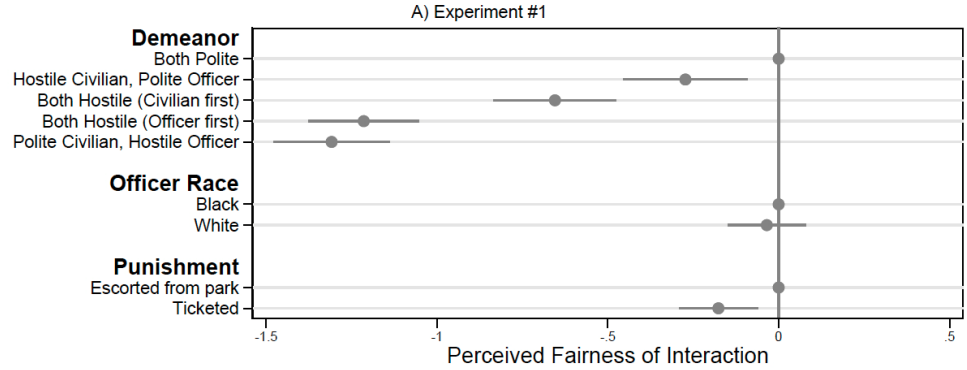
Perceived fairness ranges from 1 to 5.
Predicted perceived fairness for White officer/White civilian/Both polite/Escorted from park is 3.60



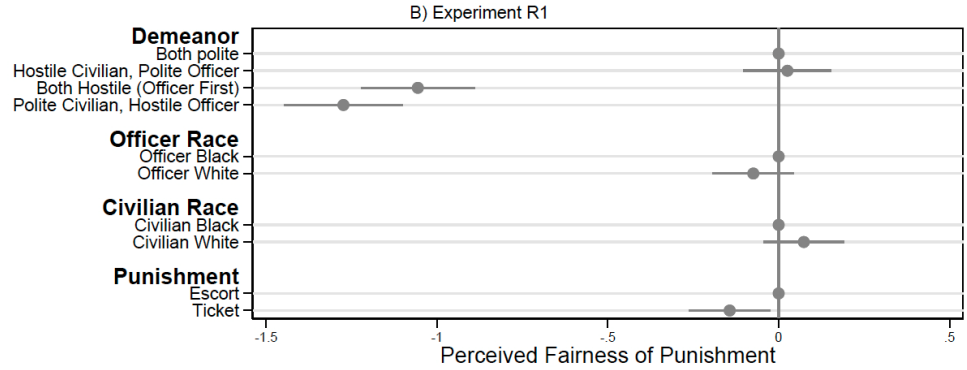
Perceived fairness ranges from 1 to 5.
Predicted perceived fairness for Black officer/Both polite/No weapon/Ticketed is 3.15

Point estimates displayed with 95% confidence intervals. Complete regression results appear in Tables A1-A6.

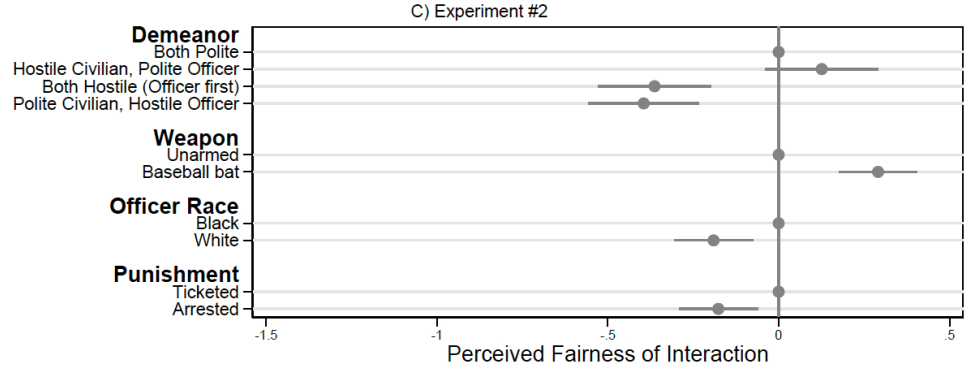
FIGURE 4 The effect of vignette characteristics on perceived fairness of the interaction



Overall fairness ranges from 1 to 5.
Predicted perceived fairness for Black officer/Both polite/Escorted from park is 4.16



Perceived fairness ranges from 1 to 5.
Predicted perceived fairness for White officer/White civilian/Both polite/Escorted from park is 3.50



Overall fairness ranges from 1 to 5.
Predicted perceived fairness for Black officer/Both polite/No weapon/Ticketed is 3.45

Point estimates displayed with 95% confidence intervals. Complete regression results appear in Tables A1-A6.

