

CHAPTER 6

The Personnel Economics of the Developing State

F. Finan^{*}, B.A. Olken^{§,1}, R. Pande^{¶,a}

^{*}University of California, Berkeley, Berkeley, CA, United States

[§]Massachusetts Institute of Technology, Cambridge, MA, United States

[¶]Harvard University, Cambridge, MA, United States

¹Corresponding author: E-mail: bolken@mit.edu

Contents

1. Introduction	468
2. Stylized Facts on the Architecture of the State and the Role of Individuals	470
2.1 Key features of the state	470
2.2 Evidence from household surveys	472
3. The Selection and Recruitment of Public Officials	482
3.1 Financial incentives	485
3.1.1 <i>Effects of financial incentives on the applicant pool</i>	485
3.1.2 <i>Effects of financial incentives on recruitment</i>	487
3.2 How should governments screen?	488
4. Using Incentives to Improve Performance	491
4.1 Financial incentives	492
4.1.1 <i>Incentives for agents of government authority</i>	492
4.1.2 <i>Incentives for frontline service providers</i>	493
4.2 Nonfinancial incentives	497
4.2.1 <i>Transfers and postings</i>	497
4.2.2 <i>Intrinsic motivation</i>	498
4.3 Summary	499
5. Monitoring Mechanisms and Public Service Delivery	500
5.1 Overview	500
5.2 Information flows and monitoring	500
5.3 Government monitoring	500
5.3.1 <i>Does more information on performance improve outcomes?</i>	500
5.3.2 <i>Who collects information and does that matter?</i>	503

^a The authors thank Nils Enevoldsen and Joyce Hahn for exceptional research assistance, and Abhijit Banerjee, Alan Gerber, and Adnan Khan for comments. The authors acknowledge financial support from the William and Flora Hewlett Foundation and the UK Department for International Development. The views expressed here are those of the authors alone.

5.4 Information flows and monitoring by citizens	504
5.4.1 Does information on program performance matter?	504
5.5 Summary	505
6. Towards Smart(er) Governance: the Promise of e-Governance and Other Avenues	505
7. Concluding Thoughts	507
Appendix	509
References	511

Abstract

Governments play a central role in facilitating economic development. Yet while economists have long emphasized the importance of government quality, historically they have paid less attention to the internal workings of the state and the individuals who provide the public services. This chapter reviews a nascent but growing body of field experiments that explores the personnel economics of the state. To place the experimental findings in context, we begin by documenting some stylized facts about how public sector employment differs from that in the private sector. In particular, we show that in most countries throughout the world, public sector employees enjoy a significant wage premium over their private sector counterparts. Moreover, this wage gap is largest among low-income countries, which tends to be precisely where governance issues are most severe. These differences in pay, together with significant information asymmetries within government organizations in low-income countries, provide a *prima facie* rationale for the emphasis of the recent field experiments on three aspects of the state-employee relationship: selection, incentive structures, and monitoring. We review the findings on all three dimensions and then conclude this survey with directions for future research.

Keywords

Financial incentives; Public sector; Public sector wage gap; Public service delivery or public goods; State capacity

JEL Codes

H40; M50; O10; O43

1. INTRODUCTION

Countries vary in their quality of governance, and by almost any measure, governance is significantly worse in low-income countries. For instance, the World Bank's Worldwide Governance Indicators project rank low-income countries substantially lower than the high-income countries that are members of the Organization for Economic Cooperation and Development (OECD) on government effectiveness (average percentile rank of 17.3 compared to 87.9 in 2014). For frontline service providers, such as teachers and nurses, a well-known study of six developing countries found that 19 percent of public primary school teachers and 35 percent of public health care workers were absent at the time of random unannounced visits to schools and clinics. The absenteeism was worse in poorer countries and in poorer states within India, with the worst absenteeism rates approaching 40 percent (Chaudhury et al., 2006). Given the natural role governments play in facilitating development, whether by providing public goods, addressing

externalities, or providing the foundation for private property and private enterprise, improvements in government performance are likely to lead to significant economic development gains.

In this chapter, we examine a particular determinant of government performance: the individuals who perform government functions. We focus on two groups of public employees: appointed civil servants, which we broadly construe to include administrators with effectively permanent government appointments and frontline service providers (e.g., teachers, nurses, firefighters, and trash collectors), who may have either permanent appointments or temporary contracts¹.

In many developing countries, policy actors and researchers attribute poor governance to public employees being lazy, corrupt, or both. Yet, it may be that the poor institutional structure within which public employees work in these countries is the dominant cause. The two may also interact: poor institutional structures may cause the lazy and the corrupt to select into public service. Since both institutional structures and personnel selection are endogenously determined, establishing causality is hard. By providing a clean empirical method for identification, field experiments can help cut the Gordian knot and identify problems in personnel selection and management, elucidate the causes of these problems, and suggest potential solutions. These contributions of field experiments are the focus of this chapter.

While the focus of this chapter is the “personnel economics” of the government sector—in particular, selection and recruitment of personnel, incentives, and monitoring—there are many aspects of the government as an organization that are also clearly important but not covered by this review.² In particular, a recent literature has emphasized the role of management skills and techniques, which are conceptually distinct from the people who implement them (e.g., Bloom and Van Reenen, 2007; Bloom et al., 2013; Rasul and Rogger, 2015). While these are key issues in determining the overall “Total Factor Productivity” (TFP) of the government sector, they are not the issues we explicitly consider here.

The chapter is organized as follows. In Section 2, we situate the problem by documenting several stylized facts of how the public sector as an employer differs from the private sector. We use survey microdata from 32 countries to establish some facts about the relative compensation and fringe benefits of public and private sector workers. The per capita Gross Domestic Product (GDP) of our study countries ranges from \$264 to \$45,710 (constant 2005 USD). The World Bank classifies 19 of our sample countries as low income or lower middle income. We use these surveys to document the stark difference between the public–private gaps in worker pay and worker tenure in developing

¹ The occupational categories of contract workers are typically varied, but important categories in our empirical sample include teachers, nurses, office cleaners, and helpers.

² For a review of personnel economics with a focus on the private sector, see, for example, Lazear and Oyer (2012).

versus developed countries: the public sector enjoys a large wage premium in poor countries but only a small—or, in some cases, zero—premium in developed countries. Public sector jobs are also more likely to provide fringe benefits like health insurance and pensions; again, the difference is much starker in poor as opposed to rich countries.

These stylized facts point to a large pay premium for public sector employees in low-income countries. The premium could reflect the more complex nature of public sector jobs in low-income countries, elite capture of the public sector, efficiency wages designed to prevent corruption à la [Becker and Stigler \(1974\)](#), differences in job preferences across societies, or some combination of these factors. Whatever the underlying cause may be, this pay premium—in combination with weak information flows within and across government agencies—has important implications for how individuals select into the public sector and their subsequent performance and incentive structure. Against this background, we argue that field experiments can play an important role in helping us understand the links between governance outcomes and one dimension of state capacity: the traits and behavior of public employees.

[Sections 3–5](#) then describe how field experiments have informed our understanding of the current problems in three arenas (selection and recruitment, incentives, and monitoring) and helped identify potential civil service reforms. In [Section 6](#), we discuss whether the changing nature of technology-driven aids for service delivery may help poor countries create smart governance structures, perhaps by constraining or eliminating human interaction. We conclude with directions for future research.

2. STYLIZED FACTS ON THE ARCHITECTURE OF THE STATE AND THE ROLE OF INDIVIDUALS

2.1 Key features of the state

The state, the world over, consists of a set of interlinked institutions staffed by officials. Typically, a mix of constitutional acts and legislative and executive orders defines the fiscal and regulatory powers of these institutions. In the Weberian model, adopted in virtually all countries (though to varying degrees; see [Evans and Rauch, 1999](#)), apolitical civil servants, in turn, are responsible for implementing the state mandate.

Several key features of the state distinguish its personnel practices from those of the private sector, particularly in developing countries, and here, we identify five features that are relevant for our analysis.

First, the state has a long horizon. Absent significant civil conflict, most states anticipate collecting and spending revenues indefinitely. This allows states to make long-lived promises to its employees, such as pensions, which may be difficult for the private sector to make. The long-lived reputation of the state as an employer also means that it may be reluctant to renege on such promises; in fact, many promises of the state, again such as pensions, survive radical regime change, at least in nominal terms.

Second, the set of contracts a state can offer its employees is limited. Whereas shareholders can create strong incentives for CEOs to maximize returns, the mechanisms that the ultimate principals—citizens—have at their disposal, namely elections, are coarser and more limited. Thus, politicians may seek to use jobs and the wages associated with them, to reward their political supporters, cronies, and friends. Politicians could also be tempted to use promotions or incentives to exert undue influence on civil servants. These issues are more prevalent in the public sector than in the private sector because of the lack of discipline from the profit motive: the politician only indirectly bears the cost of inefficiency to the extent that voters are less likely to re-elect him because of it, whereas the owner of a firm directly feels the financial losses associated with these types of inefficiencies (Shleifer and Vishny, 1994; Boycko et al., 1996). To counteract the tendency for each new politician to replace large numbers of government employees with his political supporters, over time governments have enacted rigid civil service rules that restrict the discretion politicians have over hiring and firing (Evans, 1995 AER).³ In fact, these civil service systems are typically much more rigid than their private sector counterparts, with strict formulas defining the hiring criteria, promotion patterns, and wage levels. The need to isolate the employment decision from political influence—which underlies a substantial amount of public sector personnel policy—suggests that the personnel economics of the state are likely to substantially differ from those of the private sector.

Another restriction on contracts is that public sector compensation usually does not include pay for performance. Performance pay for bureaucrats can create severe multitasking problems, where bureaucrats focus on the incentivized dimension of their job at the expense of the nonincentivized dimension (Holmstrom and Milgrom, 1987). While multitasking is an issue in many contexts, it can be particularly severe in public sector contexts where agents wield substantial authority (e.g., police and judges), and it is hard to find an objective measure of the “truth” on which to incentivize them. In practice, while financial incentives for government workers were historically quite common, they had a tendency to lead to overzealous and unpopular bureaucrats who were perceived to abuse their positions in order to overextract from the population (e.g., Parrillo, 2013), which led the populace to demand less strongly incentivized civil servants.

Third, the nature of goods exchanged between the state and citizens is substantially different than with the private sector. Very often, services provided by the state—like schooling and health—are heavily subsidized, thus limiting the competition the state faces

³ The United States at the turn of the 19th century provides a nice illustration. Prior to the Pendleton Act, federal employment was an important source of patronage. Under this system, these employees did not have tenure and turnover rates were high during changes in administration. After the passage of the Pendleton Act, which restricted the number of patronage positions, civil service reform started to take hold and federal employees began to get hired based on merit and public service exams. With these reforms, the job took on a different form. Federal employees were granted tenure and dismissals became more difficult and costly. Compensation became more formulaic which resulted in more wage compression.

from other providers. Because there is little competition from other providers, it becomes harder to base worker incentives on simple metrics like volume of services, and the lack of competitive pressure makes direct monitoring of service providers more important for the public sector (relative to the private sector where competition will naturally weed out less productive firms).

Fourth, government careers differ from nongovernment careers in the mission of the organization: government organizations often aspire to public service and private sector ones to profit. This, arguably, implies that different types of individuals are potentially drawn to the public and private sector careers, and personnel practices should account for this. The state may also seek different types of individuals who are able to balance the multifaceted objective functions inherent to a public service organization. [For a theoretical discussion of how matching mission organization to agent preferences can improve efficiency, see [Besley and Ghatak \(2005\)](#).]

Finally, a fifth key feature of the state is that it self-regulates to a much larger extent than the private sector. Monitoring structures are often embedded within the bureaucracy, and there are relatively few instances of third-party private auditing of government services. Moreover, in several cases, workers often switch between service delivery and monitoring roles, which potentially leads to conflicts of interest.

2.2 Evidence from household surveys

To quantify some of these differences between the public and private sector, we obtained household survey microdata from 32 countries around the world. [Appendix Table 1](#) lists these countries and the data sources.

We classify each working adult as a public or private sector employee using information provided in the survey specifying the type of employer. For these adults, we also know their wages and usually their occupations and the number of years at their primary jobs. In terms of other forms of employer-provided compensation, we have information on pensions for 13 countries and on health insurance for 14 countries. (We exclude state-provided benefits which may be available to all citizens, irrespective of work status.) In terms of worker demographics, in addition to occupation, we know gender, age, and education level.

We begin by estimating the public sector wage premium, separately for each country c , as follows:

$$y_{irc} = \alpha_c + \beta_c \text{Public}_{irc} + \gamma_c X_{irc} + \varepsilon_{irc}$$

where i indexes an individual, c indexes a country, r indexes a region within a country, Public_{irc} is a dummy for public sector, and X_{irc} are vector of controls (age, gender, secondary education, tertiary education, and occupation dummies). We begin with log wages as the outcome variable but also consider other aspects of compensation, such as pensions, health benefits, and job tenure. The key coefficient of interest is β_c ,

which captures, separately for each country c , the differential return for being in the public sector.

Table 1 reports the results where the outcome variable is log wages. Each cell reports the coefficient β_c from a separate regression. Column (1) reports the basic model with a single constant term for each country (i.e., no regional fixed effects) and no control variables. Column (2) adds regional (usually, province level) dummies for each country to capture geographic differences; for example, government workers may be disproportionately located in the capital city, which may have different wage levels.⁴ Column (3) adds as covariates age, gender, and education. Column (4) adds occupation fixed effects (e.g., in Uganda, commonly observed occupations included mechanics, nurses, and midwives, managing supervisors, transport laborers, restaurant service workers, and machinery mechanics and fitters). For many occupations there is substantial overlap between public and private sector. Teachers, for example, work in both sectors, as do many types of service workers. We show the results separately for each country.

In order to facilitate interpretation of the results, Fig. 1 plots the estimated coefficient β , the public sector premium, against each country's 2010 per capita GDP, measured in terms of Purchasing Power Parity (PPP). The left panel plots the coefficients from column (1), the unadjusted public–private wage difference. The right panel plots the coefficients from column (4), that is, after including region fixed effects, occupation fixed effects, and demographic controls. Since the public sector was substantially different in communist economies, we plot two regression lines: the red line shows a regression line for current and former communist countries, and the blue line shows the regression line for all other countries.

The left graph shows that, in almost all countries unadjusted average pay in the public sector exceeds that in the private sector. The difference is much starker in poor countries: in the poorest countries in our sample, such as Malawi, Niger, Tanzania, and Kenya, public sector workers earn more than double the average wage in the private sector. This declines to the point where for the rich countries in our sample—Korea, the United Kingdom, and the United States—the pay difference, while present, is on the order of 4–20 percent.

