Bureaucrats Driving Inequality in Access: Experimental Evidence from Colombia

Tara Slough*

September 10, 2018

Job market paper. Please read the latest version of the paper and the appendix.

Abstract

Bureaucrats produce and distribute public goods and services, with wide scope to influence "who gets what." Under what conditions do bureaucrats' actions create inequalities in access to public services? I contend that citizens' principal mechanism of control over bureaucrats is the complaint to a politician. When politicians respond to complaints by tightening oversight of bureaucrats, differences in citizens' access to complain induce bureaucrats to devote more effort to groups with the loudest voices. I test this theory using a national-scale factorial audit experiment of Colombia's two largest national social welfare programs to measure bureaucratic effort behaviorally. I find that bureaucrats provide less information about social welfare programs to poor citizens and internal migrants. Consistent with the theory, this bias manifests most strongly in places with greater inequalities in citizens' ability to access the state and on tasks where oversight from politicians is most likely. These results are unlikely to reflect taste-based discrimination or screening. This paper shows that inequality in access to public goods and services can emerge even when politicians' budget allocations to public goods are equitable.

*Ph.D. Candidate, Columbia University, tls2145@columbia.edu. This project is supported in part by an NSF Graduate Research Fellowship, DGE-11-44155; NSF Doctoral Dissertation Research Improvement Grant 1729334; and grants from J-PAL, Evidence in Governance and Politics (EGAP), the Center for the Study of Developement Strategies, and the Columbia Department of Political Science. The experiment was conducted with IRB approval from Columbia University (#AAAR3617) and Innovations for Poverty Action (#14442). The administrative data was collected with IRB approval from Columbia University (#AAAR0174). I am indebted to Juan Sebastian Leiva, Sofia Jaramillo, and Yuliana Silva Contreras for outstanding research assistance. I thank Innovations for Poverty Action, the Departamento Administrativo de la Función Pública, Prosperidad Social, and Departamento Nacional de Planeación for their contributions to the development and execution of this project. Thanks to Lindsay Dolan, Natalia Garbiras-Díaz, Don Green, Kolby Hanson, Shigeo Hirano, John Huber, Macartan Humphreys, Kimuli Kasara, Horacio Larreguy, Evan Lieberman, Yotam Margalit, John Marshall, Vicky Murillo, Suresh Naidu, Gareth Nellis, Cyrus Samii, Jake Shapiro, Jack Snyder, Mike Ting, and Joonseok Yang; seminar participants at WZB Social Science Center, Universidad de los Andes, and Universidad del Rosario (Political Science and Economics); policy audiences at Prosperidad Social, Departamento Nacional de Planeación, and Departamento Administrativo de la Función Pública; and conference attendees at the Harvard Experiments Working Group Conference and MPSA 2018 for comments. This project was pre-registered with EGAP prior to the realization of outcome data.

1 Introduction

Inequality in access to public goods and services represents a central challenge for economic development. What drives these inequalities in access to services? Conventional answers to this question emphasize the role of budgets for public goods and services. States require funds to provide services: where the capacity, coordination, or political will to collect taxes and raise revenues is lacking, there is less ability to deliver services. If corruption diverts funds, public services may suffer. Political targeting of funds to certain groups engenders inequalities in access. I argue that the strategic role of bureaucrats in producing public goods and services provides another explanation for inequality in access.

As "producers of public goods," bureaucrats map politicians' budget allocations into outputs, here public goods or services. I focus on the setting of service provision, and specifically on interactions between street-level bureaucrats and citizens.[1] Indeed, these interactions are the most common form of citizen-government interactions. Across 50 hybrid and democratic countries in Latin America, the Caribbean, and sub-Saharan Africa, 56 (Trinidad and Tobago) to 93 (Malawi) percent of citizens report interacting with a bureaucrat within the last twelve months.[2]

In delivering public services to citizens, bureaucrats' actions have distributional consequences. Citizens engage the bureaucrats to gain access to state benefits, from permits to subsidies. While the mode of interaction ranges across jurisdictions and services, bureaucrats' role in providing service does not. I contend that the provision of different levels of service to different citizens generates inequality in citizen access to public services. This inequality can emerge whether or not funds allocated to public goods and services are equal.

Bureaucrats' role in distributing public goods and services emerges because political principals (elected politicians) delegate program administration to bureaucrats. With this delegation comes oversight. Politicians monitor the work of bureaucrats, doling out punishments – from admonish-

[1]Following Lipsky (2010), I define street-level bureaucrats as those individuals that interface directly with citizens to implement policies that they do not create.

[2]Data comes from AfroBarometer, Round 6 (2016) and AmericasBarometer (2014). See Appendix A2.1 for details on coding.

ments to termination – and rewards – including bonuses (in some contexts) and promotion – upon observation of bureaucrats' performance. While politicians ultimately oversee bureaucrats, I consider the role of citizen complaints in directing politicians' oversight. Such complaints to political principals represent the primary form of citizen control over bureaucrats. Complaints function as a means of control by incentivizing the politician to target monitoring to specific decisions of the bureaucrat. Empirically, laws regarding citizen complaints and responses are particularly common in developing democracies and enshrine this form of citizen control over bureaucrats as a right. This paper focuses on differences in citizen propensity to complain as a driver of unequal treatment by bureaucrats that generates inequality in access to services.

This paper makes three innovations. First, I advance a theory that complaints, along with the politician's and bureaucrat's tastes drive bias in bureaucratic effort. The theory then emphasizes how such variation maps onto inequality in service provision across groups in society. Second, I develop and implement a research design capable of measuring variation in bureaucratic effort behaviorally. Finally, I leverage original administrative data to test the bias mechanisms implied by the theory in order to identify the conditions under which bias emerges.

A stylized formal model of service provision underpins my argument about how bias in bureaucratic effort leads to inequality in service provision. The model suggests that bias, or differences in the average treatment of citizens from different groups, emerges from three sources. Following Prendergast (2003, 2007), I argue that citizens' direct mechanism of control over bureaucrats is the complaint to the bureaucrat's principal, here an elected politician. Departing from existing work, I posit that the costs of engaging the bureaucracy through complaints can vary substantially across groups in the population. Where politicians exercise oversight on the basis of such complaints, a rational bureaucrat should exert more effort to provide service to those most likely to complain effectively, inducing statistical bias in effort (Becker, 1957; Phelps, 1972). I refer to this form of bias as *complaint-driven bias*.[3] The model incorporates this complaint-driven bias alongside the bureaucrat's and politician's tastes for providing service to a citizen, the remaining two sources

[3] Specifically, this bias is induced by different probabilities of citizen complaint.

of bias. I identify that two of three types of bias – the complaint-driven and politician's taste-driven bias – occur precisely *because* there is oversight of bureaucrats. In other words, unlike the bureaucrat's taste-driven biases, these biases are strategic.

I measure bias in effort with novel measures of bureaucrats' behavior in a preregistered national-scale phone audit experiment in Colombia. I audited two national social welfare programs in consultation with three national agencies overseeing the bureaucracy and these programs. The audited programs, a conditional cash transfer program (CCT) and a means testing service, have stakes: the CCT program alone confers benefits estimated at 13-17% of household consumption in the median recipient household (Fiszbein et al., 2009). These programs are implemented (in part) by bureaucrats within each local government (*alcaldía*), the petitioned entities.[4] The experiment employs a factorial design that varies both characteristics of petitioners (socioeconomic class, regional accent, and internal migrant status) and the difficulty of the petitions. The experimental design identifies bias in effort. The use of a phone audit offers rich measurement of bureaucrats' behavior, capturing access to officials and provision of information (the service).

I then leverage the national scale of the experiment to study the conditions under which bias in bureaucratic effort emerges. This allows me to probe the new mechanism that I propose – complaint-driven bias – and rule out two alternative explanations for bias: taste-based bias and screening. Specifically, I examine the role of political oversight in the administration of these programs. I do this by investigating experimental variation in the difficulty of the petitions and non-experimental variation in welfare program characteristics to understand sensitivity of bias to oversight. I also collect on an original dataset of bureaucrats and contractors in Colombia to test how bureaucrats' incentives relate to the exercise of bias. I measure incentives by comparing civil servants and contractors. At the municipal level, my data includes over 100,000 civil servants and 1,000,000 unique personnel contracts for individuals working in public administration. Finally, I draw on observational demographic and administrative data on welfare services to understand how the experimental petitioner characteristics relate to the local markets for social welfare services.

[4]For the rest of the paper, I use the Spanish word for local government, *alcaldía*, to refer to the government entities that I study.

This allows me to tease apart complaint-driven bias from politician taste-driven bias.

I find robust evidence of bias in effort: lower class individuals and internal migrants received substantially less information than their lower-middle class and resident counterparts, respectively. However, this bias is limited to specific tasks and locales. In particular, anti-poor bias is stronger in poorer places where the ostensible differences between experimental petitioners' abilities to complain is greatest. Moreover, this bias is strongest for the tasks and program most likely to receive political oversight.

The evidence is inconsistent with alternative explanations of the differential treatment. Some theories of misallocation of public services suggest that differences in treatment ("bias") could simply be efforts to screen intended from unintended potential beneficiaries. However, in contrast to screening explanations of differential treatment of citizens by bureaucrats, observed bias in effort cuts *against* the target population for these programs. Further, original measures of "red tape," oft-theorized as a screening mechanism, suggest that the type of red tape applied in this setting favors the middle class, the opposite of the target population. While I lack a direct test to eliminate the possibility of taste-based bias by the bureaucrat or politician, there is no apparent bias in favor of in-region (socially embedded) petitioners; confederates perceived no differences in respect or affect; bias emerges only with respect to specific tasks; and bias is attenuated to zero as oversight becomes less likely. As such, bias is unlikely to be driven by taste-based bias alone.

Theoretically, this paper unites and extends two foundational approaches to the analysis of agency problems in public service provision. First, it brings the moral hazard problem facing bureaucrats into a large literature on distributive politics focused on relationships between politicians and voters (Dixit and Londregan, 1995; Lindbeck and Weibull, 1987). In so doing, I argue that the strategic behavior of intermediary bureaucrats distorts politicians' desired distribution of public services. My findings imply that, even if politicians were to allocate funds equally across voters or constituencies, the lower probability that certain voters would complain to politicians induces bureaucrats to allocate services unequally.

Second, it brings citizens into models of moral hazard of bureaucrats between politicians (principals) and bureaucrats (agents). In this regard, I endogenize the information that politicians receive to guide oversight by emphasizing the role of citizen complaints to a principal (Prendergast, 2007). While citizen complaints to a political principal increase service provision, they generate inequality when some citizens are less able to engage politicians via complaint, inducing biased oversight by the principal. I find that oversight can either increase or reduce bias in bureaucratic effort, countering existing assumptions that oversight always deters bureaucratic bias (Hemker and Rink, 2017; White, Nathan, and Faller, 2015).

Empirically, this study provides novel measurement of bureaucrats' behavior as it relates to service provision. The experiment is, to my knowledge, the first experimental audit study of street-level bureaucrats in Latin America. It was purposefully conducted at national scale in a highly unequal middle-income country, making it uniquely capable of speaking to the distributive impact of bureaucrats' behaviors on service provision in a context where service provision is highly unequal and underprovision represents a central policy problem. Critically, I demonstrate that observed class-based biases in the audit outcomes are strongest in municipalities in which one of the audited programs is administratively under-provided, providing the first empirical link between informational outcomes in an audit experiment and service outputs.

Taken together, the theory and empirics suggest that socioeconomic inequalities generate political inequalities in citizens' ability to extract oversight over bureaucrats responsible for service provision. These inequalities in voice engender inequality in access to poverty reduction programs intended to mitigate existing disparities. This analysis thus reveals a new mechanism for understanding how inequalities in political voice map onto inequalities in policy outcomes, rooted in seemingly benign everyday interactions between citizens and bureaucrats. By showing that the administration of social programs by bureaucrats can reinforce inequality traps in developing contexts, I highlight the magnitude of the challenge in the design of large scale programs to effectively reduce poverty in light of recent proliferation of these programs in Latin America (De la O, 2015; Garay, 2016).

2 Theory

2.1 Model

The model consists of three actors: a citizen (or client), a street-level bureaucrat, and a politician, indexed by C, B, and P, respectively. Citizens are differentiated into two groups, $g \in \{x, y\}$ on the basis of observable ascriptive characteristics. Citizens vary in perceived costs to accessing the state. These costs are some function of physical distance, familiarity with bureaucratic procedures, and education.[5] Costs, c_C, are common knowledge and drawn from the random variable C_g which is indexed by group. F_g and f_g denote the cdf and density of C_g, respectively, and $F_g(0) = 0$. Without loss of generality, assume that $F_y(c_C) \leq F_x(c_C)$, or that F_y first order stochastically dominates F_x. Both bureaucrats and politicians, indexed by i, may have some bias toward providing the citizen of group g with the service. These tastes are represented as $\gamma_i^g \in [0, 1]$, realizations of the random variables Γ_i^g. This bias is strictly taste-based (Becker, 1957).

A bureaucrat responds to an exogenous citizen request for service by allocating effort, $e \in [0, 1]$. The service is provided with probability e. Effort is costly and is proportional to the difficulty of the task, $c_B \geq 1$. The citizen observes whether she received the service. She subsequently decides whether to complain to the politician, $q \in \{0, 1\}$ at cost c_C. Thus, the costs of complaining are non-trivial and vary across the population as a function of access to the state.

In the subgame in which the service is not realized (provided), the politician receives or does not receive a complaint from a citizen and subsequently chooses a level of effort to invest in auditing the work of the bureaucrat, $\alpha \in [0, 1]$. Politicians audit underprovision as opposed to misallocation of services. This setting characterizes many service provision settings where all citizens have a right to request service. In this sense, the model speaks clearly to the majority of tasks or programs in which equal treatment is a mandate or objective.[6]

[5] Rizzo (2018) argues that these barriers are largely psychological; this interpretation is also consistent with my argument.

[6] Even in the case of social programs with rigorously defined target populations, the first stages of enrollment are generally open to all. If the initial application for, e.g. food stamps were inaccessible, it would be impossible to ensure the program is reaching the entire target population.

