Supporting Information

"A Behavioral Theory of Discrimination in Policing"

Ryan Hübert
UC Davis

Andrew T. Little
UC Berkeley

March 2, 2020

FOR ONLINE PUBLICATION

A Racial Profiling and the Geography of Policing

The analysis in the main text assumes that police officers decide to allocate their time between policing two groups of people. In the United States, the prevailing law is unclear about whether such group-based profiling is permissible (for an extended discussion, see Knowles, Persico, and Todd 2001). However, under the U.S. Constitution, policies that explicitly treat members of protected categories differently are subject to strict scrutiny (see *Brown v. Board of Education of Topeka*, 1954). A policy of explicitly using group membership to allocate policing resources is not likely to survive a strict scrutiny legal analysis.

We focus on this simple, but potentially illegal, decision-making process in text because it allows us to more clearly focus on our core arguments. However, it can be microfounded with a more complex model where a police chief decides how many policing resources to devote to two neighborhoods: 1 and 2. Formally, assume he devotes n_1 of his time to policing neighborhood 1 and $n_2 = 1 - n_1$ of his time to policing neighborhood 2. Also assume that each neighborhood is comprised of members of the two groups, A and B. Within a neighborhood i, we assume that police interact with a member of group A with probability α_i and a member of group B with probability $1 - \alpha_i$. If police encounters with residents are random and iid, then one way to interpret α_i is that

it represents the proportion of neighborhood i that is comprised of members of group A. However, our flexible specification allows for the possibility that police come into contact with members of one group at a rate disproportionate to that group's share of the local population. (Although note that if α_i does not reflect the demographic makeup of neighborhood i, then we simply reintroduce concerns about racial profiling that motivate this microfoundation, just at a different point in the analysis.)

Conditional on a choice about how intensely to police each neighborhood, the share of group A individuals the police encounters is $\eta_A = n_1\alpha_1 + (1-n_1)\alpha_2 = \alpha_2 + (\alpha_1 - \alpha_2)n_1$ and the share of group B individuals the police encounters is $\eta_B = n_1(1-\alpha_1) + (1-n_1)(1-\alpha_2) = 1 - \eta_A$. Recall from the main text that w_A is defined as the share of time that the police officer devotes to policing group A, and $w_B = 1 - w_A$ is the corresponding share of time that the police officer devotes to policing group B. Then, η_A is equivalent to w_A and η_B is equivalent to w_B, and n_1 is a perfect proxy for w_A. More specifically, if police come into contact with group A more than group B in neighborhood 1 (alt. neighborhood 2), $\alpha_1 > \alpha_2$ (alt. $\alpha_1 < \alpha_2$), then increasing n_1 (alt. n_2) linearly increases w_A. Notice that in the extreme cases where $n_1 = 0$ and $n_1 = 1$, then $\eta_A = \alpha_2$ and $\eta_A = \alpha_1$, respectively. Then, α_1 and α_2 correspond the maximum and minimum possible allocations: $\underline{w} = \min\{\alpha_1, \alpha_2\}$ and $\overline{w} = \max\{\alpha_1, \alpha_2\}$.

In a model where police choose n_1 (and not w_A), the analysis in the main text is identical after substituting $\eta_A = \alpha_2 + (\alpha_1 - \alpha_2)n_1$ for w_A.

B Stability Conditions

The main text makes informal references to a stability condition in the single-officer model (which, as we will prove, always holds at the unique equilibrium for this version). More consequentially, Proposition 4 makes reference to a stability condition which, while standard, is not yet defined. We first discuss the condition in the simpler single-officer setting and then the extension

to two officers.

B.1 Single Officer

The main intuition for the single officer equilibrium to be stable is as follows. Let w_A^* be an equilibrium allocation. If it were the case that $w_A^{\text{br}}(r_t, \tilde{r}_p(w_A^* + \epsilon)) > w_A^* + \epsilon$ for some small $\epsilon > 0$, then a small upward perturbation in the policing of group A would lead the officer to want to police this group even more. Conversely, if $w_A^{\text{br}}(r_t, \tilde{r}_p(w_A^* - \epsilon)) < w_A^* - \epsilon$ for some small $\epsilon > 0$, a downward perturbation would lead the officer to police group A even less. We would like a stability condition which ensures that neither happens: i.e., small deviations would "return" to the equilibrium when iterating the best response function.

As $\epsilon \to 0$, since w_A^{br} is continuous, a formal statement of the stability requirement is:

Definition 5. *Let $F(w_A) = w_A^{br}(r_t, \tilde{r}_p(w_A)) - w_A$. An equilibrium allocation $w_A^* \in [\underline{w}, \overline{w}]$ is stable if and only if $\frac{\partial F}{\partial w_A}\Big|_{w_A = w_A^*} < 0$.*

To tie this to the geometric discussion surrounding Figure 1, this is equivalent to $\frac{\partial w_A^{\text{br}}}{\partial w_A}\Big|_{w_A = w_A^*} < 1$, i.e, the slope of the optimal allocation curve is less than that of the 45 degre line where they intersect.

