

ARTICLE

DOI: 10.1057/s41599-018-0110-z

OPEN

Resolution of apparent paradoxes in the race-specific frequency of use-of-force by police

Cody T. Ross¹, Bruce Winterhalder² & Richard McElreath¹

ABSTRACT Analyses of racial disparities in police use-of-force against unarmed individuals are central to public policy interventions; however, recent studies have come to apparently paradoxical findings concerning the existence and form of such disparities. Although anti-black racial disparities in U.S. police shootings have been consistently documented at the population level, new work has suggested that racial disparities in encounter-conditional use of lethal force by police are reversed relative to expectations, with police being more likely to: (1) shoot white relative to black individuals, and (2) use non-lethal as opposed to lethal force on black relative to white individuals. Encounter- and use-of-force-conditional results, however, can be misleading if the rates with which police encounter and use non-lethal force vary across officers and depend on suspect race. We find that all currently described empirical patterns in the structuring of police use-of-force—including the “reversed” racial disparities in encounter-conditional use of lethal force—are explainable under a generative model in which there are consistent and systemic biases against black individuals. If even a small subset of police more frequently encounter and use non-lethal force against black individuals than white individuals, then analyses of pooled encounter-conditional data can fail to correctly detect racial disparities in the use of lethal force. In more technical terms, statistical assessments of racial disparities conditioned on problematic intermediate variables, such as encounters, which might themselves be a causal outcome of racial bias, can produce misleading inferences. Population-level measures of use-of-force by police are more robust indicators of the overall severity of racial disparities than encounter-conditional measures—since the later neglect the differential morbidity and mortality arising from differential encounter rates. As such, population-level measures should be used when evaluating the local-level public health implications of racial disparities in police use-of-force. Research on encounter-conditional use-of-force by police can also fruitfully contribute to public policy discussions, since population-level measures alone cannot address whether racial disparities are driven by disparities in encounters or disparities in use-of-force conditional on encounters. Tests for racial biases in the encounter-conditional use of lethal force, however, must account for individual-level variation across officers in terms of race-specific encounter rates or risk falling to Simpson’s paradox.

¹Department of Human Behavior, Ecology and Culture. Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. ²Department of Anthropology, University of California, Davis, USA. Correspondence and requests for materials should be addressed to C.T.R. (email: cody_ross@eva.mpg.de)

Introduction

Two recent publications on anti-black, racial disparities in police shootings in the United States Ross (2015) and Fryer (2016) have received media attention (e.g., Cox, 2016; Li, 2016; DeVega, 2016), in part because they appear to reach opposite conclusions concerning police behavior. Ross (2015) shows that *at the population-level* individuals are about 3.5 times more likely to be black, unarmed, and shot-by-police, than white, unarmed, and shot-by-police, adjusting for relative differences in population size. Fryer (2016), however, finds that *conditional on being encountered by police*, black civilians relative to white civilians are *less* likely to be shot by police. Such apparently conflicting results have generated confusion in the public presentation of these two studies. The key difference between them, however, is that Ross (2015) looks at population-level relative risk, whereas Fryer (2016) looks at relative risk conditional on encounters or use-of-force. These studies thus tell us different things; failure to detect racial disparities in police shootings conditional on encounters does not imply absence of racial disparities in police shootings at the population level. Even in light of the well-acknowledged shortcomings in the data used in each of these studies, each exposes complementary elements of a coherent picture of the dynamics of police use-of-force.

In order to gain a more fine-grained understanding of why the encounter-conditional results of Fryer (2016) seem to run so contrary to what might be expected given the population-level findings of Ross (2015), we derive a generative model of police use-of-force outcomes, and then analyze the conditions under which this model can generate data consistent with both analyses. We use both analytic and simulation methods to draw attention to some counter-intuitive properties of this model. Specifically, we draw a parallel to classic findings on Simpson's Paradox in applied statistics (Simpson, 1951; Bickel et al., 1975; Pearl, 2014), and demonstrate that pooled analyses of encounter-conditional data as in Fryer (2016) will fail to find true encounter-conditional anti-black racial disparities over a wide range of parameter values when there is heterogeneity in the rates with which police encounter and use non-lethal force as a function of suspect race. For example, if even a small subset of police have propensities to more frequently encounter black relative to white individuals, then analyses of pooled encounter-conditional data will fail to detect systemic anti-black racial disparities in the encounter-conditional use of lethal force by the larger subset of police. Likewise, if even a small subset of police are more likely to non-lethally assault black individuals than white individuals e.g., with tasers in contexts when such force is not actually justified, and then report that such force was justified, analyses of pooled use-of-force-conditional data will suggest that lethal force is more likely to be used against white relative to black individuals.

In the analysis that follows, we demonstrate three important conclusions: (1) Ross (2015) and Fryer (2016) document quantitatively similar population-level anti-black racial disparities in the use of lethal force by police, (2) analyses of encounter-conditional rates of the use of lethal and non-lethal force by police are insufficient to address predictions about racial disparities in policing, since they neglect the relative risk of being encountered by police in the production equation of population-level racial disparities; and, most critically, (3) the findings of Fryer (2016) suggesting null or anti-white disparities in the encounter-conditional rates of the use of lethal force by police are actually consistent with a situation in which all police have elevated encounter-conditional rates of the use of lethal force against black individuals, but a small subset of police encounter and assault black individuals sub-lethally at elevated rates. In other words, apparent anti-white racial disparities in encounter-

Table 1 Population-level police shooting data (as proportions by race/ethnicity) from Ross (2015) and Fryer (2016, p. 14)

Source	Armed status	Black	Hispanic	White
US Census (2010)	–	0.13	0.29	0.58
Fryer (2016)	All shootings	0.46	0.30	0.24
Ross (2015)	All shootings	0.38	0.38	0.24
Ross (2015)	Unarmed individuals	0.36	0.38	0.26

To increase comparability to the Fryer (2016) study, census data and the police shootings data from Ross (2015) have been taken only from counties where Fryer (2016) also had data. These counties are: Dallas, TX; Harris, TX; Travis, TX; Suffolk, MA; Camden, NJ; Pinellas, FL; Lee, FL; Orange, FL; Palm Beach, FL; Duval, FL; Brevard, FL; and Los Angeles, CA. The population data show that in these counties, about 13% of residents are black, 29% of residents are of Hispanic origin, and 58% of residents are white. When considering the proportion of police shooting victims by race, the data from Fryer (2016) and Ross (2015) are very similar; both studies find that black individuals are killed by police in greater proportions than would be expected given their race-specific population levels

conditional rates of the use of lethal force by police may arise not from bias against white individuals, but rather from elevated rates of unjustifiable encounters with black individuals. In order to disentangle which of these possibilities best explains the empirical data, future research will need to better account for variation across officers in terms of: (1) their race-specific, exposure-time conditional encounter rates with civilians, (2) their race-specific, exposure-time conditional rates of use-of-force, and (3) their race-specific probabilities of use-of-force conditional on encounter.