Seeing the State in Action: Public Preferences about and Judgements of Common Police-Civilian Interactions

(Online Appendix)

Online Appendix

Contents

A Figures and Tables Referenced in Main Text

A.1 Regression Results

TABLE A1 Experiment 1 main regression results

	Outcome Measure					
Vignette Characteristic	**Severity of Expected Civilian Punishment (Prior to Learning Punishment)**	**Preferred Punishment**	**Fairness to Civilian of Assigned Punishment**	**Fairness to Civilian of Assigned Punishment (Demeanor-Punishment Interactions)**	**Overall Fairness to Civilian of Interaction**	**Overall Fairness to Civilian of Interaction (Demeanor-Punishment Interactions)**
Officer Race (baseline=Black)						
Officer White (1=yes)	0.04	-0.021	-0.03	-0.029	-0.035	-0.029
	[0.056]	[0.052]	[0.051]	[0.051]	[0.059]	[0.058]
Demeanor (baseline=Both Polite)						
Demeanor=Hostile Civilian, Polite Officer	1.707	0.707	0.462	0.068	-0.273	-0.681
	[0.087]***	[0.081]***	[0.079]***	[0.101]	[0.092]***	[0.128]***
Demeanor=Both Hostile (Civilian first)	1.578	0.565	0.314	0.029	-0.655	-1.111
	[0.092]***	[0.081]***	[0.080]***	[0.101]	[0.092]***	[0.124]***
Demeanor=Both Hostile (Officer first)	1.704	0.236	0.011	-0.159	-1.215	-1.434
	[0.093]***	[0.079]***	[0.082]	[0.104]	[0.082]***	[0.112]***
Demeanor=Polite Civilian, Hostile Officer	0.979	0.01	-0.415	-0.417	-1.309	-1.535
	[0.102]***	[0.075]	[0.082]***	[0.106]***	[0.087]***	[0.115]***
Punishment (baseline=escorted)		0.365	-0.58	-0.917	-0.176	
Ticketed (1=yes)		0.365	-0.58	-0.917	-0.176	-0.707
		[0.053]***	[0.052]***	[0.113]***	[0.059]***	[0.110]***
Hostile Civilian, Polite Officer X Ticketed				0.775		0.82
				[0.155]***		[0.183]***
Both Hostile (Civilian first) X Ticketed				0.59		0.942
				[0.161]***		[0.181]***
Both Hostile (Officer first) X Ticketed				0.353		0.464
				[0.163]**		[0.163]***
Polite Civilian, Hostile Officer X Ticketed				-0.013		0.47
				[0.165]		[0.172]***
Constant	2.596	3.216	4.141	4.302	4.161	4.412
	[0.078]***	[0.060]***	[0.065]***	[0.075]***	[0.067]***	[0.074]***
Observations	2013	2012	2010	2010	2008	2008
R-squared	0.223	0.081	0.113	0.129	0.138	0.151
Sample mean	3.799	3.686	3.911		3.372	
FX Civ. responds to hostility vs remains polite	0.725	0.226	0.426		0.094	
p. diff 1	0	0.006	0		0.29	
FX Civ. rather than officer starting hostility	-0.125	0.329	0.303		0.56	
p. diff 2	0.117	0	0		0	

OLS Coefficients with robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%

TABLE A2 Experiment 1 regression results with covariates

	Outcome Measure		
Vignette Characteristic	**Preferred Punishment (Covariates)**	**Fairness to Civilian of Assigned Punishment (Covariates)**	**Overall Fairness to Civilian of Interaction (Covariates)**
Officer Race (baseline=Black)			
Officer White (1=yes)	-0.036	-0.053	-0.052
	[0.051]	[0.049]	[0.058]
Demeanor (baseline=Both Polite)			
Demeanor=Hostile Civilian, Polite Officer	0.743	0.501	-0.25
	[0.078]***	[0.074]***	[0.091]***
Demeanor=Both Hostile (Civilian first)	0.575	0.332	-0.644
	[0.079]***	[0.075]***	[0.090]***
Demeanor=Both Hostile (Officer first)	0.247	0.02	-1.21
	[0.077]***	[0.078]	[0.082]***
Demeanor=Polite Civilian, Hostile Officer	0.02	-0.414	-1.314
	[0.075]	[0.082]***	[0.086]***
Punishment (baseline=escorted)	0.352	-0.599	-0.175
Ticketed (1=yes)	0.352	-0.599	-0.175
	[0.051]***	[0.049]***	[0.058]***
Age in years	0.015	0.017	0.008
	[0.002]***	[0.002]***	[0.002]***
Female (1=yes, else=0)	0	0.053	0.033
	[0.053]	[0.051]	[0.059]
Education (1=HS or less, 4=Post BA)	-0.041	-0.037	-0.021
	[0.028]	[0.027]	[0.032]
Race refused (1=yes)	0.22	-0.102	0.035
	[0.193]	[0.179]	[0.185]
Race = Black/African American (1=yes)	-0.198	-0.31	-0.279
	[0.094]**	[0.085]***	[0.094]***
Race = Other (1=yes)	-0.008	0.068	0.033
	[0.087]	[0.085]	[0.097]
Hispanic (1=yes)	-0.121	-0.108	-0.073
	[0.088]	[0.081]	[0.089]
Household Income (1=<15K, 24=250K+, 25=RF)	0.005	0.011	0.01
	[0.004]	[0.004]**	[0.005]**
Income refused	-0.082	-0.103	-0.017
	[0.158]	[0.134]	[0.162]
Constant	2.629	3.373	3.779
	[0.119]***	[0.120]***	[0.130]***
Observations	1993	1991	1989
R-squared	0.146	0.209	0.161