Recall that w_A^{br} may not be differentiable at points where it switches from a corner to an interior solution. However, this does not pose any problems for this definition. If there is a corner equilibrium where $w_A^{\text{br}}(r_t, \tilde{r}_p(\underline{w} + \epsilon)) = \underline{w}$, for some $\epsilon > 0$, then the derivative of the best response at $\underline{w}$ is zero, and hence the equilibrium is stable. Similarly, $w_A^{\text{br}}(r_t, \tilde{r}_p(\underline{w} - \epsilon)) = \overline{w}$, the the corner solution at $\overline{w}$ is stable.

If there is a corner equilibrium at $\underline{w}$ but $w_A^{\text{br}}(r_t, \tilde{r}_p(\underline{w} + \epsilon)) > \underline{w}$ for any $\epsilon > 0$, then the right derivative of w_A^{br} is well-defined at $\underline{w}$ and the stability condition is defined with respect to this derivative (which may or may not hold). Similarly, if there is a corner equilibrium at $\overline{w}$ but $w_A^{\text{br}}(r_t, \tilde{r}_p(\underline{w} + \epsilon)) < \overline{w}$ for any $\epsilon > 0$, the stability condition can be defined with respect to the

well-defined left-derivative of w_A^{br} at $\overline{w}$.

B.2 Multiple Officers

The equilibrium condition for the two officer model can be written:

$$F_1(w_{A,1}, w_{A,2}) \equiv w_A^{\text{br}}(r_{t,1}, \tilde{r}_{p,1}(w_{A,1}, w_{A,2}, \nu_1)) - w_{A,1} = 0$$
$$F_2(w_{A,1}, w_{A,2}) \equiv w_A^{\text{br}}(r_{t,2}, \tilde{r}_{p,2}(w_{A,1}, w_{A,2}, \nu_2)) - w_{A,2} = 0$$

Now we put a little more structure on what it means for an equilibrium to be stable. More specifically, close to an equilibrium, we want that for any "small" perturbation to both players' strategies, if the officers iteratively choose best responses given their new beliefs, then the joint allocation would move back to the equilibrium. By standard results in the study of dynamic systems (e.g., Theorem 11.4 in Gintis 2009), this can be expressed by conditions on the matrix of the partial derivatives of the F_i functions:

Definition 6. *Let*

$$D(w_{A,1}, w_{A,2}) = \begin{bmatrix} \frac{\partial F_1}{\partial w_{A,1}} & \frac{\partial F_1}{\partial w_{A,2}} \\ \frac{\partial F_2}{\partial w_{A,1}} & \frac{\partial F_2}{\partial w_{A,2}} \end{bmatrix}.$$

an equilibrium in the two-officer model is is stable if:

(i) $trace(D(w^*_{A,1}, w^*_{A,2})) < 0$*, and*

(ii) $det(D(w^*_{A,1}, w^*_{A,2})) > 0$*.*

The first condition simplifies to

$$\left.\frac{\partial F_1}{\partial w_{A,1}}\right|_{w_A = w_A^*} + \left.\frac{\partial F_2}{\partial w_{A,1}}\right|_{w_A = w_A^*} < 0$$

Note that if both derivatives are negative (as required in the single officer model), this is always true.

The second condition becomes:

$$\left[\frac{\partial F_1}{\partial w_{A,1}}\frac{\partial F_2}{\partial w_{A,2}} - \frac{\partial F_1}{\partial w_{A,2}}\frac{\partial F_2}{\partial w_{A,1}}\right]_{w_A = w_A^*} > 0$$

To provide a more easily interpretable version of these conditions, define:

$$Y_i = \left.\frac{\partial w_A^{\text{br}}}{\partial \tilde{r}_{p,i}}\right|_{\tilde{r}_{p,i}=\tilde{r}_{p,i}(w_{A,1}^*, w_{A,2}^*)}$$
$$Z_i = \left.\frac{\partial \tilde{r}_{p,i}(w_{A,1}, w_{A,2})}{\partial w_{A,1}}\right|_{w_A = w_A^*} = \left.\frac{\partial \tilde{r}_{p,i}(w_{A,1}, w_{A,2})}{\partial w_{A,2}}\right|_{w_A = w_A^*}.$$

Then:

$$\left.\frac{\partial F_i}{\partial w_{A,i}}\right|_{w_A = w_A^*} = (Y_i Z_i - 1) \qquad \left.\frac{\partial F_i}{\partial w_{A,-i}}\right|_{w_A = w_A^*} = Y_i Z_i$$

Plugging these into first stability condition gives:

$$(Y_1 Z_1 - 1) + (Y_2 Z_2 - 1) < 0 \iff Y_1 Z_1 + Y_2 Z_2 < 2 \tag{11}$$

and the second:

$$(Y_1 Z_1 - 1)(Y_2 Z_2 - 1) - (Y_1 Z_1)(Y_2 Z_2) > 0$$
$$\iff Y_1 Z_1 + Y_2 Z_2 < 1$$

which is stronger than condition (11) and hence the binding constraint.

An intuition for this condition is that due to the complementarities between action and belief, the deviations that are most apt not to return to an equilibrium are those where both officers increase

or both officers decrease their allocations. And $Y_1Z_1 + Y_2Z_2$ is the marginal change in the best response as *both* officers increase their allocation to group A. So, this condition states that if both officers were to allocate slightly more time to group A or both allocated slightly less, their best responses would move back toward the equilibrium allocation.