The right graph in Fig. 1 plots the difference after we include all the controls discussed above. For nonformer communist countries, we still see a negative relationship between income and public sector wages, but it is muted substantially; the poorest four countries in the sample have a positive public sector premium of around 0.1, whereas the richest three countries in the sample have a negative public sector premium of around -0.03 . Comparing the results in each column, the key difference from the unadjusted results

⁴ Whenever possible, we used enumeration area codes or primary sampling unit codes for geographic fixed effects. For the countries that did not have those, we grouped multiple geographic identifiers together. For example, for the case of Argentina, we grouped region and agglomeration identifiers. For South Africa, we grouped province and district council. For the United States, we grouped region, FIPS state code, and metropolitan CBSA FIPS code.

Table 1 Log pay on public sector: four models of increasing specification

	(1)	(2)	(3)	(4)
	Basic model	Add region fixed effects	Add individual demographic adjustments	Add occupation fixed effects
Albania	0.077* (0.044)	0.124* (0.068)	0.017 (0.080)	-0.001 (0.095)
Argentina	0.3468*** (0.012)	0.373*** (0.011)	0.241*** (0.012)	0.041*** (0.013)
Armenia	-0.098** (0.050)	-0.157** (0.062)	-0.061 (0.061)	-0.073 (0.074)
Bolivia	0.202*** (0.078)	0.176* (0.095)	-0.042 (0.101)	-0.077 (0.126)
Bosnia and Herzegovina	0.037 (0.025)	0.043* (0.025)	-0.051** (0.025)	-0.119*** (0.027)
Bulgaria	-0.076*** (0.017)	-0.061*** (0.019)	-0.043** (0.019)	-0.059*** (0.021)
Colombia	0.633*** (0.109)	0.457*** (0.137)	0.302** (0.134)	0.127 (0.148)
Egypt	-0.150*** (0.017)	-0.102* (0.054)	-0.274*** (0.061)	-0.227* (0.124)
Georgia	-0.000 (0.061)	0.083 (0.062)	0.137** (0.060)	0.243*** (0.078)
Ghana	0.937*** (0.089)	0.956*** (0.129)	0.588*** (0.136)	0.764*** (0.187)
India	1.113*** (0.009)	1.101*** (0.010)	0.712*** (0.010)	0.641*** (0.012)
Indonesia	0.847*** (0.032)	0.953*** (0.033)	0.462*** (0.036)	0.546*** (0.039)
Iraq	0.162*** (0.009)	0.127*** (0.011)	0.108*** (0.012)	-0.039** (0.018)
Kenya	0.886*** (0.107)	0.770*** (0.125)	0.473*** (0.116)	0.419*** (0.140)
Korea, Rep.	0.040 (0.101)	0.093 (0.112)	0.110 (0.095)	-0.224 (0.208)
Laos	-0.265*** (0.053)	-0.182*** (0.063)	-0.375*** (0.079)	-0.483*** (0.123)
Malawi	0.800*** (0.039)	0.700*** (0.046)	0.234*** (0.049)	0.126** (0.054)
Mexico	0.477*** (0.005)	0.496*** (0.005)	0.433*** (0.006)	0.220*** (0.007)
Nicaragua	0.502*** (0.028)	0.455*** (0.028)	0.262*** (0.081)	0.153 (0.130)
Niger	0.986*** (0.088)	0.676*** (0.120)	0.249** (0.118)	-0.056 (0.149)

Table 1 Log pay on public sector: four models of increasing specification—cont'd

	(1)	(2)	(3)	(4)
	Basic model	Add region fixed effects	Add individual demographic adjustments	Add occupation fixed effects
Nigeria	0.719*** (0.059)	0.669*** (0.083)	0.355*** (0.086)	0.359*** (0.106)
Pakistan	0.540*** (0.013)	0.541*** (0.014)	0.258*** (0.013)	0.322*** (0.014)
Panama	0.637*** (0.020)	0.574*** (0.022)	0.465*** (0.022)	0.388*** (0.023)
Peru	0.296*** (0.030)	0.188*** (0.034)	0.065** (0.032)	0.068* (0.038)
Serbia	0.419*** (0.019)	0.333*** (0.019)	0.249*** (0.019)	0.096*** (0.022)
Sri Lanka	0.437*** (0.047)	0.341*** (0.061)	0.347*** (0.069)	0.252*** (0.102)
Tajikistan	0.061 (0.048)	-0.201*** (0.056)	-0.193*** (0.057)	-0.317*** (0.091)
Tanzania	0.962*** (0.057)	0.962*** (0.066)	0.483*** (0.078)	0.285*** (0.097)
Timor Leste	0.325*** (0.075)	0.159* (0.096)	0.008 (0.101)	-0.086 (0.101)
Uganda	0.546*** (0.088)	0.611*** (0.108)	0.051 (0.111)	0.035 (0.139)
United Kingdom	0.086*** (0.015)	0.094*** (0.015)	0.036** (0.014)	0.025 (0.017)
United States	0.216*** (0.019)	0.216*** (0.019)	0.113*** (0.019)	0.108*** (0.020)
Vietnam	0.120** (0.052)	0.040 (0.060)	-0.082 (0.058)	-0.066 (0.060)

Standard errors in parentheses.

Each table cell contains a within-country estimate of the difference in log pay between the public sector and the private sector. Ambiguous sectors such as NGO are omitted. Estimates are obtained with Ordinary Least Squares (OLS) regression. Column 1 is the basic specification. To column 1, column 2 adds region fixed effects, such as district or municipality. To column 2, column 3 adds covariates for age, gender, and indicators for completion of secondary education and for completion of any tertiary education.

To column 3, column 4, the fully specified model, adds occupation fixed effects, as determined by occupation codes.

Each country's data come from a single household-level survey, as listed in the [Appendix](#).

*, $p < 0.10$; **, $p < 0.05$; ***, $p < 0.01$.

comes when we add controls for education in column (3), an issue we will return to in more detail in the following section. For ex-communist countries, there is an upward slope, so that the government workers in poor former or current communist countries appear to have a substantial negative wage premium relative to their private sector counterparts.

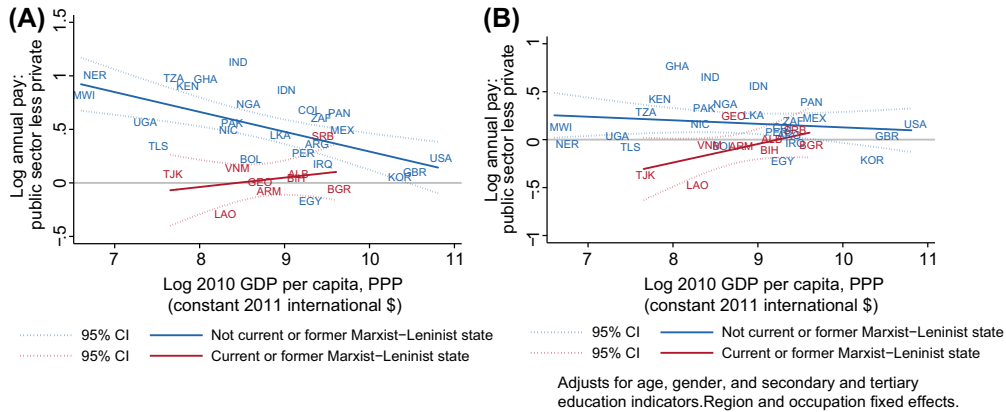


Figure 1 Public sector pay premium by GDP per capita. (A) Public sector pay premium, basic model. (B) Public sector pay premium, fully specified model. *CI*, Confidence interval.

Table 2 and Figs. 1–3 show the results for other outcome variables, notably whether the worker receives health insurance, whether he receives a pension, and his tenure on the job. The public sector looks remarkably different from the private sector in these other benefits in both rich and poor countries. In particular, on average, public sector workers are about 20 percentage points more likely to receive health insurance than private sector workers, even conditional on other job characteristics, and this does not systematically differ between rich and poor countries. In some countries, these differences are even greater; in India, for example, the public sector is 48 percentage points more likely to receive health insurance and 55 percentage points more likely to receive pensions, even conditional on job characteristics. These fringe benefits, which may reflect the government’s ability to honor commitments across states of the world (e.g., if a worker falls ill) or over time (i.e., when a worker grows old) given its much longer time horizons, are a notable difference between the public and private sector.

A different type of benefit, which we focus on in Fig. 4, is job tenure. We observe a significant public sector premium: in the unadjusted model public sector workers report having had that job for 5 years longer than their private sector counterparts. In the adjusted model, we continue to see a significant positive premium of roughly 3 years. There is no trend in this premium by country income.

Given the differences between the adjusted and unadjusted wage regressions, we next examine the differences in observable demographic characteristics. In Fig. 5, we examine worker gender. (The figure plots country-wise coefficients on the public sector dummy where the estimation equation is of the form in column (1) of Table 1.) On average, countries have a relatively higher fraction of women in the public sector relative to the private sector, and feminization of the public sector relative to private sector is

Table 2 Job benefits on public sector

	(1)	(2)	(3)
	Log pay	Health benefits	Pension
Argentina	0.041*** (0.013)	0.155*** (0.008)	0.146*** (0.009)
Bosnia and Herzegovina	-0.119*** (0.027)	0.194*** (0.022)	0.208*** (0.022)
Bulgaria	-0.059*** (0.021)	-0.084*** (0.011)	-0.084*** (0.011)
Egypt	-0.227* (0.124)	0.334*** (0.064)	—
India	0.641*** (0.012)	0.483*** (0.007)	0.545*** (0.007)
Indonesia	0.546*** (0.039)	0.330*** (0.017)	0.479*** (0.016)
Iraq	-0.039** (0.018)	0.346*** (0.011)	0.603*** (0.011)
Mexico	0.220*** (0.007)	0.133*** (0.004)	0.133*** (0.004)
Nicaragua	0.153 (0.130)	0.176** (0.085)	-0.214** (0.084)
Niger	-0.056 (0.149)	0.249*** (0.061)	0.286*** (0.076)
Nigeria	0.359*** (0.106)	—	0.555*** (0.055)
Peru	0.068*** (0.038)	0.048 (0.033)	—
Serbia	0.096*** (0.022)	0.190*** (0.012)	0.402*** (0.026)
South Africa	0.185*** (0.015)	0.374*** (0.006)	0.259*** (0.005)
Uganda	0.035 (0.139)	-0.040 (0.063)	0.047 (0.071)
United States	0.108*** (0.020)	0.157*** (0.012)	0.204*** (0.012)

Standard errors in parentheses.

Each table cell contains a within-country estimate of the difference in benefits between the public sector and the private sector. Ambiguous sectors such as NGO are omitted. Estimates are obtained with OLS regression, using the fully specified model described in Table 1. Estimates have region and occupation fixed effects, as well as covariates for age, gender, and indicators for completion of secondary education and for completion of any tertiary education. The dependent variable in column 1 is log pay and is identical to column 4 of Table 1. The dependent variable of column 2 is an indicator for employer-provided health benefits, such as insurance. The dependent variable for column 3 is an indicator for employer-provided pension. Each country's data come from a single household-level survey, as listed in the Appendix.

Albania and Mexico have single variables for health and pension; these estimates are duplicated across health and pension columns. Nicaragua's pension variable indicates if employer offers any pension benefits, while the health variable asks about health benefits in addition to but not separate from pension benefits.

*, $p < 0.10$; **, $p < 0.05$; ***, $p < 0.01$.

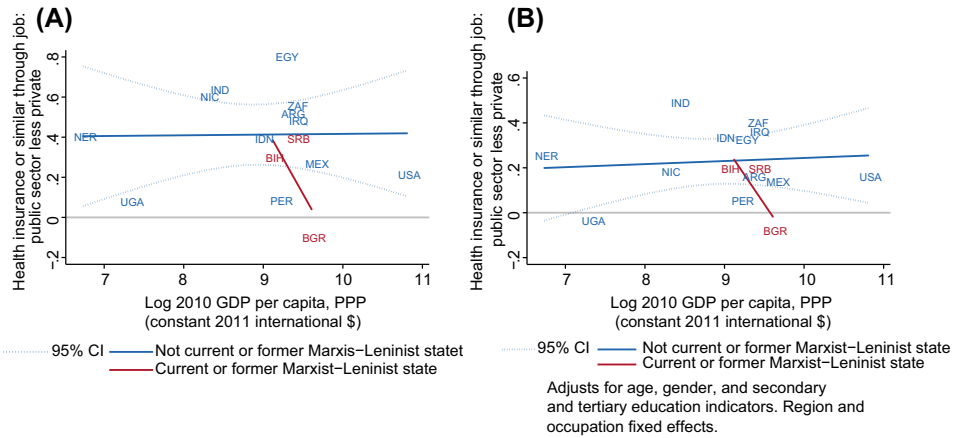


Figure 2 Public sector health benefit premium by GDP per capita. (A) Public sector health benefit premium, basic model. (B) Public sector health benefit premium, fully specified model. *CI*, confidence interval.

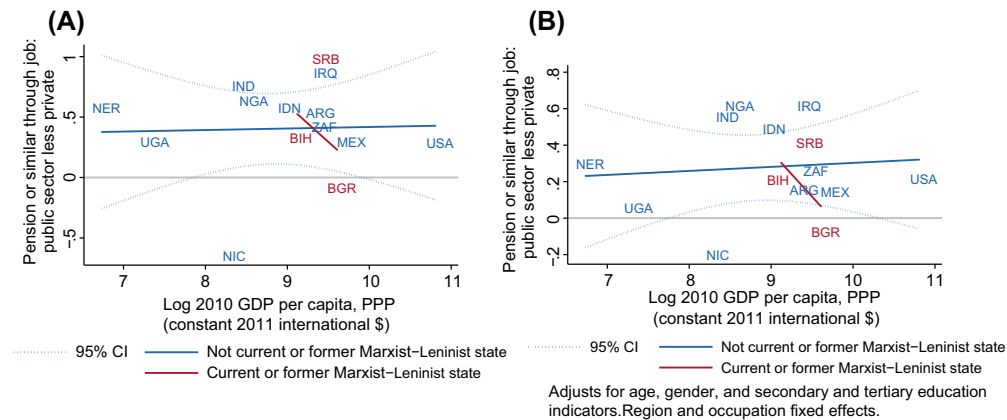


Figure 3 Public sector pension premium by GDP per capita. (A) Public sector pension premium, basic model. (B) Public sector pension premium, fully specified model. *CI*, confidence interval.

increasing with income. Fig. 6 considers age, and we find that public sector workers tend to be, on average, 5 years older. There is no significant gradient with country income.