Population-level analyses. In Table 1 we present comparable, population-level data from Ross (2015) and Fryer (2016); the outcomes are very similar. Both studies find that black individuals are killed by police in greater proportions than would be expected given their relative population size. In fact, the data mobilized by Fryer (2016) show greater evidence of racial disparities in police shootings at the population level: $\frac{46\%}{13\%} = 3.5x$, than do the data presented by Ross (2015) in comparable counties: $\frac{38\%}{13\%} = 2.9x$. See also a similar comparison including data from The Guardian, The Washington Post, and VICE (Fryer, 2017). Hispanic individuals are shot in direct proportion to their population levels in the data provided by Fryer (2016): $\frac{30\%}{29\%} = 1.0x$, and at rates slightly above their population levels in the data provided by Ross (2015): $\frac{38\%}{29\%} = 1.3x$. Finally, both studies agree numerically that white individuals are shot at a rate less than that which would be expected given their population levels: $\frac{24\%}{58\%} = 0.4x$, despite the fact the authors use different data collection methodologies over different time intervals. As such, the analyses of Ross (2015) and Fryer (2016) are in general agreement concerning the existence and magnitude of population-level racial disparities in police shootings.

The Ross (2015) analysis is resolved to the level of county, and shows significant geographical diversity in the levels of racial disparities in police shootings variation that has the potential to advance explanations of the locally contextualized causes of police shootings and to identify the locales that may benefit most from policy review. However, Ross (2015) has no data on race-specific encounter rates between police and civilians and therefore cannot address the race-specific probabilities of being shot by police conditional on being encountered by police. Fryer (2016), by contrast, has compiled a much more detailed dataset for a more limited sample of places and departments; as such, the data provided by Fryer (2016) allow for a more thorough micro-level analysis of the generative pathways of racial disparities in police use-of-force in specific locations.

The data in Ross (2015) could be biased by selective reporting, for instance were individual engaged in the crowd-sourcing effort to have preferentially reported police shootings of unarmed black civilians, rather than following the random sampling methodology procedure requested by Wagner (2014). Likewise, the data in Fryer (2016) could come from a biased sample of departments. Nevertheless, both studies show remarkably similar findings when comparable, population-level measures are used. This initial comparison serves to cross-validate, to some extent, both data sources for studying population-level disparities in police use-of-force.

Encounter-conditional analyses. Given the general agreement of the population-level results between these two studies, at first pass the encounter-conditional results of Fryer (2016) are striking. Drawing on his micro-level data, Fryer (2016) finds consistent, strong, and significant levels of anti-black racial disparities in police use of *non-lethal* force in graded categories from hands on, to handcuffing, display of a weapon, use of pepper spray, and use of a baton. These racial disparities remain unexplained even after accounting for 112 (Fryer, 2016, p. 17) context-dependent covariates. But, in a result he describes as a “stark contrast to non-lethal uses of force” (p. 5), he finds that use of *lethal* force by police actually shows reversed, anti-white, racial disparities; black individuals are less likely than white individuals to be shot by police, conditional on being encountered by police.

This result has been emphasized in media accounts and is surprising in light of general findings on racial disparities in police use of both lethal and non-lethal force spanning several decades and literatures i.e., research on implicit psychological biases (Plant et al., 2005; Correll et al., 2006, 2007), sociological drivers of racial disparities (Jacobs, 1998; Smith, 2003; Bergesen, 1982; Gilbert and Ray, 2016), structural disparities established by the existing social order (Harring et al., 1977; Jacobs and Britt, 1979; Holmes, 2000), proximate responses by police to areas of high crime and risk (Jacobs, 1998; Fyfe, 1980), racial bias in profiling and encountering individuals (Warren et al., 2006; Tomaskovic Devey et al., 2004; Gelman et al., 2007), blatant racism (Goldkamp, 1976; Martinot and Sexton, 2003; Doane, 2006), and social dominance orientation (Sidanius and Pratto, 2001), over many geographic areas: United States (Ross, 2015; Hirschfield, 2015), Canada (Wortley and Tanner, 2003; Wortley and Owusu Bempah, 2012), Brazil (Cano, 2010; French, 2013), and South Africa (Brogden and Shearing, 2005). Nonetheless, the widely publicized interpretation of Fryer (2016) one that he himself highlights is that police do not show anti-black racial bias in the use of lethal force. While we do not doubt that this pattern is apparent in the analyzed data, we take issue with interpretation, given that other aspects of Fryer’s analysis suggest that Simpson’s paradox could be driving this effect (Simpson, 1951; Pearl, 2014; Bickel et al., 1975, and see Discussion). In the following sections, we show that the encounter-contingent results of Fryer (2016) are consistent perhaps counter-intuitively so with systemic and consistent anti-black racial disparities in police encounter rates and in the subsequent rates of the use of both lethal and non-lethal force upon encounter.

A generative model of racial disparities in use-of-force by police

We resolve the counter-intuitive results by using a generative model to demonstrate that increased racial disparity in the rates at which a subset of police engage civilians with non-lethal force will generate all patterns found in the population-level results of Ross (2015) and the encounter-contingent results of Fryer (2016), including:

1. Elevated population-level rates of police encounters with black individuals relative to white individuals see Fryer (2016, Table 1A, Panel A) and Fryer (2016, Table 1D, Panel A).
2. Elevated population-level rates of lethal force against black individuals relative to white individuals see Ross (2015) and Fryer (2016, Table 1C, Panel A).
3. Elevated encounter-conditional rates of non-lethal force against black individuals relative to white individuals see Fryer (2016, Table 2A) and Fryer (2016, Table 2B).
And the so-called paradoxes in otherwise consistent patterns:
4. Elevated encounter-conditional rates of lethal force against white individuals relative to black individuals see Fryer (2016, Table 5).
5. Elevated probability ratios of the use of non-lethal force to lethal force against black individuals relative to the same ratio against white individuals see Fryer (2016, Table 5).

Informal model description. Informally, the model we use entails a homogeneous police force, and sub-populations of black and white citizens. The model makes explicit assumptions about police behavior concerning race-specific rates of police encounters with citizens, and the probabilities of the use of no force, non-lethal force, or lethal force by police conditional upon encounter. The model provides a demonstration of the manner in which an apparently paradoxical result (the simultaneous truth of all five of the observations listed above) is fully consistent with police use-of-force being systematically biased against black individuals.

Formal model description. Formally, our model assumes that we have a population of N individuals, with black and white sub-populations N_B and N_W , where: $N = N_B + N_W$. Note that we will continue to use subscripts of B and W to refer to black and white sub-populations, and that all parameters are defined in Table 2 for ease of reference. We assume that over some interval of time, E_B and E_W individuals are encountered by police, at rates: $\phi_B \in [0, 1]$ and $\phi_W \in [0, 1]$. As such, we expect the counts of encounters to follow:

$$E_B \sim \text{Binomial}(N_B, \phi_B) \quad (1)$$

$$E_W \sim \text{Binomial}(N_W, \phi_W) \quad (2)$$

We then assume that conditional upon encountering a civilian, police engage in the use of either no force, non-lethal force, or lethal force. Let the parameters: $\theta_B \in [0, 1]$ and $\theta_W \in [0, 1]$ define the probability of the use of lethal force conditional upon encounter, and the parameters: $\gamma_B \in [0, 1 - \theta_B]$ and $\gamma_W \in [0, 1 - \theta_W]$ define the probability of the use of non-lethal force conditional upon encounter. The probability of police using no force on white and black individuals upon encounter is then given by: $1 - (\theta_W + \gamma_W)$ and $1 - (\theta_B + \gamma_B)$, respectively.