C Proofs

Proof of Proposition 1. Using Definition 3, an equilibrium policing allocation w_A solves

$$w_A^* = w_A^{\text{br}}(r_t, \tilde{r}_p^*(w_A^*))$$

At any interior solution, $w_A^{\text{br}}(r_t, r_p) = \frac{r_t^2 \tilde{r}_p(w_A)}{1+r_t^2 \tilde{r}_p(w_A)}$. Substituting (5) and solving this equation for w_A gives a unique solution $\widehat{w}_A$, defined by equation (7) in the main text. Thus when $\widehat{w}_A$ lies in $[\underline{w}, \overline{w}]$ it meets the condition for a unique equilibrium allocation, $w_A^* = \widehat{w}_A$.

To prove that the corner solutions lie where the proposition claims, it helps to first describe the shape of the function which in turn describes how the allocation would change if playing an unconstrained best response starting at w_A,

$$F(w_A) = \frac{r_t^2 \tilde{r}_p(w_A)}{1 + r_t^2 \tilde{r}_p(w_A)} - w_A,$$

on the full range of $[0, 1]$. This function is continuous and differentiable. It is immediate that $F(0) = 0$ and $F(1) = 0$,[14] and by the analysis above $F(\widehat{w}_A) = 0$. So, when $\widehat{w}_A \in (0, 1)$, there are three zeroes on $[0, 1]$, and when $\widehat{w}_A$ lies outside of this interval the only zeroes are at the endpoints

[14]This implies that if we did not restrict the range to $[\underline{w}, \overline{w}]$, there would always be an equilibrium only policing either group, though this would not meet the stability condition whenever an interior equilibrium exists.

(and hence the function must be always positive or negative). Recall that:

$$\widehat{w}_A = \frac{r_t^2 r_p}{1 + r_t^2 r_p} + \frac{\nu(r_t^2 r_p - 1)}{(1-\nu)(1 + r_t^2 r_p)}$$

Rearranging and simplifying gives:

$$0 < \widehat{w}_A < 1 \iff \nu < r_t^2 r_p < 1/\nu$$

In order to see whether F is positive or negative as $w_A \to 0$ and $w_A \to 1$, we need to check F' at these two points. Taking the first derivative of F yields:

$$F'(w_A) = \frac{\nu r_p r_t^2 \left(2(1-\nu)w_A^2 - 2(1-\nu)w_A + 1\right)}{\left(\nu w_A^2 \left(r_p r_t^2 + 1\right) - 2\nu w_A + \nu + (1 - w_A) w_A \left(r_p r_t^2 + 1\right)\right)^2} - 1$$

Evaluating at 0 and 1 gives:

$$F'(0) > 0 \iff r_t^2 r_p > \nu \qquad\qquad F'(1) > 0 \iff r_p r_t^2 < \frac{1}{\nu}$$

Since $\nu < 1/\nu$, there are three cases we must consider, corresponding to three possible shapes of the F function. In case (I), $r_t^2 r_p \geq 1/\nu$. When the inequality is strict, this implies F is increasing at 0, decreasing at 1, and has no interior root, and hence $F(w_A) > 0$ for $w_A \in (0, 1)$. When $r_t^2 r_p = 1/\nu$, the only difference is that $F'(1) = 0$, but F is decreasing for w_A approaching 1, and this does not affect the rest of the argument. In case (II), $\nu < r_t r_p^2 < 1/\nu$, and so F is increasing at 0 and at 1, with an interior zero at $\widehat{w}_A$, and hence $F(w_A) > 0$ for $w_A \in (0, \widehat{w}_A)$ and $F(w_A) < 1$ for $w_A \in (\widehat{w}_A, 1)$. In case (III) $r_t^2 r_p \leq \nu$, and F is decreasing at 0 (or, in the case where $r_t^2 r_p = \nu$, flat at 0 but decreasing for small w_A), increasing at 1, and has no interior root, and hence $F(w_A) < 0$ for $w_A \in (0, 1)$. Note that there can only be an interior equilibrium in case (II), and it must be the case that $F'(\widehat{w}_A) < 0$, i.e., the stability condition for this version of the model described in

Appendix B.1.

Now we can complete proving where the equilibrium lies and uniqueness when the domain of the allocation choice is restricted to $[\underline{w}, \overline{w}]$. If $\widehat{w}_A \leq 0$ then the F function is in case (III) above, and so $F(\underline{w}) < 0$. If $0 < \widehat{w}_A < \underline{w}$, it is in case (II), but since $\underline{w} \in (\widehat{w}_A, 1)$ it must also be the case that $F(w_A) < 0$ for all $w_A \in [\underline{w}, \overline{w}]$. And returning to the definition of w_A^{br}, $F(\underline{w}) < 0$, implies $w_A^{\text{br}}(r_t, \tilde{r}_p(\underline{w})) = \underline{w}$, meaning there is an extreme equilibrium at $\underline{w}$. $F(w_A) < 0$ also implies there is no interior equilibrium or equilibrium at $\overline{w}$ since $F(\overline{w}) < 0$, so this equilibrium is unique. If $\widehat{w}_A = \underline{w}$, then it is immediate that $F(\underline{w}) = \underline{w}$, and hence there is an extreme equilibrium at this bound, and this equilibrium is unique since $F(w_A) < 0$ for $w_A \in (\underline{w}, \overline{w}]$.

