

Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy and Training*

Abhijit Banerjee¹, Raghavendra Chattopadhyay², Esther Dufllo¹, Daniel Keniston³, and Nina Singh⁴

¹Massachusetts Institute of Technology

²Indian Institute of Management, Calcutta

³Yale University

⁴Rajasthan Police

Abstract

The role of good management practices in organizations has recently been emphasized. Do the same principles also apply in government organizations, even the most bureaucratic and hierarchical of them? And can skilled, motivated managers identify how to improve these practices, or is there a role for outsiders to help them in this task? Two unique large-scale randomized trials conducted in collaboration with the state police of Rajasthan, India sought to increase police efficiency and improve interactions with the public. In a sample of 162 police stations serving almost 8 million people, the first experiment tested four interventions recommended by police reform panels: limitations of arbitrary transfers, rotation of duty assignments and days off, increased community involvement, and on-duty training. Field experience motivated a novel fifth intervention: “decoy” visits by field officers posing as citizens attempting to register cases, which gave constables incentives to behave more professionally. Only two of these, training and decoy visits, had robust impacts. The other three, which would have reduced middle managers’ autonomy, were poorly implemented and ineffective. Building upon these findings, we designed a second experiment that provided explicit incentives to police officers to carry out sobriety traffic checkpoints and did not rely on middle managers. Linking good performance with the promise of a transfer from the reserve barracks to a desirable police station posting, these incentives worked within existing organizational constraints and had very large effects on performance.

*We thank Clément Imbert, Selvan Kumar, Dhruva Kothari, Neil Buddy Shah, and Pankaj Verma for their outstanding contributions as research assistants on this project. We are indebted to M.K. Devarajan, and M. L. Lather for their insights and comments, and A.S. Gill, formerly Director General of Police for Rajasthan Police, for initiating and constantly supporting this research. The findings and opinions expressed in this paper are those of the authors and do not in any way engage the responsibility of the Rajasthan Police. The paper was written in complete independence and the Rajasthan Police, as an institution, did not weigh in on the analysis of the data or the drafting of the report. All the data collected in the course of this project is available for download and public use (including, but not limited to, replication of our results) at <http://dvn.iq.harvard.edu/dvn/dv/jpal>.

We are grateful for funding from the Will and Flora Hewlett Foundation and the UNODC. The Hewlett Foundation was not involved in the drafting of this paper.

1 Introduction

An extremely important line of work in empirical organizational economics in the last few years was initiated by Bloom and Van Reenen (2007) (henceforth BVR), who argue that the quality of management has an independent role in explaining firm productivity (“management as technology”). In a proof-of-concept experiment in India, Bloom et al. (2013) (BEMMR) demonstrate that management consultants, by providing specific, customized advice, can improve practices and profits even in large and successful firms operating in a competitive environment. We investigate what kind of management reforms can work in a very different type of institution—the police department of the state of Rajasthan, India—by conducting two consecutive large-scale experiments, covering 162 then 123 police stations serving several million citizens.

The police, a rigidly hierarchical organization, may seem an unlikely setting for standard management principles to apply. More generally, the question is still quite open on whether the broad principles of good management that are identified in Bloom and Van Reenen’s 2007 work apply to public organizations or government bureaucracies.¹ Bloom et al. (2010) find that, in public hospitals, competition improves management quality, and in turn financial and health outcomes. However, a recent paper by Rasul and Rogger (2013) uses a unique dataset from Nigeria to examine whether the BVR measures of management quality explain productivity differences across different government departments that do the same thing, and thereby throws a different light on this issue. The results show that (middle) managerial autonomy, one of the study’s main measures of management quality, has a positive effect on performance, just as in BVR. However, the aspect of management related to the provision of incentives, which is associated with good performance in the private sector in BVR, turns out to be *negatively* correlated with performance on project implementation in the Nigerian bureaucracy.

One possible interpretation of the negative effect of incentives is that strong incentives are just inappropriate in government jobs. The seminal paper by Holmstrom and Milgrom (1991) highlights the problem of multitasking: the nature of a civil servant’s job may be such that he has to attend to multiple differing objectives, some of which are much more easily measured than others. This makes it difficult to reward or punish civil servants based on performance, as such incentives will attract their attention away from the tasks that not easily measured, potentially with counter productive effects.²

¹There is also a long line of work showing that incentives (an important principle of management according to BVR) do work in non-commercial settings. Di Tella and Schargrodsky (2003) show that both higher wages (and therefore a higher cost of job loss) and tighter scrutiny reduces what government hospitals in Buenos Aires pay for supplies, and Olken (2007) demonstrate that audits reduce corruption in local government construction projects. In India, Hanna and Dhaliwal (2013) find that while incentives were implemented in a government hospital, health outcomes did improve.