OLS Coefficients with robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

TABLE A3 Experiment R1 main regression results

	Outcome Measure											
Vignette Characteristic	Preferred Punishment	Preferred Punishment (Civilian-Officer Racial Dyads)	Preferred Punishment (Both Police)	Preferred Punishment (Hostile Civilian, Police Officer)	Preferred Punishment (Both Hostile - Officer First)	Preferred Punishment (Police Civilian, Hostile Officer)	Fairness to Civilian of Assigned Punishment	Fairness to Civilian of Assigned Punishment (Civilian-Officer Racial Dyads)	Fairness to Civilian of Assigned Punishment (Demeanor-Punishment Interactions)	Overall Fairness to Civilian of Treatment	Overall Fairness to Civilian of Treatment (Civilian-Officer Racial Dyads)	Overall Fairness to Civilian of Treatment (Demeanor-Punishment Interactions)
Officer Race (baseline=Black)												
Officer White (1=yes)	0.020						-0.080		-0.065	-0.074		-0.068
	[0.050]						[0.053]		[0.053]	[0.060]		[0.060]
Civilian Race (baseline=Black)												
Civilian White (1=yes)	0.150						0.045		0.050	0.074		0.070
	[0.050]***						[0.053]		[0.053]	[0.060]		[0.060]
Both Civilian and Officer White		0.175	0.008	0.382	0.232	0.068		-0.031			0.006	
		[0.069]**	[0.135]	[0.130]***	[0.153]	[0.134]		[0.072]			[0.083]	
Civilian Black and Officer White		-0.092	-0.166	-0.034	-0.149	-0.038		-0.207			-0.228	
		[0.069]	[0.124]	[0.150]	[0.141]	[0.139]		[0.076]***			[0.086]***	
Civilian White and Officer Black		0.034	0.009	-0.007	0.010	0.126		-0.086			-0.085	
		[0.070]	[0.138]	[0.147]	[0.142]	[0.130]		[0.074]			[0.084]	
Demeanor (baseline=Both Polite)												
Demeanor=Hostile Civilian, Polite Officer	0.517	0.507					0.283	0.272	0.066	0.026	0.012	-0.135
	[0.069]***	[0.069]***					[0.070]***	[0.069]***	[0.077]	[0.065]	[0.065]	[0.086]
Demeanor=Both Hostile (Officer first)	0.410	0.400					0.002	-0.009	-0.132	-1.056	-1.069	-1.342
	[0.071]***	[0.071]***					[0.076]	[0.076]	[0.087]	[0.085]***	[0.085]***	[0.114]***
Demeanor=Polite Civilian, Hostile Officer	-0.020	-0.035					-0.358	-0.375	-0.270	-1.274	-1.294	-1.461
	[0.068]	[0.068]					[0.082]***	[0.082]***	[0.091]***	[0.088]***	[0.088]***	[0.119]***
Punishment (baseline=escorted)												
Officer gave ticket (1=yes, 0=no)	0.405	0.405	0.602	0.511	0.241	0.308	-0.615	-0.614	-0.740	-0.143	-0.143	-0.446
	[0.050]***	[0.050]***	[0.093]***	[0.101]***	[0.107]**	[0.098]***	[0.052]***	[0.052]***	[0.106]***	[0.060]**	[0.060]**	[0.093]***
Hostile Civilian, Polite Officer X Ticketed									0.405			0.301
									[0.134]***			[0.129]**
Both Hostile (Officer first) X Ticketed									0.261			0.555
									[0.150]*			[0.168]***
Polite Civilian, Hostile Officer X Ticketed									-0.177			0.351
									[0.160]			[0.175]**
Constant	2.568	2.632	2.588	3.023	3.119	2.645	3.604	3.675	3.661	3.504	3.590	3.665
	[0.062]***	[0.067]***	[0.098]***	[0.113]***	[0.110]***	[0.098]***	[0.065]***	[0.069]***	[0.068]***	[0.068]***	[0.074]***	[0.069]***
Observations	1529	1529	375	398	376	380	1528	1528	1528	1528	1528	1528
R-squared	0.098	0.101	0.105	0.078	0.030	0.029	0.121	0.124	0.131	0.206	0.210	0.211
Sample mean	3.092						3.251			2.858		
FX Civ. responds to hostility vs remains polite	-0.430											
p. diff 1	0.000											

Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

TABLE A4 Experiment R1 regression results with covariates

	Outcome Measure								
Vignette Characteristic	**Preferred Punishment (Covariates)**	**Preferred Punishment (Black Respondents)**	**Preferred Punishment (White Respondents - Incl. Hispanic**	**Fairness to Civilian of Assigned Punishment (Covariates)**	**Fairness to Civilian of Assigned Punishment (Black Respondents)**	**Fairness to Civilian of Assigned Punishment (White Respondents - Incl. Hispanic)**	**Overall Fairness to Civilian of Treatment (Covariates)**	**Overall Fairness to Civilian of Treatment (Black Respondents)**	**Overall Fairness to Civilian of Treatment (White Respondents - Incl. Hispanic)**
Officer Race (baseline=Black)									
Officer White (1=yes)	0.014	-0.027	-0.013	-0.092	-0.079	-0.115	-0.092	0.003	-0.035
	[0.050]	[0.131]	[0.059]	[0.053]*	[0.153]	[0.061]*	[0.060]	[0.153]	[0.072]
Civilian Race (baseline=Black)									
Civilian White (1=yes)	0.155	0.132	0.162	0.050	0.097	0.042	0.074	-0.131	0.130
	[0.050]***	[0.133]	[0.059]***	[0.053]	[0.148]	[0.061]	[0.060]	[0.148]	[0.072]*
Demeanor (baseline=Both Polite)									
Demeanor=Hostile Civilian, Polite Officer	0.516	0.602	0.550	0.278	0.580	0.274	0.028	0.195	0.064
	[0.069]***	[0.175]***	[0.080]***	[0.069]***	[0.195]***	[0.080]***	[0.065]	[0.172]	[0.077]
Demeanor=Both Hostile (Officer first)	0.404	0.750	0.346	-0.008	0.328	-0.025	-1.065	-0.767	-1.089
	[0.071]***	[0.173]***	[0.086]***	[0.074]	[0.207]	[0.089]	[0.084]***	[0.219]***	[0.103]***
Demeanor=Polite Civilian, Hostile Officer	-0.012	0.174	-0.061	-0.348	-0.129	-0.388	-1.254	-0.960	-1.304
	[0.068]	[0.203]	[0.077]	[0.080]***	[0.226]	[0.094]***	[0.087]***	[0.227]***	[0.104]***
Punishment (baseline=escorted)									
Officer gave ticket (1=yes, 0=no)	0.398	0.352	0.399	-0.627	-0.722	-0.649	-0.152	-0.349	-0.131
	[0.050]***	[0.132]***	[0.059]***	[0.051]***	[0.145]***	[0.061]***	[0.060]**	[0.149]**	[0.072]*
Age in years	0.004			0.007			0.006		
	[0.002]***			[0.002]***			[0.002]***		
Female (1=yes)	-0.061			-0.021			-0.012		
	[0.051]			[0.054]			[0.061]		
Non-binary (1=yes)	-0.270			-0.125			-0.038		
	[0.227]			[0.226]			[0.229]		
Education (1-4)	0.018			0.054			0.067		
	[0.030]			[0.031]*			[0.038]*		
Income (1-11, 12=refused)	0.000			0.004			0.008		
	[0.012]			[0.014]			[0.015]		
Income refused (1=yes)	-0.111			0.125			-0.051		
	[0.196]			[0.191]			[0.227]		
Ethnicity=Black (1=yes)	0.046			-0.078			0.149		
	[0.076]			[0.083]			[0.087]*		
Ethnicity=Asian (1=yes)	0.156			0.170			0.191		
	[0.108]			[0.112]			[0.136]		
Ethnicity=Other (1=yes)	0.028			-0.101			-0.206		
	[0.095]			[0.101]			[0.113]*		
Hispanic (1=yes)	0.057			0.096			0.248		
	[0.081]			[0.076]			[0.092]***		
Partisanship (1=Strong Dem.; 7=Strong Rep)	0.007			-0.028			-0.035		
	[0.021]			[0.021]			[0.023]		
Ideology (1=Extremely Lib.; 7=Extremely Consv.)	0.048			0.097			0.097		
	[0.022]**			[0.023]***			[0.025]***		
Constant	2.128	2.461	2.602	2.885	3.288	3.690	2.749	3.548	3.451
	[0.136]***	[0.161]***	[0.074]***	[0.148]***	[0.185]***	[0.075]***	[0.173]***	[0.167]***	[0.084]***
Observations	1529	245	1058	1528	245	1057	1528	245	1057
R-squared	0.116	0.117	0.108	0.157	0.128	0.142	0.230	0.166	0.224
Sample mean		3.086	3.092		3.118	3.280		2.947	2.846

Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

TABLE A5 Experiment 2 main regression results

	Outcome Measure							
Vignette Characteristic	**Severity of Expected Civilian Punishment (Prior to Learning Punishment)**	**Preferred Punishment**	**Perceived Threat to Victim by Civilian**	**Perceived Threat to Police by Civilian**	**Fairness to Civilian of Assigned Punishment**	**Fairness to Civilian of Assigned Punishment (Demeanor-Punishment Interactions)**	**Overall Fairness to Civilian of Treatment**	**Overall Fairness to Civilian of Treatment (Demeanor-Punishment Interactions)**
Officer Race (baseline=Black)								
Officer White (1=yes)	-0.016	-0.147	-0.014	-0.086	-0.076	-0.084	-0.191	-0.195
	[0.057]	[0.052]***	[0.046]	[0.050]*	[0.058]	[0.058]	[0.059]***	[0.059]***
Demeanor (baseline=Both Polite)								
Demeanor=Hostile Civilian, Polite Officer	0.265	0.307	0.159	0.263	0.243	0.280	0.126	0.206
	[0.079]***	[0.074]***	[0.066]**	[0.071]***	[0.081]***	[0.121]**	[0.085]	[0.124]*
Demeanor=Both Hostile (Officer first)	0.424	0.074	-0.033	-0.018	-0.155	-0.152	-0.363	-0.327
	[0.080]***	[0.072]	[0.065]	[0.072]	[0.082]*	[0.116]	[0.084]***	[0.120]***
Demeanor=Polite Civilian, Hostile Officer	0.475	-0.002	-0.164	-0.139	-0.117	-0.100	-0.395	-0.327
	[0.080]***	[0.072]	[0.064]**	[0.071]*	[0.082]	[0.119]	[0.083]***	[0.120]***
Weapon Reference: Unarmed								
Civilian has weapon (1=yes)	0.208	0.332	0.580	0.288	0.313	0.137	0.291	0.195
	[0.057]***	[0.051]***	[0.046]***	[0.050]***	[0.058]***	[0.083]*	[0.059]***	[0.083]**
Punishment (baseline=ticketed)		0.000			0.000	0.000	0.000	0.000
Arrested (1=yes)		0.206			-0.012	-0.163	-0.177	-0.186
		[0.052]***			[0.058]	[0.129]	[0.059]***	[0.134]
Hostile Civilian, Polite Officer X Arrested						-0.077		-0.155
						[0.163]		[0.170]
Both Hostile (Officer first) X Arrested						-0.016		-0.073
						[0.165]		[0.168]
Polite Civilian, Hostile Officer X Arrested						-0.030		-0.132
						[0.163]		[0.166]
Weapon X Arrested						0.353		0.194
						[0.116]***		[0.118]*
Constant	3.021	2.995	3.197	2.250	3.154	3.235	3.446	3.452
	[0.065]***	[0.068]***	[0.055]***	[0.062]***	[0.076]***	[0.100]***	[0.080]***	[0.105]***
Observations	2147	2147	2146	2145	2147	2147	2147	2147
R-squared	0.025	0.041	0.081	0.033	0.028	0.032	0.047	0.048
Sample mean	3.412	3.293	3.478	2.382	3.264		3.257	

Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

TABLE A6 Experiment 2 regression results with covariates

Vignette Characteristic	Outcome Measure Preferred Punishment (Covariates)	Perceived Threat to Victim by Civilian (Covariates)	Perceived Threat to Police by Civilian (Covariates)	Fairness to Civilian of Assigned Punishment (Covariates)	Overall Fairness to Civilian of interaction (Covariates)
Officer Race (baseline=Black)					
Officer White (1=yes)	-0.151	-0.016	-0.075	-0.073	-0.189
	[0.051]***	[0.046]	[0.049]	[0.058]	[0.059]***
Demeanor (baseline=Both Polite)					
Demeanor=Hostile Civilian, Polite Officer	0.310	0.169	0.257	0.251	0.132
	[0.074]***	[0.065]***	[0.071]***	[0.082]***	[0.085]
Demeanor=Both Hostile (Officer first)	0.062	-0.034	0.002	-0.150	-0.364
	[0.071]	[0.065]	[0.070]	[0.082]*	[0.083]***
Demeanor=Polite Civilian, Hostile Officer	-0.008	-0.158	-0.126	-0.114	-0.389
	[0.071]	[0.064]**	[0.070]*	[0.081]	[0.082]***
Weapon Reference: Unarmed					
Civilian has weapon (1=yes)	0.316	0.579	0.300	0.315	0.295
	[0.051]***	[0.045]***	[0.049]***	[0.058]***	[0.059]***
Punishment (baseline=ticketed)	0.000			0.000	0.000
Arrested (1=yes)	0.213			-0.007	-0.161
	[0.051]***			[0.058]	[0.059]***
Age in years	0.009	0.005	-0.009	0.003	0.004
	[0.002]***	[0.001]***	[0.001]***	[0.002]	[0.002]**
Female (1=yes, else=0)	0.053	-0.096	-0.237	-0.158	-0.135
	[0.055]	[0.049]*	[0.054]***	[0.063]**	[0.063]**
Education (1=HS or less, 4=Post BA)	-0.009	-0.010	-0.015	-0.003	0.008
	[0.028]	[0.025]	[0.027]	[0.032]	[0.032]
Race refused (1=yes)	-0.021	0.073	0.151	-0.139	0.119
	[0.202]	[0.218]	[0.171]	[0.233]	[0.245]
Race = Black/African American (1=yes)	-0.249	-0.143	-0.081	-0.098	-0.238
	[0.083]***	[0.076]*	[0.078]	[0.090]	[0.091]***
Race = Other (1=yes)	-0.055	0.001	0.051	0.014	0.025
	[0.088]	[0.077]	[0.084]	[0.093]	[0.095]
Hispanic (1=yes)	-0.097	-0.142	0.045	-0.031	-0.133
	[0.083]	[0.076]*	[0.077]	[0.086]	[0.088]
Household Income (1=<15K, 24=250K+, 25=RF)	-0.013	-0.002	0.008	0.002	0.000
	[0.004]***	[0.004]	[0.004]**	[0.005]	[0.005]
Income refused	0.354	0.196	-0.017	0.399	0.083
	[0.127]***	[0.114]*	[0.129]	[0.155]**	[0.170]
Constant	2.755	3.090	2.738	3.096	3.354
	[0.127]***	[0.109]***	[0.120]***	[0.140]***	[0.142]***
Observations	2143	2142	2141	2143	2143
R-squared	0.076	0.100	0.064	0.040	0.061
Sample mean					

Robust standard errors in brackets

* significant at 10%; ** significant at 5%; *** significant at 1%

A.2 Detailed Discussion of Treatment Manipulations on Expected Punishments

In experiments 1 and 2, we measured some outcomes before presenting the respondent with the punishment that the officer assigned (if any). We wanted to understand how these vignette features affected perceptions of the interaction but were also concerned that any treatment effects could operate through two distinct causal pathways. The first is simply that the vignette features affect perceptions of the interaction directly, an assumption compatible with modeling the outcomes as affected by the vignette manipulations and interpreting them as being caused by those manipulations. A second distinct potential causal pathway, however, and one that has gone unacknowledged in prior research, is that the manipulated features affect expectations about how the interaction will end (i.e., expected punishment) and that it is through changes in expected punishments that differences in outcomes arise. For example, people might perceive that a hostile officer is more likely to assign a more punitive punishment and might react negatively in expectation of that punishment rather than because of the officer's hostility per se.