When $\underline{w} < \widehat{w}_A < \overline{w}$, $\widehat{w}_A$ is an interior equilibrium, and there can't be another interior state since there is no other point on $[\underline{w}, \overline{w}]$ where $F(w_A) = 0$. The F function is in case (II), which implies $F(\underline{w}) > 0$ and $F(\overline{w}) < 0$, so there is no equilibrium at the extremes. Thus the equilibrium is unique.

By a similar argument to the $\widehat{w}_A \leq \underline{w}$ case, if $\widehat{w}_A \geq \overline{w}$, then $w_A^{\text{br}}(r_t, \tilde{r}_p(\overline{w})) = \overline{w}$, and there can't be an equilibrium at $\underline{w}$ or on the interior. ∎

Proof of Proposition 2. Let $\nu \in (0, 1)$. Note that $\widehat{w}_A$ in the main text simplifies to

$$\widehat{w}_A = \frac{r_t^2 r_p - \nu}{(1-\nu)(1 + r_t^2 r_p)} \tag{12}$$

It follows that $\widehat{w}_A < 1/2$ for $r_t^2 r_p < 1$ and $\widehat{w}_A > 1/2$ for $r_t^2 r_p > 1$. To see this, note that using (12), $\widehat{w}_A < 1/2$ reduces to $r_t^2 r_p < 1$ and $\widehat{w}_A > 1/2$ reduces to $r_t^2 r_p > 1$.

We now consider the two cases in the statement.

Case 1. First suppose $r_t^2 r_p = 1$. Then, using the fact that $0 < \underline{w} \leq 1/2 \leq \overline{w} < 1$ (from Assumption 2), it is direct to see that $w_A^* = w_A^\dagger = 1/2$ for all $\nu \in (0, 1)$. Then, $\Delta^\dagger = \Delta^* = 0$. Moreover, since $w_A^* = 1/2$, using (5) it follows that the equilibrium belief is correct, $\tilde{r}_p^* = r_p$.

Case 2. Suppose $r_t^2 r_p \neq 1$. Under Assumption 2, $\underline{w} < w_A^\dagger < 1/2$ if $r_t^2 r_p < 1$ and $1/2 < w_A^\dagger <$

$\underline{w}$ if $r_t^2 r_p > 1$. It is immediate to see that $\Delta^\dagger = |w_A^\dagger - 1/2| > 0$.

Next, consider an interior equilibrium so that $w_A^* = \widehat{w}_A$. Since $0 < \nu < 1$, then it is immediate to see that $\widehat{w}_A \neq w_A^\dagger$ and therefore, $\Delta^* = |\widehat{w}_A - w_A^\dagger| > 0$.

We next consider a corner solution so that $w_A^* \in \{\overline{w}, \underline{w}\}$. If $r_t^2 r_p < 1$, then $w_A^* \neq \overline{w}$. To see this, assume by contradiction that $w_A^* = \overline{w}$. Then, since there is a corner solution, $\widehat{w}_A \geq \overline{w}$. We showed above that $\widehat{w}_A < 1/2$ if $r_t^2 r_p < 1$, which means that $\overline{w} < 1/2$. However, this contradicts our assumption that $\overline{w} \geq 1/2$ (from Assumption 1). By the same logic, if $r_t^2 r_p > 1$, then $w_A^* \neq \underline{w}$.

We then consider the only possible corner solutions. If $r_t^2 r_p < 1$, then a corner solution involves $w_A^* = \underline{w}$. Then, using Assumption 2, $w_A^* = \underline{w} < w_A^\dagger < 1/2$, and $\Delta^* = |w_A^* - w_A^\dagger| = w_A^\dagger - \underline{w} > 0$. If $r_t^2 r_p > 1$, then a corner solution involves $w_A^* = \overline{w}$. Then, using Assumption 2, $1/2 < w_A^\dagger < w_A^* = \overline{w}$, and $\Delta^* = |w_A^* - w_A^\dagger| = \overline{w} - w_A^\dagger > 0$.

Finally, note that if $w_A^* \neq 1/2$, then using (5) it follows that his equilibrium belief is incorrect, $\tilde{r}_p^* \neq r_p$. ∎

Proof of Proposition 3. Suppose that $r_t^2 r_p \neq 1$. First assume that $w_A^* = \widehat{w}_A$. Then,

$$\Delta^* = |\widehat{w}_A - w_A^\dagger| = \left| \frac{\nu(r_t^2 r_p - 1)}{(1-\nu)(1 + r_t^2 r_p)} \right|$$

Since $0 < \nu < 1$, this may be rewritten

$$\Delta^* = \frac{\nu}{1-\nu} \left| \frac{r_t^2 r_p - 1}{1 + r_t^2 r_p} \right|$$

Then, since $\frac{\nu}{1-\nu}$ is strictly increasing in ν, it follows that Δ^* is also strictly increasing in ν.