With probability α, the politician is able to deliver the service to the citizen. The politician benefits reputationally and thus electorally from the increase in service provision when she detects underprovision, parameterized as $S > 0$. A biased politician will also gain utility from providing the service to a favored citizen. Failing to remedy a complaint induces a separate reputational cost of q. Finally audits are costly, which constrains the intensity of auditing; the marginal cost of an audit on a given task is c_P. To avoid corner solutions, I assume that $c_P > S + 2$. In addition, $c_P > c_B$ implying that it is costlier for politicians to recover the service than for bureaucrats to provide it in the first place. This assumption is consistent with standard arguments about bureaucratic expertise. The politician's expected utility can thus be expressed as:

$$E[U_P(\alpha)] = \alpha(S + \gamma_P^g) - (1-\alpha)(q) - \frac{c_P\alpha^2}{2} \tag{1}$$

The citizen receives a utility of $b > 0$ if she receives the service. The citizen's expected utility conditional on not having received the service from the bureaucrat is a function of the probability that the oversight process process will recover the service and her decision to complain (q), as expressed in Equation 2.

$$E[U_C(q)] = \alpha b - qc_C \tag{2}$$

Finally, consider the bureaucrat's utility.[7] He gains utility proportional to γ_B^g by (directly) providing a favored citizen with service. If a decision is reversed during the course of an audit, bureaucrats incur a penalty of $r \in [0,1]$. In practice, these costs range from a reprimand, to transfer, or even termination. The bureaucrat's expected utility is thus:

$$E[U_B(e)] = e\gamma_B^g + (1-e)(-r\alpha) - \frac{c_B e^2}{2} \tag{3}$$

2.1.1 Sequence

The game proceeds as follows:

[7] One can assume that the bureaucrat receives a fixed wage that satisfies his participation constraint; importantly, in this public sector setting, the wage does not depend on the effort exerted.

1. The bureaucrat chooses an effort level e to provide the service to the citizen.

2. The citizen decides whether or not to lodge a complaint to the politician.

3. The politician chooses the intensity of audits, α. With probability α she overturns the bureaucrat's decision.

4. Payoffs are realized.

I characterize the unique subgame perfect Nash equilibrium (SPNE) in pure strategies. The bureaucrat's allocation strategy sets $e \in [0, 1]$. The citizen's complaint strategy maps the realization of the service provided into a binary decision whether to complain to the politician $q : \{0, 1\} \to \{0, 1\}$. The politician's audit strategy then maps receipt of a complaint into auditing intensity, $\alpha : \{0, 1\} \times \{0, 1\} \to [0, 1]$.

2.2 Results

The main results characterize equilibrium effort, which allows for derivation of levels of bias. I solve the model by backward induction, beginning with the politician's decision whether or not to audit the bureaucrat's allocation. The politician's objective is clearly concave in α; differentiating Equation (1) with respect to α yields an interior optimal audit intensity of:

$$\alpha^* = \frac{S + \gamma_P^g + q}{c_P} \tag{4}$$

Note that α^* includes two types of oversight. S and γ_P^g represent "police patrols" for underprovision of the service while a complaint, q, represents a "fire alarm" (McCubbins and Schwartz, 1984). The optimal audit intensity allows for analysis of the citizen's optimal complaint strategy. In the subgame in which service is not provided, citizens will lodge a complaint if:

$$\frac{(S + \gamma_P^g + 1)}{c_P} b - c_C > \frac{S + \gamma_P^g}{c_P} b \tag{5}$$

This yields an optimal complaint strategy of:

$$q^* = \begin{cases} 1 & \text{if } c_C < \frac{b}{c_P} \\ 0 & \text{if } c_C \geq \frac{b}{c_P} \end{cases} \tag{6}$$

This implies that for higher c_C, citizens are effectively "frozen out" of contesting the bureaucrat's service provision. The audit and complaint strategies map directly into the bureaucrat's initial decision on whether to exert effort. Substituting equations 4 and 6 into the bureaucrat's objective and maximizing, the bureaucrat's optimal effort is:

$$e_g^* = \min\left\{\frac{\gamma_B^g}{c_B} + \frac{r}{c_B c_P}\left(S + \gamma_P^g + \mathbb{I}\left[c_C < \frac{b}{c_P}\right]\right), 1\right\} \tag{7}$$

where $\mathbb{I}$ is an indicator function.

Collectively e_g^*, q^*, and α^* characterize the SPNE of the game. In Appendix A1.3, I endogenize the citizen request for service by assuming that citizens pay a cost proportional to c_C to request the service in the first place. Including this cost introduces two mechanisms through which service provision changes from the baseline results. Clearly, if costs are sufficiently large relative to the benefits of receiving the service, some citizens opt out, receiving no service. Less obviously, it changes the composition of the portion of each group that requests service. This increase in the conditional probability that a citizen that requests service will complain increases the expectation of equilibrium effort across the population.

2.3 Defining and Measuring Bureaucratic Bias

There are two measures of bias implied by the model: bias in effort and inequality in outputs, defined in Definition 1. Bias in effort corresponds to different equilibrium levels of effort across groups. Inequality in outputs corresponds to different levels of ultimate service provision by group (at the conclusion of the game). I derive these quantities formally in Appendix A1.2. I assume that the effort and service afforded to each citizen is independent of the effort and service afforded to other citizens. In the context of service provision, if citizens request services at different times or different days, this assumption is plausible. Even in environments in which bureaucrats face unmanageable caseloads such that more effort for one citizen implies less effort for another, so long as citizens receive service on a first-come-first-served basis and order is not correlated with group, bias at the aggregate level can be captured by treating cases independently.

Definition 1. *Bias in effort. Bias in effort refers to the difference in expectation of equilibrium effort devoted to a citizen from each group, formally,* $\Delta = \mathbb{E}[e_x^*] - \mathbb{E}[e_y^*]$.

For the purposes of characterizing bias empirically or considering the distributional implications of bias, it is useful to define bias between groups in the aggregate. I focus on the case in which effort is interior, i.e. $e_g^* < 1$ for all citizens. I characterize bias in terms of aggregate differences by group. Define differences in the expectation of bureaucrat's tastes as $\eta_B = \mathbb{E}[\gamma_B^x] - \mathbb{E}[\gamma_B^y]$; differences in the expectation of politician's tastes as $\eta_P = \mathbb{E}[\gamma_P^x] - \mathbb{E}[\gamma_P^y]$; and differences in the probability of complaint as $\eta_Q = F_x(c_C) - F_y(c_C)$.

Proposition 1. *Between-group bias in effort. The aggregate level of bias between groups x and y evaluates to:*

$$\Delta = \underbrace{\frac{\mathbb{E}[\gamma_B^x] - \mathbb{E}[\gamma_B^y]}{c_B}}_{\textit{Bureaucrat's Tastes}} + \frac{r}{c_B c_P}\left[\underbrace{\mathbb{E}[\gamma_P^x] - \mathbb{E}[\gamma_P^y]}_{\textit{Politician's Tastes}} + \underbrace{F_x\left(\frac{b}{c_P}\right) - F_y\left(\frac{b}{c_P}\right)}_{\textit{Complaint-Driven}}\right] = \frac{\eta_B}{c_B} + \frac{r(\eta_P + \eta_Q)}{c_B c_P} \tag{8}$$

(Proof in appendix.)

Bias in effort between groups x and y can be decomposed into bias that enters through the probability of oversight, Δ_O and bias from the bureaucrat's tastes, Δ_B:

$$\Delta_O = \frac{r(\eta_P + \eta_Q)}{c_P c_B} \tag{9}$$

$$\Delta_B = \frac{\eta_B}{c_B} \tag{10}$$

Proposition 1 implies three mechanisms that drive the bias in effort and outcomes. The differences η_P and η_B capture taste-driven biases of the politician and bureaucrat, respectively. The model also implies the potential for complaint-driven bias, a form of statistical bias, parameterized as η_Q. Note that, in contrast to standard models of statistical bias in which group membership is observable and correlates with some latent trait, I show that this bias emerges even with perfect information. When one group is more able to complain, bureaucrats anticipate the increased probability of oversight by giving better service ex-ante. This is captured through a comparison of the distribution of costs for each group. The stochastic dominance assumption serves as a sufficient condition for complaint-driven bias to emerge on average (in the aggregate).

Of the three sources of bias, the complaint-based bias and the politician's taste-based bias are driven by oversight of the bureaucrat by the politician. In this sense, both forms of bias are *strategic*. Oversight is biased if $\eta_P + \eta_Q \neq 0$. This implies that the politician exerts more effort in auditing members of one group than another when service is not provided by the bureaucrat. In focusing on bureaucratic behavior, I emphasize that the probability of audit conditions the bureaucrat's level of effort. Without loss of generality, assume that oversight is biased in favor of group x, e.g. $\eta_P + \eta_Q > 0$.

Proposition 2. *Bias and the likelihood of oversight. Given $\eta_P + \eta_Q > 0$, a higher probability of audits for citizens of group x citizens increases the magnitude of the bureaucrat's bias in effort if and only if $\frac{r(\eta_P+\eta_Q)}{2c_P} > -\eta_B$. The higher probability of audits for citizens of group x will only reduce the magnitude of bias in effort if $\frac{r(\eta_P+\eta_Q)}{2c_P} < -\eta_B$. (Proof in appendix.)*

Thus, oversight can increase or decrease the level of bias in effort exerted by bureaucrats. Critically, in order for oversight to decrease bias, if oversight-driven biases favor group x, the bureaucrat's tastes must favor y ($\eta_B < 0$) and be sufficiently large in magnitude. This result emerges because the politician optimizes service provision (possibly with some preference to one group), not equality in access. I extend this analysis to investigate the relationship between oversight and bias in outputs in Appendix A1.4. Importantly, I find that if oversight is biased, bias in effort is a sufficient condition to generate inequality in outputs.

2.4 Testable Implications

The model posits three types of bureaucratic bias in effort. These biases emerge in the bureaucrat's original decision to devote effort to provide service. The distributional consequences of bias in effort for "who (ultimately) gets what" services depend on what is driving these biases. For this reason, it is important to disentangle the mechanisms underlying any observed patterns of bias. While the model implies no direct econometric test of the mechanism, Proposition 3 derives several testable implications that I use to discriminate between types of bureaucratic bias.

Proposition 3. *Tests of the mechanism. Decomposing oversight-driven and non-oversight-driven bias:*

1. *Bias in effort varies in the politician's cost of auditing if and only if oversight is biased:* $\frac{\partial \Delta}{\partial c_P} \neq 0$*, if and only if* $\Delta_O \neq 0$*. For sufficient increases in* c_P*, bias in effort attenuates toward zero.*

2. *The magnitude of bias in effort increases in the strength of bureaucratic incentives if and only if oversight is biased:* $\frac{\partial \Delta}{\partial r} > 0$ (< 0)*, if and only if* $\Delta_O > 0$ (< 0)*.*

Discriminating politician's tastes from complaint-driven bias:

1. *The magnitude of bias in effort increases in the between-group differences in ability to complain:* $\frac{\partial \Delta}{\partial \eta_Q} > 0$ (< 0) *if* $\eta_Q > 0$ (< 0)*.*

2. *The magnitude of bias in effort increases in the between-group differences in the politician's tastes:* $\frac{\partial \Delta}{\partial \eta_P} > 0$ (< 0) *if* $\eta_P > 0$ (< 0)*.*

Proposition 3 guides efforts to test the mechanisms described in the model and in Table 1. I proceed in two steps. First, I test for evidence of oversight-driven bias. This distinguishes bias coming from bureaucrats' tastes (Δ_B) from the bias that comes from different probabilities of oversight (Δ_O). As in Table 1, oversight-driven bias incorporates both politician tastes and complaint-driven bias. To do this, I examine variation in bias with respect to the two parameters that should only

	Classification			Testable Implications			
Bias Mechanism	Bias Type	Oversight-driven	Case	$\frac{\partial \Delta}{\partial c_P}$	$\frac{\partial \Delta}{\partial r}$	$\frac{\partial \Delta}{\partial \eta_Q}$	$\frac{\partial \Delta}{\partial \eta_P}$
Complaint-driven: Citizens from group x are more likely to complain than from group y which draws a higher likelihood of auditing by the politician. The bureaucrat devotes more effort to x in anticipation of higher probability of audit.	Statistical	Yes	$\Delta > 0$	$\neq 0^*$	> 0	> 0	0
			$\Delta < 0$	$\neq 0^*$	< 0	< 0	0
Politician's tastes: The politician prefers to audit service to group x more than to group y. The bureaucrat devotes more effort to x in anticipation of higher probability of audit.	Taste-based	Yes	$\Delta > 0$	$\neq 0^*$	> 0	0	> 0
			$\Delta < 0$	$\neq 0^*$	< 0	0	< 0
Bureaucrat's tastes: Bureaucrat prefers providing service to group x over group y.	Taste-based	No	$\Delta > 0$	0	0	0	0
			$\Delta < 0$	0	0	0	0

Table 1: Summary of the bias mechanisms implied by the theory and the testable implications for distinguishing the mechanisms. Bias is defined as a difference between two groups: when $\Delta > 0$ x is preferred to y and when $\Delta < 0$, y is preferred to x, so testable implications should be seen as *magnitudes*. *Note that sufficient increases in c_P attenuate bias toward zero.

drive variation in the oversight mechanism, the politician's cost of effort, c_P, and the "bite" of possible punishment, r. If bias varies in these two measures, there is evidence of oversight-driven bias.

Conditional on finding evidence of oversight-driven bias, I aim to distinguish between politician tastes and citizen propensity to complain, the two components of oversight-driven bias. To do this, I examine variation in citizens' cost of complaint (enters through η_Q). As the "distance" between petitioner types' costs of complaint, η_Q, increases, so too should the relative magnitude of complaint-driven bias. In particular, if the magnitude of bias increases as η_Q increases, there is evidence that bias comprises complaint-driven bias. This test is able to distinguish between politician tastes and complaint-driven bias when $Cov(\eta_P, \eta_Q)$ is small. I also consider a parallel test with regard to politician incentives that may drive politician tastes, η_P with a parallel logic.

3 Case Context

I measure variation in bureaucratic discretion at national scale in Colombia. Writings on state capacity in Colombia have long focused on a two-century history of civil wars as a cause or consequence of state weakness (Centeno, 2003; González González, 2014). Nevertheless, studies of bureaucracy typically characterize Colombia's national bureaucracy as comparatively "Weberian" by regional standards (Evans and Rauch, 1999; Mayka, 2016). Analyses of the World Bank's Worldwide Governance Indicators echo these findings (see Appendix A2.2). Aside from measures of political violence, the other governance indicators approximate the world median, rank in the highest tercile in Latin America, and rank in the lowest decile of OECD countries.[8] At the national level, bureaucratic capacity in Colombia is believed to vary with the relative concentration of "technocrats" and patronage employees (e.g., Schmidt, 1974; Dargent, 2016) and the presence of national bureaucrats across the territory (Acemoglu, García-Jimeno, and Robinson, 2015). Less is known about municipal bureaucrats – the subjects of this investigation – though dozens of interviews with national bureaucrats and participant observation in *alcaldías* suggest tremendous variance in professionalism, competence, and outputs.

3.1 Municipal Politics in Colombia

Since political decentralization in the late 1980s, Colombia's 1102 municipalities have assumed responsibility for most services, ranging from roads to education.[9] Decentralization gave rise to larger and relatively more professional municipal public administration (Fizbein, 1997). Important for this project, some national programs are implemented "on the ground" within municipalities by municipal bureaucrats.