Then, to define the counts of lethal force by police, S_B and S_W , and non-lethal force by police, G_B and G_W , we write:

$$(S_B, G_B, O_B)' \sim \text{Multinomial}(E_B, (\theta_B, \gamma_B, 1 - (\theta_B + \gamma_B)))' \quad (3)$$

$$(S_W, G_W, O_W)' \sim \text{Multinomial}(E_W, (\theta_W, \gamma_W, 1 - (\theta_W + \gamma_W)))' \quad (4)$$

where: $O_B = E_B - (S_B + G_B)$ and $O_W = E_W - (S_W + G_W)$, meaning that the outcome vectors contain the predicted counts of lethal

Table 2 Model variables and relationships. Subscripts *B* and *W* refer to black and white sub-populations, respectively

Variable	Definition	Formula(s)	Domain
<i>N</i>	Population under consideration	$N = N_W + N_B$	$\in \mathbb{N}^+$
<i>E</i>	Count of individuals subjected to a police encounter	Eqs. (1) and (2), text	$\in \{1, \dots, N\}$
<i>S</i>	Count of lethal police actions	Eqs. (3) and (4), text	$\in \{1, \dots, E\}$
<i>G</i>	Count of non-lethal police actions	Eqs. (3) and (4), text	$\in \{1, \dots, E\}$
ϕ	Probability of encounter with police		$\in [0, 1]$
θ	Probability of receiving lethal force, conditional on encounter		$\in [0, 1]$
γ	Probability of receiving non-lethal force, conditional on encounter		$\in [0, 1 - \theta]$
ψ	Population-level rate of lethal police actions	$\psi = \frac{S}{N}$	$\in [0, 1]$
δ	Population-level rate of non-lethal police actions	$\delta = \frac{G}{N}$	$\in [0, 1]$
Φ	Relative population-level risk of being subjected to a police encounter	$\Phi = \frac{\phi_B}{\phi_W}$	$\in [0, \infty)$
Θ	Relative risk of being the victim of a lethal police action conditional upon encounter	$\Theta = \frac{\theta_B}{\theta_W}$	$\in [0, \infty)$
Γ	Relative risk by being of being the victim of a non-lethal police action conditional upon encounter	$\Gamma = \frac{\gamma_B}{\gamma_W}$	$\in [0, \infty)$
Ψ	Relative population-level risk of being the victim of a lethal police action	$\Psi = \frac{\psi_B}{\psi_W}$	$\in [0, \infty)$
Δ	Relative population-level risk of being the victim of a non-lethal police action	$\Delta = \frac{\delta_B}{\delta_W}$	$\in [0, \infty)$

The “hat” symbol used in the text indicates parameter values unique to the *routine over-escalation* sub-model; the parameters without the “hat” symbol refer to the *standard policing* model. We note that the term “non-lethal” includes sub-lethal violent engagements, but excludes non-violent engagements

Table 3 Conditions for the presence or absence of racial disparities and their direction if present

Form of bias	Absent	Present	
		Against black individuals	Against white individuals
Encounter	$\phi_B = \phi_W \Leftrightarrow \Phi = 1$	$\phi_B > \phi_W \Leftrightarrow \Phi > 1$	$\phi_B < \phi_W \Leftrightarrow \Phi < 1$
Population-level, non-lethal force	$\delta_B = \delta_W \Leftrightarrow \Delta = 1$	$\delta_B > \delta_W \Leftrightarrow \Delta > 1$	$\delta_B < \delta_W \Leftrightarrow \Delta < 1$
Population-level, lethal force	$\psi_B = \psi_W \Leftrightarrow \Psi = 1$	$\psi_B > \psi_W \Leftrightarrow \Psi > 1$	$\psi_B < \psi_W \Leftrightarrow \Psi < 1$
Encounter-conditional, non-lethal force	$\gamma_B = \gamma_W \Leftrightarrow \Gamma = 1$	$\gamma_B > \gamma_W \Leftrightarrow \Gamma > 1$	$\gamma_B < \gamma_W \Leftrightarrow \Gamma < 1$
Encounter-conditional, lethal force	$\theta_B = \theta_W \Leftrightarrow \Theta = 1$	$\theta_B > \theta_W \Leftrightarrow \Theta > 1$	$\theta_B < \theta_W \Leftrightarrow \Theta < 1$

Variables are defined in Table 2; the double arrow (\Leftrightarrow) is read as “is equivalent to”

force, non-lethal force, and no force, given the number of encounters and the probability parameters for each use-of-force category.

Finally, we can write the total population-level rates of the use of lethal force as: $\psi_B = S_B/N_B$ and $\psi_W = S_W/N_W$, and the total population-level rates of non-lethal force as: $\delta_B = G_B/N_B$ and $\delta_W = G_W/N_W$. The population-level relative risk of being the victim of a lethal police action, Ψ , is then given as: $\Psi = \psi_B/\psi_W$. The population-level relative risk of being the victim of a violent but non-lethal police action, Δ , is given as: $\Delta = \delta_B/\delta_W$. The relative risk of being encountered by police, Φ , is defined as: $\Phi = \phi_B/\phi_W$. The relative risk of being the victim of a lethal police action conditional on being encountered by police, Θ , is given as: $\Theta = \theta_B/\theta_W$. The relative risk of being the victim of a violent but non-lethal police action conditional on being encountered by police, Γ , is given as: $\Gamma = \gamma_B/\gamma_W$.

Assessing sources of disparities. We define racial disparities as the disproportionate representation of a racial sub-group in a particular type of police action category. That is, we seek here to avoid any prejudicial connotations of the term, as we do not know, or need to know for the narrow purposes of our analysis, if such disparities are justifiable in some manner or not. We recognize five ways in which racial disparities may occur or be documented as absent: in encounters, and in the population-level and encounter-conditional use of non-lethal and lethal force (see Table 3).

For example, we would consider disparity in encounters to be absent if race-specific rates of encounters are equal: $\phi_B = \phi_W$. This is the same as saying that the relative risk of encounter by race is equal to one: $\Phi = 1$, or that the log of this value is equal to zero: $\log(\Phi) = 0$. We would consider encounters to be structured against black individuals if: $\phi_B > \phi_W$, or equivalently if: $\Phi > 1$, or finally, if: $\log(\Phi) > 0$. Reversing the inequalities would indicate disparities against white individuals. The remaining four types of disparities are described in the same fashion.

The standard policing model. We name the full generative model described above the *standard policing* model, and take it to define the universe of standard police interactions, although there may be either unjustifiable (implicit or explicit bias) or justifiable (a response to crime) racial disparities in the rates of encounters or use-of-force in this model.