Next, assume that $w_A^* \in \{\underline{w}, \overline{w}\}$. We prove for $w_A^* = \underline{w}$, but the same logic applies for the other case. Then,

$$\Delta^* = |w_A^* - w_A^\dagger| = |\underline{w} - w_A^\dagger|$$

Since $\underline{w} < w_A^\dagger$ by Assumption 2, this may be rewritten

$$\Delta^* = w_A^\dagger - \underline{w} = \frac{r_t^2 r_p}{1 + r_t^2 r_p} - \underline{w}$$

Moreover, from the proof of Proposition 2, if $w_A^* = \underline{w}$, it must follow that $w_A^\dagger < 1/2$ Then, the total disparity is

$$\Delta^\dagger + \Delta^* = \left[\frac{1}{2} - \frac{r_t^2 r_p}{1 + r_t^2 r_p}\right] + \left[\frac{r_t^2 r_p}{1 + r_t^2 r_p} - \underline{w}\right] = 1/2 - \underline{w}.$$

Finally, we show that there exists a threshold $0 < \widehat{\nu} < 1$ such that $w_A^* \in \{\underline{w}, \overline{w}\}$ for $\nu \geq \widehat{\nu}$. First, by Assumption 2, $\widehat{w}_A \in (\underline{w}, \overline{w})$ for $\nu = 0$. Second, examining (7), as $\nu \to 1$, $\widehat{w}_A \to \infty$ for $r_t^2 r_p > 1$ and $\widehat{w}_A \to -\infty$ for $r_t^2 r_p < 1$. Since $0 < \underline{w} < w_A < \overline{w} < 1$, then for $\nu = 1$, $w_A^* \in \{\underline{w}, \overline{w}\}$. Finally, $\widehat{w}_A$ is continuous and increasing in ν for $r_t^2 r_p > 1$ and decreasing in ν for $r_t^2 r_p < 1$, so there exists some $\widehat{\nu} < 1$ such that for all $\nu \geq \widehat{\nu}$, $w_A^* \in \{\underline{w}, \overline{w}\}$. Then, from above, for all $\nu \geq \widehat{\nu}$, the policing disparity is at its maximum. ∎

Proof of Proposition 4. To prove the existence of an equilibrium allocation, define a function $G : [\underline{w}, \overline{w}]^2 \to [\underline{w}, \overline{w}]^2$ given by

$$G(w_{A,1}, w_{A,2}) \equiv (w_A^{\text{br}}(r_{t,1}, \tilde{r}_{p,1}(w_{A,1}, w_{A,1}), w_A^{\text{br}}(r_{t,2}, \tilde{r}_{p,2}(w_{A,1}, w_{A,1})).$$

This is a continuous mapping from a compact and convex set to itself, so by the Brouwer fixed point theorem there must be a $(w_{A,1}^*, w_{A,2}^*)$, such that $G(w_{A,1}^*, w_{A,2}^*) = (w_{A,1}^*, w_{A,2}^*)$, which is an equilibrium allocation, with corresponding equilibrium beliefs given by $\tilde{r}_{p,i}^* = \tilde{r}_{p,i}(w_{A,1}^*, w_{A,2}^*)$.

We now show the comparative static results. First, recall we can write the equilibrium condi-

tions as the following system of equations:

$$F_1(w_{A,1}, w_{A,2}; r_{t,1}, \nu_1) = w_A^{\text{br}}(r_{t,1}, \tilde{r}_{p,1}(w_{A,1}, w_{A,2})) - w_{A,1} = 0$$
$$F_2(w_{A,1}, w_{A,2}; r_{t,1}, \nu_1) = w_A^{\text{br}}(r_{t,2}, \tilde{r}_{p,2}(w_{A,1}, w_{A,2})) - w_{A,2} = 0$$

For part (i), we prove the result as $r_{t,1}$ changes, but identical logic holds for $r_{t,2}$.

To implicitly differentiate the equilibrium conditions with respect to $r_{t,1}$, take the total derivative of F_1 and F_2 (at w_A^*, accounting for the fact that $w_{A,i}$ are a function of $r_{t,1}$:

$$\left.\frac{dF_1}{dr_{t,1}}\right|_{w_A=w_A^*} = \left(\left.\frac{\partial w_A^{\text{br}}}{\partial r_{t,1}}\right|_{w_A=w_A^*} + Y_1\left(Z_1\left.\frac{\partial w_{A,1}}{\partial r_{t,1}}\right|_{w_A=w_A^*} + Z_1\left.\frac{\partial w_{A,2}}{\partial r_{t,1}}\right|_{w_A=w_A^*}\right)\right) - \left.\frac{\partial w_{A,1}}{\partial r_{t,1}}\right|_{w_A=w_A^*} = 0 \tag{13}$$

$$\left.\frac{dF_2}{dr_{t,1}}\right|_{w_A=w_A^*} = \left(Y_1\left(Z_2\left.\frac{\partial w_{A,1}}{\partial r_{t,1}}\right|_{w_A=w_A^*} + Z_2\left.\frac{\partial w_{A,2}}{\partial r_{t,1}}\right|_{w_A=w_A^*}\right)\right) - \left.\frac{\partial w_{A,2}}{\partial r_{t,1}}\right|_{w_A=w_A^*} = 0 \tag{14}$$

where as in section B.2 we define:

$$Y_i = \left.\frac{\partial w_A^{\text{br}}}{\partial \tilde{r}_{p,i}}\right|_{\tilde{r}_{p,i}=\tilde{r}_{p,i}(w_{A,1}^*, w_{A,2}^*)}$$
$$Z_i = \left.\frac{\partial \tilde{r}_{p,i}(w_{A,1}, w_{A,2})}{\partial w_{A,1}}\right|_{w_A=w_A^*} = \left.\frac{\partial \tilde{r}_{p,i}(w_{A,1}, w_{A,2})}{\partial w_{A,2}}\right|_{w_A=w_A^*}.$$