In Fig. 7, we turn to education. The variable of interest is an indicator of whether the worker has completed at least secondary education. We observe both a public sector education premium, wherein more educated workers are attracted to the public sector at a higher rate and a significant negative gradient in this education premium by county

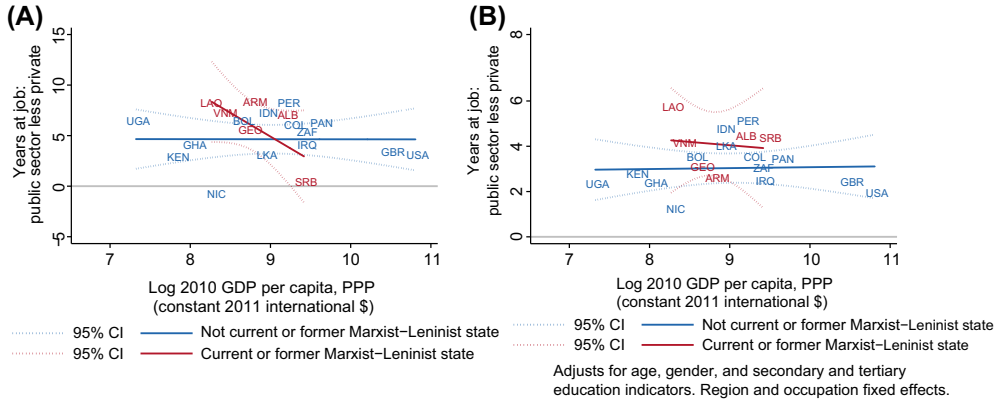


Figure 4 Public sector tenure premium by GDP per capita. (A) Public sector tenure premium, basic model. (B) Public sector tenure premium, fully specified model. *CI*, confidence interval.

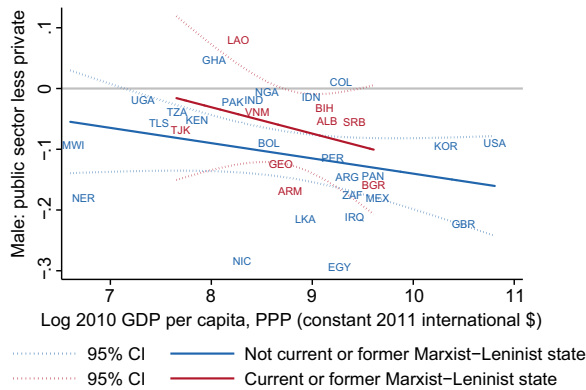


Figure 5 Public sector gender difference by GDP per capita. *CI*, confidence interval.

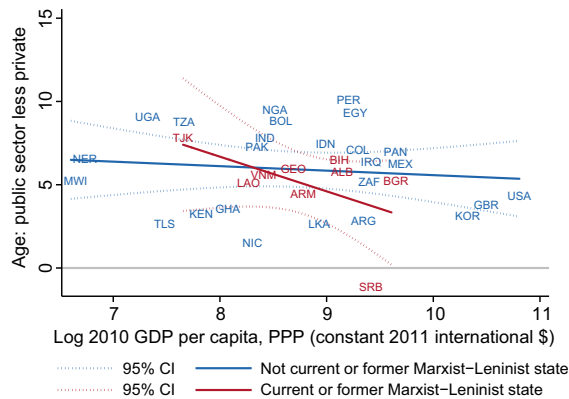


Figure 6 Public sector age difference by GDP per capita. *CI*, confidence interval.

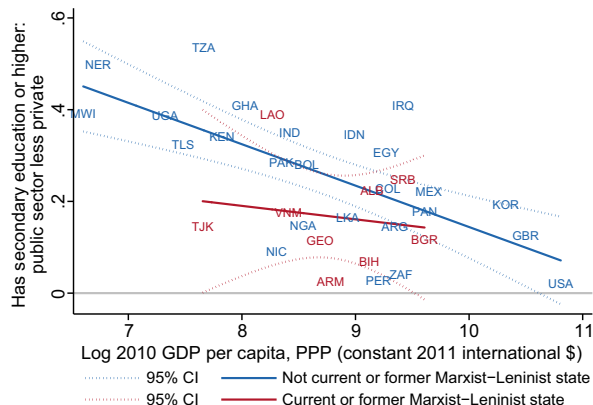


Figure 7 Public sector education difference by GDP per capita, unadjusted. *CI*, confidence interval.

income. This strong gradient in the education premium, together with its much weaker fully specified counterpart in Fig. 8, underlies the significant difference in the income gradient in public sector wage premium across panels (A) and (B) of Fig. 1.

Finally, in Fig. 9, we report estimates based on a basic Mincer wage regression where we report the differential public sector wage premium for years of education. Differential Mincerian educational premia are low and show no income gradient.

The survey data paint an intriguing picture: In poorer countries, the public sector attracts relatively more educated individuals, who in turn are paid an equivalent wage and, in almost all places, higher wage premium. Yet, qualitative evidence suggests that the quality of government services is lower in developing countries than in developed ones. While there are a myriad of explanations for poor government performance, the evidence presented here suggests that—assuming similar job descriptions across countries—one can likely rule out the explanation that government workers are underpaid relative to their private sector counterparts.⁵ Rather, it is likely that, in settings with weak information flows within the government and between the government and citizens, high wages per se are insufficient to motivate performance (though we suggest further direct evidence on this relationship in the following section).

The key point established here is that the public sector is substantially different, in terms of the level of wages, composition of the labor force, fringe benefits, and tenure. But understanding the impact of these personnel policies—and ways they can be altered

⁵ We, of course, do not know the counterfactual. It may be that government performance in developing countries would be even worse if pay was lower, if high pay is acting as an efficiency wage to prevent corruption. Alternatively, it may be that the job description is more taxing in low-income countries, causing the effective wage to be lower. While we cannot directly rule this out, the evidence on, for instance, teacher and health worker absenteeism across rich and poor countries would go against such an explanation.

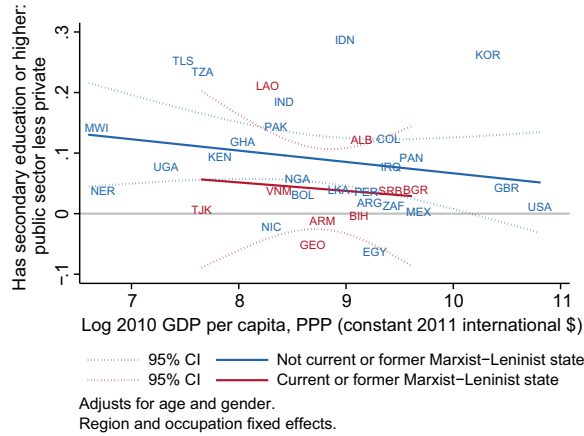


Figure 8 Public sector education difference by GDP per capita, adjusted. *CI*, confidence interval.

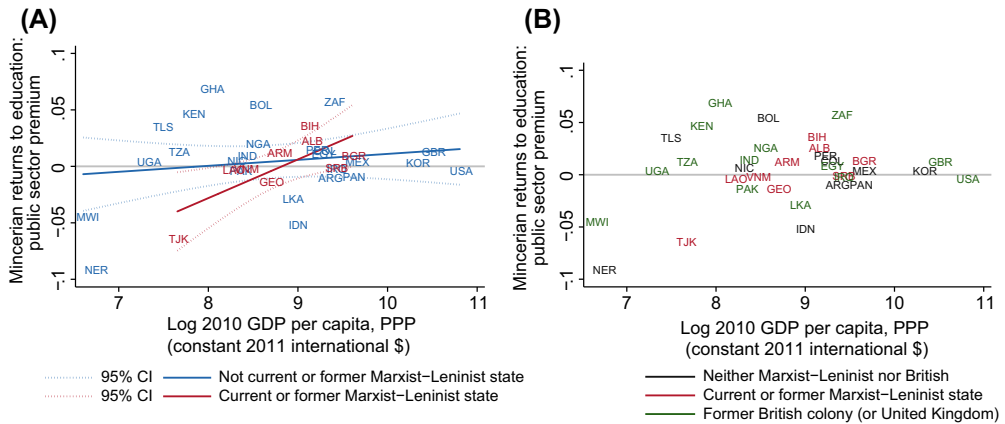


Figure 9 Mincerian returns to education, public sector premium. (A) Mincerian returns to education, public sector premium. (B) Mincerian returns to education, public sector premium. *CI*, confidence interval.

to improve government performance—is challenging. Field experiments offer an attractive way of providing evidence on these outcomes, by examining what happens when these practices are altered or changed on various dimensions. To explore these issues in detail and shed light on how to think about various aspects of personnel policy in the government sector, in the remaining sections, we explore the evidence on three dimensions through which performance of government employees is determined and potentially can be improved: selection and recruitment (Section 3), incentives (Section 4), and monitoring (Section 5). In each case, we describe existing evidence (focused, in particular, on field experiments) and point out important open questions for future

research. The first section (selection and recruitment) considers how to improve the selection of public employees with the set of attributes that best contribute to public sector performance; the next two sections (incentives and monitoring) consider how to improve the performance of a given set of individuals, holding their basic characteristics fixed.

3. THE SELECTION AND RECRUITMENT OF PUBLIC OFFICIALS

The first questions we consider are: who are government employees, how are they recruited, and are there ways of improving the recruitment process? After all, individuals choose their career paths given the options available to them. In addition to the differences in compensation and job tenure documented previously, a key difference between government and nongovernment careers that we identified earlier is mission differences: government organizations often aspire to public service and private sector ones to profit. On the other hand, the rampant corruption in the public sector in developing countries may attract those who are interested in pursuing corrupt activities. More generally, [Table 3](#) highlights several descriptive differences between the public and private sector across countries.

Do mission-driven organizations, such as public bureaucracies or private nonprofit organizations, attract employees with high levels of prosocial motivation? The idea is that some individuals care about benefiting others and thus feel drawn to organizations that provide them with the opportunity to do so. In general, the literature in public administration and economics supports this idea. For example, [Cowley and Smith \(2014\)](#) use data from the World Values Survey to measure the intrinsic motivation of public and private sector employees in 52 countries. They find that public sector workers are on average much more intrinsically motivated than private sector workers, even after adjusting for differences in basic socioeconomic characteristics. Similarly, [Banuri and Keefer \(2013\)](#) sampled about 1700 individuals from the government and private sectors in Indonesia and had them play a dictator game. They also find that subjects in the governmental sector are more prosocial.⁶

Besides the prosocially motivated, public bureaucracies can also attract individuals with less desirable personality traits. Organizations that offer low-powered incentives or are unable to hold their employees accountable can attract individuals with limited aspirations and a poor work ethic. Widespread corruption may attract dishonest or venal individuals. Recent evidence also supports this view. In the study by [Cowley and Smith \(2014\)](#), for example, although public sector workers tend to be more intrinsically

⁶ Other examples include: [Dohmen and Falk \(2010\)](#) find that German teachers are more trusting and less negatively reciprocal than employed nonteachers. [Lagarde and Blaauw \(2014\)](#) find in an adapted dictator game that giving to patients predicts student nurses' subsequent decisions to take rural hardship posts in South Africa. See [Perry and Hondeghe \(2008\)](#) for additional studies examining the role of prosocial motivation in selection in public service.

Table 3 Demographic differences of public sector

	(1)	(2)	(3)	(4)
	Male	Age	Secondary or tertiary education	Tertiary education
Albania	-0.039* (0.024)	5.748*** (0.528)	0.224*** (0.018)	0.320*** (0.021)
Argentina	-0.147*** (0.008)	2.832*** (0.203)	0.145*** (0.006)	0.258*** (0.008)
Armenia	-0.169*** (0.037)	4.426*** (0.973)	0.024* (0.014)	0.249*** (0.034)
Bolivia	-0.090** (0.044)	8.816*** (0.978)	0.279*** (0.025)	0.448*** (0.032)
Bosnia and Herzegovina	-0.028 (0.023)	6.460*** (0.502)	0.068*** (0.019)	0.175*** (0.017)
Bulgaria	-0.159*** (0.017)	5.473*** (0.374)	0.117*** (0.011)	0.274*** (0.016)
Colombia	0.010 (0.090)	6.900*** (1.807)	0.228*** (0.046)	0.359*** (0.084)
Egypt	-0.295*** (0.009)	9.283*** (0.218)	0.306*** (0.009)	0.314*** (0.010)
Georgia	-0.125*** (0.038)	5.918*** (0.932)	0.113*** (0.030)	0.125*** (0.034)
Ghana	0.047 (0.049)	3.395*** (1.021)	0.409*** (0.040)	0.450*** (0.048)
India	-0.019*** (0.005)	7.785*** (0.122)	0.350*** (0.005)	0.301*** (0.005)
Indonesia	-0.015 (0.015)	7.442*** (0.315)	0.346*** (0.010)	0.413*** (0.014)
Iraq	-0.211*** (0.004)	6.647*** (0.162)	0.408*** (0.006)	0.342*** (0.006)
Kenya	-0.048 (0.060)	3.234*** (1.064)	0.340*** (0.028)	0.366*** (0.059)
Korea, Rep.	-0.087 (0.068)	3.144* (1.642)	0.192*** (0.060)	0.192*** (0.060)
Laos	0.079* (0.042)	5.134*** (0.890)	0.389*** (0.035)	0.029 (0.043)
Malawi	-0.094*** (0.018)	5.226*** (0.456)	0.391*** (0.016)	0.274*** (0.017)
Mexico	-0.181*** (0.004)	6.287*** (0.084)	0.221*** (0.003)	0.084*** (0.003)
Nicaragua	-0.284*** (0.017)	1.495*** (0.395)	0.090*** (0.013)	0.477*** (0.041)
Niger	-0.181*** (0.028)	6.537*** (0.773)	0.498*** (0.027)	0.254*** (0.032)

Continued

Table 3 Demographic differences of public sector—cont'd

	(1)	(2)	(3)	(4)
	Male	Age	Secondary or tertiary education	Tertiary education
Nigeria	−0.015 (0.029)	8.820*** (0.749)	0.147*** (0.027)	0.391*** (0.027)
Pakistan	−0.022*** (0.006)	7.256*** (0.196)	0.285*** (0.008)	0.162*** (0.007)
Panama	−0.153*** (0.014)	6.708*** (0.330)	0.178*** (0.010)	0.300*** (0.014)
Peru	−0.115*** (0.025)	9.617*** (0.566)	0.027*** (0.009)	0.330*** (0.022)
Serbia	−0.055*** (0.013)	−1.147*** (0.311)	0.248*** (0.010)	0.226*** (0.011)
South Africa	−0.175*** (0.005)	5.184*** (0.101)	0.041*** (0.003)	0.138*** (0.004)
Sri Lanka	−0.216*** (0.039)	2.624*** (0.823)	0.165*** (0.020)	0.326*** (0.035)
Tajikistan	−0.067** (0.026)	7.794*** (0.680)	0.145*** (0.018)	0.361*** (0.025)
Tanzania	−0.039 (0.030)	8.792*** (0.748)	0.536*** (0.028)	0.118*** (0.020)
Timor Leste	−0.056** (0.028)	2.658*** (0.746)	0.323*** (0.035)	0.044** (0.018)
Uganda	−0.020 (0.041)	9.086*** (0.934)	0.388*** (0.038)	0.360*** (0.042)
United Kingdom	−0.224*** (0.005)	3.787*** (0.137)	0.126*** (0.004)	0.253*** (0.005)
United States	−0.085*** (0.012)	4.314*** (0.302)	0.020*** (0.002)	0.159*** (0.010)
Vietnam	−0.023 (0.031)	5.589*** (0.660)	0.175*** (0.018)	0.305*** (0.030)

Standard errors in parentheses.