At this point, we note a key dependency in the model. The overall counts of lethal and non-lethal force depend on the rates of encounters given by ϕ_W and ϕ_B , via the counts E_W and E_B . This gives us our first insight on the apparent paradox that population-level analyses of U.S. data show anti-black racial disparities in police shootings, while encounter-contingent analyses show reversed, anti-white disparities. A difference in race-specific encounter rates could lead to population-level racial disparities in police shootings, even in the absence of racial disparities in police shootings conditional on encounter.

More formally, the log of the expected population-level extent of racial disparities in police shootings is approximately:

$$\log(\mathbb{E}[\Psi]) \approx \log(\Theta) + \log(\Phi) \quad (5)$$

for N_W sufficiently large and $\theta_W \phi_W$ not too small (see [Supplementary Materials](#) for derivation). Even if police are more likely to use lethal force against white people than black people conditional on encounter (i.e., $\log(\Theta) < 0$, which might be expected if police preferentially encounter white individuals engaging in crimes, but encounter black individuals more frequently or haphazardly), then we will still see population-level anti-black racial disparities in police shootings on average so long as the relative risk of encounters (i.e., $\log(\Phi)$) is large enough to compensate. This might also occur if the threshold of suspicion leading to an encounter is lower for black individuals than for white individuals. Note also that the final population-level metric is the more appropriate indicator of the public health issue at hand; regardless of whether differential rates of encounters or differential use-of-force conditional on encounters drives differential morbidity and mortality rates, the end result is the same greater overall morbidity and mortality from police per unit population. This being said, the results of Fryer (2016) should they replicate under individual-level analyses that control for the possible dynamics discussed in this paper provide the key insight that reduction in population-level racial disparities in police use-of-force might depend more on attenuation of racial disparities in encounter rates than on changes in how officers respond to suspects conditional on their race.

The routine over-escalation model. The above result suggests that sub-structure in the behavioral patterns of police which increases their rate of non-lethal encounters with black individuals will have the effect of diminishing the apparent rate of lethal outcomes conditional on encounters, while leaving the population-level rates of racial disparities in police shootings unchanged. To formally explore this idea, we introduce a second generative model below.

Assume that there is heterogeneity in the types of encounters with police e.g., i) those resulting from police operating under standard protocols and ii) those resulting from police engaging in more routine and excessive use of non-lethal force. If this is the case, then the probability that white individuals will be more likely than black individuals to receive lethal force conditional on encounter grows as the small subset of police that engage in excessive use of non-lethal force target black people at increasing rates. This effect holds even when all parameter values (i.e., encounter rates, and rates of the use of lethal and non-lethal force conditional on encounter) are structured against black individuals in both subsets of police officers.

We demonstrate this result by defining an additional sub-model, the *routine over-escalation* model. In this model, some small subset of police officers are more prone to engage in additional illegitimate and unnecessary encounters with individuals. While it is not necessary to provide a reason for why we might see such heterogeneity in officer propensities, a body of research in social dominance theory (e.g., Sidanius et al., 1994; Sidanius and Pratto, 2001) provides some coherent explanations. Since the individuals encountered in these cases are not necessarily encountered on reasonable suspicion of a crime, we assume that use of lethal force conditional on encounter is less likely than under the *standard policing* model, but that use of non-lethal force is elevated.

We assume that over the same interval of time as in the *standard policing* model, \hat{E}_B and \hat{E}_W individuals in each respective sub-population get encountered by the subset of police who

engage in excessive use of non-lethal force, at rates given by the parameters: $\hat{\phi}_B \in [0, 1]$ and $\hat{\phi}_W \in [0, 1]$. As such, we expect the counts of these encounters to follow:

$$\hat{E}_B \sim \text{Binomial}(N_B, \hat{\phi}_B) \quad (6)$$

$$\hat{E}_W \sim \text{Binomial}(N_W, \hat{\phi}_W) \quad (7)$$

The counts of lethal and non-lethal force events resulting from these interactions are then given by:

$$(\hat{S}_B, \hat{G}_B, \hat{O}_B)' \sim \text{Multinomial}(\hat{E}_B, (\hat{\theta}_B, \hat{\gamma}_B, 1 - (\hat{\theta}_B + \hat{\gamma}_B)))' \quad (8)$$

$$(\hat{S}_W, \hat{G}_W, \hat{O}_W)' \sim \text{Multinomial}(\hat{E}_W, (\hat{\theta}_W, \hat{\gamma}_W, 1 - (\hat{\theta}_W + \hat{\gamma}_W)))' \quad (9)$$

where $\hat{O}_B = \hat{E}_B - (\hat{S}_B + \hat{G}_B)$ and $\hat{O}_W = \hat{E}_W - (\hat{S}_W + \hat{G}_W)$. Equations (6), (7), (8), and (9) are of the same form as Eqs. (1), (2), (3), and (4) in the *standard policing* model. The parameters also are the same in definition, with the “hat” symbol indicating that they refer to the *routine over-escalation* model and can take on different values than in the *standard policing* model. The logical conditions establishing the presence or absence of racial disparities in this model are the same as those given in Table 3.

If the population of officers is composed of some police operating under the standard parameters, and some operating under the routine over-escalation parameters, then the joint mixture model is able to produce the aggregate appearance of anti-white disparities in lethal force conditional on encounter (i.e., $\frac{\hat{S}_B + \hat{S}_B}{\hat{E}_B + \hat{E}_B} < \frac{\hat{S}_W + \hat{S}_W}{\hat{E}_W + \hat{E}_W}$), even when there are no anti-white racial disparities in the use of lethal force conditional on encounter in either subgroup of police (i.e., $\frac{\hat{S}_B}{\hat{E}_B} > \frac{\hat{S}_W}{\hat{E}_W}$ and $\frac{\hat{S}_B}{\hat{E}_B} > \frac{\hat{S}_W}{\hat{E}_W}$). The easiest sufficient condition to understand arises when set $\frac{\hat{S}_B}{\hat{E}_B} = \frac{\hat{S}_W}{\hat{E}_W} \lambda$, with $\lambda > 1$, and then make the assumption that the count of lethal force incidents against white individuals in the *routine over-escalation* sub-model declines to zero (i.e., $\hat{S}_W \rightarrow 0$). Even under these conditions, we show (in the [Supplementary Materials](#)) that this model produces an apparent absence of encounter-conditional anti-black racial disparities actually anti-white racial disparities in the use of lethal force whenever:

$$\lambda \frac{E_B}{(E_B + \hat{E}_B)} < \frac{E_W}{(E_W + \hat{E}_W)} \quad (10)$$

This expression indicates that as long as there are sufficient anti-black racial disparities in the frequency of encounters arising from the *routine over-escalation* model to offset the anti-black racial disparities in the use of force conditional on encounter in the *standard policing* model, we will observe an apparent absence of anti-black disparities in the use of lethal force conditional on encounter in the pooled data. In the language of the model, if \hat{E}_B is sufficiently large relative to other terms, then it can offset a $\lambda > 1$. For example, if the encounter-conditional black to white relative risk of being shot by police were $\lambda = 1.11$, then if just 1 in 20 white encounters, but 3 in 20 black encounters occur under the *routine over-escalation* model, we would still perhaps paradoxically see aggregate-level anti-white racial disparities in encounter-conditional use of lethal force by police.