In experiment 1, for the outcomes measured before the punishment outcome was revealed to the respondent, we included indicators for officer race (officer White, baseline category is officer Black) and indicators for each demeanor condition (baseline category is both always polite). In experiment 2, for the outcomes measured before the punishment was revealed to the respondent, we included indicators for officer race (officer Black, baseline category is officer White), indicators for each demeanor condition (baseline category is both always polite), and indicators for incident severity (armed with baseball bat, baseline category is unarmed).

In figure A1 panels A (experiment 1) and B (experiment 2) points indicate regression coefficients for each treatment condition with associated 95% confidence intervals, with full regression results listed above in tables A1 and A2. In the figure, the horizontal axis measured perceived punishment expectations compared to the baseline condition in which the officer is Black and both the civilian and officer remain polite throughout the interaction (panel A) and the baseline condition in which the officer is Black, both the civilian and officer remain polite throughout the interaction, and the civilian is unarmed (panel B).

In experiment 1, the mean level of expected punishment in the officer Black and both polite condition is 1.03 on the 4-point scale, which corresponds to an average expected punishment in this condition of "civilian was escorted from the park." In experiment 2, the mean level of expected punishment in the officer White, both polite, and unarmed condition is 2.02 on the 5-point scale, which corresponds to an average expected punishment in this condition of "officer released the civilian from handcuffs, escorted him from the baseball field, and gave him a verbal warning."

Panels A (experiment 1) and B (experiment 2) show that compared to an interaction involving a Black officer, there is no effect of the officer being White on expected punishment ($b= 0.04$, $se=0.06$, $p>0.10$ for experiment 1; $b=-0.02$, $se=0.06$, $p>0.10$ for experiment 2). By contrast, officer and civilian demeanor have large and significant effects on punishment expectations in both experiments. Panel A demonstrates that regardless of an officer's conduct, when a civilian is hostile, respondents predict the civilian's punishment will be more severe, by roughly similar amounts. Compared to both individuals being polite, a hostile civilian's punishment is expected to be 1.7 units (officer polite), 1.6 units (officer hostile after civilian), and 1.7 units (officer

hostile first) greater. These differences are roughly similar in magnitude. Notably, if the civilian remains polite but the officer is hostile, respondents still expect a more severe punishment, but the effect is much smaller than the expected effect of both being hostile (b=0.98, se=0.05, p<0.001; difference with all other treatments significant at p<0.01).

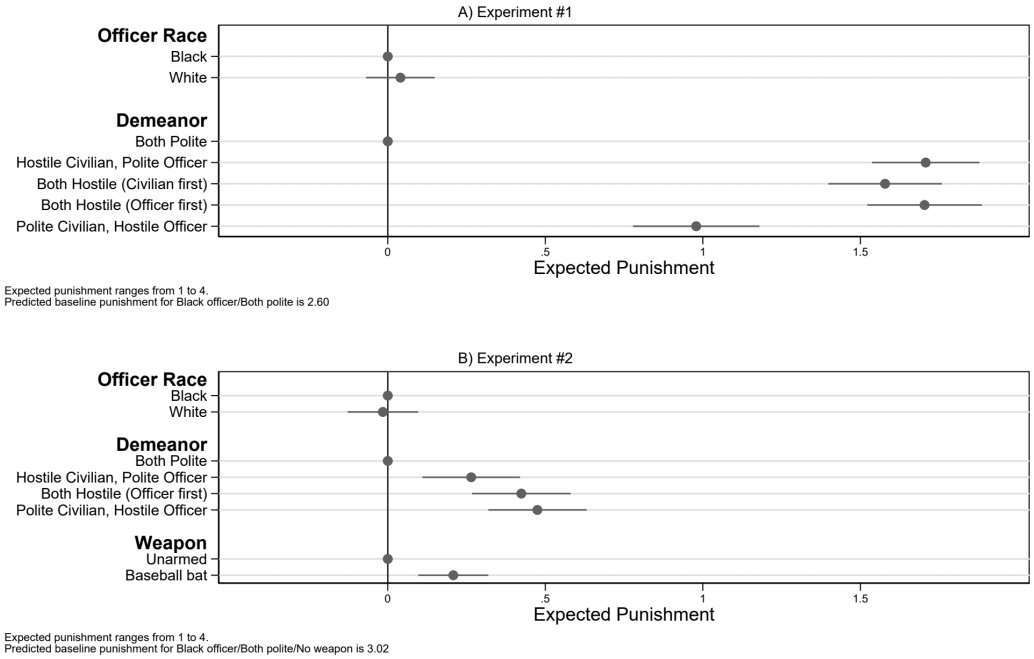


FIGURE A1 The effect of vignette characteristics on expected punishment

In experiment 2, which includes a weapon treatment, we also find that an armed suspect is more likely than an unarmed suspect to be punished, all else equal. The key implication of this result is that when the punishment an officer will assign is not specified, respondents are nonetheless forming beliefs about it based on both demeanor and whether a weapon is present. Thus, when punishment is left unspecified, it is very difficult to interpret any treatment effect estimates because multiple causal pathways may explain these effects. For this reason, the remainder of our analysis focuses on outcomes gathered *after* all treatment manipulations were provided to the respondent, including how the interaction concluded and the specific punishment that the officer assigned. Once a specific punishment has been assigned, expectations are fixed across respondents about the material outcome of the interaction, allowing us to make clearer theoretical inferences about other treatment effects.