Municipalities are governed by a mayor and local council (*concejo*) of seven to 45 councilors, according to population.[10] Mayors are elected every four years by plurality vote and are limited to serving one consecutive term. In contrast, councilors are elected by optional open list PR without

[8]Colombia joined the OECD in July 2018.

[9]There are 1122 territorial units, however 20 are *corregimientos* as opposed to municipalities. There were elections in 1100 municipalities 2015; it is from these municipalities that the sample of experimental municipalities was drawn.

[10]Bogotá has a local council of 45 councilors; the next largest councils have 21 members.

term limits. In these elections, parties are weak and the role of ideology in elections is limited. According to the Colombian party classification by Fergusson et al. (2018), 232 current mayors represent parties classifiable as “left” or “right;” 543 mayors represent parties without an identifiable ideology; and 325 mayors ran without a party (as their own party). In these contexts, the distribution of public and private goods arguably constitutes the basis of political competition.

3.2 Bureaucratic Hiring

At the municipal level, bureaucrats are hired and overseen by local politicians. Politicians staff the public sector via two hiring mechanisms: civil service (called *empleados de planta*) and contracting. The degree to which mayors delegate staffing the *alcaldía* varies predictably with the size of the municipality. In small municipalities, contracts are signed directly by the mayor (on behalf of the *alcaldía*); in larger municipalities, high-level political appointees sign off.

On average, contractors are less expensive to employ than civil servants. From a purely practical standpoint, contractors have higher powered incentives than are typical in the public sector setting (Dixit, 2002). Contracts are short term, on average less than five months, whereas civil service employees empirically enjoy longer tenure and thus higher job security. The processes of contracting are relatively lax and the share of contractors that could reasonably be considered “patronage hires” is certainly higher than the share of of civil servants, though this is an admittedly imperfect proxy.

Municipalities are legally constrained to a maximum budget share that they can devote to remuneration of civil servants. These constraints are a function of a municipal classification based on revenues and population (*Ley 617 de 2000*). Municipal workforces are thus supplemented by contractors. From the perspective of a politician, contracting provides a relatively flexible means of delivering jobs. Yet, limited tenure provides few opportunities to develop expertise and, as practiced, contracting yields high fluctuation in actual staffing levels within the *alcaldías* throughout the year, possibly reducing productivity of the bureaucracy.

3.3 Complaints and Oversight

The model focuses on the fundamental role of citizen complaints as a means of seeking oversight over bureaucrats. As in much of Latin America, Colombia provides substantial legal rights for making complaints. The Colombian Constitution of 1991 mandates the right to access public information for all citizens (Article 74) and statutory law allows Colombians to request "recognition of rights, intervention of a government entity or official, legal resolution, service provision, information, copies of documents, consultations, [various forms of] complaints, and claims" by written petition (translated from *Ley 1755 de 2015*, Article 14). In turn, the government has three weeks (fifteen business days) to respond to the petition. Complaints also emerge through less formal channels; the distinction is not relevant for the purposes of the theory so long as costs vary across the population.

Some complaints are ostensibly handled by other higher-level bureaucrats, while others rise to local mayors. While such instances are, in principle, hardly newsworthy, mayors do audit the local administration of social programs. Yet even in this context, there are regular news reports of mayors responding to complaints about the function of social programs, typically by auditing local rolls of beneficiaries.[11] For such action to influence bureaucrats' behavior, threats of reprimand must be perceived. An original survey of street-level bureaucrats in *alcaldías* in Bogotá and Cundinamarca finds that 78% (57/73) of these bureaucrats perceive that a mistaken decision would be punished (with varying severity) and decisions would be reversed (see Appendix A3 for survey information).

While nationwide data on complaints to local entities is not collected, analysis of over 440,000 complaints filed in public entities in Bogotá from January 2017 through June 2018 and compiled by the Veeduría Distrital provides two stylized facts of note. First, virtually all complaints relate to service provision and approximately 125,000 complaints (28 percent) explicitly relate to bureau-

[11] Recent newsworthy investigations include investigation of how a councilor in Mosquera, Cundinamarca made it onto a list of means-tested beneficiaries for social programs (SISBÉN); a scam to stuff the rolls for Adultos Mayor, a subsidy for senior citizens, in Florencia, Caquetá; and a general audit of the SISBÉN rolls in Pitalito, Huila. The first two investigations occurred in response to citizen complaints.

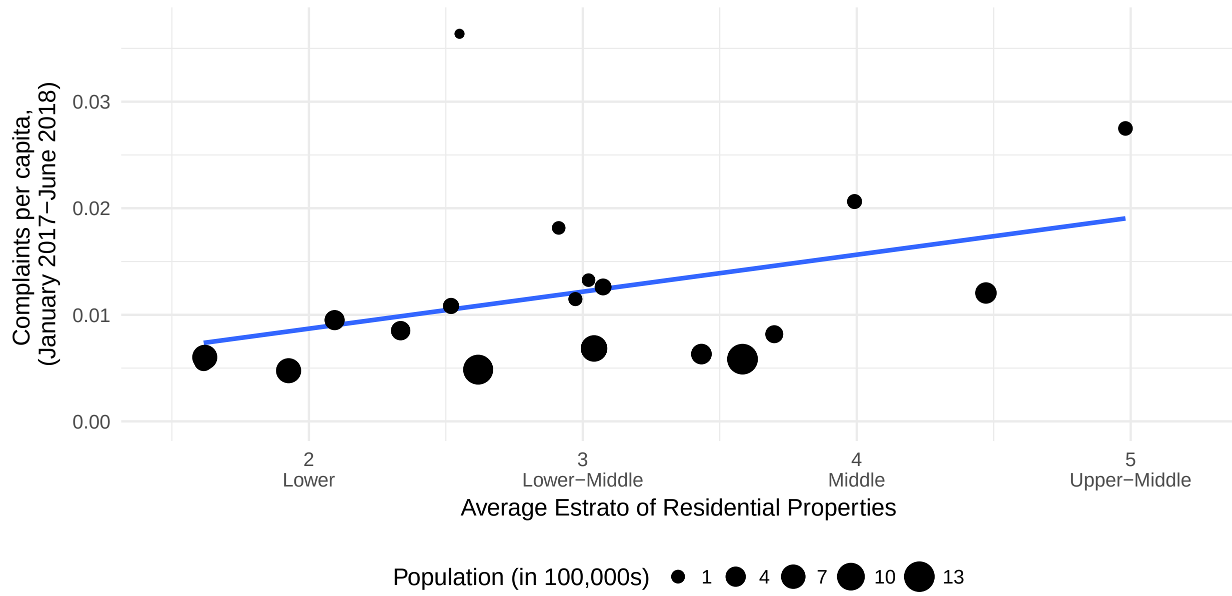

Figure 1: Rate of complaints filed by locality in Bogotá by average wealth of the locality.

crats' actions in service provision. This represents the most common class of complaints. Second, leveraging Bogotá's geographic segregation by class, analysis of complaints at the locality level suggests that there are substantially *more* complaints per capita in richer localities. Figure 1 shows a positive association between the the average class designation (*estrato*) of residential properties in each locality and the per-capita rate of complaint submission to each local *alcaldía*. This pattern emerges despite the fact that service provision is thought to be better in richer localities – those where complaints are most frequent. (See Appendix A4 for details.)

3.4 Social Welfare Services

I audit two nationwide social welfare programs that are administered, in part, by officials embedded in every municipal government. Specifically, the national government agencies that oversee these programs maintain agreements (*convenios*) with each municipality that mandates that local *alcaldías* hire the local program officials. The programs, the System for the Identification of Potential Beneficiaries of Social Programs (SISBÉN) and Más Familias en Acción (MFA), provide coverage on a nationwide geographic scale. These programs provide access to or transfers to low-income Colombians. Colombia is among world's most unequal countries with a very large low-income population, indicating a large but not universal pool of potential beneficiaries for each program. Existing analyses of the programs suggest very different levels of politicization between the two programs.

Created in 1995, SISBÉN is a household index of assets used for qualification for means-tested social programs.[12] SISBÉN is a prerequisite to access subsidized health insurance and most social programs, including Más Familias en Acción. The municipal service associated with SISBÉN is the administration and readministration of household survey of assets. The survey is then sent to the National Department of Planning (DNP), which generates a score on an index through a private formula. At present, over 36 million Colombians are registered in the SISBÉN system, representing approximately 70 percent of the population. SISBÉN is thought to be manipulated by citizens (i.e. by hiding assets during the survey) or local bureaucrats. Past iterations of the formula have been changed to reduce local discretion. I identify many municipalities with far more SISBÉN registrations than the ostensible population, supporting longstanding views of politicization (Camacho and Conover, 2011).

Created in 2002, MFA (formerly *Familias en Acción*), is Colombia's national conditional cash transfer program that provides subsidies to (mostly) mothers on the basis of compliance with their children's health and educational requisites. Each municipality has one MFA official (*enlace*), though the program is overseen by a much larger office in larger municipalities. These individuals provide information to recipients and monitor compliance with community participation aspects of the program. MFA enrolls between 15-20 percent of Colombian households, providing subsidies estimated at approximately 15 percent of median household consumption among recipients (Fiszbein et al., 2009). In contrast to SISBÉN, MFA is designed and implemented as per transparent and uniform programmatic guidelines, at least relative to other conditional cash transfer programs in the region (De la O, 2015).

[12] SISBÉN is not a social program in the conventional sense; it does not confer a direct benefit on registrants. However, this is the means by which Colombians access social benefits and is widely regarded as a central piece of the social policy landscape.

4 Research Design

To test the implications of my theory of bureaucratic biases, the research design must be able to (a) elicit bias and (b) measure bureaucratic effort, the outcome of interest. I utilize phone audits to facilitate direct behavioral measurement of bureaucrats' response to requests for service (information). I implement these audits at national scale in Colombia. In order to elicit bias, I randomly assign petitioner characteristics. The random assignment permits direct identification of bias, Δ. I also randomize the several characteristics of the petition itself to test theoretical claims about the sensitivity of bias to the cost of effort and to reduce detection of the audits.

4.1 Audit Experiment

The unit of random assignment is the petition: a call with a request for a service. I utilize a factorial experiment to randomly assign characteristics of petitions. The treatments provide exogenous manipulations in the bureaucrat's and politician's marginal cost of effort, c_B and c_P, as well as in observable attributes of the petitioner, g. Several aspects of petitions are constant across calls. Given the composition of MFA recipients – mostly mothers – all petitioners are female. While all calls were made from Bogotá, the outgoing phone numbers appeared as standard cell phone numbers. In Colombia, cell phone numbers do not convey geographic information.

4.1.1 Treatments

In order to induce experimental variation in bias, I manipulate identity-based characteristics of the petitioners. These characteristics are rooted in the Colombian social and political context, and serve as the analogue to the groups in the theoretical model. First, I assign the socioeconomic class (*estrato*) of the caller. Since independence (and before), class has represented an organizing feature of Colombian society and political life (Martz, 1997; Sanders, 2004). Given the focus of the social programs audited, I differentiate between low- and lower middle-class callers.[13] Focus groups with Colombians of different socioeconomic classes and observation of calls in a government call

[13] For readers familiar with Colombia's class categorization system, lower class refers to *estratos* 1 and 2. Middle class refers to *estrato* 3. Politically-defined class categories (*estratos*) range from 1 to 6. Appendix A5.2 reflects the distribution of individuals by class in the population as of 2005.

center suggest several avenues in which the class of a caller can be immediately distinguished by phone. Specifically, callers in the two groups vary in their vocabulary, salutations for figures of authority (bureaucrats), and the framing of questions. While class communicates a variety of features, I assume that on average lower middle-class individuals have relatively greater ability to engage (access) the bureaucracy than lower-class individuals.

A second identity-based manipulation relies on regionalism in Colombia. Due to colonial settlement patterns, a rugged topography, and limited central government penetration, social and political life flourished within regions in the 19th century (González González, 2014; Uribe de Hincapié and Álvarez Gaviria, 1998). Two centuries later, regional accents remain quite distinct and regions maintain relatively distinct cultures and political organization (Ocampo, 2014). I randomly assign Bogotano (*rolo*), Paisa, and Costeño regional accents. These accents represent the most widely-spoken accents in Colombia and are collectively spoken by $\approx$60 percent of the Colombian population. Appendix A5.1 includes maps of the geographical coverage of these accents.

The accents probe concepts of embededness of petitioners. Recent arguments have emphasized the embeddedness of bureaucrats within state or local governments, using bureaucrat region (resp. state) of origin to measure "embeddedness" (Pepinsky, Pierskalla, and Sacks, 2017; Bhavnani and Lee, 2018). Extending this logic, in a context where the vast majority of local bureaucrats are drawn from local communities, regional accent should provide a signal of embeddedness or lack thereof. There is not an obvious ranking of costs of access within the accents, though long-held stereotypes and data on outcomes hold that service is substantially better in the highlands (home of the Bogotano and Paisa accents) than on the Caribbean coast (home of the Costeño accent).

The third identity-based treatment focuses on migrant status. This treatment is communicated during the petition through a statement that the individual in need of the service is a recent (internal) migrant. The "resident" condition does not provide this information. Rates of internal migration have long stood among the highest in Latin America and encompass both ordinary and conflict-induced migration (Martine, 1975; Ibáñez and Vélez, 2008).[14] See Appendix A5.3 for estimates

[14]For example, 7.4 million Colombians are internally displaced, representing approximately 15% of the Colombian population.

of rates of internal migration in Colombia. The migrant condition signals two potentially relevant features. First, migrants are apt to have less familiarity with a municipality, suggesting higher costs of access. Second, following Gaikwad and Nellis (2017), migrant status indicates a much lower likelihood of voting, as there is no absentee voting and re-registering in a new municipality is cumbersome and occurs during relatively short temporal windows prior to elections.

The final manipulation varies the technical specificity of the petition.[15] For both programs, the "easy" version of the question simply asks how a non-enrolled/registered citizen could enter SISBÉN or MFA, respectively. The "technical" version of the questions poses a question about a situation with specific technical program requirement. This manipulation allows me to test sensitivity of bias to costs to the bureaucrat which aids in discriminating between bias mechanisms.

Collectively, all four factors are fully crossed, yielding a $2 \times 3 \times 2 \times 2$ factorial design summarized in Table 2. This yields 24 distinct treatments for each of two programs, though I analyze along the margins (by attribute). Note that the twelve confederates were actresses. All confederates voiced both low- and middle-class petitions. To maximize authenticity, actresses voiced only their own regional accent and calls were divided between four actresses per region of origin. Calls were randomly assigned to each confederate.