Implications of our model for statistical analysis of police shooting data. To emphasize the implications of our findings for empirical analyses, we present in the [Supplementary Materials](#) an additional analysis in which we simulate data on police

encounters under the union of the *standard policing* and *routine over-escalation* models. We then analyze the aggregate data with a statistical model that ignores the fact that the use-of-force outcomes arise through a generative model structure with heterogeneous decision making units (i.e., the above-listed sub-types of police officers). The results again confirm that the “paradoxical” results with which we began are only apparent. Furthermore, we show that a wide range of parameter values are sufficient to produce the seemingly paradoxical results and demonstrate that the phenomenon does not strongly depend on the particular reducing assumptions used to reach an analytical solution.

Discussion

The data and methodological approach presented in Fryer (2016) represent the most important advance to date in the analysis of race-specific use-of-force by police. Fryer (2016) is able to investigate aspects of police use-of-force (like encounter-conditional, race-specific rates of the use of lethal force) that have thus far been opaque. However, this new ground brings with it some counter-intuitive terrain for analysts and for the journalists who summarize research for the public. Specifically, we show that the presentation of Fryer’s results as indicating an absence of anti-black racial disparities and perhaps even the presence of anti-white racial disparities in police use of lethal force is unwarranted. Both Ross (2015) and Fryer (2016) show population-level anti-black racial disparities in the use of lethal force by police—disparities that are indicative of disproportionate morbidity and mortality per unit population. The theoretical model presented here allows us to decompose these disparities into the effects of differential encounters and differential use of lethal force conditional on encounters. Either of these statistical biases may be justifiable on the basis of differential crime rates or contextualizing details. The narrow objective of our analysis, however, does not require we take a position as to whether or not differential encounters and differential use of force are justified. Nonetheless, the existence of population-level disparities in police use-of-force raises social and policy issues that must be addressed.

Taking the analysis of Fryer (2016, 2017) at face value, it would appear that race-based differences in the use of lethal force conditional on encounter can be entirely explained by contextualizing details. But, as we show, if police behavior is heterogeneous, with most officers following standard protocol and a small subset of officers engaging in unwarranted use of excessive non-lethal force (like use of tasers), racial disparities in who this subset of officers target can complicate interpretation of encounter-conditional data. It is ironic that elevated levels of sub-lethal assault against innocent black individuals by a subset of police would have the effect of diminishing the *apparent* severity of anti-black racial disparities in lethal force conditional on encounter in the full set of officers. Nevertheless, this is a key finding of our model. Increased encounter rates and excessive use of violent but non-lethal force by a subset of police against black individuals can mask the existence of anti-black racial disparities in the encounter-conditional use of lethal force by police in pooled data.

Given that the existence of racial disparities in police use-of-force is a serious public policy issue (e.g., USDOJ, 2016), it is critical that these dynamics be understood. Although the work of Fryer (2016) presents one of the most detailed empirical analyses of racial disparities in police use-of-force, only individual-level estimates of officer parameters will allow a convincing demonstration of his arguments. Though it is impossible to demonstrate strictly through our modeling effort here, it seems improbable that there would be such consistent evidence of racial disparities in rates of: (i) encounters with police, (ii) population-level and

encounter-conditional use of non-lethal force by police, and (iii) population-level use of lethal force by police, only to see a complete reversal in encounter-conditional rates of lethal force. Our model results, however, establish that precisely this set of racial disparities in rates of encounters and use of non-lethal force can lead to the principle finding of Fryer (2016).

Our findings parallel classic results in applied statistics related to Simpson’s paradox (e.g., Simpson, 1951; Bickel et al., 1975); analysis of aggregate data (i.e., data arising from pooling the outcomes of heterogeneous decision-making units) can show discrimination against a given class of individuals, even when there is actually discrimination in the opposite direction in every decision-making unit. Disparities in the rates with which individuals interact with each decision-making unit are typically found in cases where Simpson’s paradox is demonstrated. For example, in the case of the Berkeley graduate admissions data (Bickel et al., 1975), women tended to apply to departments that were more difficult for applicants of either sex to enter, while men tended to apply to departments that accepted a greater majority of applicants of either sex; as a result, in the aggregate, men were accepted in greater proportion than women, despite the fact that there was actually evidence of a small average bias in admitting women at the level of the decision-making unit.

The case of racial disparities in police shootings conditional on encounter holds a similar logical structure. However, the variable that structures decision-making (akin to the academic department in the Berkeley graduate admissions data) is much harder to empirically identify in the case of police shootings. We show that racial disparities in encounter rates can generate Simpson’s paradox, and we provide an example of a kind of decision-structuring variable (whether a given officer has *standard policing* or *routine over-escalation* parameters) which produces model outcomes that are similar to the empirically observed data. It will likely be difficult to empirically evaluate if Simpson’s paradox truly explains the results of Fryer (2016) since the longitudinal, individual-level data needed to estimate more finely-resolved model parameters will be very hard to acquire. Nonetheless, this point is largely eclipsed by a more important point—namely, that the public health implications of racial disparities in police shootings and non-lethal use-of-force (e.g., Miller et al. 2017; DeVyllder et al., 2017; Ross, 2017), in the end, come down to the population-level relative risk not the encounter-conditional risk.

Pathways leading to population-level racial disparities in police use-of-force. While the use of lethal force against unarmed individuals by police is quite rare, it does occur and with racial disparities in population-level rates. Non-lethal force also is disproportionate at the population level (Fryer, 2016; Miller et al., 2017). These observations are not contradicted by the observation that the encounter-conditional probabilities of lethal outcomes by race are reversed in aggregate data. While Fryer (2016) and Ross (2015) address somewhat different questions, each using variables unique to their datasets, they arrive at complementary, not paradoxical, conclusions. Those conclusions reinforce the prevailing consensus that there are significant racial disparities in police behavior at the population level, and together they provide an expanded basis for appreciating that this is a complex problem at the local level Fryer (2016), with significant differences among locales nationally Ross (2015).

Within each locale, the research strategy advanced by Fryer (2016) is important, as it allows us to move beyond simply demonstrating the presence of disparities and instead investigate how and why such disparities arise when and where they do. Without being able to identify the causal drivers of racial disparities operating in a given location we are unable to offer

useful policy recommendations (Ross, 2016). It is important to determine if racial disparities in a particular locale arise due to officer behavior conditional on encounter, or via the policies and social contexts leading to racial disparities in encounter rates. The combination of explicit theoretical modeling of the generative pathways of use-of-force outcomes, wariness concerning Simpson's paradox, and mixed use of both qualitative and quantitative methods will help researchers better understand the empirical data.