It is useful to examine whether there are differences in punishment expectations for cases with hostility started by civilians versus cases involving officers who started the hostility. Punishment expectations are dependent on the person who starts the hostile interaction, but the differences are only significant at the .10 level (b=-0.12, se=0.06, p<0.10). We find that incidents involving

civilians who initiated hostility are expected to receive less punishment than ones involving officer-initiated hostility, suggesting that officer demeanor is expected to affect punishment. Punishment expectations also vary by a civilian's reaction (polite, hostile) to a hostile officer. Specifically, the findings from experiment 1 demonstrate that incidents involving civilians who become hostile toward an already hostile officer are expected to be given significantly harsher punishment than incidents involving civilians who stay polite in interacting with a hostile officer ($b=1.70$, $se=0.09$, $p<0.01$ and $b=0.98$, $se=0.10$, $p<0.01$, respectively).

The above results reveal that the concern that the vignette features we manipulated affected expected punishment is well-founded. While officer race does not significantly influence expected punishment in either experiment, we find large effects of both officer and civilian demeanor on expected punishment. When either an officer or a civilian is hostile, respondents expect more punitive punishments.

In experiment 2, officer and civilian demeanor and incident severity have effects on expected punishment. The results mirror the results from the first experiment, showing that regardless of an officer's demeanor, when a civilian is hostile, respondents predict the civilian's punishment will be more severe, by roughly similar amounts. Compared to both individuals being polite, a hostile civilian's punishment is expected to be significantly greater when the officer is polite ($b=0.27$, $se=0.08$, $p<0.05$) and the officer is hostile ($b=0.42$, $se=0.08$, $p<0.01$). If the civilian remains polite but the officer is hostile, respondents also expect a more severe punishment ($b=0.47$, $se=0.08$, $p<0.01$).

Experiment 2 also tests the effect of incident severity on punishment expectations. The results show that civilians who are armed (as opposed to unarmed) are expected to receive significantly harsher punishments ($b=0.21$, $se=0.06$, $p<0.01$), suggesting that civilians expect more severe incidents to be met with more severe punishments.

Together, these results have important implications. Specifically, they show that while officer race on its own does not significantly influence punishment expectations, differences in demeanor and incident severity influence expected punishment. In particular, people believe that hostile civilians are more likely to be punished regardless of an officer's demeanor, although we do not ask whether these expected punishments are fair. Further, people believe that armed civilians are more likely than unarmed civilians to be punished.

A.3 Detailed Discussion of Interaction Effects on Preferred Punishment

In the main text, most of our analysis of preferred punishment considers the effect of treatment manipulations in isolation. We also tested whether there are theoretically interesting interactions among assigned treatments to understand whether more saturated models are appropriate. We began by interacting officer race with demeanor (and weapon status in experiment 2) to determine whether they affect preferred punishment in combination. The block F-tests of the joint significance of the officer race x demeanor interactions were statistically insignificant for experiments 1 (Likelihood ratio chi^2= 4.17, p=0.38) and R1 (Likelihood ratio chi^2= 3.89, p=0.27). The block F-tests of the joint significance of the officer race x demeanor and officer race x weapon interactions were also statistically insignificant for experiment 2 (Likelihood ratio chi^2= 1.26, p=0.87). Individuals therefore do not appear to react differently to officer and civilian demeanor or weapon status depending on officer race.

Experiment R1 also allowed us to interact civilian race with demeanor to determine whether they work together to shape punishment preferences. To determine whether individuals react differently to officer and civilian demeanor depending on whether a civilian is Black or White, we re-estimated the experiment R1 specification with interactions for demeanor x civilian race. The block F-test of the joint significance of the demeanor x civilian race interactions was statistically insignificant (Likelihood ratio chi^2= 0.96, p=0.81), suggesting that the public does not react more or less punitively to officer and civilian demeanor depending on the race of the civilian involved in the interaction.

For the second experiment, we also tested whether demeanor and weapon status interact to shape preferred punishment. It could be that demeanor matters more when a suspect is armed because the impact of being threatening is larger, or that it matters less because a weapon alone is sufficient to warrant greater punishment. To test for these possibilities, we re-estimated the experiment 2 specification with interactions for demeanor x weapon. The block F-test of the joint significance of the demeanor x weapon interactions was statistically insignificant (Likelihood ratio chi^2= 4.62, p=0.20), although the point estimates imply that the effect of a citizen being hostile when an officer is polite drops by about half when a weapon is present (from 0.40 to 0.21, difference not significant), consistent with the notion that a weapon or hostile civilian behavior act as partial substitutes in generating preferences for greater punishment.