Equations (13) and (14) are a system of two equations where we want to solve for $\frac{\partial w_{A,1}}{\partial r_{t,1}}$ and $\frac{\partial w_{A,1}}{\partial r_{t,1}}$. Define the following:

$$T_1 = \left.\frac{\partial w_{A,1}}{\partial r_{t,1}}\right|_{w_A=w_A^*} \qquad T_2 = \left.\frac{\partial w_{A,2}}{\partial r_{t,1}}\right|_{w_A=w_A^*} \qquad X = \left.\frac{\partial w_A^{\text{br}}}{\partial r_{t,1}}\right|_{w_A=w_A^*}$$

Then, we can rewrite this system of equations as

$$(X + Y_1 Z_1(T_1 + T_2)) - T_1 = 0$$
$$Y_1 Z_1(T_1 + T_2) - T_1 = 0$$

and goal is to solve for T_1 and T_2. This gives:

$$T_1 = X + \frac{XY_1Z_1}{1 - Y_1Z_1 - Y_2Z_2}$$
$$T_2 = \frac{XY_2Z_2}{1 - Y_1Z_1 - Y_2Z_2}.$$

Since we know that $X > 0$, $Y_i > 0$, and $Z_i > 0$, both of these are strictly positive if and only if $1 - Y_1Z_1 - Y_2Z_2 > 0$, which is exactly the stability condition for an interior equilibrium derived in section B.2. Finally, since $\Delta^* = |w^*_{A,i} - w^\dagger_{A,i}|$ and $w^\dagger_{A,i}$ is constant in $r_{t,1}$, then for each $i \in \{1, 2\}$, Δ^* increases in $r_{t,1}$.

For part (ii), we prove the result as ν_1 changes, but identical logic holds for ν_2. We now define the following:

$$N_1 = \left.\frac{\partial w_{A,1}}{\partial \nu_1}\right|_{w_A = w^*_A} \qquad\qquad N_2 = \left.\frac{\partial w_{A,2}}{\partial \nu_1}\right|_{w_A = w^*_A}$$

To implicitly differentiate the equilibrium conditions with respect to ν_1, take the total derivative of the equilibrium conditions at w^*_A, accounting for the fact that $w_{A,i}$ is a function of ν_1:

$$Y_1\left(Z_1N_1 + Z_1N_2 + \frac{\partial \tilde{r}_{p,1}}{\partial \nu_1}\right) - N_1 = 0 \qquad\qquad Y_2\left(Z_2N_1 + Z_2N_2\right) - N_1 = 0$$

Our goal is to solve for N_1 and N_2, which gives:

$$N_1 = \frac{\partial \tilde{r}_{p,1}}{\partial \nu_1}\left(\frac{Y_1(1 - Y_2Z_2)}{1 - Y_1Z_1 - Y_2Z_2}\right) \qquad\qquad N_2 = \frac{\partial \tilde{r}_{p,1}}{\partial \nu_1}\left(\frac{Y_1Y_2Z_2}{1 - Y_1Z_1 - Y_2Z_2}\right)$$

Again since we know that $X > 0$, $Y_i > 0$, and $Z_i > 0$, both of these are strictly positive at an interior equilibrium if and only if the stability condition is met and $\frac{\partial \tilde{r}_{p,1}}{\partial \nu_1} > 0$. This latter condition holds if $w_A = w_{A,1} + w_{A,2} > 1$ (i.e., group A receives a higher allocation than group B). Similarly, if $w_A < 1$, then $\frac{\partial \tilde{r}_{p,1}}{\partial \nu_1} < 0$ and hence both officers police group B more as ν_1 increases. ∎

D Generalizing the Model

D.1 More General Utility Function

A simple generalization is to write the utility function as:

$$u = t_A(w_A p_A)^\alpha + t_B(w_B p_B)^\alpha \tag{15}$$

where $\alpha \in (0, 1)$. So the formulation in the main text is the $\alpha = 1/2$ case. With $w_B = w - w_A$, the FOC is now:

$$\alpha t_A p_A (w_A p_A)^{\alpha-1} - \alpha p_B t_B ((w - w_A) p_B)^{\alpha-1} = 0$$

Rearranging gives:

$$1 = r_t r_p^\alpha \left(\frac{w_A}{w - w_A} \right)^{\alpha-1}$$

Without even explicitly characterizing the solution, we can see that this is just a function of r_t and r_p (rather than the primitive p_J and t_J parameters) and the solution is increasing in r_t and r_p. Solving fully:

$$w_A = \left(\frac{\left(r_t r_p^\alpha\right)^{\alpha-1}}{1 + \left(r_t r_p^\alpha\right)^{\alpha-1}} \right) w$$

which is increasing in r_t and r_p as in the main formulation.