All calls were recorded. I hired Colombian coders to listen to all of the recordings to double code call characteristics and responses. Given that the coders were blinded to treatment assignment, this yields one measure of compliance with treatment assignment. I define compliance as a measure of whether coders reported hearing the assigned factors (i.e. if they heard a Costeña petitioner on a call assigned to a Costeño accent). The rates of compliance are reported in the final column of Table 2. A more detailed analysis of compliance is reported in Appendix A11.1. While I cannot know what bureaucrats intuited, rates of compliance in the double coding exercise are quite high across all factors and levels, alleviating major concerns.

[15]Under the assumptions of the model, that $c_P > c_B$ this may or may not also increase the politician's cost of effort c_P commensurately.

Factor	Levels	Mode of Administration	Compliance Rate
EFFORT (COSTS)			
Difficulty of Request	• Easy • Difficult	Technical specificity of request to petition, as defined by national government partners	99.3% 99.2%
BIAS			
Regional Accent	• Bogotá • Paisa • Costeño	Regional accent of caller employed in interaction with bureaucrats.	99.7% 98.4% 98.7%
Socioeconomic Class	• Low • Lower Middle	Vocabulary, salutations, and framing of the interaction.*	76.7% 79.3%
Stated Migrant Status	• Migrant • Resident	One statement in delivery of petition (migrant). No reference to internal migration in resident's call.	97.3% 95.0%

Table 2: Factors and levels employed in the factorial design. Compliance rates are calculated as the proportion of calls correctly classified by double coders out of the number of calls assigned to each level for which the factor was revealed (see Section 4.3 for details on the rollout and estimation). *Note that while the framing of the interaction varied across class but the statement of the question itself was stated identically for both classes.

4.1.2 Sampling, Assignment

The sample of *alcaldías* was selected with two opposing objectives. First, by maximizing the number of petitions made to the same *alcaldía*, I increase statistical efficiency and allow the estimation of within-*alcaldía* treatment effects. Second, I seek to minimize the probability of detection. In order to achieve both objectives, I stratify municipalities into three groups by estimated 2018 population.[16] Note that Bogotá provides services at the level of 20 localities. The entities are thus municipal *alcaldías* outside of Bogotá and local *alcaldías* in Bogotá. The number of petitions varies by stratum. In the large stratum, six petitions were assigned, three each for SISBEN and MFA. In the medium stratum, four petitions were assigned, two per program. In the small stratum, one petition was assigned per program. The distribution and number of petitions is depicted in Table 3.

Blocking by *alcaldía* was used in order to ensure maximal within variation and avoid detection. The blocking procedures are detailed in Appendix A9.1. The blocking ensures that each *alcaldía*

[16]The last Colombian census was conducted in 2005. I use population estimates from the Departamento Administrativo Nacional de Estadística (DANE) to make this determination.

				n Petitions per Entity			
Stratum	Stratum Size	Population threshold	Sample	SISBÉN	MFA	Total	Total Petitions
Large	80	> 100,000	All	3	3	6	480
Medium	140	[35,000, 100,000)	All	2	2	4	560
Small	898	< 35,000	398	1	1	2	796
Total	1118		618				1836

Table 3: Sample of municipalities (or localities) and number of petitions. Note that in the small stratum, localities are selected proportionally to population size. All population data from 2018 estimates from DANE.

received equal numbers of low- and middle-class petitioners; equal numbers of easy and difficult questions; and received half the petitions from migrants. To minimize the likelihood of detection, the more specific technical questions were never repeated within an *alcaldía*. This implies that the ratio of easy to technical questions in the large stratum was 2:1. The estimation strategy accounts for these differential probabilities of assignment to treatment. Further, no *alcaldía* received more than one call from the same class/accent combination or was asked the same question more than once.

The order of calls was randomly assigned to space out calls to the same *alcaldía* over approximately four weeks. The assignment process for this rollout procedure is documented in Appendix A9.3. In general, first attempts of each call were consistent with the assigned ordering (within morning or afternoon), but repeated attempts complicate this mapping. Finally, the time of day – morning or afternoon– within each *alcaldía's* hours of service was randomly assigned. Each *alcaldía* received equal numbers of calls at each time. Ultimately, just 7 calls were detected. Lack of systematic patterns in these calls suggests that the threat of detection was minimal (see Appendix A11.2).

4.1.3 Outcomes

The audits measure a rich set of behavioral outcomes relating to service provision through the course of the call. Appendix 3 clarifies the sequencing of calls and outcome measurement. To measure service provision, all enumerators filled out an instrument to document the trajectory, outcomes, and information conveyed in each call. Further, all calls were recorded. Two trained research assistants listened to every recording and double entered all data, including additional measures of compliance and qualitative observations of each call.

I focus on three classes of outcomes. For the *alcaldías* reached by phone, I provide a mapping of the call through the *alcaldía*. Since dispatchers who answer are not generally program officers, I measure whether a petitioner was provided access to a program officer in order to make the petition. I map the mode of transmission through the bureaucracy to measure the accessibility and navigability of service providers within local bureaucracies. In particular, I measure four outcomes dichotomously: (1) whether the dispatcher identified himself/herself; (2) whether the petitioner was able to make (state) the petition; (3) whether the petitioner was connected to at least a second official; and (4) whether a program officer for SISBÉN or MFA from an ex-ante pre-treatment list was identified.

Most important, I measure agents' responses to the petition. I focus on the amount and veracity of information provided relative to the benchmark (correct) answers specified by the national government agencies that oversee each program. Outcomes at this stage also include a measure of red tape: whether an official asked for *extra* requirements not specified by program guidelines and whether petitioners were asked to come "in person" without further guidance. I measure five pre-registered outcomes of interest: (1) whether the correct, complete answer was provided; (2) whether partial information was provided; (3) whether any actionable information was provided (a sum of #1 and #2); (4) whether the petitioner was asked to come to the *alcaldía* in person without further instruction; and (5) whether red tape was solicited. The "come to the *alcaldía*" response merits some clarification. All services require an eventual trip to the *alcaldía*. Arriving without the requisite documents imposes additional costs on the petitioner, regardless of the bureaucrat's intent.[17]

Finally, I use confederate ratings of service as a benchmark to the behavioral measures of service provision. Here I examine whether the perceptions of the petitioner align with experiences of

[17] Two plausible interpretations of the "come to the *alcaldía*" response include: (a) political capture is more likely to occur in person than on the phone; or (b) the bureaucrat believes that the petitioner will only understand in person. I remain agnostic between these interpretations but maintain that failure to provide information imposes an additional cost to petitioners.

service. A z-score index includes assessments of competence, knowledge, respect, trustworthiness, and satisfaction.

4.1.4 Ethical Considerations

Government audit experiments generally raise three ethical concerns: the use of deception, the protection of subjects, and the waste of time and public resources. I address the concern of deception through a novel model of collaboration with national government agencies. The collaboration included consultation throughout the research design process with the agency overseeing the Colombian bureaucracy at the national level (the Administrative Department of Public Administration) as well as the agencies overseeing SISBÉN (National Department of Planning) and MFA (Department of Social Prosperity). These agencies provided guidance on the programs to be audited, the content of the audits, the correct answers to the audits, and some administrative data. In exchange, I conducted the experiment independently with external funding and produced and presented a policy report to each agency in June 2018.

Notably, these agencies conduct their own "mystery shopper" (*cliente incógnito*) audits of employees and contractors periodically, though my collaborators do not recall randomizing any components. By conducting the audits independently, I provide additional privacy protections to subjects (audited bureaucrats) in a manner that cannot be guaranteed in government audits.

In terms of wasting of time and resources, the costs to public entities in Colombia should be weighed against the benefits of this original data and report. The upper bound on the costs to these entities can be quantified quite simply. The answered calls (i.e. those that occupied the time of public employees) total under 200 hours. At the maximum monthly salary for the maximum rank of employee ("*Profesional*") that would have engaged with a caller, the upper bound on the cost of these calls totals $\$2,644$ USD.[18] This totals less than 10 months for one employee at the official minimum wage, a common local benchmark.

[18] Calculated from Decreto No. 309 de 2018. Maximum public sector salaries are benchmarked by municipal "category," a measure of population and local development. This calculation uses the highest salary in the "special" category of municipality (highest paying) and is thus a strict upper bound.

4.2 Administrative Data

In order to understand the relationship between bureaucratic organization and behavioral measures of service provision by street-level bureaucrats, I leverage several original administrative datasets on public sector personnel in Colombia. The first datasets contains individual public employees working as civil servants in Colombia with self-entered name, position, work experience, and education. Outside of Bogotá, I use the data from the *Sistema de Información y Gestión del Empleo Público* (SIGEP) and inside Bogotá I use the city-level equivalent (SIDEAP). This provides data on public employees hired under the law for public employment.[19]

Second, I generate a list of contractors working for municipal governments using data from Colombia Compra Eficiente, the national government entity that oversees public procurement. This source contains data on public sector contractors working in all government entities. While contracting is generally cheaper than hiring civil servants, it concurrently serves as a means to preserve patronage in the face of civil service laws. As such, not all contractors are patronage employees, but my measurement relies on the assumption that contractors are more likely than civil servants to be patronage employees and that aggregate patterns of contracting by *alcaldía* measure the use of contracts for patronage.

Both datasets are entered and maintained by officials within each *alcaldía*. In full, 82 percent of the employees reached in the experiment and program officers appear in this combined list. In the cases in which I am unable to identify the employee, 4.7 percent come from municipalities that do not use one or both datasets.

I also leverage additional demographic and electoral data. Demographic data on the characteristics of municipalities allow me to contextualize the identities portrayed in the calls to local constituencies. In terms of the theoretical model, such data provides some information about the shape of the distribution of costs, f_g, within a given population of citizens.

With the electoral data compiled at Universidad de los Andes, I seek to measure political competition, a feature which should increase the politician's incentives to provide public goods, S, and

[19] *Ley 80 de 2003.*

may covary with the tastes of elected politicians, γ_P^g. Standard measures of political competition are complicated by features of municipal politics in the Colombian context. First, at the municipal level, party labels do not signal ideology and high rates of party switching suggest that analyses at the party level contain little meaningful information. Further, measures based on raw electoral data such as mayoral margin of victory (distance between the winner and runner up) exhibit very little serial correlation.[20] Thus, observing a close election at time t provides essentially no information about competitiveness at $t + 1$. As a result, I develop other measures of political competition. In particular, I look at the frequency with which individuals are re-elected and which family names concentrate among local council members (*concejales*) over a twenty year panel (six electoral cycles). These measures build upon those used by Acemoglu et al. (2008) to measure political inequality in the department of Cundinamarca, Colombia in the nineteenth century.

4.3 Estimation

The estimation of causal quantities in the experiment accounts for the process of selection and the delivery of treatment during the course of the interactions with local government officials. Post-treatment selection represents a threat to inference in existing audit experiments (Coppock, 2018). In the present experiment, if a Costeña petitioner was more likely to be able to state the petition, conditioning the sample on having made a petition may induce bias in estimates of the effect of accent on informational responses.

To overcome this limitation, the attributes (factors) in the factorial design were revealed at three distinct points in the call, as depicted in Table 4. This defines three relevant samples: all attempted calls, all answered calls, and all calls in which the petition was delivered. Factors not yet revealed in a given sample are referred to as *placebos*; factors revealed within the sample are referred to as *treatments*; and factors revealed prior to revelation of the a sample are regarded as *pre-treatment covariates*. Point estimates on the treatment variables (in the relevant sample) are causally identified. Taking advantage of the rollout of factors during the course of the call

[20]The quality of electoral data is generally high. As such, lack of serial correlation is ostensibly driven by characteristics of electoral competition, not data limitations.

increases statistical efficiency and while avoiding the threat of bias induced by post-treatment sample selection.

I seek to estimate the Average Marginal Component Effect (AMCE) of the randomly-assigned treatments. This effect is the marginal effect of each factor, averaged over the joint distribution over other factors. I account for the differential probabilities of assignment to easy and technical questions across the strata of municipalities with two estimators. I estimate the sample AMCE using inverse probability weighting (IPW) or *alcaldía* fixed effects.[21] The latter strategy examines differential treatment of petitioners leveraging variation only from within the same *alcaldía*.

I estimate the AMCE with either estimator with regressions of the form of Equation 11 using OLS with heteroskedasticity-robust standard errors. Note that these standard errors correspond to the level of treatment assignment: the petition. The set of indicators in the regression model corresponds to the factor levels in the design, here $\mathbf{Z} = \{\text{Afternoon}_i, \text{Technical}_i, \text{Lower Middle Class}_i,$ $\text{Bogotá accent}_i, \text{Costeño accent}_i, \text{ and Resident}_i\}$. In Equation 11, $\boldsymbol{\psi}_m$ indicates municipality fixed effects; IPW specifications do not include this term.[22] κ_p indicates a vector of program (SISBÉN or MFA) fixed effects that are included in all specifications.

$$Y_{ipm} = \sum_{j \in \mathbf{Z}} \beta_j Z_i^j + \kappa_p + \boldsymbol{\psi}_m + \epsilon_{ipm} \tag{11}$$

In order to estimate the conditional AMCE with respect to institutional, demographic, or political covariates, I estimate Equation 12 using OLS with heteroskedasticity-robust standard errors. In this equation, moderators and covariates are represented by the variable X_i (resp. X_m). The conditional AMCEs estimated in are causally identified under the conditions specified above for the AMCE. The conditional AMCEs are estimated by β_j and $\beta_j + \gamma_j$, where j indexes the treatment level. The difference in conditional AMCEs (γ_j) is not causally identified absent additional

[21]The implied AMCE estimand coming from the two estimators is subtly different. For the IPW estimator, the sample AMCE (for a factor Z) is $\mathbb{E}[Y_{ipm}(Z_i = 1, \mathbf{Z}) - Y_{ipm}(Z_i = 0, \mathbf{Z})]$, where Z_i is the factor of interest, $\mathbf{Z}$ is a vector of all other attributes. For the FE estimator, the AMCE is given by $\sum_{m \in M} w_m \mathbb{E}[Y_{ipm}(Z_i = 1, \mathbf{Z}) - Y_{ipm}(Z_i = 0, \mathbf{Z})]$, where w_m is a weight proportional to the inverse of the variance within the block.

[22]Fixed effects are highly prognostic of outcomes. Regressions of the main outcomes on a vector of municipality fixed effects yield R^2's of 0.37 to 0.55.