Each table cell contains a within-country difference of means of demographic characteristics of public sector and private sector employees. Ambiguous sectors such as NGO are omitted. Each country's data come from a single household-level survey, as listed in the [Appendix](#).

*, $p < 0.10$; **, $p < 0.05$; ***, $p < 0.01$.

motivated than private sector workers on average, this difference depends on the corruption level of the country. In countries with high levels of corruption, intrinsically motivated individuals are not more likely to join the public sector. Two recent laboratory studies are consistent with this association. [Hanna and Wang \(2014\)](#) have students from a university in India play a series of experimental games designed to measure various personality traits, such as cheating and prosocial behavior. The authors find that students

who cheated in a random dice game are more likely to express interest in a public sector job. In a corruption experiment with private sector job aspirants and aspirants of the Indian bureaucracy, [Banerjee et al. \(2015\)](#) examine embezzlement of resources in which “supervisors” evaluate the performance of “workers” and then pay them. They find that aspirant bureaucrats are more corrupt than private sector aspirants, but their likelihood of being corrupt is similar across sectors.

This tradeoff in vocational profiles has also received theoretical attention regarding how best to design personnel policy. In addition to the intrinsically motivated, governments and other mission-driven organizations value individuals of high quality as well. But if higher quality candidates demand more compensation, then higher wages may be needed to attract these individuals. But as a small theoretical literature in economics [e.g., [Delfgaauw and Dur \(2007\)](#), [Francois \(2000\)](#), and [Prendergast \(2007\)](#)] has pointed out, offering higher wages may come at the cost of attracting individuals who are more corruptible or care less about the mission. Whether this tradeoff exists is ultimately an empirical question, which has recently led several scholars to explore the extent to which financial incentives can affect a government’s ability to recruit publicly motivated and high-quality individuals. In the following section, we review a nascent experimental literature on how certain job attributes, including compensation, affect two aspects of the recruitment process into the public sector: who applies for the job and who accepts the job. We then conclude with a short discussion of the empirical evidence on whether intrinsic motivation does in fact lead to higher job performance in the public sector.

3.1 Financial incentives

3.1.1 *Effects of financial incentives on the applicant pool*

If higher quality candidates, as priced by the market, demand higher compensation, then higher wages in the public sector are necessary to attract those candidates. Does offering higher wages come with the cost of attracting candidates with weaker public service motivation? This question has motivated three recent experimental studies. [Dal Bo et al. \(2013\)](#) implemented a field experiment as part of an official program of Mexico’s federal government called the Regional Development Program (RDP). The program, which sought to enhance state presence in 167 of Mexico’s most marginalized municipalities, conducted a recruitment drive to hire 350 community development agents who were tasked with the responsibility of identifying areas where public good provision is deficient and working with existing public programs and local authorities to remedy such deficiencies.

A unique feature of this recruitment drive was the exogenous assignment of wage offers across recruitment sites. Two different wage offers were randomly assigned across 106 recruitment sites. In one set of recruitment sites, the program offered 5000 pesos per month, while in the other sites, the program offered a wage of 3750 pesos. Candidates who were interested in this position were then required to undertake a screening exam that was designed to measure various dimensions of quality and motivation.

Dal Bó et al. (2013) find that higher wages do help attract a higher quality candidate pool. In the places that announced a higher salary, the average applicant was smarter, had better personality traits, had higher earnings, and had a better occupational profile (e.g., more experience and white collar background). Moreover, contrary to theoretical concerns, these effects do not come at the cost of attracting less publicly motivated candidates, as measured by their performance on a public service motivation inventory.⁷

An important design feature of this study was its ability to offer two different wages for the exact same position, which was in large part due to the size of the program and its expansive geographical coverage. In other settings, where offering different wages has not been feasible, researchers have had to adopt alternative, creative approaches to addressing this question. One such example is a study by Ashraf et al. (2015). In this study, the authors partnered with the Government of Zambia to hire approximately 330 community health care workers. Instead of offering different wages, the authors introduced experimental variation in the how the position was advertised. In 24 of the 48 districts, potential candidates saw a job advertisement that highlighted the job's promotion prospects and the opportunity for career advancement. In the other districts, applicants saw a poster that emphasized the social importance of the job.

This recruitment process led to over 2400 applicants. In the districts where the job ads stressed career incentives, applicants were much more qualified as measured by their high-school test scores and past performance in their natural science courses. These applicants also displayed a high degree of prosocial motivation, with levels that were similar to the applicants that applied under the social incentive treatment. While the applicants who applied under the career incentives treatment did place a higher weight on career benefits, the authors conclude that making career versus social incentives salient did not induce a tradeoff between a higher quality applicant pool and a prosocially motivated one.

In contrast to these two studies, Deserranno (2015) finds that financial incentives can lead to a less socially motivated applicant pool. Her field experiment was conducted in rural villages of Uganda in collaboration with the NGO, BRAC. The recruitment drive was for health promoters, which was a position that did not previously exist and whose remuneration was uncertain since it depended on the sales of health products. The experiment exploits these two features of the position to introduce variation in how the financial aspects of the job were advertised. In one treatment arm, the job advertisement

⁷ Perry's Public Service Motivation Index (Perry, 1996) is the most commonly used measure of intrinsic motivation in the public sector. This index is constructed based on a questionnaire in which the subject must express agreement or disagreement with each of 32 statements. The questionnaire elicits opinions on the attractiveness of politics, public service, and prosocial activities. The questionnaire is subdivided into six modules labeled "Attraction to Policy Making," "Commitment to Public Interest," "Social Justice," "Civic Duty," "Compassion," and "Self-Sacrifice". Each dimension is an average of responses to several statements that are measured on a Five-point Likert scale, where a five represents strong agreement with the statement, and a one denotes strong disagreement.

mentioned the minimum amount that health promoter was expected to earn (low-pay treatment) and, in another treatment arm, advertised the maximum amount a health promoter was expected to earn (high-pay treatment). A third treatment advertised the mean of the expected earnings distribution (medium-pay treatment).

The study finds that while the high-pay treatment attracted 30 percent more applicants relative to the low-pay treatment, the applicants had much less experience as health volunteers and were much more likely to state “earning money” as the most important feature of the job. Applicants under the medium- and high-pay treatments were also 24 percentage points less likely to make a donation to a public health NGO in the context of a dictator game. Although the author finds large effects on these various measures of intrinsic motivation, she does not find treatment effects on candidate quality, as measured by the applicant’s education and income.

In sum, the conclusions from a scant experimental literature are mixed, but this is not at all surprising. Putting aside differences in the actual treatments, as [Dal Bo et al. \(2013\)](#) point out in their model, whether financial incentives crowd out the intrinsically motivated will depend on how these personal traits (e.g., intrinsic motivation versus quality) are correlated within the broader population—correlations which the literature has yet to document in a systematic way. These three studies, which were conducted in very different contexts, represent a step forward in the literature, but without more information on how personality traits vary across broader populations, it is difficult to make general conclusions about the exact tradeoffs that financial incentives induce.

3.1.2 Effects of financial incentives on recruitment

The power of wages is not limited to attracting a larger and better applicant pool. Higher wages also increase an organization’s ability to fill vacancies. In the [Dal Bo et al. \(2013\)](#) study, the authors found that the Mexican government was 35.2 percent more likely to fill the vacancy when offering the higher wage, which corresponds to a short-run labor supply elasticity of 2.15. This elasticity is similar to other quasi-experimental estimates found in the literature (e.g., [Manning, 2011](#)), as well as one reported in [Deserranno \(2015\)](#). Even though the study by [Deserranno \(2015\)](#) manipulates earning expectations (as opposed to actual earnings), the author finds an experimental elasticity of 1.8 when comparing takeup in the low-pay treatment group to takeup in the high-pay treatment group.

Part of the reason why higher wages lead to higher recruitment rates is because they help to compensate for aspects of the job that a candidate dislikes. This mechanism was on clear display in the study by [Dal Bo et al. \(2013\)](#). Although the applicants for the RDP position were all applying for the same job, the jobs were located in different municipalities throughout the country. At the time of the application, the candidates did not know where the job was located and were only told this information during the offer stage. As a result, jobs that were ex-ante quite similar became quite different ex-post depending on

where the job was located and the characteristics of the municipality. The authors show that distance to the municipality (from their current residence) and attributes such as the level of drug violence and the lack of public goods in the municipality were all important hurdles to filling the vacancies. Fortunately, however, higher wages proved to be an effective instrument in clearing these hurdles.

3.2 How should governments screen?

Wage offers affect who applies for government jobs, but government jobs are typically oversubscribed, so government—like all employers—needs to winnow down the set of applicants to those they hire. Governments vary in the way they screen their public servants. Some rely on the passage of civil service exams or attainment of university degrees, while others adopt more discretionary approaches that, while permitting them more flexibility, can also be prone to corruption and patronage. These different screening strategies have important implications for not only the quality and performance of the bureaucracy but also for the type of person who applies.

An important consideration for any organization when screening and selecting personnel is match quality. If employers and employees share a common vision and objectives, then this positive match quality increases organizational efficiency and diminishes the need for high-powered incentives. For governments, who are responsible for providing public goods that are difficult to price in the market, the ability to recruit public service—motivated individuals might be especially beneficial (Besley and Ghatak, 2005).⁸

Arguably, the recruitment of publicly motivated individuals has other benefits as well. Individuals with high levels of intrinsic motivation are less likely to shirk in an environment where incentives are low powered and/or when noncontractible elements of the service provision exist (Francois, 2000).

In support of these theoretical arguments, a large empirical literature in public administration shows that intrinsic motivation—and specifically public service motivation—is associated with higher levels of performance in government work (Perry and Hondeghem, 2008).⁹ Recently, economists have begun to contribute to this literature.

⁸ Before we can answer the question whether or not government should screen on intrinsic motivation, we must take a step back and ask whether it even measurable and quantifiable. There is a rich and growing literature in psychology and economics that suggest that personality traits including intrinsic motivation can be measured (see, e.g., Amlund et al., 2011).

⁹ In the public administration literature, recent meta-studies suggest that public-service motivation is positively correlated with job performance in the public sector, broadly defined (Petrovsky, 2009). Naff and Crum (1999) use a sample of over 8000 U.S. federal employees and find that public-sector motivation correlates with individuals' last performance evaluations. Park and Rainey (2008) analyze data from 22 federal agencies in the United States and find that public service motivation is positively correlated with self-reported measures of job productivity and quality of work. Similar results are found using government data from Switzerland (Ritz, 2009) and the Netherlands (Steijn, 2008).

For example, as part of a monitoring experiment of health clinics in the district of Punjab, [Callen et al. \(2015\)](#) examine the job performance of clinic doctors. They find that those who scored higher on the public service motivation index are much less likely to shirk and falsify health reports. In the same study discussed previously, [Deserranno \(2015\)](#) also finds, in the case of the Ugandan health promoters, that prosocial motivation is a strong predictor of job performance. Health promoters who had donated a greater share of their endowment to a local NGO visited a larger number of households, provided more prenatal checks, and organized more public presentations. [Dizon-Ross et al. \(2015\)](#) conduct a survey on nurses of antenatal care centers in Uganda, Ghana, and Kenya as part of an audit study on bednet distribution programs. They find that nurses not only exhibit high levels of prosocial motivation but that it is predictive of job performance.

Despite the mounting evidence linking public service motivation to job performance, establishing causality has proven challenging. Given the difficulties in directly generating experimental variation in a person's level of intrinsic motivation, researchers have had to rely on indirect approaches. One approach has been to introduce experimental variation in who applies for the same job. The experiment induces a selection effect, while keeping any potential incentive effect constant. Both the studies by [Ashraf et al. \(2015\)](#) and [Deserranno \(2015\)](#) provide examples of this approach. In the study by [Ashraf et al. \(2015\)](#), the authors used the two different recruitment strategies to create variation in the type of health promoters that were recruited. Once employed, all the health workers were tasked with the same responsibilities and faced the same incentives. Based on this design, they find that health workers attracted by career incentives are much more effective at delivering health services, as measured by home visits and the organization of community meetings. These health promoters were also more likely to remain in their posts over the course of 18 months. Although these results imply a negative relationship, if any, between intrinsic motivation and performance, it is worth noting that the level of prosocialness among the health promoters who were recruited in the career incentives treatment was also quite high.

The study by [Deserranno \(2015\)](#) provides stronger evidence in support of the relationship between prosocialness and job performance. Among the agents recruited under the low-pay treatment who were measured to be more prosocial, she finds higher aggregate performance in the first year of work. Compared to the high-pay treatment, the health promoters recruited under the low-pay treatment visited a larger number of households, organized more public presentations in the village, and provided more natal checks. She also finds that they were more likely to target the most vulnerable households.

While these studies can credibly identify the effects of "selection" on job performance, what this selection effect comprises is not entirely clear. Prosocial motivation is frequently found to be correlated with various other personal traits, including the Big Five. Short of randomly assigning individuals based on a specific attribute, it is difficult to separate the effects of intrinsic motivation from other positive personality traits.

Another experimental approach that studies have explored has been to test whether a particular intervention is more effective among individuals with high levels of intrinsic motivation. For example, the goal of the experiment in the study by Callen et al. (2015) was to reduce high levels of absenteeism among clinic doctors and staff in Punjab.¹⁰ One source of this absenteeism was the fact that these clinics were rarely inspected, and when inspections did occur, doctors and inspectors would collude and falsify the report. In collaboration with senior health officials of the Department of Health, the authors introduced a new monitoring program in 18 of the 35 districts constituting their experimental sample. In the treatment districts, the traditional paper-based monitoring system for clinic utilization and worker absence was replaced with a smartphone application. The new system allowed health system inspectors to upload the results of their assigned visit to a central dashboard which instantly updated reports at different levels of aggregation. The data, which included geo-tagged, time-stamped facility staff photos, made it difficult for the inspector to falsify his report. While the study finds that the monitoring technology did increase the number of inspections, there was significant heterogeneity in the treatment effects by the personality type of the inspector. Higher quality inspectors responded much more positively to the treatment. A Big Five index one standard deviation higher, for example, is associated with a differential 35 percentage point treatment effect in terms of health inspections.

In another example, Bellé (2012) conducted two experiments with nurses at a large public hospital in Italy. He was interested in understanding whether public service motivation interacts with two interventions that social psychologists have found to be effective at stimulating job performance: (1) beneficiary contact (Grant et al., 2007) and (2) self-persuasion interventions (Aronson, 1999). The first intervention is based on the idea that contact with customers, clients, and beneficiaries outside of the organization can help to motivate employees to perform more effectively. The second experiment is based on the premise that employees are most likely to be influenced by credible and trustworthy sources, and one such source is the person himself. Based on this theory, researchers have shown that employees find public service more important after they were asked to reflect on the importance of public service and then made to publicly advocate for it both in writing and in person.