D.2 More General Best Response

A more general way to think about the officer utility is to make weaker assumptions, but directly built into the best response function. In particular, suppose that the officer still choose an allocation $w_A \in [\underline{w}, \overline{w}]$, where:

Assumption 3. *For any belief about the relative prevalence of crime among members of the two groups, $\tilde{r}_p \in \mathbb{R}_+$ (formed by equation 5), and relative animus parameter r_t, the officer has a unique best response allocation $w_A^{br}(r_t, \tilde{r}_p) \in [\underline{w}, \overline{w}]$ such that: (i) w_A^{br} is continuous and weakly increasing in both arguments, and (ii) where $w_A^{br} \in (\underline{w}, \overline{w})$, w_A^{br} is differentiable in both arguments and strictly increasing in r_t and $\tilde{r}_p$.*

Implicit in this definition is the fact that (1) the officer's animus can be captured by a single parameter r_t (rather than the primitive t_A and t_B in the main utility function) and that the best response is only a function of the ratio of the crime rates rather than the individual crime rates. Both are also true in the main model, as well as in the more general utility function given by equation (15).

Our main technical result in this section is that a stable equilibrium allocation always exists:

Proposition 5. *For any officer utility meeting Assumption 3, there exists a stable equilibrium allocation.*

Proof of Proposition 5. If $w_A^{\text{br}}(r_t, \tilde{r}_p(\underline{w} + \epsilon)) = \underline{w}$ for some $\epsilon > 0$ or $w_A^{\text{br}}(r_t, \tilde{r}_p(\overline{w} - \epsilon)) = \overline{w}$ for some $\epsilon > 0$, then there is a stable corner equilibrium allocation. To complete the proof we need to show that if neither of these hold, there is an interior equilibrium. Let

$$F(w_A) = w_A^{\text{br}}(r_t, \tilde{r}_p(w_A)) - w_A$$

That is, $F(w_A)$ represents how he would change his allocation if starting from w_A, and an equilibrium is a point where $F(w_A^*) = 0$. If there is no stable corner solution, then it must be the case that $w_A^{\text{br}}(r_t, \tilde{r}_p(\underline{w} + \underline{\epsilon})) > \underline{w}$ for some small $\underline{\epsilon} \in (0, 1/2)$, and hence $F(\underline{w} + \underline{\epsilon}) > 0$. There must also be a $\overline{\epsilon} \in (0, 1/2)$ such that and $w_A^{\text{br}}(r_t, \tilde{r}_p(\overline{w} - \overline{\epsilon})) > 0$ and similarly $F(\overline{w} - \overline{\epsilon}) < 0$. By the continuity of w_A^{br} in $\tilde{r}_p$ and the continuity of $\tilde{r}_p$ in w_A, F is continuous in w_A, and so the intermediate value theorem implies there must be a $w_A^* \in (\underline{\epsilon}, \overline{\epsilon})$ such that $F(w_A^*) = 0$, where $F'(w_A) < 0$. Finally, since $F'(w_A) = \frac{\partial w_A^{\text{br}}}{\partial w_A} - 1$, then $F'(w_A^*) < 0 \iff \frac{\partial w_A^{\text{br}}}{\partial w_A}\Big|_{w_A = w_A^*} < 1$, and w_A^* is stable. ∎

There is no guarantee of uniqueness in this more general formulation. There may be multiple stable solutions, but there must be at least one. Also, note that since there is only one equilibrium in the main model which is a special case encompassed by Assumption 3, it must be stable. Further, the core result of the main model that behavioral policing amplifies disparities holds in this general formulation.

Proposition 6. *Under Assumption 3, in any stable interior equilibrium allocation:*
(i) w_A^* *is strictly increasing in* r_t*, and* $\frac{w_A^*}{\partial r_t} > \frac{\partial w_A^\dagger}{\partial r_t}$*, and*
(ii) If $w_A^* \neq 1/2$*, then* Δ^* *is strictly increasing in* ν*.*

Proof of Proposition 6. The general equilibrium condition for the single-officer model is as follows:

$$F(w_A^*; r_t, \nu) = w_A^{\text{br}}(r_t, \tilde{r}_p(w_A^*)) - w_A^* = 0 \tag{16}$$

where we explicitly write F to be a function of exogenous parameters of interest (here, r_t and ν).

Implicitly differentiating 16 with respect to r_t gives:

$$\frac{\partial w_A^*}{\partial r_t} = \frac{\frac{\partial w_A^{\text{br}}}{\partial r_t}}{1 - \frac{\partial w_A^{\text{br}}}{\partial \tilde{r}_p} \frac{\partial \tilde{r}_p}{\partial w_A^*}}$$

The denominator of the right hand side is $-\frac{\partial F}{\partial w_A^*}$ which at a stable solution must be positive, so the expression is positive. Further, we know that $\frac{\partial w_A^{\text{br}}}{\partial \tilde{r}_p}\frac{\partial \tilde{r}_p}{\partial w_A^*} > 0$, and so $\frac{\partial w_A^*}{\partial r_t} > \frac{\partial w_A^{\text{br}}}{\partial r_t} = \frac{\partial w_A^\dagger}{\partial r_t}$.