Call Made 1836 Calls 618 Municipalities		$\longrightarrow$	**Call Answered** 1194 Calls 466 Municipalities		$\longrightarrow$	**Petition Made** 911 Calls 424 Municipalities	
(Time of Day)	✓		(Time of Day)	Not point identified		(Time of Day)	Not point identified
Accent	Not revealed		Accent	✓		Accent	Not point identified
Class	Not revealed		Class	✓		Class	Not point identified
Difficulty	Not revealed		Difficulty	Not revealed		Difficulty	✓
Migrant Status	Not revealed		Migrant Status	Not revealed		Migrant Status	✓

Table 4: Timing of treatment delivery during the process of a call. Attributes that are "not yet revealed" serve as placebos in the outcomes prior to their revelation. The timing of treatment delivery defines the relevant sample upon which effects are estimated. In endogenous samples where treatment effects are "not point identified," the attributes are included as covariates but (point) estimates are not causally identified.

assumptions.

$$Y_{ipm} = \sum_{j \in \mathbf{Z}} \beta_j Z_i^j + \sum_{j \in \mathbf{Z}} \gamma_j Z_i^j X_i + \kappa_p + \sum_{p \in P} \alpha_p X_i + \epsilon_{ipm} \tag{12}$$

With multiple outcomes, high dimensional treatments, and covariates, the design gives rise to some concerns of limited power, particularly for interaction terms, and of multiple comparisons problems. To alleviate these concerns and make inferences on more theoretically-relevant concepts, I seek to aggregate "up" from the basic AMCE estimates presented here. To test for bias, I make inferences on the basis of F-tests (or the equivalent) on the joint significance of relevant coefficients. To estimate these models, I specify the subset of relevant estimators (β's in Equation 12) and implement an F-test to test the null hypothesis that all β's in the subset are equal to zero. I refrain from the use of high-dimensional interactions which are underpowered in the present design. Importantly, note that the inclusion of interactions between identity-based characteristics does not improve the predictive power of the models. Joint tests of interactions between the identity treatments reported in Appendix A11.3 provide no evidence that identity characteristics serve as complements or substitutes in terms in bureaucrats' responses to petitioners.

4.4 On Identification

Section 2.4 shows that bias in effort, Δ, is identifiable. The research design identifies Δ by manipulation (random assignment) of petitioner type. The model further implies that there is no direct test of the sources of bias through the manipulated attributes in the factorial design. As such, I test the broader theory about the sources of bureaucratic bias through examination of the testable implications of the model. With one exception, the tests of these implications rely upon treatment-by-covariate interactions with institutional and societal features that I cannot manipulate. While I use a flexible, non-parametric, and interactive covariate adjustment strategy to probe the robustness of these inferences, tests of the comparative statics measured using administrative data remain observational.[23]

5 Identifying Bias in Bureaucratic Effort

I begin by estimating the magnitude of bureaucratic bias by socioeconomic class, migrant status, and regional accent. Bias in effort, Δ, is identified by *differences* across petitioner treatment conditions. Where this difference is zero, there is no evidence of bias. In addition, baseline *levels* of service provision are also relevant for interpreting these differences.

This analysis focuses on how observed effort varies with randomly assigned petitioner identity characteristics. Given that the regional accent and socioeconomic class (ideally) were revealed as soon as the call was answered, I consider outcomes of process and access as well as the information provision outcomes. I analyze these outcomes on the full sample of answered calls ($n = 1,194$). Logically, the "unrevealed" factors (class, accent, migrant status, and petition difficulty) should be orthogonal to whether or not a call was answered or not. Reassuringly, F-tests of the joint significance of these factors provide no evidence of selection (imbalance) across unrevealed factors on the probability that a call was answered in Appendix A15.

[23]Note that the conditional AMCEs by subgroup are identified. The difference in conditional AMCES – the tests implied by the comparative statics from the model – are not causally identified without imposing additional assumptions.

5.1 Bias in Access to the *Alcaldía*

First, I investigate whether petitioner characteristics influence process of the petition through the *alcaldías* in Table 5. Column 1 examines whether the dispatcher (original official) identified themselves; if they did not identify themselves ex-ante, callers asked for a name. Levels of identification were high (≈ 0.85) and there are no apparent differences across callers per the randomly assigned petitioner characteristics.

Column 2 examines whether the caller was able to ask the question. In general, confederates were unable to make petitions when the dispatcher passed the call or referred the caller to a second official and the second official did not answer within two attempts. The lack of (robust) differences across across identity characteristics is therefore not surprising. Further, it provides no evidence that the dispatcher's handling of calls varied by class or regional accent of the callers.

Column 3 measures whether the petitioner was successful in speaking to (at least) a second official, a measure of access. The mean of this outcome is lower than with the petitions because some petitions were made to the dispatcher directly at the dispatcher's request. There is no bias on the basis of class or accent on this measure of access.

Column 4 examines whether any official identified herself as one of the officials on the pre-treatment administrative lists of MFA and SISBÉN officials collected from government partners. The results indicate somewhat higher levels of access to these program administrators for the middle-class petitioners relative to lower class petitioners, a difference of approximately 4.9 percentage points. Taken with Column 3, this finding likely emerges from higher levels of identification (by a second official) to middle-class petitioners. The joint test of coefficients on class and accent, however, is only marginally significant. Collectively, these analyses suggest limited, if any, bias in navigating the *alcaldías* within an initial interaction on account of class or accent.

The lack of evidence of bias by dispatchers is important for several reasons. First, high rates of identification, petition making, and access to a second official indicate statistical power to identify even modest amounts of bias. The power of the tests combined with the null findings suggest that for this class of tasks, there is no evidence of bias within the present research design. Second, the lack of differences in Columns 1-3 (outcomes measuring dispatcher behavior) provide no evidence that bureaucrats were "differentially confused" by some petitioner or script characteristics.[24]

5.2 Bias in Information Provided

Columns 5-10 examine bias in the responses to the petitions. Note that these responses are not conditional on making a petition; thus failing to receive information comprises both wrong responses and no response. Column 5 provides no evidence of bias in the probability that a petitioner receives a complete, correct response on the basis of the identity attributes. Note, however that baseline levels of correct responses are quite low. To the extent that bias represents the *withholding* of effort or information, there is limited scope to move this outcome. In this context, note that the (small) treatment effects on on lower-middle class and resident represent effect sizes of around 20% of this baseline.

There is notable bias in the likelihood of receiving a partial response or any information (Column 6). Lower middle-class petitioners are substantially more likely to receive a partial response or any information relative to lower class petitioners. In Column 7, the point estimate on receipt of any information is 8.1 percentage points and represents a 16 percent increase in the probability of receiving any information relative to the baseline (lower class). There is noisy evidence of a penalty against migrants.

Columns 8 and 9 track two outcomes in which information was not provided. Column 8, "no information" includes any response that did not provide individuals information or invite them to come to the *alcaldía*.[25] These responses included hang-ups, "don't know"-type responses, and situations in which the bureaucrat stated that they did not want to provide information. It is a relatively rare outcome and disproportionately impacts lower-income callers, though the point estimates and F-tests are not significant at conventional thresholds.

Column 9 measures whether or not individuals were simply told to "come to the *alcaldía*" without further information. While all services require the person to come to the *alcaldía* with

[24] One concern is that because the lower class petitioner scripts were less direct, they may have confused the bureaucrats that answered the phones. There is no evidence that this was the case.

[25] This outcome was not prespecified.

	Access/Process				Response to Petition					
	Dispatcher Gave Name	Petition Made	Second Official	Program Officer	Complete	Incomplete	Any Info.	No Info.	*Alcaldía* Only	Red Tape
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
PANEL A: IPW ESTIMATES										
Lower-Middle Class	0.011	0.014	−0.014	0.043	0.021	0.048*	0.069**	−0.019	−0.036**	0.003
	(0.020)	(0.025)	(0.028)	(0.026)	(0.018)	(0.029)	(0.029)	(0.021)	(0.017)	(0.024)
Bogotá Accent	0.029	0.038	0.030	0.027	0.002	−0.005	−0.003	0.026	0.015	−0.056*
	(0.026)	(0.031)	(0.034)	(0.033)	(0.022)	(0.035)	(0.035)	(0.025)	(0.019)	(0.030)
Costeño Accent	0.034	0.078**	−0.020	−0.038	−0.007	0.033	0.026	0.020	0.032	−0.027
	(0.026)	(0.031)	(0.035)	(0.032)	(0.022)	(0.036)	(0.036)	(0.026)	(0.021)	(0.030)
Resident	0.015	−0.018	0.015	−0.002	0.027	0.003	0.029	0.007	−0.054***	0.036
	(0.020)	(0.025)	(0.028)	(0.026)	(0.018)	(0.029)	(0.029)	(0.021)	(0.017)	(0.024)
F-test, p-value	0.472	0.071*	0.445	0.078*	0.462	0.352	0.100*	0.698	0.001***	0.197
Placebo F-test, p-value	0.400	0.761	0.643	0.891						
PANEL B: ESTIMATES WITH ENTITY FIXED EFFECTS										
Lower-Middle Class	−0.015	0.016	−0.021	0.046*	0.024	0.057**	0.081***	−0.027	−0.039**	0.002
	(0.019)	(0.025)	(0.025)	(0.024)	(0.019)	(0.028)	(0.028)	(0.021)	(0.017)	(0.026)
Bogotá Accent	0.023	0.002	−0.023	−0.016	−0.006	−0.044	−0.050	0.039	0.014	−0.098***
	(0.026)	(0.034)	(0.035)	(0.034)	(0.026)	(0.038)	(0.038)	(0.028)	(0.022)	(0.036)
Costeño Accent	0.021	0.038	−0.063*	−0.058*	−0.005	−0.016	−0.021	0.025	0.034	−0.071**
	(0.026)	(0.034)	(0.034)	(0.031)	(0.026)	(0.038)	(0.038)	(0.029)	(0.023)	(0.035)
Resident	0.021	−0.020	0.022	−0.010	0.020	0.021	0.041	0.00005	−0.061***	0.041
	(0.019)	(0.025)	(0.025)	(0.024)	(0.019)	(0.028)	(0.028)	(0.022)	(0.017)	(0.026)
F-test, p-value	0.612	0.495	0.193	0.060*	0.597	0.153	0.009***	0.447	0.000***	0.015**
Placebo F-test, p-value	0.173	0.693	0.481	0.840						
PANEL C: ESTIMATES WITH ENTITY + ENUMERATOR FIXED EFFECTS										
Lower-Middle Class	−0.016	0.019	−0.026	0.045*	0.028	0.053*	0.081***	−0.023	−0.039**	0.004
	(0.019)	(0.025)	(0.025)	(0.024)	(0.019)	(0.028)	(0.028)	(0.021)	(0.017)	(0.026)
Resident	0.020	−0.017	0.017	−0.014	0.022	0.019	0.041	0.003	−0.061***	0.040
	(0.019)	(0.025)	(0.025)	(0.024)	(0.019)	(0.028)	(0.028)	(0.021)	(0.017)	(0.026)
F-test, p-value	0.468	0.263	0.020**	0.061*	0.033**	0.306	0.035**	0.130	0.001***	0.567
Placebo F-test, p-value	0.411	0.246	0.019**	0.166						
Mean DV, Lower Class	0.859	0.758	0.685	0.276	0.104	0.396	0.5	0.153	0.104	0.234
Mean DV, Paisa Accent	0.842	0.723	0.675	0.296	0.119	0.414	0.533	0.124	0.066	0.266
Mean DV, Resident	0.856	0.771	0.668	0.302	0.101	0.428	0.529	0.134	0.108	0.219
Program	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
All Factors	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
DV range	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}	{0, 1}
Observations	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194

Note: *p<0.1; **p<0.05; ***p<0.01

Table 5: The AMCEs of identity-based characteristics on access to bureaucrats and responses to petitions. All specifications are estimated in OLS with heteroskedasticity-robust standard errors. Panel C omits the accent manipulation because regional accents were voiced by natives of each region (not acted); joint tests in this panel refer only to class and migrant manipulations. The F-tests test the joint significance of coefficients on class and accent factors in Columns 1-4 (prior to the petition) and class, accent, and migrant factors in Columns 5-9. The placebo F-test test the joint significance of attributes of the petition, migrant status and technical question, in Columns 1-4.

documents, failure to specify these requirements by phone passes the cost onto the citizen. The estimates suggest that lower-middle class individuals are 37.5 percent less likely to receive this response than lower class individuals, while residents are half as likely as migrants to receive the response.

The results in Column 10 indicate disproportionate use of red tape – a request for *extra* requirements – against Paisas relative to both Bogotanas and Costeñas, with sizable point estimates of 0.071 and 0.098, respectively. These differences are not driven by individual enumerators or pairs of enumerators across the groups (Appendix 16.2). It is unclear why differences emerge on this outcome specifically.

The observed biases in information provision on the basis of petitioner class merit some additional discussion. It does seem that the class treatment was recognizable; independent coders identified the assigned coding in 77.5% of calls, as reported in Table 2.[26] While class is necessarily a compound treatment in the Colombian context, analysis of the magnitude of the "complier" AMCE relative to the intent-to-treat AMCE in Appendix A16.1 suggests that bias enters through what blinded coders perceive to indicate social class within the calls.[27]

There are several explanations for the generally null effects of regional accent. First, it could be that bureaucrats did not hear regional accents. This seems highly unlikely as blinded coders listening to recordings of the calls correctly identified over 99% of calls (Appendix A11.1). Second, it could be that a Bogotá accent means something different in Bogotá than in other parts of the country. To this end, I report the results of a prespecified analysis on the subsample of calls from the regions in which these accents are local in Appendix A16.3. Here the treatment is defined as an "in-region" accent. In the full subsample, estimates are near-zero and confidence intervals bound zero for all outcomes. This masks some heterogeneity between the three regions, however. While I cannot reject the null hypothesis that any of the conditional AMCEs is zero, there is suggestive evidence that Costeñas are *punished* in their home region relative to outsiders, while there is mild

[26]The blinded coders were given an "I don't know" option in addition to the two class categories; another 13.5% of calls fell into this category. Only 9% of calls were incorrectly classified.

[27]The estimates of the complier AMCE can be seen as an informal test of the excludability assumption as applied to the social class treatment.

evidence of an in-region bonus for both Bogotanas and Paisas in their respective home regions.

Collectively the differences in treatment of lower-middle class versus lower class petitioners track those of residents versus migrants, though the class effects are substantively stronger. However, as indicated in Table 4, migrant status was not revealed until the petition was made. Appendix A16.4 reveals that these estimates are conservative and less efficient than estimates of migrant status on the sample of petitions alone. To the extent that these groups are relatively marginalized at least within the experimental comparisons, these comparisons provide some evidence about the dynamics of bias that I explore in the next section.

Beyond the behavioral measures, confederates evaluated their interactions with bureaucrats after each call. These results, reported in Appendix A16.5, suggest that perceptions largely aligned with the behavioral outcomes. Within enumerator and *alcaldía*, enumerators perceived slightly worse treatment when calling as low-income petitioners. The alignment between the behavioral measurements and perceptions of the calls increases confidence in the behavioral measures.