The study by Bellé (2012) was based on a sample of 90 nurses, randomly assigned across the two treatment interventions and a control. The nurses were tasked with assembling surgical kits that were being shipped to a former war zone that was facing a humanitarian emergency. Based on this task, the study examined three principal performance measures: (1) the number of minutes each participant contributed to task, (2) the number

¹⁰ Based on unannounced visits to clinics at baseline, they found that only 56 percent of clinics had been inspected in the prior 2 months and that 32 percent of clinics had no doctor present.

of surgical kits that each participant assembled during her shift, and (3) the average number of surgical kits each participated completed per minute. The author finds that the effects of both interventions were stronger for employees who had a higher prosocial motivation level at baseline.

While these studies establish that the characteristics of individuals are an important determinant of performance and they suggest how governments could change the applicant pool, they do not necessarily tell us how governments should screen among the candidates who apply to further improve selection. [Hanna and Wang \(2014\)](#), for example, show that current Indian civil service—type screening exams would not eliminate the negative selection on dishonesty they find in their setting. Given the large public sector premia we observe for low-income countries, an important friction in improving the human capabilities of the state may lie in its screening technologies. To understand how to most effectively screen seems an important direction for future research.

4. USING INCENTIVES TO IMPROVE PERFORMANCE

Once selected, it may be possible to use incentives to further improve workers' performance. Public sectors careers, however, typically feature a relatively flat incentive structure. As discussed previously, in the standard civil service model, adopted almost universally since the early 20th century to limit politician discretion over appointments and salaries, the public sector is staffed by salaried civil servants whose salaries are based on rigid and formulaic pay scales. These pay scales feature compressed wages relative to that in the private sector [e.g., [Borjas \(2015\)](#) for the United States]. The combination of formulaic pay systems and wage compression limits the degree to which financial incentives can be used to reward the performance of public servants: the formulas are largely based on seniority and position, allowing little room for discretion, and the wage compression is such that even promotions within the civil service are less of an incentive than in the private sector.

While this type of salary structure may be appropriate for governments in many contexts, it may not be effective in developing countries where government officials are often thought to have poor job performance. Therefore, scholars have begun to examine the costs and benefits of providing additional incentives to government workers in developing countries. In this section, we review recent experimental evidence that seeks to shed light on these issues in a variety of sectors. We begin first with evidence on using financial incentives to reward good performance. The evidence not only sheds light on the degree to which such incentives can improve performance but also highlights the challenges with using such incentives in practice, particularly those that arise when using them in the public sector.

Given the constraints in financial payments imposed by civil service systems, we then go on to consider nonfinancial incentives, which are prevalent in government contexts.

We examine one type, in particular, that is quite common in government practice: using transfers to more or less desirable postings as an incentive device. We then examine other types of nonfinancial incentives.

4.1 Financial incentives

Government officials do many types of jobs, and some are easier to incentivize than others. In some cases, what we refer to as “agents of government authority” are those who are tasked with ensuring citizens comply with government laws and regulations. For such officials—such as police, judges, prosecutors, tax inspectors, building inspectors, and so on—there is a natural tension between what the government would like the agent to do (e.g., to make people pay taxes that are due under the law) and what the targets of government enforcement would like the agent to do (e.g., to allow them to avoid paying taxes). This tension invites opportunities for corruption between the agent and the citizen (e.g., reducing taxes in exchange for a bribe) and, as we will see, complicates the incentive problem for the government. In other cases, such as frontline government service providers, the government and the citizen’s incentives are aligned: both would like the agent (e.g., the teacher) to provide more or better services. Providing incentives may therefore be more straightforward in the second case. We consider incentives in both contexts in turn.

4.1.1 *Incentives for agents of government authority*

4.1.1.1 Incentives for tax collection

Several recent experiments explore the risks and rewards of financial incentives in the public service. [Khan et al. \(2014\)](#) conduct a field experiment in urban Punjab, Pakistan, to study performance pay for tax inspectors. The experiment involved high-powered financial incentives for property tax inspectors who are in charge of assessing properties, collecting property taxes, and levying sanctions on those who fail to pay. The basic treatment gave the team of tax staff in an area, which consisted of three people, an incentive payment equal to an average of 30 percent of tax revenues collected above a historically predicted benchmark, enough to double their baseline wages. Tax inspectors are exactly the sort of government worker where one might be concerned ex-ante about the efficacy of incentives; while there is substantial scope for improvement through either increased effort or reduced corruption, incentives also have the potential to increase bribes by raising the bargaining power of tax inspectors (who now must be paid a higher bribe to compensate them for their foregone incentive payment) or to lead to overtaxation as was thought to be the case historically.

[Khan et al. \(2014\)](#) find evidence for both the positive and negative aspects of incentives. On the plus side, the incentives raised revenue substantially. On average, treated areas had revenue growth that was 9.3 log points greater than control, which translates to a 46-percent higher growth rate in revenue. Incentive schemes that only rewarded

on revenue did best, increasing revenue growth by 12.8 log points (62-percent higher growth), whereas those that attempted to control multitasking problems through incentive schemes that also rewarded taxpayer satisfaction and accuracy of tax assessments had less impact on revenue yet did not improve these dimensions. The revenue gains substantially exceeded the costs of the incentives. The incentives did not appear to reduce taxpayer satisfaction, in part, because the increased tax collection was concentrated among a small number of taxpayers. On the negative side, however, they find evidence that bribe rates did increase in incentive areas, potentially to compensate incentivized tax inspectors for foregone incentive payments.¹¹

4.1.1.2 Incentives for policing and justice

The study by [Khan et al. \(2014\)](#) is relatively rare in focusing on the tax sector, where the government has a potentially adversarial role against the taxpayer, which leads to opportunities for collusion and where incentives can have perverse effects, such as the increase in bribes they document. The police force is another area with agency problems, where one might be concerned that financial incentives (e.g., on the number of citations issued, arrests made, or the like) could lead to overzealous or inaccurate enforcement or simply a reallocation of resources from nonincentivized to incentivized tasks. [Baicker and Jacobson \(2007\)](#), for example, document that when police agencies in the United States are allowed to keep the revenue they obtain from assets they seize in drug arrests, they increase drug arrests but do so by reducing enforcement of other petty crimes, suggesting that multitasking is an important issue. A commonly voiced concern about these laws that their paper does not address is whether these type of laws lead to unjustified seizures and abuses ([Miller and Selva, 1994](#)). Other similar areas where financial incentives have been tried, but not rigorously evaluated, include incentives for prosecutors in New York City to ensure speedy disposition of cases ([Church and Heumann, 1989](#)), as well as historical examples from the United States where prosecutors were paid incentives based on conviction rates ([Meares, 1995](#)). Exploring the impact of incentives in these areas in a more rigorous and careful way, and seeing whether they are effective in developing countries with more corrupt and generally less-effective police forces, seems an important area for future work.

4.1.2 Incentives for frontline service providers

A more common area of focus has been incentives for health and education service providers. These have taken two broad forms, incentives on outcomes (e.g., test scores, immunizations given) and incentives on inputs (e.g., provider attendance).

¹¹ Other nonexperimental studies of tax agencies also find increases in revenue (e.g., [Kahn et al., 2001](#); [Burgess et al., 2010](#)) but are unable to examine the potential downside in terms of over-enforcement or bribery.

4.1.2.1 Incentives on outcomes: test scores

One commonly considered type of outcome-based incentive is teacher incentives based on student test scores.¹² Muralidharan and Sundararaman (2011) report the results of a large-scale teacher incentive program run by the Indian state of Andhra Pradesh as a school-level randomized trial. Public school teachers were paid incentives based on test scores, with both group and individual incentives considered. Incentives were substantially smaller than in the study by Khan et al. (2014) discussed previously, as they were calibrated to be around three percent of a typical teacher's annual salary. They find that the incentives were effective in promoting learning: after 2 years, students in incentivized schools had test scores that were 0.27 standard deviations higher in math and 0.17 standard deviations higher in language. They find no evidence of multitasking; in fact, students also do better in nonincentivized subjects, such as science and social studies. Incentives appear to have worked by increasing effort conditional on attendance not by increasing teacher attendance. The individual incentives outperformed the group incentives by the end of the second year. Muralidharan (2012) reports that the effects increase even more with time: after 5 years, students in treatment schools had test scores 0.54 standard deviations higher in math and 0.35 standard deviations higher in language and still had higher test scores in nonincentivized subjects.

On the other hand, Glewwe et al. (2010) find somewhat less encouraging results. They conduct a randomized trial of teacher incentives in Kenya, where an NGO provided in-kind prizes to teachers in Kenyan government schools on the basis of school-level performance on district exams, where those who did not take the exam were imputed a low score. They find that incentivized schools had more people taking the government exam and higher scores on the government exam used for the incentives. However, unlike the example by Muralidharan and Sundararaman (2011) that found positive spillovers to nonincentivized subjects, they find no evidence of higher scores on an independent exam administered by the NGO that was not linked to performance incentives. The authors conclude that multitasking was a real issue in their context and that teachers may have emphasized test-taking skills, as opposed to general instruction, in response to the incentives. In both cases, incentives improved targeted indicators, but understanding why there were positive spillovers to nonincentivized contexts in India but not in Kenya seem an important area for future research.

4.1.2.2 Incentives on outcomes: health

While incentives based directly on health outcomes are rare, one notable example is the study by Miller et al. (2012). Miller et al. (2012) conducted a randomized trial in 72 Chinese primary schools in which school principals received performance payments based on

¹² In addition to the experimental studies reviewed here, there is also a large literature on teacher performance pay in the United States. See Neal (2011) for a review.

reduction in anemia among their students. Specifically, principals were paid 150 RMB per student who changed from anemic to nonanemic over the course of the intervention. This implied a payment of roughly 2 months' salary for reducing anemia by half. Comparison groups were given the same information and subsidies as the incentive treatment, but no direct financial incentives. They find that the incentives reduced anemia compared with the pure control group by about 5 percentage points (23 percent). The nonincentivized comparison groups did not achieve statistically detectable reductions in anemia, suggesting a role for incentives, but confidence intervals are such that they cannot statistically distinguish between the incentive group and the nonincentive information and subsidy groups. They report that incentivized school principals were more likely to use subsidies for iron-focused supplements, whereas nonincentivized school principals used subsidies for supplements that could affect both iron and overall calorie intake. Depending on one's perspective, this could be considered a multitasking issue as well to the extent that one is interested in both types of supplementation.

4.1.2.3 Incentives on service delivery

In health, incentives for providers have tended to be focused instead on measures of service delivery, such as the number of immunizations given. These service delivery metrics can be thought of as somewhere between ultimate outcomes (e.g., learning, lack of disease) and provider inputs (e.g., attendance). One reason for focusing on this level is that for ultimate health outcomes, the signal to noise ratio may be high; that is, in a given context, most of the variance in health outcomes is idiosyncratic rather than due to provider effort. If one believes there is a clear mapping from health service delivery to health, these types of incentives may make sense.

Basinga et al. (2011) and Gertler and Vermeersch (2012) examine this approach. The intervention they study in their experiment took place in Rwanda and provided incentives to primary care facilities, which were in turn used to compensate facility personnel. The incentives were based on the quantity of visits to the facility for various services (e.g., childbirth in the facility, prenatal care) and the content of services provided in those visits (e.g., pregnant women receiving tetanus vaccines and malaria prophylaxis during prenatal care, immunizations given during postnatal care, etc.), weighted by an overall quality index of the facility. The incentive payments were substantial, equal to a 38-percent increase in total compensation for facility personnel. They found that the incentives led to a substantial increase in prenatal and postnatal services, which translated into increased health: infant weight-for-age increased by 0.53 standard deviations and height-for-age for children aged two to five by 0.25 standard deviations, with increased breastfeeding and reductions in infant illness hypothesized to be important channels.

Olken et al. (2014) report the results of a large-scale field experiment in Indonesia in which villages were provided with incentive payments based on health service delivery (similar to that in Rwanda), school enrollment, and education. Specifically, villages

received a block grant each year that they could use for any purpose related to health or education. In incentivized areas, 20 percent of the total amount set aside for block grants in a subdistrict was allocated to villages based on their performance on the targeted health and education indicators; in nonincentivized areas, the block grant was allocated based only on population. The incentive was to the community as a whole and, unlike the previous examples, was generally not passed on to service providers but was instead used for programming (e.g., nutritional supplements, subsidies for childbirth, etc.). Comparing the incentivized to nonincentivized areas, they found the incentivized areas performed better on the targeted health indicators. On average, the eight targeted health indicators were about 0.04 standard deviations higher in the incentivized than nonincentivized areas. These effects were about twice as large in areas with low initial levels of performance, but the relative gain of incentivized to nonincentivized areas declined over time as nonincentivized areas improved. The main health reduction was a 15-percent (2.6 percentage points) decline in malnutrition rates, though again this effect became more muted over time. There were no detectable differences between incentivized and nonincentivized areas on educational outcomes. The program suggests that the incentives sped up improvements on the targeted health outcomes, with no detectable multitasking effects.

4.1.2.4 Incentives on inputs: provider attendance

The final category of financial incentives that we consider is incentives based on attendance. Given the problems with provider attendance highlighted by [Chaudhury et al. \(2006\)](#), this is clearly an important issue—but key questions are whether attendance responds to financial incentives and, if so, if it translates into ultimate outcomes. Here the evidence is mixed. In an experimental study for teacher attendance at single-teacher schools run by NGOs in India, [Duflo et al. \(2012\)](#) provided linear incentives based on the number of days (above 10 per month) that teachers could submit time-stamped photos of themselves with students to prove they had been attending. They found that the incentives not only increased attendance but also led to increased learning, with students in schools where teachers were incentivized to attend having test scores about 0.17 standard deviations higher than in control schools after 1 year.

On the other hand, a similar study by [Banerjee et al. \(2008\)](#) of financial incentives for nurses' attendance in India provides a more cautionary note. In that experiment, the incentives for attendance were broadly similar to those in [Duflo et al. \(2012\)](#): nurses who were recorded absent more than 50 percent of the days in a month would have their pay reduced by the number of days they were recorded absent, and nurses who were absent more than 50 percent of the days in two consecutive months would be suspended from government service. Nurses used a protected time/date stamp machine to verify attendance. In their study, while there was initially a substantial treatment effect, the effect diminished over time and was zero at the end of their study. Although they do not

have the data to confirm this, anecdotal evidence suggests the decline was due to nurses learning how to exploit loopholes in the systems and recording more exempt absences over time. One possible difference is that the [Duflo et al. \(2012\)](#) schools were run by an NGO, which may have had more independence in enforcing the incentives than the government. A key question then is understanding which of these effects is more likely to generalize: the positive long-run effects found in the study by [Duflo et al. \(2012\)](#) or the rapid decline in effectiveness found in the study by [Banerjee et al. \(2008\)](#).