Officer behavior. Police officers and police departments are diverse, and it is inappropriate to characterize such institutions in monolithic terms. There is no shortage of cases of police officers engaging in overt racism in recent years (e.g., cases in Los Angeles, (Stack, 2016; Whaley, 2013), Oakland, CA (Shoichet et al., 2016), San Francisco, CA (Serna and Romney, 2015; Serna, 2016), San Diego, CA (Perry, 2015), Miami, FL (Buncombe, 2015), Fort Lauderdale, FL (Barszewski, 2016), Clatskanie, OR (Park, 2015), Seattle, WA (Campbell, 2015), New York, NY (Sapien, 2015), Camden, NJ (Walsh, 2016), Edison, NJ (Amaral, 2015), Ferguson, MO (Swaine, 2014), Saint Louis, MO (Hudson, 2014), Cleveland, OH (Hensley, 2014), Detroit, MI (Murphy, 2013), New Orleans, LA (McCarthy, 2012), Baton Rouge, LA (Alejandro, 2014), Chicago, IL (Stahl, 2008), Baltimore, MD (USDOJ, 2016), and other locations). Excessive media and public focus on the sensationalized cases mentioned above, however, has the potential to wrongfully impute such attitudes to police officers in general, trigger in-group out-group identity politics, and to distract from the dialog needed to implement effective policies addressing the problem. While is important to end impunity for the small subset of officers who abuse their positions of power and responsibility, or display outward signs of racist hatred like some of those mentioned above, it is also important to identify the less visible, institutional drivers of racial disparities in police use-of-force.

There is increasing evidence that the population-level presence of racial disparities in police shootings might have less to do on average with officer behavior conditional on encounter, and more to do with differences in encounter rates (Fryer, 2016; Miller et al., 2017). This information may be key to creating more effective policy change. Selby et al. (2016) have compiled a detailed data set of officer caused deaths of unarmed individuals in the United States in 2015. At the population level, their data closely echo those of Ross (2015) and Fryer (2016) black individuals represent about 42% of the victims of known race, despite being only about 12% of the population ($\frac{42\%}{12\%} = 3.5x$), while non-Hispanic white individuals represent about 39% of the victims of known race, but make up about 62% of the population ($\frac{39\%}{62\%} = 0.63x$). Selby et al. (2016), however, investigate the contextualizing details of each case, and find that there is no evidence to suggest that overt or even unconscious racial prejudice plays a *systematic* role in the officer-involved deaths of unarmed citizens which is not to say that cases to the contrary never occur. Instead, their findings point to other characteristics associated with the victim, including acute drug intoxication, mental and physical illness, violent behavior, and perceived threats to innocent civilians as being critical predictors of the deaths of unarmed civilians due to police intervention (Selby et al. 2016). Further, Selby et al. (2016) and Ross (2015) both note that most civilians who are shot by police were armed at the time of the incident.

Additionally, Selby et al. (2016) mention that in about 88% of the lethal police encounters not beginning with traffic stops, officers engage suspects only after being requested to do so by the community (i.e., via 911 calls). As such, these encounters are not

initiated at the officers' own discretion. In the data compiled by Selby et al. (2016), 70% of the unarmed people killed by police were only encountered because they were reported to police by other civilians as being probable threats to the community, which the police were obliged to investigate, often after being primed that the indicated persons posed a threat.

Black Americans are disproportionately likely to be encountered by police, be the victims of violent but non-lethal use-of-force by police, and be the victims of lethal use-of-force by police. However, these observations are fully consistent with the observation that the great majority of individual officers do not disproportionately target black people based on their own personal discretion and do not show racial bias in the use of lethal or non-lethal force conditional on encountering a given individual. Social context is likely to play a significant role in explaining differential encounter rates and their downstream effects on population-level racial disparities in police use-of-force.

Social context, policies, and encounter rates. Recent data show that black individuals are more likely to be stopped by police than white individuals (Fryer, 2016; Miller et al., 2017; USDOJ, 2016). Miller et al. (2017) argue that the differential population-level rates of the use of lethal force by police are almost entirely explainable by these disparities in encounters an interpretation which is not inconsistent with our present analysis. There are two major lines of argument concerning the causes of differential encounter rates: (1) that differential encounter rates are driven by differential rates of crime, and (2) that the 'criminalization of blackness' leads to racial disparities in encounters for reasons having little to do with differential crime rates.

In the aggregate in the United States, there appear to be higher per capita rates of violent crime committed by black individuals relative to white individuals (e.g., USDOJ, 2014) the caveat is that these arrest/conviction records may themselves be an outcome of racial disparities in policing intensity and conviction rates. But, there is no evidence to suggest that the counties with relatively high black to white crime rate ratios are those with disproportionately high rates of racial disparities in police use of lethal force against unarmed individuals (Ross 2015); however, the analysis of Ross (2015) linking county-level crime rates and racial disparities in police shootings is itself very susceptible to the ecological inference fallacy. This being said, other more detailed, geographically localized studies have also found racial disparities in rates of encounters and use-of-force that are not fully explainable by differential crime rates or related variables (e.g., Fryer, 2016; Gelman et al. 2007; USDOJ, 2016).

Explanations for differential encounter rates based on the 'criminalization of blackness' tend to hold strong narrative weight but are more difficult to confirm with empirical data. The principle idea is that black individuals are more likely to be reported by community members to the police than white individuals especially in socio-economically unequal areas for innocuous activities like smoking a cigarette at night, jogging, walking or driving in 'the wrong kind of neighborhood', wearing hoodie-style sweat shirts, driving nice cars, driving not-so-nice cars, etc. (Beer, 2016); there is a body of qualitative literature of this topic, and arguments linking community behaviors and biases in the representations of black people by the U.S. media (see further discussion in the [Supplementary Materials](#)). However, fine-grained, geographically representative quantitative data on the disproportionate rates at which black individuals are reported to the police for innocuous activities appear to be unavailable. Production of this kind of data would be a fruitful direction for future research.

Other dynamics might lead to disproportionately high rates of encounters between black individuals and police. Racial disparities in access to health care and mental health care, differential rates of poverty and homelessness, and differential intensities of policing effort in different geographic locales are all candidates. There are many possible and potentially overlapping reasons why black individuals are stopped more often than white individuals, and these reasons are likely to be variable in space and time. Successful policy interventions are likely to require local-level investigation, and should be adaptive to the shifting circumstances affecting local communities.

Conclusions

We establish that: (1) the analyses of Ross (2015) and Fryer (2016) are in general agreement concerning the existence and magnitude of population-level anti-black, racial disparities in police shootings; (2) because of racial disparities in rates of encounters and non-lethal use-of-force, the encounter-conditional results of Fryer (2016) regarding the relative frequency of the use of lethal force by police are susceptible to Simpson's paradox. They should probably not be interpreted as providing support for the idea that police show no anti-black bias or even an unexpected anti-white bias in the use of lethal force conditional on encounter; and, (3) even if police do not show racial bias in the use of lethal force conditional on encounter, racial disparities in encounters themselves will still produce racial disparities in the population-level rates of the use of lethal force, a matter of deep concern to the communities affected.