A.4 Discussion of Interaction Effects on Threat

We also examined the interaction between demeanor and weapon status. This allowed us to understand if the demeanor of participants can reduce or amplify the threat posed by an armed civilian. To test for this possibility, we interacted demeanor with the weapons indicator in a more saturated model. (We also tested whether officer race moderated the effect of demeanor or weapon status and found no evidence to support either set of interactions. Both sets of analysis are available upon request.)

Including these interactions of incident severity with demeanor significantly improves fit in models predicting civilian threat to the victim (Likelihood ratio $chi2=8.78$, $p<0.05$), but not civilian threat to the officer (Likelihood ratio $chi2=5.15$, $p>0.05$). While one set of results is significant and the other is not, the pattern is the same across both outcomes. First, being armed always increases perceived threat, so an armed civilian who is polite when confronting a hostile officer does not neutralize the perceived threat associated with being armed. (We do find that the effect of civilian hostility when the officer is polite on threat to the other civilian is smaller when the civilian is armed (Hostile Citizen, Polite Officer x Weapon is negative and statistically significant). Thus, being armed or hostile increases perceived threat, but the cumulative effect is diminishing relative to the component effects in isolation.) Second, when the civilian is unarmed, they are only perceived as more threatening when they are hostile and the officer is polite (compared to both being polite). The remainder of the demeanor conditions (when unarmed) are statistically insignificant. Overall, this shows that a civilian's hostile demeanor can substitute for a weapon as a source of threat, but not the opposite: A positive demeanor in the face of officer hostility does not remove the threat posed by being armed.

A.5 Sample Descriptives

Table A7 shows information about experiments 1 and 2, and Table 8 shows information about experiment R1.

TABLE A7 Sample descriptive statistics for experiments 1 and 2

Variable (level)	Experiment 1 Sample Means	Experiment 2 Sample Means
Female:		
Yes	0.521	0.524
No	0.479	0.477
Region:		
South	0.417	0.399
West	0.189	0.239
Midwest	0.216	0.191
Northeast	0.178	0.171
Race/Ethnicity:		
AAPI	0.041	0.033
Black	0.118	0.116
White	0.663	0.656
Hispanic	0.134	0.162
Other race	0.044	0.034
Household income:		
$24,999 or less	0.302	0.288
$25,000 - $44,999	0.201	0.184
$45,000 - $64,999	0.150	0.128
$65,000 - $84,999	0.097	0.089
$85,000 - $124,999	0.100	0.158
$125,000 +	0.098	0.114
Unanswered	0.052	0.40
Education:		
High school diploma or less	0.302	0.273
Some college/ Associate's degree	0.346	0.322
Bachelor's degree	0.227	0.216
Graduate degree	0.126	0.190
Partisanship:		
Democrat	0.458	0.487
Independent	0.162	0.150
Republican	0.380	0.363
Age:		
Age	46.4	46.8

TABLE A8 Sample descriptive statistics for experiment R1

	Variable (level)	Experiment R1 Sample Means (Standard Deviations)
Gender:		
	Female	0.502
	Male	0.498
	Non-binary	0.021
Race:		
	AAPI	0.063
	Black	0.161
	Other race	0.110
Ethnicity:		
	Hispanic	0.177
Household income:		
	<$25,00	0.214
	$25,000 - $49,999	0.276
	$50,000 - $74,999	0.186
	$75,000 - $99,999	0.130
	$100,000 - $124,999	0.070
	$125,000 +	0.098
	Unanswered	0.026
Education:		
	High school diploma or less	0.233
	Some college/ Associate's degree	0.450
	Bachelor's degree	0.213
	Graduate degree	0.105
Partisanship:		
	1=Strong Democrat.; 7=Strong Republican	4.00(1.92)
Ideology		4.15(1.80)
	1=Extremely Liberal; 7=Extremely Conservative	
Age:		
	Age	47.8(16.84)

A.6 Ordered Probit Model Results

TABLE A9 Ordered probit model results

	Experiment 1	**Experiment R1**	**Experiment 2**
	Preferred Punishment	**Preferred Punishment**	**Preferred Punishment**
Officer Race (baseline=Black)	[0.000]	[0.000]	[0.000]
Officer White (1=yes)	-0.018	0.030	-0.133
	[0.046]	[0.054]	[0.046]***
Civilian Race (baseline=Black)		[0.000]	
Civilian White (1=yes)		0.180	
		[0.055]***	
Demeanor (baseline=Both Polite)			
Demeanor=Hostile Civilian, Polite Officer	0.665	0.606	0.276
	[0.074]***	[0.076]***	[0.067]***
Demeanor=Both Hostile (Civilian first)	0.526		
	[0.073]***		
Demeanor=Both Hostile (Officer first)	0.223	0.479	0.066
	[0.069]***	[0.077]***	[0.064]
Demeanor=Polite Civilian, Hostile Officer	0.012	-0.020	-0.001
	[0.064]	[0.071]	[0.064]
Weapon Reference: Unarmed			[0.000]
Civilian has weapon (1=yes)			0.297
			[0.046]***
Punishment (baseline=escorted)			
Ticketed (1=yes)	0.337	0.486	
	[0.048]***	[0.058]***	
Punishment (baseline=ticketed)			
Arrested (1=yes)			0.189
			[0.046]***
Observations	2012	1529	2147

Ordered Probit Coefficients with robust standard errors in brackets
* significant at 10%; ** significant at 5%; *** significant at 1%