4.2 Nonfinancial incentives

While the majority of work has focused on financial incentives, nonpecuniary incentives are potentially important. While civil service regimes typically place many restrictions on hiring and firing, they have much more flexibility in assigning bureaucrats to postings within the civil service and these postings can be used as reward and punishment devices. Many bureaucracies informally recognize high achievers (i.e., “employee of the month” type awards). Public sector jobs, in particular, may seek to take advantage of the fact that their employees may be public spirited and use this as a way of creating rewards. While much less extensively studied than pecuniary incentives, several studies suggest that it is a promising direction for further exploration.

4.2.1 *Transfers and postings*

Civil service regimes typically feature much more flexibility in where people are posted than in whether people are fired or how much they are paid. This is perhaps natural, in that there are a wide variety of positions that need to be filled and these positions are heterogeneous in many dimensions, both in terms of the skills needed to complete the job effectively and their desirability as a place to work. In both cases, there can be a mix of common and idiosyncratic rankings. For example, as we discussed in the case of Mexico RDP program, civil servants were much more willing to work in a safe community, as opposed to one with high incidences of drug violence; this would be a common preference. The civil servants also preferred to work near their place of residence; since people are from different places, this creates idiosyncratic preferences. The same can be said for job attributes: a common attribute would be the need to put the cleverest tax inspectors on the most complicated corporate tax cases; an idiosyncratic attribute would be the need to match police with areas where they have social connections they can use to gather information. To the extent that there are common components to preferences, this creates scarcity for posting in the most desirable locations, and such plum postings can be used as an incentive device.

One problem with transfers as an incentive device is that politics often gets in the way. [Iyer and Mani \(2012\)](#) examined a comprehensive data set that tracks the careers of elite Indian Administrative Service personnel. They show that transfers are likely right after a new Chief Minister is elected, particularly for those bureaucrats who were not at the very

top of their initial class in terms of performance. However, even though ability is predictive of future success, caste affinity to the politician also plays an important role. Bureaucrats who share the same caste as the chief minister's party are just as likely to be assigned to important posts as the high-ability bureaucrats. These results, while not definitive, suggest that while transfers are quite common, they are not entirely based on performance, which may dampen their usefulness as a performance tool.

Banerjee et al. (2014) explore several aspects of these issues in the context of the police force in Rajasthan, India, a context in which transfers are frequent: one-third of all policemen were transferred during a typical 18-month period. As in the elite civil servants examined by Iyer and Mani (2012), anecdotal evidence suggests that police transfers are frequently imposed by politicians, often for reasons that may reduce their use as incentive device (e.g., for partisan or corrupt motives). They explore a treatment where all transfers were frozen during a 2-year period, except for well-documented cases of police misconduct. The idea was to remove arbitrary transfers and leave only transfers being used as an incentive device. They find that this freeze had no effect on outcomes such as whether decoy surveyors were treated differently or community satisfaction. One potential reason for the lack of effect is that, if the exceptions were sufficiently difficult to implement, the freeze could have eliminated both transfers used as incentives as well as politically motivated transfers. The elimination of the transfer-as-incentive could have then offset the positive effect of removing political transfers.

Consistent with this idea, a second treatment suggests that transfers have the potential to be used as an incentive device. In a second experiment, Banerjee et al. (2014) examined an anti-drunk-driving reform, where police were supposed to conduct sobriety effects. Two groups of police ran sobriety checks. In the first group, they worked with police in the central reserve "police lines" group, who are outside of typical station assignments and was given the incentive that they would be transferred back to the regular police unit if they performed well. The second group consisted of police from normal stations, who had no transfer incentives. They randomized which units were sent to which areas and found that those in the first group performed better in terms of whether the roadblock actually occurred, the number of people stopped, and so on. Of course, the composition of personnel was different in the two groups, so one cannot know if it is the transfer incentive per se that is driving the results or some other factor (e.g., maybe the police line teams had nothing else to do with their time, while the regular police teams were juggling many other tasks), but the results are suggestive that this could be important.

4.2.2 Intrinsic motivation

Beyond explicit incentives, it may be possible to use other types of intrinsic rewards as a motivational tool. One experimental study that examines this idea is the study by Ashraf

et al. (2014). In their study, public health extension workers who are tasked with selling condoms are randomly assigned to either different financial rewards (margins of 10 percent or 90 percent on each condom sale) or a nonfinancial reward that gives agents additional stars for each sale on a thermometer-type display. They find remarkable evidence of the effectiveness of nonfinancial rewards: the thermometer treatment agents sell twice as much as those in the financial rewards treatment.

Another recent, nonrandomized study also suggests that dimensions other than incentives may be important to job performance. *Rasul and Rogger* (2015) use a survey to measure management practices in the Nigerian bureaucracy and find that autonomy is positively correlated with job performance, whereas performance incentives are negatively correlated. Of course, there could be endogeneity problems: one might choose to give performance incentives to those bureaucrats who behave badly and to reward high performers with autonomy.

4.3 Summary

Several themes emerge from the evidence reviewed in this section. First, there is robust evidence that financial incentives matter: across a wide variety of settings, financial incentives in a government context seem to increase performance on the incentivized dimension. This is not particularly surprising. In fact, given these robust results, the question becomes why do governments not use financial incentives more often? Part of the reason can be attributed to the simple fact that unlike in the private sector where firms can contract on profits, performance in the public sector can be hard to measure. Also, as we discussed, there is some evidence that multitasking issues can be a problem—yet we do not have a clear understanding of when multitasking issues will or will not be present. Finally, as *Benabou and Tirole* (2006) have highlighted, financial incentives may even reduce effort among the prosocially motivated. While a number of studies have documented such adverse effects in such activities as volunteer work (e.g., *Gneezy and Rustichini*, 2000) or blood donation (e.g., *Mellstrom and Johannesson*, 2008), it remains to be seen whether this is a first-order issue in the context of governments. Developing a clearer understanding of when these issues will be present seems an important direction for future research.

Finally, research on financial incentives has primarily focused on front-line service providers, where the agent's principal (e.g., bureaucrats at the central government ministry) and the citizens served by the front-line agent have aligned incentives. With just a few exceptions, there has been much less work on the more complex case where these interests are unaligned, such as tax, police, procurement, and so on. Understanding the degree to which incentives can be effective in this context without further empowering these officials to collect more bribes or overenforce the law seems an important area for future work.

5. MONITORING MECHANISMS AND PUBLIC SERVICE DELIVERY

5.1 Overview

Incentives focus on tying rewards—typically financial rewards—to easily observable and verifiable measures of performance. Taxes is a canonical example: the state easily observes the amount of taxes collected by each tax inspector and so can base rewards on that. In many cases, however, monitoring performance itself requires costly effort on the part of either state or nonstate actors. We now turn to whether improved monitoring can improve the performance of civil servants.

Increased monitoring can improve program performance via multiple channels. First, in cases where outcomes are not observed without some effort, increased monitoring can allow managers to directly enforce punishments and rewards based on program outcomes (e.g., firing or transferring poor performers). Second, monitoring can play an important deterrence role. Third, access to monitoring results can empower citizens to demand and obtain better services by threatening to report on or vote out poor performers.

However, there are also reasons to believe that information alone may not suffice. First, in situations where state capacity is weak, managers' or regulators' ability to impose punishments is limited, and improving information flows may do relatively little by itself. Second, those in charge of collecting information or monitoring based on available information may themselves be susceptible to corruption and misuse this information. One may worry that allowing discretion to managers in collecting and using information may have the perverse effect of increasing rather than reducing program leakage. Thus, a key dimension of heterogeneity surrounding the role of information will be the extent to which those who receive the information—be they supervisors or workers—have the incentives and ability to act on it.

5.2 Information flows and monitoring

Information on project and intermediaries' performance arises in multiple ways. The classic method remains via government auditing and inspection units that are required to monitor government programs. More recently, the rise of e-governance has meant that government agencies have access to large administrative data sets on funds flow, intermediaries' behavior (typically attendance), and monitored program outcomes. These data directly allow managers to obtain better real-time data on program performance and, in many cases, public availability of these data (aided in part by the rise of freedom of information acts) increases citizen monitoring.

In the following section, we first discuss findings from experiments that evaluate government monitoring processes and then we turn to citizen monitoring.

5.3 Government monitoring

5.3.1 *Does more information on performance improve outcomes?*

Studies of audits. Several studies examine the role of government audits. [Olken \(2007\)](#) conducted a field experiment in Indonesia where a local village body implemented a

road construction program, and audits were conducted by the government agency. The key finding is that audits have a significant deterrence impact. Before villages began building road projects, some were randomly selected for a high-audit intensity group, where they faced an audit by the government agency with 100-percent probability as opposed to a 4-percent probability in the control group. [Olken \(2007\)](#) found substantial effects of the government audits, reducing unaccounted for expenditures by about 8-percent points or about 30 percent from the baseline level.

[Ferraz and Finan \(2008\)](#) examined audits of municipal accounts in Brazil where small municipalities were randomly chosen to be audited by government auditors. They examine the impact of the timing of auditing on the probability that the mayor is reelected. They find that, conditioning on the actual number of corruption violations found by the auditors, those audited before the election were less likely to be reelected than those who were audited after the election for those with an intermediate number of violations.

An open question is whether a higher likelihood of punishment also had a deterrence effect in this setting. [Bobonis et al. \(2015\)](#) examine this question in the context of Puerto Rico. Puerto Rico has established an independent body that systematically conducts municipal government audits, the findings of which are made publicly available and disseminated to media sources. [Bobonis et al. \(2015\)](#) exploit two features of the audit process. First, municipalities are audited in a preestablished order, making the timing of audits and their dissemination predetermined. Second, audits are “timely audits,” such that reports released in the period leading up to an election are more likely to inform on the incumbent mayor’s activities than those reports published shortly after an election due to a high independent turnover rate of politicians. They find that timely audits induce a significant short-term reduction in municipal corruption levels of approximately 67 percent, as well as an increase in incumbent mayors’ electoral accountability. However, in contrast to these desirable short-run consequences of the audits, municipal corruption levels in the subsequent round of audits are, on average, the same in municipalities audited preceding the previous election and those whose audits became publicly available afterward. They also find that incumbent reelection rates in the subsequent election are significantly higher in municipalities in which there was an earlier timely audit. The presence of selection effects in future reelection rates, but not in corruption, is *prima facie* evidence in favor of the view that the information contained in the audits helps voters select competent but opportunistic politicians, rather than honest or virtuous ones.

There is, however, the potential for monitoring to backfire. [Lichand et al. \(2015\)](#) studies the introduction of the municipal audits as in [Ferraz and Finan \(2008\)](#) in a differences-in-differences framework. They show that in municipalities that expected they might be audited, procurement went down, with negative consequences for health. To the extent that the problems we observe are due to incompetence or laziness rather than corruption, as suggested by [Bandiera et al. \(2009\)](#), too much of a focus on corruption could backfire. We regard continued explorations of this issue as an important area for future work.

Biometric/time stamp studies. As e-governance becomes more widespread, countries have increased their investment in, and use of, e-monitoring systems. Several studies described in our discussion of incentives based on provider attendance (Section 4) utilize such systems. [Duflo et al. \(2012\)](#) based incentives on information obtained via time-stamped photographs. However, as the study by [Banerjee et al. \(2008\)](#) highlights, the robustness of such monitoring systems is sensitive to how tamper proof the monitoring mechanism is. If nurses can destroy the monitoring system (here, by literally breaking the time stamping mechanism), then they will do so. However, more importantly, if the incentive system allows for loopholes, then improved information may do little—a general lesson that exists above and beyond the nature of monitoring mechanism.

Two recent studies that expand our understanding of the issues at stake are [Dhaliwal and Hanna \(2014\)](#) and [Callen et al. \(2015\)](#). [Dhaliwal and Hanna \(2014\)](#) studied the rollout of biometric monitoring of the staff at primary health centers in South India. Health worker attendance increased by 14.7 percent in clinics with improved monitoring and was driven by lower-level staff in these centers (nurses and pharmacists), rather than by doctors. Improved monitoring had important health impacts: there was a 16-percent increase in the delivery of infants by a doctor and a 26-percent reduction in the likelihood that infants are born under 2500 grams. However, they also find lower staff satisfaction and widespread attempts by the staff to circumvent the system. Taken together, the results show both how improved monitoring can improve service delivery and also the rents at stake in the system. To the extent that staff dissatisfaction and delays in implementation tend to be more visible, this study also points to the importance of measuring impacts. In the absence of careful measurement of health outcomes, it would have been easy to focus on the partial implementation and to deem the system a failure.

If poor monitoring opens the door for shirking and potentially outright corruption by service providers, then we may expect these impacts to vary by the personality of the service provider. As discussed previously, [Callen et al. \(2015\)](#) report on a monitoring experiment where the traditional paper-based monitoring system for clinic utilization, resource availability, and worker absence was replaced by an Android smartphone application. In the new system, data generated by health inspections are transmitted to a central database using a mobile data connection in real time. Data are then aggregated and summary statistics, charts, and graphs are presented in a format designed in collaboration with senior health officials to effectively communicate information about health facility performance. Especially relevant given the study by [Dhaliwal and Hanna \(2014\)](#), these authors find that senior health inspectors who score higher on the Big Five personality inventory are more likely to respond to a report of an underperforming facility by compelling better subsequent staff attendance. More surprisingly, they also find that inspectors who score higher on personality tests are more likely to reduce absenteeism when dashboards are implemented. This paper provides one way into understanding the observed heterogeneity in responsiveness to better monitoring: individual personality characteristics.

5.3.2 *Who collects information and does that matter?*

A common concern in the literature is the veracity of information collected by monitors. For instance, both [Banerjee et al. \(2008\)](#) and [Dhaliwal and Hanna \(2014\)](#) report on how service providers seek to reduce the functionality of monitoring devices.

Incentives to provide poor-quality information may also arise if inspectors and auditors are corruptible by those they are intended to monitor. This possibility is, arguably, particularly stark in the case of private-sector auditors paid by the firms or institutions they audit. [Duflo et al. \(2013\)](#) examine the implications of corrupted information flows for regulatory efficacy. In a large field experiment conducted with the environmental regulator in Gujarat, they altered the assignment and payment mechanism for third-party environmental auditors of industrial plants. Under the status quo, the auditors were hired and paid for by the plant they audited. In the treatment group, auditors were instead randomly assigned to plants and paid a fixed salary from a central pool of funds. The experiment demonstrated that the status quo system was largely corrupted, with auditors systematically reporting plant emissions just below the standard, although true emissions were typically higher. Second, the treatment caused auditors to report more truthfully and significantly lowered the fraction of plants that were falsely reported as compliant with pollution standards. Third, treatment plants, in turn, reduced their pollution emissions.

A different margin of potential corruption in information acquisition is providing higher-level officials discretion in whom to inspect or audit and when. Another is the choice of intermediaries who have monitoring responsibilities.