For part (ii), implicitly differentiating 16 with respect to ν gives:

$$\frac{\partial w_A^*}{\partial \nu} = \frac{\frac{\partial w_A^{\text{br}}}{\partial \tilde{r}_p}\frac{\partial \tilde{r}_p}{\partial \nu}}{1 - \frac{\partial w_A^{\text{br}}}{\partial \tilde{r}_p}\frac{\partial \tilde{r}_p}{\partial w_A^*}}$$

By the logic from above, the denominator of the right hand side is strictly positive. Then, this condition shows that $\frac{\partial w_A^*}{\partial \nu} > 0$ if and only if $\frac{\partial \tilde{r}_p}{\partial \nu} > 0$. However, from equation (5) in the main text, there are three cases: (1) $\frac{\partial \tilde{r}_p}{\partial \nu} > 0$ if $w_A^* > 1/2$, (2) $\frac{\partial \tilde{r}_p}{\partial \nu} < 0$ if $w_A^* < 1/2$ and (3) $\frac{\partial \tilde{r}_p}{\partial \nu} = 0$ if $w_A^* = 1/2$. Moreover, note that that if $\nu = 0$, then $w_A^* = w_A^\dagger$. Combining these observations: (1) if $w_A^* > 1/2$, then $|w_A^* - w_A^\dagger| = w_A^* - w_A^\dagger$ is increasing in ν since w_A^* is increasing away from $w_A^\dagger$ as ν increases, and (2) if $w_A^* < 1/2$, then $|w_A^* - w_A^\dagger| = w_A^\dagger - w_A^*$ is increasing in ν since w_A^* is decreasing away from $w_A^\dagger$ as ν increases. Finally, since (3) $w_A^* = w_A^\dagger$ for all ν if w_A^*, then we have shown that Δ^* is strictly increasing in ν if and only if $w_A^* \neq 1/2$. ∎

D.3 More General Beliefs (Multiple Officer Model)

There are several ways one could extend the definiton of non-conditioning bias to the multiple officer model. One potentially realistic change would be to assume that officers may do a better (or worse) job of adjusting for their own behavior than others' behavior when forming inferences about the p_J parameters. Formally, we could define the officer belief as:

$$\tilde{r}_{p,i}(w_A) = \frac{\frac{c_A}{\nu_i^s + (1-\nu_i^s)w_{A,i} + \nu_i^o + (1-\nu_i^o)w_{j,2}}}{\frac{c_B}{\nu_i^s + (1-\nu_i^s)w_{B,i} + \nu_i^o + (1-\nu_i^o)w_{B,j}}} \tag{17}$$

where the $\nu_i^s \in [0,1]$ represents how well the officer conditions for his own allocation and $\nu_i^o \in [0,1]$ represents how well he conditions on the other officer choice. A key feature of this more general belief is that as long as $\nu_i^s > 0$, it is increasing in $w_{A,i}$, meaning the officer's belief about

A's relative crime rate increases in how much he polices this group. Similarly, as long as $\nu_i^o > 0$, the officer's belief about the relative crime rate of group A increases in how much the other officer polices this group. So, while the the analysis is more complicated with this belief formation, the general feedback loop and spillover dynamics are present here as well.

E Nonlinear Returns to Policing

Returning to the original utility function, recall an additional way to motivate the diminishing returns assumption is that the marginal rate of crimes caught among group J decreases as w_J increases. Suppose the number of crimes caught is equal to $c_J = f(p_J w_J)$ where f is an increasing and concave function. Assume that the officers knows this functional form, but not the p_J parameters.

Knowing c_J and w_J, a fully Bayesian officer could then infer p_J by inverting the f function: $p_J = f^{-1}(c_J)/w_J$. The officer would then form a correct inference about the relative crime "rates" of the group, where the scare quotes highlight that the p_J parameters no long have a simple interpretation as the average crime rates of the groups:

$$\tilde{r}_p(0) = \frac{f^{-1}(c_A)/w_A}{f^{-1}(c_B)/w_B} = p_A/p_B$$

Note that if the officer now beliefs the relative crime rates are equal to c_A/c_B, he is making two mistakes: not adjusting for w_J, and also not accounting for the nonlinear effect of policing effort. In this case his belief about the relative prevalence of crime among members of each group (as a function of the allocation decision) becomes:

$$\frac{f(p_A w_A)}{f(p_B(w - w_A))}$$

Which, as long as f is increasing, is increasing in w_A. Unfortunately with this notion of naivety

there is not a natural way to come up with an "intermediate" form of the bias.

One potentially instructive special case is if f is a power function: $f(p_A w_A) = (p_A w_A)^\alpha$, $\alpha \in (0, 1)$. In this case the fully naive belief simplifies to:

$$\frac{(p_A w_A)^\alpha}{(p_B(w - w_A))^\alpha} = r_p^\alpha \left(\frac{w}{w - w_A}\right)^\alpha$$

If $r_p = 1$ this belief will be correct when $w_A = 1/2$ (and, so with no animus, the officer will again pick a correct allocation). Now when $r_p > 1$, $1 < r_p^\alpha < r_p$. So, if the officer were to allocate his time evenly between the groups, he would now *underestimate* the relative prevalence of crime among members of the group with the higher crime rate. In other words, "not understanding diminishing returns" could lead to the opposite effect as the bias we study.

Another way to model a naive officer is that he is able to "invert" the f function but does not account for the differential policing rate. Such an officer's belief becomes:

$$\frac{p_A w_A}{p_B(w - w_A)}$$

as in the baseline, so we can again define the intermediate form of naivety identically.

References

Gintis, Herbert. 2009. *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Behavior.* 2nd ed. Princeton University Press.

Knowles, John, Nicola Persico, and Petra Todd. 2001. "Racial Bias in Motor Vehicle Searches: Theory and Evidence." *Journal of Political Economy* 109 (1): 203–229.