These findings have specific public policy implications: (1) population-level measures of the use of lethal and non-lethal force by police are more robust indicators of the overall severity of racial disparities in use-of-force by police than encounter-conditional measures. Population-level results should thus be used when evaluating the local-level public health implications of racial disparities in police use-of-force. We can improve our understanding of the broader effects of racial disparities in police use-of-force if we avoid assessments conditioned on problematic intermediate variables such as encounters, which might occur for different reasons within the black and white sub-populations and might themselves be a causal outcome of racial bias; (2) the population-level relative risk of being the victim of lethal force by police can be roughly decomposed into the relative risk of encounters multiplied by the relative risk of receiving lethal force conditional on encounter, and empirical data show strong differences in race-specific encounter rates. Together, these observations might suggest that policies and social programs designed to minimize racial disparities in encounters may be as (or even more) important to consider as policies aimed at changing police behavior conditional on encounter; and finally, (3) we agree with Fryer (2016) that departments might wish to "...increase the expected price of excessive force on lower level uses of force" (p. 36). Our simulation-based analysis also supports this recommendation, inasmuch as unjustified use of non-lethal force is both more common than unjustifiable use of lethal force and even low levels of such actions have the potential to obscure our understanding of racial disparities in more extreme levels of use-of-force.

Received: 14 December 2017 Accepted: 10 April 2018

Published online: 05 June 2018

References

- Alejandro R (2014) Baton Rouge police officer resigns after sending racist text-messages. *Afro*. <http://www.afro.com/baton-rouge-police-officer-resigns-after-sending-racist-text-messages>
- Amaral B (2015) Edison cop who sent racist texts 'not credible' as witness. *NJ Media*. http://www.nj.com/middlesex/index.ssf/2015/07/prosecutor_cop_who_sent_racist_texts_not_credible.html
- Barszewski L (2016) Arbitrator upholds firing of Fort Lauderdale officer over racist texts. *Sun Sentinel*. <http://www.sun-sentinel.com/local/broward/fort-lauderdale/flauderdale-police-slurs-arbitrator-ruling-20160524-story.html>
- Beer T (2016) "But Blacks commit more crimes": scholars discuss conservative logic. *Sociology Toolbox*. https://thesocietypages.org/toolbox/blacks_commit_more_crimes_1/
- Bergesen A (1982) Race riots of 1967: an analysis of police violence in Detroit and Newark. *J Black Stud* 12(3): 261–274.
- Bickel PJ, Hammel EA, O'Connell JW et al. (1975) Sex bias in graduate admissions: data from Berkeley. *Science* 187(4175):398–404
- Brogden M, Shearing CD (2005) *Policing for a new South Africa*. Routledge, London and New York
- Buncombe A (2015) Miami beach cops sent more than 200 racist, homophobic and pornographic emails, says senior officer: probe launched amid concerns hundreds of criminal prosecutions could be jeopardised. *Independent*. <http://www.independent.co.uk/news/world/americas/miami-beach-cops-sent-more-than-200-racist-homophobic-and-pornographic-emails-says-senior-officer-10255914.html>
- Campbell A (2015) Police officer fired for racial bias after falsely claiming Black man attacked her with golf club. *The Huffington Post*. http://www.huffingtonpost.com/entry/officer-fired-racial-bias-golfclub-cane_us_55f9a334e4b0e333e54c2d56
- Cano I (2010) Racial bias in police use of lethal force in Brazil. *Police Pract Res* 11 (1):31–43
- Correll J, Park B, Judd CM, Wittenbrink B, Sadler MS, Keesee T (2007) Across the thin blue line: police officers and racial bias in the decision to shoot. *J Pers Soc Psychol* 92(6):1006
- Correll J, Urland GR, Ito TA (2006) Event-related potentials and the decision to shoot: the role of threat perception and cognitive control. *J Exp Soc Psychol* 42(1):120–128
- Cox A (2016) Roland Fryer answers reader questions about his police force study. *The New York Times*. <http://www.nytimes.com/2016/07/13/upshot/roland-fryer-answers-reader-questions-about-his-police-force-study.html>
- DeVega C (2016) Sorry conservatives, new research from Harvard shows a profound amount of racism by police...not less of it. *Salon* 26 (5):466–477. http://www.salon.com/2016/07/14/sorry-conservatives_new_research_from_harvard_shows_a_profound_a-mount_of_racism_by_police_not_less_of_it/
- DeVylder JE, Oh HY, Nam B, Sharpe TL, Lehmann M, Link BG (2017) Prevalence, demographic variation and psychological correlates of exposure to police victimisation in four US cities. *Epidemiol Psychiatr Sci* 26(5):466–477
- Doane A (2006) What is racism? Racial discourse and racial politics. *Crit Sociol* 32 (2-3):255–274
- French JH (2013) Rethinking police violence in Brazil: unmasking the public secret of race. *Lat Am Polit Soc* 55(4):161–181
- Fryer RG (2016) An empirical analysis of racial differences in police use of force. Technical report. National Bureau of Economic Research
- Fryer RG (2017) Reconciling results on racial differences in police shootings. Technical report. Harvard university and the National Bureau of Economic Research, Harvard
- Fyfe JJ (1980) Geographic correlates of police shooting: a microanalysis. *J Res Crime Delinquency* 17(1):101–113
- Gelman A, Fagan J, Kiss A (2007) An analysis of the New York City police department's stop-and-frisk policy in the context of claims of racial bias. *J Am Stat Assoc* 102(479):813–823
- Gilbert KL, Ray R (2016) Why police kill black males with impunity: applying public health critical race praxis (phcrp) to address the determinants of policing behaviors and "justifiable" homicides in the USA. *J Urban Health* 93 (1):122–140
- Goldkamp JS (1976) Minorities as victims of police shootings: Interpretations of racial disproportionality and police use of deadly force. *Justice Syst J* 2(2): 169–183
- Harring S, Platt T, Speigman R, Takagi P (1977) The management of police killings. *Crime Soc Justice* 8:34–43
- Hensley N (2014) Ohio cops suspended for years of racist texts leaked to civil rights group. *New York Daily News*. <http://www.nydailynews.com/news/crime/ohio-cops-suspended-racist-texts-leaked-nonprofit-article-1.2033713>
- Hirschfield PJ (2015) Lethal policing: making sense of American exceptionalism. *Sociol Forum* 30(4)