[Duflo et al. \(2015\)](#) examine this issue in the context of environmental inspections in India. They examine whether raising the frequency of inspections in a rule-bound manner changes regulator behavior and improves plant compliance. They find that more inspections lead the regulator to send more warnings but not to increase incidence of punishments. They use detailed information on regulator plant interactions to show that the regulator uses his discretion to target information collection and punishment efforts at a smaller subset of highly polluting plants. In this case, regulatory discretion is valuable as it allows the regulator to best target his scarce inspection resources.

Finally, a different way to improving monitoring is by increasing private-sector incentives to report the truth. A commonly cited example is value added tax (VAT), which generates paper trails on transactions between firms. [Pomeranz \(2015\)](#) analyzes the role of third-party information for VAT enforcement through two randomized experiments among over 400,000 Chilean firms. She shows that announcing additional monitoring has less impact on transactions that are subject to a paper trail, indicating the paper trail's preventive deterrence effect. This leads to enforcement spillovers up the VAT chain. We return to the theme of how technology of service provision or revenue collection can be harnessed for better monitoring when we discuss the promise of e-governance.

5.4 Information flows and monitoring by citizens

The last decade has seen increased interest in monitoring undertaken directly by citizens. In part, this reflects the increasing incidence of freedom of information acts and in part the greater ease of obtaining already digitized data on program performance.

5.4.1 Does information on program performance matter?

Björkman and Svensson (2010) found that informing Ugandan citizens of the dismal state of local health service delivery and holding meetings between citizens and health workers to agree on action plans significantly reduced provider absenteeism, increased utilization, and improved health. In a second randomized evaluation, Björkman et al. (2014) examined a less expensive version of the program where they did not provide information on health worker performance and found no impacts, suggesting that access to information was key.

More recently, Banerjee et al. (2015) examine the impact of mailing cards with program information to beneficiaries of a subsidized rice program in Indonesia. They found that this increased the extent of subsidy receipt from the program. Beneficiaries received 26 percent more subsidy in treated villages. Ineligible households received no less, suggesting reduced leakage. The impact appears to be driven by citizens bargaining with local officials. Experimentally, adding the official price to the cards increased the subsidy by 21 percent compared to cards without price information. Additional public information increased higher-order knowledge about eligibility, leading to a 16-percent increase in subsidy compared to just distributing cards.

In contrast to these papers, Banerjee et al. (2010) report limited results from a report card intervention in which village volunteers prepared a report card on student learning. They interpret this as reflecting the absence of mechanisms to hold teachers accountable. This study points to the importance of understanding how monitoring mechanisms interact with underlying incentives for government workers. These ideas are explored in the study by Pradhan et al. (2014) who examine the role of school committees in improving education quality in Indonesia. The study is a randomized evaluation covering 520 schools in Central Java from 2007 to 2008. They have four main treatments. The first treatment facilitated democratic elections of school committee members. The second treatment linked school committees to the village council by facilitating joint planning meetings (which they describe as linkage). They benchmark these two treatments against more common treatments: providing block grants and providing training. Two years later, test scores increased by 0.17 standard deviations for linkage and 0.23 standard deviations for linkage plus elections. In contrast, training did not impact learning, and the effect of grants, while positive, was typically statistically indistinguishable from zero. Taken together, the contrasting results from India and Indonesia point strongly to the

importance of institutional reforms that lead to positive interactions between monitoring and incentive mechanisms.

5.5 Summary

The evidence reviewed in this section leads to several conclusions. First, monitoring can to help resolve two agency problems. The first of these is the asymmetrical information that exists between the employer and the employee. With better monitoring, employees have less of an incentive to shirk, and employers also have the flexibility to offer high-powered contracts. The second is the information asymmetry between the service provider and the citizen. With better monitoring (and the disclosure of this information) citizens can hold their service providers accountable, by applying both bottom-up pressure, as well as inducing top-down pressure. Of course, improving monitoring capabilities is unlikely to induce much change without accountability mechanisms in place. Second, third-party reporting matters. With more monitoring, employees may try to game the system. Third-party reporting creates conflicting interests; this helps establish the veracity of the information.

6. TOWARDS SMART(ER) GOVERNANCE: THE PROMISE OF E-GOVERNANCE AND OTHER AVENUES

A common theme that emerges from the body of experimental evidence is the sensitivity of individual behavior to the incentives they face. This, in turn, points to the importance of the structure within which government workers function. Below, we discuss how an emerging body of evidence suggests that e-governance and other technological changes in how the government functions may well help improve governance in low-income settings, perhaps by replacing some of the functions played by personnel—who are subject to all the various problems discussed previously—with technological solutions.

There are a number of ways in which technology can constrain the discretion of local officials and improve performance. For example, in low-income countries the rural poor—an important target group for government transfer programs—are often less informed about state services available to them. In addition, limited state presence in rural areas implies that it is often harder to deploy traditional personnel-intensive monitoring mechanisms to ensure that intended program beneficiaries get their due. As a result, traditional modes of delivery tend to provide the village-level service provider significant discretion in who gets the transfer and when. In cases where payment is supposed to be conditional on the beneficiary undertaking certain activities (e.g., working in a workfare program or children going to school), the village provider often receives funds ahead of the activity having occurred. The monitoring system in these cases often focuses on reviewing the

funds request system. Two recent papers examine how e-governance and the use of biometrics can help improve both the fund-flow system from central coffers to village-level coffers and the transfer of resources from the village-level provider to the final beneficiaries.

[Banerjee et al. \(2015\)](#) report on a field experiment which evaluated an e-governance reform of the fund-flow system for the federal workfare program in the Indian state of Bihar. The reform changed the traditional fund-flow practice by instead conditioning fund disbursement for wage payments on incurred expenditure as reflected in worker detail entry on a new electronic platform. This reform reduced the number of administrative tiers associated with wage disbursement and changed the informational requirements for requesting and disbursing program funds. It did not alter the flow of funds from the village fund to workers. The authors find that program expenditure and reported employment declined by 25% but with no discernible impact on actual employment as measured by an independent survey. Thus, the financial reform was effective in reducing corruption and program costs, but actual demand that was met by the program was unchanged.

[Muralidharan et al. \(2014\)](#) provide evidence on the last step of the payment process. They evaluate the impact of a biometrically authenticated payments infrastructure which provided “Smartcards” to beneficiaries of the federal workfare program. Their experiment randomized the rollout of Smartcards over 158 subdistricts and 19 million people in the southern state of Andhra Pradesh. They find that, while incompletely implemented, the new system delivered a faster, more predictable, and less corrupt National Rural Employment Guarantee Scheme (NREGS) payments process without adversely affecting program access. For each of these outcomes, treatment group distributions first-order stochastically dominated those of the control group. The investment was cost-effective, as time savings to NREGS beneficiaries alone were equal to the cost of the intervention, and there was also a significant reduction in the “leakage” of funds between the government and beneficiaries in both NREGS and Social Security Pension (SSP) programs.

[Barnwal \(2014\)](#) reports supportive quasi-experimental evidence for the staggered rollout and subsequent pulling back of a biometric-based scheme for fuel subsidies in India. He finds that the biometric-based transfer policy reduced fuel purchases in the domestic fuel sector by 11–14% suggesting a reduction in subsidy diversion. In addition, after the policy is terminated, fuel purchases in the domestic sector revert to levels similar to before the policy was introduced.

A different way in which technology can help delivery of transfer programs is the use of mobile money. [Aker et al. \(2014\)](#) use data from a randomized experiment of a mobile money cash transfer program in Niger and find evidence of benefits of this new system: Households receiving mobile transfers had higher diet diversity and children consumed more meals per day. These results can be partially attributed to increased time saving, as

m-transfer program recipients spent less time traveling to and waiting for their transfer, as well as increased intrahousehold bargaining power for women.

Technology can also play a role in reducing malfeasance in elections and promoting electoral accountability. [Callen and Long \(2015\)](#) implement an experiment to estimate the causal effects of photo quick count—a technology used to reduce electoral fraud—and its announcement on aggregation fraud. Photo quick count announcement reduced damaging of election materials by candidate representatives from 18.9 to 8.1 percent, and reduced votes for politically powerful candidates at a given polling location from about 21 to about 15 percent.

In a quasi-experimental study, [Fujiwara \(2015\)](#) studies the introduction of electronic voting on voter enfranchisement in Brazil. Electronic voting was introduced at scale in 1998 elections. But because of a limited supply in devices, only municipalities with more than 40,500 registered voters used the new technology. Using a regression discontinuity design, he finds that electronic voting reduced residual voting in state legislature elections by a magnitude larger than 10% of total turnout, thus enfranchising millions of voters. By enfranchising a poorer and less education population, the introduction of electronic voting led to an increase in the number of prenatal visits by health professionals and lowered the prevalence of low-weight births (below 2500 g) by less educated women.

A third area where technology shows some promise is in government procurement. Government procurement accounts for an enormous amount of government expenditures, and despite many regulations put in place to ensure that procurement is conducted fairly and with limited corruption, the fact that procurement regulations must be implemented by officials allows scope for discretion. For example, officials can withhold detailed bidding documents from bidders outside the favored cartels. [Lewis-Faupel et al. \(2016\)](#) study the introduction of electronic procurement systems for public works projects in India and Indonesia using a differences-in-differences design that takes advantage of the differential rollout of electronic procurement by states/provinces over time. These electronic procurement systems allow greater access to information for all potential bidders and ensure that the procurement rules are followed correctly. [Lewis-Faupel et al. \(2016\)](#) find that electronic procurement leads to contracts being more likely to be won by providers from outside the region where the project is being executed, suggesting that an important role for e-procurement is increasing access to information. They also find that it leads to quality improvements, though not lower prices paid by the government.

7. CONCLUDING THOUGHTS

In countries where the quality of government is low, public servants tend to be paid relatively well. While it is difficult to assign causality to this relationship, it does hint at some important directions for future research.

That public servants earn on average significantly more than their private sector counterparts does not suggest that financial incentives do not matter for bureaucratic performance. The evidence we have reviewed here suggests that they clearly do. But rather than differences in the levels, it is the nature of the incentives that perhaps matter most for performance. In most countries, wages in the public sector tend to exhibit a high degree of compression, which combined with long tenure rates can make it hard to incentivize an individual once employed. But in settings where government can offer high-powered incentives (e.g., tax administration or education), the evidence suggests that public servants do in fact perform better.

High-powered incentives come with tradeoffs: these incentive schemes have the potential to discourage effort among the prosocially motivated, as well as to create issues of multitasking, though the empirical evidence suggests that in many cases these concerns are not as strong as one might initially have thought. The literature has begun to identify some of these tradeoffs, but much more research is needed to better understand in which settings these issues are most likely to arise. For example, task complexity might provide such a setting. The multitasking concerns associated with performance pay are more likely to arise when bureaucrats are tasked with complex jobs. At the same time, complex jobs are more difficult to monitor. Whether the benefits of lower monitoring costs outweigh the costs associated with multitasking is an interesting question with important implications for how bureaucracies should be organized.

Financial incentives also matter for selection. Organizations that offer higher wages will attract more qualified applicants. Given the large public-sector wage gaps, one might be tempted to reason that selection issues are not a first-order concern. But high wages are only a necessary condition for attracting talent, not a sufficient one. The type of individuals who work in the public sector will ultimately depend on how candidates are screened. The screening mechanism can easily undo any positive selection created by higher wages, as is perhaps the case when countries hire based on patronage and not for meritocratic reasons. Governments vary in the way they screen their public servants. To understand how governments should screen among the candidates who apply to further improve selection is another exciting avenue for future research. It is of course difficult to consider optimal screening mechanisms without raising the question of what personality traits we should screen for. While there is an extensive empirical literature arguing that individuals who exhibit high degrees of prosocialness perform better in the public sector, experimental evidence of this relationship is virtually nonexistent.

The usefulness of financial incentives can also be limited if internal accountability mechanisms do not exist or function. One issue that bureaucracies typically face is the inability to perfectly monitor their employees. Fortunately advances in technology appear to be a step forward. The use of smartphone technology and e-governance platforms not only promote transparency and accountability but also serve as disciplining devices. Importantly, these technological advances may also create a feedback loop on the

compensation structure of employees. As governments increasingly adopt these new technologies, thus enabling them to better monitor and evaluate its employees, the set of contracts that it can offer its employees expands. The relationship between technology adoption and compensation scheme is another exciting area of future research.

As this survey documents, the (experimental) research on trying to understand how bureaucracies work is still in its infancy, so there is plenty to do and a lot to learn. We are excited to see what the next several years will bring for this research agenda, which has the potential to unlock some of the doors to efficient service delivery and good governance.

APPENDIX

Region	Country	Year	Survey	Source
Africa	Ghana	2013	STEP Skills Measurement Program survey	World Bank
Africa	Kenya	2013	STEP Skills Measurement Program survey	World Bank
Africa	Nigeria	2012	Living Standards Measurement survey	World Bank
Africa	Niger	2011	Living Standards Measurement survey	World Bank
Africa	Malawi	2010	Living Standards Measurement survey	World Bank
Africa	Tanzania	2010	Living Standards Measurement survey	World Bank
Africa	Timor-Leste	2007	Living Standards Measurement survey	World Bank
Asia	Lao PDR	2012	STEP Skills Measurement Program survey	World Bank
Asia	Sri Lanka	2012	STEP Skills Measurement Program survey	World Bank
Asia	Vietnam	2012	STEP Skills Measurement Program survey	World Bank
Asia	India	2011	Socio Economic survey	National Sample Survey Office, Government of India
Asia	Korea, Rep.	2011	Korean General Social Survey	ICPSR

Continued

Region	Country	Year	Survey	Source
Asia	Indonesia	2007	Indonesia Family Life Survey	RAND Corporation
Asia	Pakistan	2006	Labour Force Survey	Federal Bureau of Statistics, Government of Pakistan
Central Asia	Armenia	2013	STEP Skills measurement Program survey	World Bank
Central Asia	Georgia	2013	STEP Skills Measurement Program survey	World Bank
Central Asia	Tajikistan	2009	Living Standards Measurement survey	World Bank
Europe	United Kingdom	2014	Quarterly Labour Force Survey	UK data service
Europe	Albania	2011	Labour Force Survey	Institute of Statistics (INSTAT), Republic of Albania
Europe	Bulgaria	2007	Living Standards Measurement survey	World Bank
Europe	Serbia	2007	Living Standards Measurement survey	World Bank
Europe	Bosnia-Herzegovina	2004	Living Standards Measurement survey	World Bank
Latin America	Argentina	2014	Permanent Household Survey	National Institute of Statistics and Census, Republic Argentina
Latin America	Mexico	2014	National Survey of Occupation and Employment	National Institute of Statistics and Geography (INEGI), Government of Mexico
Latin America	Bolivia	2012	STEP Skills Measurement Program survey	World Bank
Latin America	Colombia	2012	STEP Skills Measurement Program survey	World Bank
Latin America	Peru	2011	Specialized household survey on employment levels	Ministry of Work and Employment Promotion, Government of Peru