5.3 Does Information Provision Reflect Costly Effort?

I seek to validate that information provision does indeed reflect exertion of costly effort. I consider the total amount of time spent on the call (mean: 4.83, standard deviation: 6.32 minutes). Because the scripts for the petitions varied in length of delivery across the identity characteristics, I lack the ability to identify differences across petitioner identities on this outcome.[28] This represents an excludability violation: observation of a longer call could mean the bureaucrat spent more time answering the question or that the petition took longer to make. By the same token, where script length and information provision counterbalance each other, an inference of no difference in time does not provide a clear measure of differential effort.

Instead, I show that the length of calls is increasing in the amount of information provided (correct, partial, or no information).[29] I first aim to purge differences in the length of calls due to variation in the experimental scripts. To do so, I fit a regression of $\ln$(Minutes on call) on the

[28] For example, the migrant petitions included an extra sentence that was not included in the resident petition.

[29] Incorrect includes the *alcaldía* only response from Table 5.

experimental factors, a program indicator, and enumerator fixed effects with IPW.[30] I then examine the distribution of residuals from this regression across the three types of outcomes.

Figure 2 depicts the distribution of residualized (logged) call length by the amount of information provided as empirical CDFs (ECDFs). The graph indicates that the cumulative length of contact for petitions providing no information was substantially shorter than the length of those providing some information. On average, petitions receiving no information were 1.17 minutes shorter ($p < 0.01$) than calls providing partial information and 1.21 minutes ($p < 0.01$) shorter than calls providing complete answers. These differences represent effects of approximately 25 percent of the mean for calls with no information (4.63 minutes). Further, the crossing of the ECDFs for partial and complete information provide some evidence to adjudicate the competence vs. effort distinction between the two types of answers. It suggests that the difference between the two answers is not simply differential competence and that, in the upper median of the distribution, bureaucrats spent more time to provide a more complete answer. This is consistent with qualitative observations of confederates.

6 Examining the Mechanism

The evidence of bias in information provision against lower class petitioners and, to a lesser extent, internal migrants motivates analysis of what drives these biases. I seek to distinguish between the three mechanisms suggested by the model: bureaucrats' tastes, politicians' tastes, and complaint-driven bias. The experiment measures bureaucrats' effort, meaning I do not measure oversight or biases in oversight directly. Indeed, the question is not whether biased oversight leads to biased service outcomes, but how the likelihood of oversight conditions bureaucrats' initial behavior. I use the model to identify the conditions under which oversight-driven bias should be magnified.

To conduct this analysis, I proceed in two steps, following Proposition 3. To begin, I endeavor to separate bureaucrats' taste-based bias from oversight-based bias (composed of politicians' taste-based and complaint-driven bias) on class. Then, I seek to tease apart complaint-driven bias and the politician's taste-based bias. These tests follow directly from the comparative statics presented in

[30]There was some heterogeneity in the pacing of calls between enumerators.

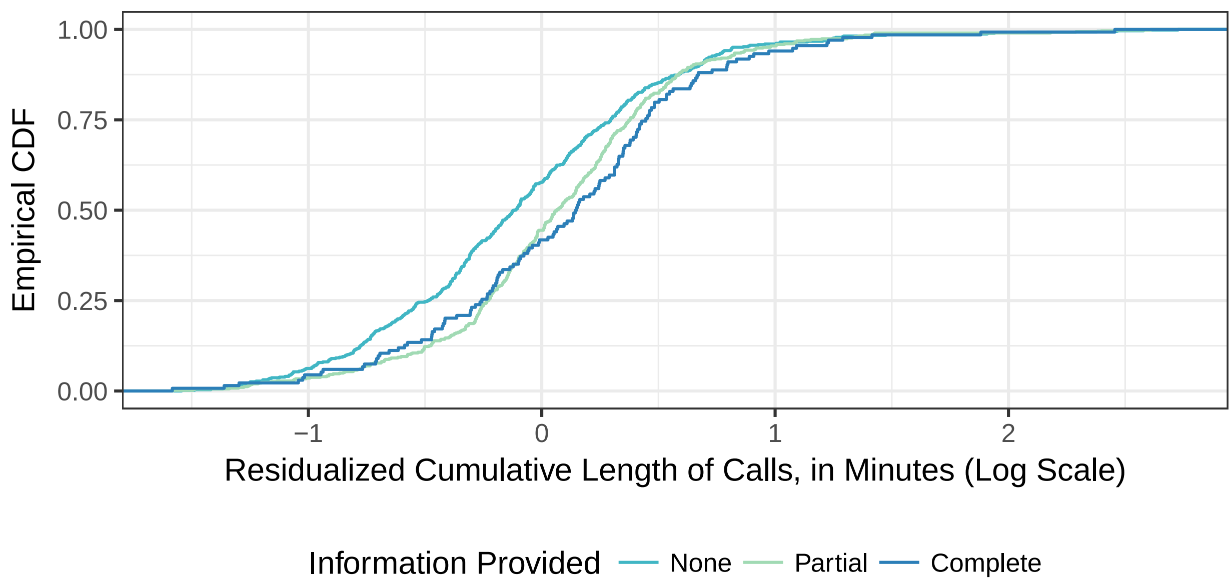

Figure 2: The distribution of residualized call length by the amount of information contained in the response. Calls that provided no information were uniformly shorter than calls with partial or complete responses. In the upper quantiles of the distribution, calls providing correct answers were longer than those providing partial information.

Table 1. I then distinguish the results from an alternate account from existing models of differential treatment in the form of screening clients for services.

6.1 Political Oversight Drives Bias

In order to disentangle bureaucratic taste-based bias from oversight-driven, bias, I consider two parameters of the model: the politician's marginal cost of effort, c_P, and bureaucratic incentives, r. Do the costs of a task reduce the level of bias in effort? As effort becomes more costly to the bureaucrat and audits become more costly for the politician, effort should decline. Increases in these costs should also attenuate bias.

Analysis of bias as a function of petition difficulty provides an experimental test of this logic. I find that class-based bias emerges on easy questions but not on technical questions, where the cost of effort is highest for the bureaucrat. The left panel of Figure 3 visualizes estimates from an interactive specification where the technical petition indicator is interacted across all other experimental factors (petitioner characteristics and program) in the design. Consistent with the theory, substantially less information is provided in response to the technical petition. Further, there is no evidence of bias, given that the solid and dotted blue lines are do not substantially diverge for any response. Bias is driven by the easy (registration) petition, as is evident from the divergence of the green lines. Bias is most strongly apparent in easy petitions for the provision of *any* information ($p < 0.002$). Indeed, the difference in the estimated bias against poor petitioners for easy and technical questions in the provision of any information is substantively large at 10.1 percentage points and statistically significant at the $\alpha = .1$ level in a two-tailed test ($p = 0.079$).

This finding is consistent with the prediction with any of the bias mechanisms. However, consider that technical petition may also induce a shock to the politician's cost of effort, c_P. Per Proposition 3, if bias varies in the politician's cost of effort, there is evidence of oversight-driven bias ($\frac{\partial \Delta}{\partial c_P} \neq 0$ if and only if $\Delta_O \neq 0$). If the task is harder for the bureaucrat, it should also be harder for the less expert politician. This (simultaneous) increase in the cost of the politician's effort should increase the relative contribution of the bureaucrat's taste-based bias to the estimated bias in effort. Stated another way, if the oversight-driven biases were entirely absent, the reduction of bias observed in the data corresponds to a very large shock to c_B. Importantly, service provision is not commensurately driven to zero. There is scope to observe taste-based bias, but I do not detect any. This test provides no evidence against oversight-driven bias.

As a further test of whether bias varies in costs to the politician (c_P), consider variation between the two audited programs as a more direct test of the oversight channel. By all accounts, one fundamental distinction between the operation of SISBÉN and MFA at the municipal level is the degree of politicization. One plausible operationalization of politicization is a lower c_P for SISBÉN and a higher c_P for MFA since it is less costly for a politician to intervene in SISBÉN. Examining variation in information provision by program thus provides one test of how bias changes when the politician is more apt to intervene.

For sufficient increases in the politician's cost of effort, she is less willing to monitor the bureaucrat. This implies a higher propensity to monitor the more politicized program (SISBÉN) than the less politicized program (MFA). Less monitoring reduces the bureaucrat's incentives to work

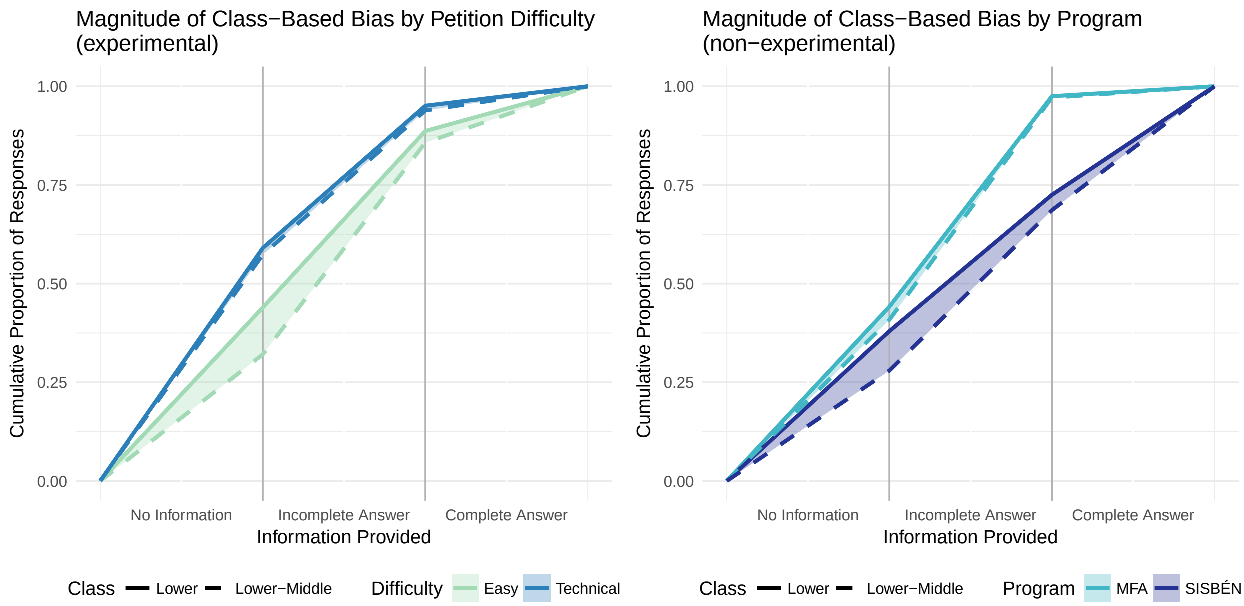

Figure 3: Sensitivity of bias in information provision to the cost of effort (difficulty of the petition) (left) and program (SISBÉN or MFA) (right). The area of the polygon corresponds to the level of bias for each subgroup and for each level of information provision. Given that the lower-middle class consistently received more information, the lower curve of all polygons represents the lower-middle class and the upper curve represents the lower middle class. All estimates come from interactive models estimated with IPW.

and reduces the contribution of oversight-driven bias to total bias. Thus, if bias attenuates substantially for MFA relative to SISBÉN, there is evidence that bias enters through oversight. The results in the right panel of Figure 3 mirror this expectation quite precisely. More information was provided for SISBÉN than MFA. There is clear class-based bias in the the provision of any information for SISBÉN ($p = 0.013$) but no evidence of bias in the administration of MFA ($p = 0.421$). The difference-in-difference estimate on the interaction between class and program is sizable at 6.7 percentage points, but is not statistically significant at conventional thresholds ($p = 0.24$). Regression tables supporting these analyses are reported in Appendix A17.1.

A final analysis seeking to disentangle bureaucratic taste-based from oversight-driven bias turns to variation in bureaucrats to understand bias. Per Proposition 3, an increase in the "bite" of punishment, r, should increase oversight-driven bias. The central characteristic of bureaucratic employment that I measure is contract type, comparing contractors to civil servants. I argue that contractors, all else equal, face higher powered incentives to exert effort given the prospect of contract non-renewal.[31] In terms of the model, I argue that r is higher for contractors than for civil servants. As such, we would expect higher levels of effort and a magnification of any oversight-induced bias.

In this analysis, I study the program officers administering the program in each municipality. Because Table 5 indicated that identifiability of program officers was related to the class of the petitioner, I must restrict analysis to the dispatcher sample or the program officers (from administrative data) to avoid conditioning on a post-treatment variable. Program offices were likelier to take the petition than dispatchers so I focus on the *ex-ante* list of program officers. Table 6 examines the conditional effects by program officer contract. It provides no evidence of differences, on average, in bias between contractors and civil servants. Specifically, there are few differences in information provision in the aggregate.[32] Moreover, rates of bias against lower class individuals cannot be distinguished between contractors and civil servants.

The null findings of this analysis are mixed with respect to the model. However, two caveats in the present analysis are in order. As documented in Appendix A12, I am unable to identify individuals in the poorest municipalities. As subsequent analyses demonstrate, these are precisely the locations where class-based bias is most pronounced. Further, the majority of calls did not reach the individual identified as a program officer in the data. In larger municipalities, they often spoke to subordinates; in smaller municipalities, they spoke to other individuals from the *alcaldía* with or without knowledge of the program. In that sense, this result is very much an "intent-to-treat" effect and attenuate differences between contractors and civil servants to zero.

Second, and more importantly, politicians' choice of hiring mechanism is, to some extent, strategic. While I lack data on the tenure of all SISBÉN program officers, approximately half of all MFA officers (whether contractors or civil servants) have been appointed in the last two years, and fewer than 19% have served more than five years. This suggests substantial scope for

[31] Interviews suggest contractors – patronage or not – work hard, sometimes to "compensate" for shirking civil servant colleagues.

[32] There is evidence that contractors provide more information in response to technical petitions. The estimates in the table refer to an easy petition given the interactions in the estimator.