- Holmes MD (2000) Minority threat and police brutality: Determinants of civil rights criminal complaints in US municipalities. *Criminology* 38(2):343–368
- Hudson R (2014) Being a cop showed me just how racist and violent the police are. There's only one fix. *The Washington Post*. <http://www.washingtonpost.com/posteverything/wp/2014/12/06/i-was-ast-louis-cop-my-peers-were-racist-and-violent-and-theres-only-onefix/>
- Jacobs D (1998) The determinants of deadly force: a structural analysis of police violence. *Am J Sociol* 103(4):837–862
- Jacobs D, Britt D (1979) Inequality and police use of deadly force: An empirical assessment of a conflict hypothesis. *Soc Probl* 26(4): 403–412.
- Li R (2016) The research is only as good as the data: a new study found no evidence of racial bias in police shootings. But last year, a study came to the opposite conclusion. Why? *Slate*. http://www.slate.com/articles/health_and_science/science/2016/07/roland_fryer_s_new_paper_came_to_the_opposite_conclusion_as_a_papar_last.html
- Martinot S, Sexton J (2003) The avant-garde of white supremacy. *Social Identity* 9(2):169–181
- McCarthy B (2012) NOPD officer under investigation in shooting is suspended after online comments. *The Times-Picayune*. http://www.nola.com/crime/index.ssf/2012/03/nopd_officer_under_investigati.html
- Miller TR, Lawrence BA, Carlson NN, Hendrie D, Randall S, Rockett IR, Spicer RS (2017) Perils of police action: a cautionary tale from US data sets. *Inj Prev* 23(1):27–32
- Murphy D (2013) Michigan police suspend cop behind humiliating, “racist” footage of Black man. *New York Daily News*. <http://www.nydailynews.com/news/national/michigan-police-suspendbehind-racist-footage-black-man-article-1.1527387>
- Park E (2015) Police chief “compared African-Americans to monkeys”. *KOIN 6 News*. <http://koin.com/2015/09/06/police-chief-compared-african-americans-to-monkeys/>
- Pearl J (2014) Comment: understanding Simpson’s paradox. *Am Stat* 68(1):8–13
- Perry T (2015) Black sergeant sues San Diego police department over racist cartoon. *LA Times*. <http://www.latimes.com/local/lanow/la-me-ln-san-diego-police-blackofficer-lawsuit-20150128-story.html>
- Plant EA, Peruche BM, Butz DA (2005) Eliminating automatic racial bias: making race non-diagnostic for responses to criminal suspects. *J Exp Soc Psychol* 41(2):141–156
- Ross CT (2017) Digging deeper: population-level racial disparities, exposure to police victimisation and psychological trauma. *Epidemiol Psychiatr Sci* 26(5):478–480
- Ross CT (2015) A multi-level Bayesian analysis of racial bias in police shootings at the county-level in the United States, 2011–2014. *PLoS ONE* 10(11):e0141854
- Ross CT (2016) Identifying the drivers of the exceptionally high rates of lethal and racially biased violence by police in the United States requires geographically resolved data and analysis. *Sociol Forum* 31:231–232
- Sapien J (2015) Racist posts on NY cop blog raise ire at time of tension. *ProPublica*. <https://www.propublica.org/article/racist-posts-on-ny-cop-blograise-ire-at-time-of-tension>
- Selby N, Singleton B, Flosi E (2016) In context: understanding police killings of unarmed civilians. Contextual Press
- Serna J (2016) More racist text messages uncovered among San Francisco police officers. *LA Times*. <http://www.latimes.com/local/lanow/la-me-ln-sfpd-racist-textmessages-20160426-story.html>
- Serna J, Romney L (2015) Racist, homophobic texts by San Francisco police trigger case reviews. *LA Times*. <http://www.latimes.com/local/lanow/la-me-ln-racist-police-textmessages-review-20150317-story.html>
- Shoichet CE, Almasy S, Cuevas M (2016) 3 chiefs in 9 days: Scandals plague police in Oakland, CA. *CNN*. <http://edition.cnn.com/2016/06/19/us/oakland-police-scandals/>
- Sidanius J, Liu JH, Shaw JS, Pratto F (1994) Social dominance orientation, hierarchy attenuators and hierarchy enhancers: Social dominance theory and the criminal justice system. *J Appl Soc Psychol* 24(4):338–366
- Sidanius J, Pratto F (2001) Social dominance: An intergroup theory of social hierarchy and oppression. Cambridge University Press, Cambridge and New York
- Simpson EH (1951) The interpretation of interaction in contingency tables. *J Royal Stat Soc B* 13(2):238–241
- Smith BW (2003) The impact of police officer diversity on police-caused homicides. *Policy Stud J* 31(2):147–162
- Stack L (2016) Los Angeles police official resigns over racist emails. *New York Times*. http://www.nytimes.com/2016/05/02/us/los-angeles-police-officialresigns-over-racist-emails.html?_r=0
- Stahl J (2008) Off duty Chicago police accused of election night hate crimes. *Huffington Post*. http://www.huffingtonpost.com/jayne-lyn-stahl/off-duty-chicagopolice-a_b_142536.html
- Swaine J (2014) Ferguson: officer relieved of duty after video of racist remarks surfaces. *The Guardian*. <http://www.theguardian.com/world/2014/aug/23/ferguson-officerrelieved-duty-video-racist-remarks>
- Tomaskovic-Devey D, Mason M, Zingraff M (2004) Looking for the driving while black phenomena: Conceptualizing racial bias processes and their associated distributions. *Police Q* 7(1):3–29
- US Census Bureau (2010). <https://www.census.gov/data.html>
- USDOJ (2014) Uniform crime reporting program data: arrests by age, sex, and race, 2012. Inter-University Consortium for Political and Social Research, Washington DC.
- USDOJ (2016) Investigation of the Baltimore city police department. <http://www.nytimes.com/interactive/2016/08/09/us/document-baltimore-P-D-Findings-Report-FINAL.html>(accessed on December 2014)
- Wagner K (2014) U.S. police shootings database. <http://goo.gl/Su60Cm> (accessed on December 2014)
- Walsh J (2016) Read racist texts that got Camco guards fired. *Courier Post*. <http://www.courierpostonline.com/story/news/2015/11/19/textscorrections-camden-jail-public-records/76040752/>
- Warren P, Tomaskovic-Devey D, Smith W, Zingraff M, Mason M (2006) Driving while black: bias processes and racial disparity in police stops. *Criminology* 44(3):709–738
- Whaley N (2013) Two Black cops write statements about racism in LAPD. *BET*. <http://www.bet.com/news/national/2013/02/14/two-black-cops-writestatements-about-racism-in-lapd.html>
- Wortley S., Owusu-Bempah A. (2012) Race, Ethnicity, Crime and Criminal Justice in Canada. In: Kalunta-Crumpton A. (eds) *Race, Ethnicity, Crime and Criminal Justice in the Americas*. Palgrave Macmillan, New York
- Wortley S, Tanner J (2003) Data, denials, and confusion: the racial profiling debate in Toronto. *Can J Criminol Crim Justice* 45(3):367–390

Data availability

The simulation code used in this study is available in the Dataverse repository: <https://doi.org/10.7910/DVN/PVPJWM>, and will be maintained at: <https://github.com/Ctross/Publications/tree/master/PalgravePolicingModels>

Acknowledgements

We thank members of the Department of Human Behavior, Ecology, and Culture at the Max Planck Institute for Evolutionary Anthropology for comments on an early draft of this paper.

Additional information

Supplementary information accompanies this paper at <https://doi.org/10.1057/s41599-018-0110-z>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://www.nature.com/reprints>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018