Region	Country	Year	Survey	Source
Latin America	Panama	2008	Living Standards Measurement survey	World Bank
Latin America	Nicaragua	2005	Living Standards Measurement survey	World Bank
Middle East	Egypt	2012	Egypt Labor Market Panel Survey	Economic Research Forum
Middle East	Iraq	2006	Living Standards Measurement survey	World Bank
North America	United States	2010	Current Population Survey	National Bureau of Economic Research (NBER)

REFERENCES

- Aker, J., Boumnijel, R., McClelland, A., Tierney, N., 2014. Payment Mechanisms and Anti-poverty Programs: Evidence from a Mobile Money Cash Transfer Experiment in Niger. Working Paper 268.
- Almlund, M., Lee Duckworth, A., Heckman, J., Kautz, T., 2011. Personality psychology and economics. In: Hanushek, E.A., Machin, S., Woessmann, L. (Eds.), *Handbook of the Economics of Education*. Elsevier, Amsterdam, pp. 1–181.
- Aronson, E., 1999. The power of self-persuasion. *Am. Psychol.* 54 (11), 875–884.
- Ashraf, N., Bandiera, O., Kelsey Jack, B., 2014. No margin, No mission? a field experiment on incentives for public service delivery. *J. Public Econ.* 120, 1–17.
- Ashraf, N., Bandiera, O., Lee, S.S., 2015. Do-gooders and Go-getters: Career Incentives, Selection, and Performance in Public Service Delivery (Working Paper).
- Baicker, K., Jacobson, M., 2007. Finders keepers: forfeiture laws, policing incentives, and local budgets. *J. Public Econ.* 91, 2113–2136.
- Bandiera, O., Prat, A., Valletti, T., 2009. Active and passive waste in government spending: evidence from a policy experiment. *Am. Econ. Rev.* 99 (4), 1278–1308.
- Banerjee, A.V., Banerji, R., Duflo, E., Glennerster, R., Khemani, S., 2010. Pitfalls of participatory programs: evidence from a randomized evaluation in education in India. *Am. Econ. J. Econ. Policy* 2 (1), 1–30.
- Banerjee, A.V., Chattopadhyay, R., Duflo, E., Keniston, D., Singh, N., 2014. Can Institutions Be Reformed from within? Evidence from a Randomized Experiment with the Rajasthan Police. NBER Working Paper 17912.
- Banerjee, A.V., Hanna, R., Kyle, J., Olken, B., Sumarto, S., 2015. The Power of Transparency: Information, Identification Cards and Food Subsidy Programs in Indonesia (HKS Faculty Research Working Paper Series).
- Banerjee, A.V., Glennerster, R., Duflo, E., 2008. Putting a band-aid on a corpse: incentives for nurses in the indian public health care system. *J. Eur. Econ. Assoc.* 6 (2/3), 487–500.
- Banerjee, R., Baul, T., Rosenblat, T., 2015. On self selection of the corrupt into the public sector. *Econ. Lett.* 127, 43–46.
- Banuri, S., Keefer, P., 2013. Intrinsic Motivation, Effort and the Call to Public Service. World Bank Policy Research Working Paper 6729.
- Barnwal, P., 2014. Curbing Leakage in Public Programs with Biometric Identification Systems: Evidence from India's Fuel Subsidies (Job Market Paper).
- Basinga, P., Gertler, P., Binagwaho, A., Soucat, A., Sturdy, J., Vermeersch, C., 2011. Effect on maternal and child health services in rwanda of payment to primary health-care providers for performance: an impact evaluation. *Lancet* 377 (9975), 1421–1428.

- Becker, G.S., Stigler, G.J., 1974. Law enforcement, malfeasance, and compensation of enforcers. *J. Leg. Stud.* 1–18.
- Bellé, N., 2012. Experimental evidence on the relationship between public service motivation and job performance. *Public Adm. Rev.* 73 (1), 143–153.
- Bénabou, R., Tirole, J., 2006. Incentives and prosocial behavior. *Am. Econ. Rev.* 96 (5), 1652–1678.
- Besley, T., Ghatak, M., 2005. Competition and incentives with motivated agents. *Am. Econ. Rev.* 95 (3), 616–636.
- Björkman, M., de Walque, D., Svensson, J., 2014. Information Is Power: Experimental Evidence on the Long-Run Impact of Community Based Monitoring. World Bank Policy Research Working Paper 7015.
- Björkman, M., Svensson, J., 2010. When is community-based monitoring effective? Evidence from a randomized experiment in primary health in Uganda. *J. Eur. Econ. Assoc.* 8 (2/3), 571–581.
- Bloom, N., Van Reenen, J., 2007. Measuring and explaining management practices across firms and countries. *Q. J. Econ.* 122 (4), 1351–1408.
- Bloom, N., Eifert, B., Mahajan, A., McKenzie, D., Roberts, J., 2013. Does management matter? Evidence from India. *Q. J. Econ.* 128 (1), 1–51.
- Bobonis, G.J., Cámara Fuertes, L.R., Schwabe, R., 2015. Monitoring Corruptible Politicians (Working Paper).
- Borjas, G.J., 2015. The Wage Structure and the Sorting of Workers into the Public Sector. NBER Working Paper 9313.
- Boycko, M., Shleifer, A., Vishny, R., 1996. A theory of privatisation. *Econ. J.* 106, 309–319.
- Burgess, S., Propper, C., Ratto, M., von Scholder, S.H.K., Tominey, E., 2010. Smarter task assignment or greater effort: the impact of incentives on team performance. *Econ. J.* 120, 968–989.
- Callen, M., Gulzar, S., Hasanain, A., Khan, Y., Rezaee, A., 2015. Personalities and Public Sector Performance: Evidence from a Health Experiment in Pakistan. NBER Working Paper 21180.
- Callen, M., Long, J., 2015. Institutional corruption and election fraud: evidence from a field experiment in Afghanistan. *Am. Econ. Rev.* 105 (1), 354–381.
- Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K., Halsey Rogers, F., 2006. Missing in action: teacher and health worker absence in developing countries. *J. Econ. Perspect.* 20 (1), 91–116.
- Church, T.W., Heumann, M., 1989. The underexamined assumptions of the invisible hand: monetary incentives as policy instruments. *J. Policy Anal. Manag.* 8 (4), 641–657.
- Cowley, E., Smith, S., 2014. Motivation and mission in the public sector: evidence from the world values survey. *Theory Decis.* 76, 241–263.
- Dal Bó, E., Finan, F., Rossi, M.A., 2013. Strengthening state capabilities: the role of financial incentives in the call to public service. *Q. J. Econ.* 128 (3), 1169–1218.
- Delfgaauw, J., Dur, R., 2007. Signaling and screening of workers' motivation. *J. Econ. Behav. Organ.* 62 (4), 605–624.
- Deserranno, E., 2015. Financial Incentives as Signals: Experimental Evidence from the Recruitment of Health Workers (Working Paper).
- Dhaliwal, I., Hanna, R., 2014. Deal with the Devil: The Successes and Limitations of Bureaucratic Reform in India. NBER Working Paper 20482.
- Dizon-Ross, R., Dupas, P., Robinson, J., 2015. Governance and the Effectiveness of Public Health Subsidies. SCID Working Paper 510.
- Dohmen, T., Falk, A., 2010. You get what you pay for: incentives and selection in the education system. *Econ. J.* 120 (546), F256–F271.
- Duflo, E., Greenstone, M., Pande, R., Ryan, N., 2013. Truth telling by third-party audits and the response of polluting firms: experimental evidence from India. *Q. J. Econ.* 128 (4), 1499–1545.
- Duflo, E., Greenstone, M., Pande, R., Ryan, N., 2015. The Value of Regulatory Discretion: Estimates from Environmental Inspections in India. Harvard Environmental Economics Program Discussion Paper 15–60.
- Duflo, E., Hanna, R., Ryan, S.P., 2012. Incentives work: getting teachers to come to school. *Am. Econ. Rev.* 102 (4), 1241–1278.

- Evans, P.B., 1995. *Embedded autonomy: states and industrial transformation*. Vol. 25. Princeton University Press, Princeton, NJ.
- Evans, P., Rauch, J.E., 1999. Bureaucracy and growth: a cross-national analysis of the effects of “Weberian” state structures on economic growth. *Am. Sociol. Rev.* 64 (5), 748–765.
- Ferraz, C., Finan, F., 2008. Exposing corrupt politicians: the effects of Brazil’s publicly released audits on electoral outcomes. *Q. J. Econ.* 123 (2), 703–745.
- Francois, P., 2000. Public service motivation’ as an argument for government provision. *J. Public Econ.* 78 (3), 275–299.
- Fujiwara, T., 2015. Voting technology, political responsiveness, and infant health: evidence from Brazil. *Econometrica* 83 (2), 423–464.
- Gertler, P., Vermeersch, C., 2012. Using Performance Incentives to Improve Health Outcomes. World Bank Policy Research Working Paper 6100.
- Glewwe, P., Ilias, N., Kremer, M., 2010. Teacher incentives. *Am. Econ. J. Appl. Econ.* 2 (3), 205–227.
- Gneezy, U., Rustichini, A., 2000. A fine is a price. *J. Leg. Stud.* 29 (1).
- Grant, A.M., Campbell, E., Chen, G., Cottone, K., Lapedis, D., Lee, K., 2007. Impact and the art of motivation maintenance: the effects of contact with beneficiaries on persistence behavior. *Organ. Behav. Hum. Decis. Process.* 103, 53–67.
- Hanna, R., Wang, S.-Y., 2014. Dishonesty and Selection into Public Service: Evidence from India. NBER Working paper 19649.
- Holmstrom, B., Milgrom, P., 1987. Aggregation and linearity in the provision of intertemporal incentives. *Econometrica* 55 (2), 303–328.
- Iyer, L., Mani, A., 2012. Traveling agents: political change and bureaucratic turnover in India. *Rev. Econ. Statistics* 94 (3), 723–739.
- Khan, A.Q., Khwaja, A.I., Olken, B.A., 2014. Tax Farming Redux: Experimental Evidence on Performance Pay for Tax Collectors. NBER Working Paper 20627.
- Kahn, C.M., Silva, E.C.D., Ziliak, J.P., 2001. Performance-based wages in tax collection: the Brazilian tax collection reform and its effects. *Econ. J.* 111 (468), 188–205.
- Lagarde, M., Blaauw, D., 2014. Pro-social preferences and self-selection into rural jobs: evidence from South African nurses. *J. Econ. Behav. Organ.* 107 (Part A), 136–152.
- Lazear, E.P., Oyer, P., 2012. Personnel economics. In: Gibbons, R., Roberts, J. (Eds.), *The Handbook of Organizational Economics*. Princeton University Press, pp. 479–519.
- Lichand, G., Lopes, M.F.M., Medeiros, M., 2015. Is Corruption Good for Your Health? (Working Paper).
- Lewis-Faupel, S., Neggess, Y., Olken, B., Pande, R., 2016. Can Electronic Procurement Improve Infrastructure Provision? Evidence from Public Works in India and Indonesia. *AEJ: Economic Policy* 8 (3), 258–283.
- Manning, A., 2011. Imperfect competition in the labor market. In: Ashenfelter, O., Card, D. (Eds.), *Handb. Labor Econ.* 4, 973–1041.
- Meares, T.L., 1995. Rewards for Good Behavior: Influencing Prosecutorial Discretion and Conduct with Financial Incentives. Faculty Scholarship Series Paper 473.
- Mellstrom, C., Johannesson, M., 2008. Crowding out in blood donation: was titmuss right? *J. Eur. Econ. Assoc.* 6 (4), 845–863.
- Miller, G., Luo, R., Zhang, L., Sylvia, S., Shi, Y., Foo, P., Zhao, Q., Martorell, R., Medina, A., Rozelle, S., 2012. Effectiveness of provider incentives for anaemia reduction in rural China: a cluster randomised trial. *BMJ* 345, e4809.
- Miller, J.M., Selva, L.H., 1994. Drug enforcement’s double-edged sword: an assessment of asset forfeiture programs. *Justice Q.* 11 (2), 313–335.
- Muralidharan, K., 2012. Long-term Effects of Teacher Performance Pay: Experimental Evidence from India (Working Paper).
- Muralidharan, K., Sundararaman, V., 2011. Teacher performance pay: experimental evidence from India. *J. Political Econ.* 119 (1), 39–77.
- Muralidharan, K., Niehaus, P., Sukhtankar, S., 2014. Building State Capacity: Evidence from Biometric Smartcards in India. NBER Working Paper 19999.

- Naff, K.C., Crum, J., 1999. Working for America: does public service motivation make a difference? *Rev. Public Personnel Adm.* 19 (4), 5–16.
- Neal, D., 2011. The design of performance pay in education. In: Hanushek, E., Machin, S., Woessmann, L. (Eds.), *Handbook of Economics of Education*, vol. 4. Elsevier, Oxford, pp. 495–550.
- Olken, B.A., 2007. Monitoring corruption: evidence from a field experiment in Indonesia. *J. Public Econ.* 115 (2), 200–249.
- Olken, B.A., Onishi, J., Wong, S., 2014. Should aid reward performance? evidence from a field experiment on health and education in Indonesia. *Am. Econ. J. Appl. Econ.* 6 (4), 1–34.
- Park, S.M., Rainey, H.G., 2008. Leadership and public service motivation in U.S. Federal agencies. *Int. Public Manag. J.* 11 (1), 109–142.
- Parrillo, N., 2013. *Against the Profit Motive: The Salary Revolution in American Government, 1780–1940*. Yale University Press.
- Perry, J.L., 1996. Measuring public service motivation: an assessment of construct reliability and validity. *J. Public Adm. Res. Theory* 6 (1), 5–22.
- Perry, J.L., Hondeghem, A., 2008. Building theory and empirical evidence about public service motivation. *Int. Public Manag. J.* 11 (1), 3–12.
- Petrovsky, N., 2009. Does public service motivation predict higher public service performance? A research synthesis. *J. Public Adm. Res. Theory* 20 (Suppl. 2), i261–i279.
- Pomeranz, D., 2015. No taxation without information: deterrence and self-enforcement in the value added tax. *Am. Econ. Rev.* 105 (8), 2539–2569.
- Pradhan, M.P., Suryadarma, D., Beatty, A., Wong, M., Gaduh, A., Alisjahbana, A., Artha, R.P., 2014. Improving educational quality through enhancing community participation: results from a randomized field experiment in Indonesia. *Am. Econ. J. Appl. Econ.* 6 (2), 105–126.
- Prendergast, C., 2007. The motivation and bias of bureaucrats. *Am. Econ. Rev.* 97 (1), 180–196.
- Rasul, I., Rogger, D., 2015. *Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service (Working Paper)*.
- Ritz, A., 2009. Public service motivation and organizational performance in Swiss federal government. *Int. Rev. Adm. Sci.* 75 (1), 53–78.
- Shleifer, A., Vishny, R.W., 1994. Politicians and firms. *Q. J. Econ.* 109 (4), 995–1025.
- Steijn, B., 2008. Person–environment fit and public service motivation. *Int. Public Manag. J.* 11 (1), 13–27.