	Complete	Incomplete	Any Info.	Alcaldía Only	Red Tape
	(1)	(2)	(3)	(4)	(5)
PANEL A: CONDITIONAL AMCE BY EMPLOYEE TYPE; IPW ESTIMATES					
Lower-Middle Class	−0.002	0.077*	0.076*	−0.044*	−0.032
	(0.026)	(0.042)	(0.041)	(0.025)	(0.035)
Contractor: Lower-Middle Class	0.043	−0.040	0.003	−0.003	0.062
	(0.042)	(0.067)	(0.067)	(0.039)	(0.057)
Conditional Effect, Contractor	0.042	0.037	0.079	-0.047	0.030
	(0.034)	(0.053)	(0.054)	(0.030)	(0.045)
PANEL B: CONDITIONAL AMCE BY EMPLOYEE TYPE; STRATUM + ENUMERATOR FE					
Lower-Middle Class	−0.002	0.084**	0.082**	−0.045*	−0.027
	(0.026)	(0.041)	(0.040)	(0.024)	(0.035)
Contractor: Lower-Middle Class	0.042	−0.054	−0.012	0.0002	0.053
	(0.041)	(0.067)	(0.066)	(0.037)	(0.057)
Conditional Effect, Contractor	0.040	0.030	0.070	-0.045	0.026
	(0.032)	(0.053)	(0.053)	(0.029)	(0.044)
Observations	1,194	1,194	1,194	1,194	1,194
Mean, Lower Class and Contractor	0.077	0.438	0.515	0.112	0.207
Mean, Lower Class and Civil Servant	0.125	0.407	0.532	0.115	0.276
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 6: Estimates of the conditional AMCE of a lower-middle class petitioner by program officer type (civil servant or contractor). The base category is civil servant. In the sample of petitions, there are $n = 611$ petitions to civil servants; $n = 354$ petitions to contractors; $n = 103$ petitions to a vacant program officer; and $n = 126$ petitions to an official whose contract type is unidentifiable. The program indicator is interacted with all factors in the experimental design and a program indicator. "Conditional effect" refers to the conditional effect of lower-middle class. Heteroskedasticity-robust standard errors in parentheses.

mayors' appointment of individuals to the position, whether or not they had previously served in the *alcaldía*. To this extent, the distribution of contract types varies across the two programs. Among the identified MFA officials 50% are contractors; just 25% of SISBÉN officials are contractors.[33] Because the effort gains of contractors relative to civil servants are driven by the ratio of r to c_P, the imbalance in contractors should theoretically attenuate differences in the strength of oversight-driven bias in the full sample.

Collectively, these analyses provide evidence that the observed bias is driven, at least in part,

[33] The fact that there were *any* civil servants running these programs in the municipalities surprised officials that oversee the program at the national level.

by oversight of political principal, e.g. $\Delta_O \neq 0$. The degree to which bias is attenuated by technical questions or less politicization, there is evidence that taste-driven bias by bureaucrats is quite limited, e.g. $\Delta_B \approx 0$. This conclusion is supported by several other facts beyond the scope of the model. First, while confederates perceived worse service on average after lower-class calls, they perceived precisely no difference in respect (see Appendix A16.5). Respect is the only measure of affect in the battery, an indicator of bureaucratic bias in existing audit studies (Einstein and Glick, 2016). Second, the lack of bias in the access outcomes in Table 5 provides additional evidence that outcomes that are less easily audited (since complaints are typically received and ratified through the dispatcher) do not exhibit class-based bias.

6.2 Bias Occurs where Differences in Access to Complaints is Greatest

In the model, bias between groups is driven by the differentiation of citizens. The differentiation of citizens could come from three sources: differences in the distribution of costs of complaint across groups (η_Q), differences in tastes of the bureaucrat (η_B), differences in tastes of the politician (η_P).[34] Bias should be greatest in magnitude in places where these distances are larger. To examine whether observed biases vary with the distribution of citizen costs of complaint, I leverage the fact that markets for social services vary substantially across Colombia.

I argue that the differences in the distribution of costs of complaint are relative. Measures of physical distance and familiarity with the bureaucracy vary with context. In particular, the relative status of the individual analogues to the experimental profiles – lower and lower-middle class petitioners – varies substantially across Colombia. The intuition is straightforward: a lower-middle class profile connotes a higher status, with more access to complain, in a place where the entire population is poor than a place with many lower-middle class (and higher) individuals.

For concreteness, consider the role of the lower and lower-middle class petitioner profiles in two hypothetical municipalities. In one municipality, the vast majority of the population is poor (lower class). In another, a plurality of the population is lower-middle class. The experimental profiles relate quite differently to the underlying population distribution. If the cost of complaint

[34] To the extent that the previous section focused on separating Δ_B from Δ_O, it implicitly addresses η_B.

is a function of relative *status*, the status differential between experimental profiles is far larger in the first (poorer) municipality than in the second.

Thus, to understand where class-based bias emerges, I examine how class differences in treatment vary with the class composition of municipalities. I assume that differences in cost of access to complain are greatest where the lower middle-class is most empowered relative to poor individuals. This occurs in places with more poor citizens, or higher poverty rates. Thus I use poverty rates to operationalize η_Q. Per Proposition 3, the magnitude of bias should increase in municipal poverty if complaint-driven bias is operative. Accordingly, this analysis should be interpreted as heterogeneous treatment effects with municipal poverty rate as the moderator.

Figure 4 examines bias in information provision as a function of a the portion of residents in poverty as per the multidimensional index of poverty, calculated from the 2005 census.[35] The figure shows that anti-poor bias emerges against poor petitioners only in poorer places. The bias is restricted to the inscriptions question (left column) and reception of partial information or the *alcaldía* only response, as described above.

To subject these graphical intuitions to a more rigorous test, I run a series of regression analyses in Appendix A17. I bin the poverty index into terciles to reduce functional form assumptions on the moderator. Because poverty and population are strongly negatively correlated ($\rho = -.61$ in the sample), I include an interactive binned population control with deciles of the estimated 2018 population in a second estimator. Both the moderator, municipal poverty, and the (demeaned) population decile bin controls are interacted across the whole design (all factors and the program indicator).

This analysis suggests that bias against lower-class individuals (the baseline) is worse in poorer places. There is no evidence of bias in the lowest tercile (the municipalities with the lowest poverty rates) for any outcome. Class bias against the poor is increasing in the middle-poverty and high-poverty terciles. I find clear, statistically significant evidence of bias in the high-poverty tercile for

[35]This index is compiled by the Departamento Administrativo Nacional de Estadísticas (DANE) at the level of rural and urban populations within each municipality. I take the weighted average where weights correspond to the share of urban and rural residents in the population.

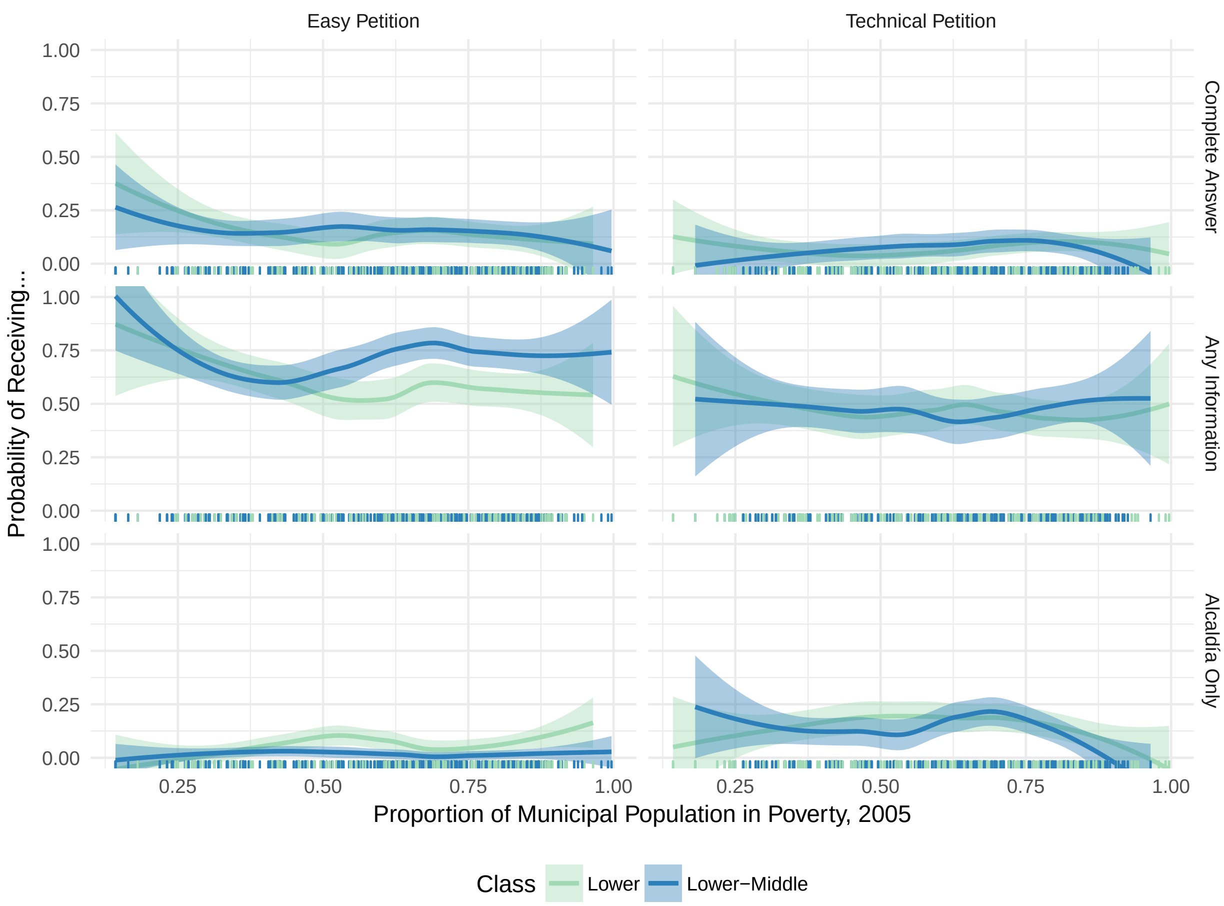

Figure 4: Heterogeneity in treatment effects by the level of municipal poverty. Column 1 examines average marginal effects on "easy" (enrollment) questions while Column 2 examines average marginal effects on technical questions. Lines are estimated by Loess regression with a span of 0.75. The shaded regions are 95% confidence intervals. Graphs exclude observations from Bogotá given the vastly disproportionate number of observations, though substantive results on bias are robust to including all observations in the analysis.

the receipt of partial information. Further there is suggestive evidence that differential application of the "*alcaldía* only" outcome against poor individuals is driven by poorer municipalities. These findings are robust to other operationalizations of poverty including rates of secondary education (2005).

Combined with the analysis of in-region accents, these findings provide suggestive evidence that differentiation of citizens correlates with bias in service provision. Attributing such patterns to tastes requires a theory of bureaucratic selection or politician incentives that yield divergent tastes. While I cannot eliminate this possibility, several findings are useful to consider. First, there is no evidence that service favors the median voter in each municipality – neither the class or regional accent analysis supports this interpretation. If service were to favor the median voter, the poor should do the *best* in the highest poverty places; these are the places that they do the *worst*. From an elected politician's perspective, providing worse service to poorer individuals in poor places works against the median voter and those most likely to turn out in Colombia (Kasara and Suryanarayan, 2015).

Other explanations of bias in terms of politician tastes do not account for these geographic patterns of bias. While there may be a disproportionate incentive to politicize social programs (outside the scope of the model) to claim credit or buy votes in poorer places, it is not clear why such opportunities to claim credit would yield *unequal* information provision, as opposed to simply less information provision. Importantly, as is evident in Figure 4, there is no evidence that less information is provided to the middle class in poor municipalities. To this extent, unless politician tastes vary systematically in unmeasured ways with the degree of poverty in a municipality, there is little evidence supportive of politicians' taste-based bias in driving the bias results. I find no evidence that political competition drives bias. If competition drives politician incentives to provide public goods, S is the relevant parameter of the model. Note that the expression for bias in effort (Equation 8) does not include S, implying S does not drive bias. However, relaxing the assumption of interior effort, bias is eliminated when S drives universal service provision. Empirically, the low rates of information provision suggest that this is not the case for the relevant services.

To the extent that competition drives the selection of different "types" of politicians, there is no evidence that this manifests in politicians with different tastes.

One final explanation concerns the selection of bureaucrats themselves. In rural areas, in particular, bureaucrats are hired from a less skilled local labor market. This could drive anti-poor bias via lower competence, perhaps accompanied by tighter oversight, or different tastes, potentially as a function of status. To this extent, a theory of bureaucratic selection consistent with these results cannot distinguish between oversight and taste explanations. Further, recall that the estimates in Appendix A17.2 suggest that the association between poverty and bias is not driven by differences in population.

I therefore argue that the most plausible interpretation of the finding is that where differentials in relative ability to complain between the treatment conditions are theoretically the strongest, levels of bias against the less able group are strongest. Structurally, this analysis suggests that $\eta_Q > 0$. In particular, lower-middle class individuals are relatively more empowered in places where a plurality of the population is poor. Suggestive patterns of an in-region penalty against Costeñas in the Coast (Caribe) are consistent with this interpretation. If Costeños are perceived as less demanding and therefore less likely to complain within the region, the observed penalty is consistent with the complaint-driven bias explanation of findings.

6.3 Alternative Explanation of Bias: Screening

The theoretical model described here is a model of underprovsion rather than misallocation of services. Some theories of misallocation suggest that the "bias" that we observe could simply be efforts to screen citizens that the program is intended to serve from unintended potential recipients (Banerjee, 1997; Ting, 2017). The intended population of beneficiaries/registrants for SISBÉN and MFA are poor individuals and households. However, I find no evidence that differences in levels of service provision are consistent with the theoretical predictions of a screening account.

Informational outcomes cut *against* lower class individuals and internal migrants, the target populations for these programs.[36] One outcome of interest for testing the screening logic is the use

[36]To the extent that migrants requesting services are associated with internally displaced persons (IDPs), a special

of red tape. Red tape is hypothesized to serve as a mechanism that induces individuals to truthfully reveal their "type" – whether or not they comprise the intended target population of the program. Table 5 indicates high levels of red tape generally, but no differences in its application by class or migrant status.

While a story of screening is inconsistent with the observed data, it may still be the case that red tape is employed to deter unintended recipients from requesting services in the first place. Yet, qualitatively, the forms of red tape requested appear to disparately impact intended recipients (the poor). The most common extra requirements were a receipt for utilities (usually electricity), a formal letter of application for the service, or extra government documents (other services). Thus, within the experimental data, there is no evidence of differential treatment due to screening. Speculatively, in a setting with endogenous requests, the type of red tape may exacerbate inequality in service provision, but in a direction *opposite* to that predicted by existing screening theories.

7 Discussion: Bias in Effort and Inequality in Outputs

To what extent does bias in information provision map onto inequality in public service outputs? Bureaucratic bias in effort is important because of its link to inequality in citizen access to public services. Recall that the theory suggests that in the presence of oversight-driven bias, bias in effort is a sufficient condition for ultimate inequality in outcomes.

While the experiment allows for measurement of bureaucratic effort in information provision, confederates did not try to obtain the service. Yet, using pretreatment data on SISBÉN registration from across Colombia's municipalities, I can examine the correspondence between rates of enrollment (outputs) and the experimental measures of bias.[37] In general, two pathologies of SISBÉN enrollment exist in the administrative data: over enrollment and under enrollment of the relevant population. Some municipalities maintain rolls that could not possibly cover the estimated poor population; other municipalities maintain rolls far larger than the population as a whole. The focus of this experiment is on the former category. To that end, I investigate whether there exists

category for both programs, these programs should also favor the migrant condition.

[37] MFA data by municipality is not publicly available. However, aside from IDPs and indigenous Colombians, SISBÉN is used to qualify for MFA. As such, under-enrollment of SISBÉN should predict under-enrollment of MFA.

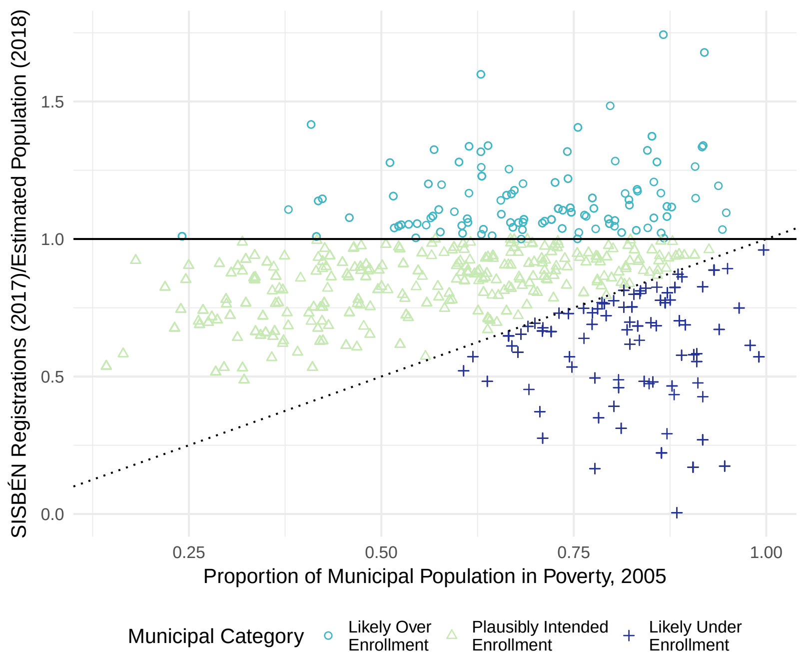

Figure 5: The relationship between municipal poverty rates and SISBÉN enrollment as a proportion of the population.

differential levels of bias in municipalities where the service is under-provided.

Figure 5 suggests substantial variation in the provision of SISBÉN across experimental municipalities in which a call was answered. In 107 municipalities, SISBÉN enrolls more individuals than the ostensible municipal population. I consider the other two categories: municipalities that could not possibly enroll all poor individuals (those below the 45-degree line) and municipalities where enrollment is plausibly aligned with intended administration. While there could be under enrollment of individuals in poverty or poor targeting of SISBÉN above the 45-degree line, the points below the line suggest a smaller list than the share of individuals in poverty.

Focusing on the latter two theoretically relevant categories, Table 7 suggests that bias is present precisely in the places in which under enrollment of plausible beneficiaries is the strongest concern.[38] There is strong evidence of bias in information provision in the base category (under-enrolled) municipalities. This bias is substantively, and for some outcomes, significantly attenuated in municipalities with ostensibly "intended" enrollment. These results are robust to redefinition of

[38]To the extent that over-enrollment represents politicization, empirically bias also emerges in these places.

	Complete	Incomplete	Any Information	*Alcaldía* Only	Red Tape
	(1)	(2)	(3)	(4)	(5)
PANEL A: CONDITIONAL AMCE ON CLASS BIAS BY ENROLLMENT TYPE					
Lower-Middle Class	0.072	0.091	0.163***	−0.039	−0.010
	(0.045)	(0.064)	(0.061)	(0.049)	(0.055)
Plausible Enrollment: Lower-Middle Class	−0.056	−0.072	−0.129*	−0.005	0.006
	(0.050)	(0.074)	(0.071)	(0.054)	(0.064)
Conditional Effect, Plausible Enrollment	0.015	0.019	0.034	-0.044**	-0.005
	(0.022)	(0.037)	(0.036)	(0.022)	(0.033)
PANEL B: CONDITIONAL AMCE ON CLASS BIAS BY ENROLLMENT TYPE WITH COVARIATES					
Lower-Middle Class	0.108	0.082	0.190**	0.008	0.076
	(0.051)	(0.092)	(0.089)	(0.058)	(0.082)
Plausible Enrollment: Lower-Middle Class	−0.094	−0.084	−0.179*	−0.062	−0.112
	(0.059)	(0.108)	(0.105)	(0.069)	(0.097)
Conditional Effect, Plausible Enrollment	0.013	-0.002	0.011	-0.053**	-0.036
	(0.027)	(0.042)	(0.040)	(0.024)	(0.036)
Interactive Poverty Decile Bins	✓	✓	✓	✓	✓
Interactive Poverty Decile Bins	✓	✓	✓	✓	✓
Mean, Lower Class and Plausible Enrollment	0.107	0.449	0.556	0.107	0.251
Mean, Lower Class and Under Enrollment	0.084	0.379	0.463	0.126	0.242
Observations	903	903	903	903	903
All Factors	✓	✓	✓	✓	✓
Program	✓	✓	✓	✓	✓
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 7: OLS estimates of the conditional AMCE of class by municipal SISBÉN enrollment type. The sample includes places that are under enrolled or plausibly enrolled as intended. Standard errors are clustered by municipality ($n = 366$) because the conditioning variable is measured at the municipal level.

the "plausible enrollment" category (see Appendix A18). While it is evident that under-enrollment occurs in poorer places, the results are robust to controlling flexibly interactively for municipal poverty and population (Panel B). This finding is consistent with the logic that bias in bureaucratic effort yields inequality in service provision. It bolsters confidence that the bias in effort measured in the experiment correlates with public service outputs. These results are also consistent with the theoretical extension of endogenous requests for service. In places with where prospects for service are lowest, lower-income Colombians may opt out of seeking SISBÉN registration altogether.

8 Conclusion

This paper argues that in order to understand inequality in public service provision, we must better understand the incentives and behaviors of the bureaucrats tasked with implementing and delivering these services. One way in which bureaucrats' behavior influences "who gets what" is through provision of different quality service to different groups in society.

In considering the strategic interactions between citizens, bureaucrats, and politicians, this paper generates two central insights on why bureaucracies generate unequal outputs. First, citizens complaints serve as a check on bureaucrats, alleviating moral hazard problems of the bureaucrat. But reliance on citizens to articulate their right to service has the potential to generate stark inequalities in service provision when some citizens are not sufficiently empowered to complain.

Second, I show that political oversight can generate or exacerbate biases in service provision. Where oversight is more responsive to some citizens than others, bureaucrats rationally anticipate different probabilities of oversight by exerting more effort serving some citizens than others. This implies that efforts to reduce bureaucratic bias through hiring (selection) or (re)training will not will not necessarily eliminate bureaucratic bias. To reduce oversight-driven biases, my theory implies that insulation of bureaucrats and empowerment of all groups of citizens to complain should also combat these inequalities.

To what extent do the results travel? The theory suggests that we should consider both societal characteristics and the institutional structure of the bureaucracy to set expectations about bureaucratic bias. On one hand, differences in citizen ability to lodge complaint suggest poorer service where fewer this strategy is less feasible. Complaint-driven bias is largest where citizens are most differentiated on this dimension. In this way, inequality in voice begets inequality in access. We should also consider the role of oversight in generating bias. Institutional features such as the insulation of bureaucrats and the competence/ability of politicians to monitor are apt to condition the magnitude of oversight-driven bias (Raffler, 2017). To the extent that these social and institutional features vary, we may expect differences in both the *magnitude* and *composition* of bureaucratic bias. While the evidence presented characterizes bureaucratic bias at scale in Colombia, the model allows for a more general characterization of bias.

The implications of the argument presented in this paper are wide-reaching. A vast literature attributes inequalities in public goods provision to differences in budget allocations. Yet, the findings presented here suggest that inequalities in budget can be exaggerated or attenuated by bureaucrats in the production of public goods. To the extent that politicians allocate funds strategically, they should anticipate how public goods will be actuated when making allocations. This poses empirical challenges to work that equates budget allocation with public goods outputs. It also opens space for theoretical development on the provision of public goods with bureaucrats as strategic actors.

References

Acemoglu, Daron, Camilo García-Jimeno, and James A. Robinson. 2015. "State Capacity and Economic Development: A Network Approach." *American Economic Review* 105 (8): 2364–2409.

Acemoglu, Daron, Maria Angelica Bautista, Pablo Querubín, and James A. Robinson. 2008. *Institutions and Economic Performance*. Number 5 1st ed. Cambridge: Harvard University Press chapter Economic and Political Inequality in Development: The Case of Cundinamarca, Colombia, pp. 181–245.

Banerjee, Abhijit. 1997. "A Theory of Misgovernance." *Quarterly Journal of Economics* 112 (3): 1289–1332.

Becker, Gary S. 1957. *The Economics of Discrimination*. 2nd edition ed. Chicago, IL: University of Chicago Press.

Bhavnani, Rikhiil R., and Alexaner Lee. 2018. "Local Embeddedness and Bureaucratic Performance: Evidence from India." *Journal of Politics* 80 (1): 71–87.

Camacho, Adriana, and Emily Conover. 2011. "Manipulation of Social Program Eligibility." *American Economic Journal: Economic Policy* 3: 41–65.

Centeno, Miguel Angel. 2003. *Blood and Debt: War and the Nation-State in Latin America*. University Park, PA: Penn State University Press.

Coppock, Alexander. 2018. "Avoiding Post-Treatment Bias in Audit Experiments." *Journal of Experimental Political Science* . Forthcoming.

Dargent, Eduardo. 2016. *Tchnocracy and Democracy in Latin America*. New York: Cambridge University Press.

De la O, Ana. 2015. *Crafting Policies to End Poverty in Latin America*. New York, NY: Cambridge University Press.

Dixit, Avinash. 2002. "Incentives and Organizations in the Public Sector: An Interpretive Review." *Journal of Human Resources* 37 (4): 696–727.

Dixit, Avinash, and John Londregan. 1995. "Redistributive Polticis and Economic Efficiency." *American Political Science Review* 89 (4): 856–866.

Einstein, Katherine Levine, and David Glick. 2016. "Does Race Affect Access to Government Services? An Experiment Exploring Street-Level Bureaucrats." *American Journal of Political Science* 61 (1): 100–116.

Evans, Peter B., and James E. Rauch. 1999. "Bureaucracy and Growth: A Cross-National Analysis of the Effects of "Weberian" State Structures on Economic Growth." *American Sociological Review* 64 (5): 748–765.

Fergusson, Leopoldo, Pablo Querubín, Nelson A. Ruiz, and Juan F. Vargas. 2018. "The Real Winner's Curse." Working paper.

Fiszbein, Ariel, Norbert Schady, Francisco H.G. Ferrerira, Margaret Grosh, Niall Keleher, Pedro Olinto, and Emmanuel Skoufias. 2009. *Conditional Cash Transfers: Reducing Present and Future Poverty*. Washington DC: The International Bank for Reconstruction and Development/The World Bank.

Fizbein, Ariel. 1997. "The Emergence of Local Capacity: Lessons From Colombia." *World Development* 25 (7): 1029–1043.

Gaikwad, Nikhar, and Gareth Nellis. 2017. "Do Politicians Discriminate Against Internal Migrants? Evidence from Nationwide Field Experiments in India." Working paper.

Garay, Candelaria. 2016. *Social Policy Expansion in Latin America*. New York, NY: Cambridge University Press.

González González, Fernán E. 2014. *Poder y Violencia en Colombia*. Bogotá, D.C.: ODECOFI-CINEP.

Hemker, Johannes, and Anselm Rink. 2017. "Multiple Dimensions of Bureaucratic Discrimination: Evidence from German Welfare Offices." *American Journal of Political Science* Forthcoming.

Ibáñez, Ana María, and Carlos Eduardo Vélez. 2008. "Civil Conflict and Forced Migration: The Micro Determinants and Welfare Losses of Displacement in Colombia." *World Development* 36 (4): 659–676.

Kasara, Kimuli, and Pavithra Suryanarayan. 2015. "When Do the Rich Vote Less Than the Poor and Why? Explaining Turnout Inequality across the World." *American Journal of Political Science* 59 (3): 613–627.

Lindbeck, Assar, and Jorgen W. Weibull. 1987. "Balanced-Budget Redistribution as the Outcome of Political Competition." *Public Choice* 52 (3): 273–297.

Lipsky, Michael. 2010. *Street-Level Democracy: Dilemmas of the Individual in Public Service*. 2nd edition ed. New York, NY: Russell Sage Foundation.

Martine, George. 1975. "Volume, Characteristics and Consequences of Internal Migration in Colombia." *Demography* 12 (2): 193–208.

Martz, John D. 1997. *The Politics of Clientelism in Colombia*. New Brunswick, NJ: Transaction Publishers.

Mayka, Lindsay R. 2016. "Colombia's Surprising Resilience." *Journal of Democracy* 27 (3): 139–147.

McCubbins, Matthew D., and Thomas Schwartz. 1984. "Congressional Oversight Overlooked: Police Patrols versus Fire Alarms." *American Journal of Political Science* 28 (1): 165–179.

Ocampo, Gloria Isabel. 2014. *Poders Regionales, Clientelismo, y Estado: Etnografía del Poder y La Política en Córdoba, Colombia*. Bogotá, D.C.: ODECOFI-CINEP.

Pepinsky, Thomas B., Jan H. Pierskalla, and Audrey Sacks. 2017. "Bureaucracy and Service Delivery." *Annual Review of Political Science* 20: 249–268.

Phelps, Edmund S. 1972. "The Statistical Theory of Racism and Sexism." *American Economic Review* 62 (4): 659–661.

Prendergast, Canice. 2003. "The Limits of Bureaucratic Efficiency." *Journal of Political Economy* 111 (5): 929–958.

Prendergast, Canice. 2007. "The Motivation and Bias of Bureaucrats." *American Economic Review* 96 (1): 180–196.

Raffler, Pia. 2017. "Does Political Oversight of the Bureaucracy Increase Accountability? Field Experimental Evidence from an Electoral Autocracy." Working paper.

Rizzo, Tesalia. 2018. "When Clients Exit: Breaking the Clientelistic Feedback Loop." Working paper.

Sanders, James E. 2004. *Contentious Republicans: Popular Politics, Race, and Class in Nineteenth-Century Colombia*. Durham, NC: Duke University Press.

Schmidt, Steffen W. 1974. "Bureaucrats as Modernizing Brokers? Clientelism in Colombia." *Comparative Politics* 6 (3): 425–450.

Ting, Michael M. 2017. "Politics and Administration." *American Journal of Political Science* 61 (2): 305–319.

Uribe de Hincapié, María Tereesa, and Jesús María Álvarez Gaviria. 1998. *Las Raíces del Poder Regional: El Caso Antioqueño*. Medellín: Editorial Universidad de Antioquia.

White, Ariel R., Noah L. Nathan, and Julie K. Faller. 2015. "What Do I Need to Vote? Bureaucratic Discretion and Discrimination by Local Election Officials." *American Political Science Review* 109 (1): 129–142.