²Consistent with this point, Bandiera, Prat, and Valletti (2009) show that an increased possibility of external monitoring of government purchases actually increases the cost to government of those purchases. Prendergast (2001) makes the related point that part of the difficulty in providing incentives to government workers comes from the nature of the task; for example, criminals are likely to be the only people who have direct evidence that the police are not pursuing them very vigorously, but they have no incentive to complain. He goes on to show that attempts to subject the LA police to external oversight led to a sharp reduction in their effort to fight crime.

However, another possible interpretation is that the problem is not with incentives per se but the specific way they get implemented in these government departments: perhaps the managers interviewed by Rasul and Rogger (2013) did not know how to design incentives in a way that avoids the multitasking pitfalls. If managers cannot always be relied upon to choose the incentives that work best, then the effect of incentives is mediated through the competence of the managers who are assigning them, and the appropriate action is not to abandon incentives, but to give managers the kind of specific advice that helps them choose the right incentive system.

The tension between the “best practice” and the “contingency” view of management is neither new, nor specific to government organizations. In BVR, the specific claim is that the relevant quality of management is reflected in whether the firm has a certain set of best practices in place, and hence quantified by asking the manager questions of general principles. For example, one of the indicators for whether the firm has a proper incentive system is the answer to the question, “If two people both joined the company five years ago and one was much better than the other, would he/she be promoted faster?” In BEMMR, on the other hand, the kind of advice offered by the consultants, while inspired by general principles such as “lean management,” was often very specific, e.g. “Spares [parts should be] stored in a systematic basis (labeling and demarcated locations).” The job of the consultant was not only to inform the managers of a given practice, but also to convince them that it was worth trying. The role of the consultant, who adapts the general principles to the specific setting of the organization, may be particularly important in government.

This paper sheds light on these issues using a suite of randomized experiments carried out in partnership with police department in the Indian state of Rajasthan. In the first set of experiments, the team implemented a series of reforms designed to improve the behavior of the police officers with the general public and crime victims. These reforms fall into four categories. The first entailed training police staff, with a strong emphasis on “soft skills” such as sensitivity and communication best practices. The second was a pair of interventions aimed at improving the working conditions of lower-level police staff by guaranteeing them a minimum set of working conditions: (a) rotation of duty assignments within the police station and guaranteed days off, and (b) a freeze on transfers of police staff for two years. The third was to install civilian observers at police stations in an effort to improve community monitoring and understanding of the police. The fourth intervention was designed to increase incentives for police to register crimes that were reported to them: surveyors posing as crime victims (henceforth referred to as “decoys”) were sent to police stations to register imaginary crimes. The decoys then revealed their role, giving the police a sense that they were being monitored.

To evaluate these reforms, we collected data through two rounds of surveys including police interviews, surprise visits by surveyors to police stations, and a large-scale representative crime victimization survey—the first of its kind in India.

In the second leg of the project we implemented an incentive scheme to improve the efficiency with which police officers conduct drunk driving prevention checkpoints. The incentives leveraged the preference among police staff for assignments to station houses rather than the central police

barracks (“police lines”) in order to provide incentives. We designed an intervention that assigned staff from the police barracks to some randomly assigned checkpoints and promised them an increased possibility of being transferred to a police station if they conducted checks regularly, and did well in terms of stopping drunk drivers. Data was collected in these and comparison checkpoints (which were manned by the usual teams from the local police station) on whether checks were performed, how many drivers were stopped in total, and how many drunk drivers were stopped and fined.

All of these interventions came out of a collaborative effort between Rajasthan police and the research team, which included a senior member of the police department. The effort was driven by the top management of the police, who were keen to improve the popular image of the police force as well as its performance. These were motivated, thoughtful, reform-minded, and powerful high-level officials. The first three interventions came out of their ideas, which were in turn influenced by the recommendations of previous Police Reform Commission reports.³ Interestingly, all of these in effect *reduced* the autonomy of the middle manager (the police station chief, in this case), violating one of the broad management principles found to be important in the literature.

In contrast, the decoy intervention and the drunk driving checkpoint intervention, both of which had to do with incentives, came out of our suggestions. Since performance pay and explicit financial incentives were totally alien to the police managerial system, even senior officers who appreciated incentives in principle thought that adopting them would require infeasible legal changes. The researchers’ contribution as pseudo-consultants was to help identify a window where it was possible to provide some incentive within the existing organization.

The basic results for the two incentive interventions were uniformly positive. The introduction of decoy visits to police stations increased the probability of a given case being registered by 16–17 percentage points. The politeness with which potential complainants were received also went up significantly as measured by our surveyors. Likewise, in the drunk driving checkpoint intervention we find that the police barracks staff, who were subject to stronger incentives, were much more effective along every dimension. They were 50% more likely to carry out the checks, stayed longer, and caught more drivers while there. Over the duration of the project, barracks teams brought to court twice as many drunk drivers as the police station teams.

The “soft skills” part of the training experiment (which sought to teach police how to deal with the public more kindly and effectively) also clearly worked. Trained station staff were more polite with crime victims who were attempting to register a crime, and the victims were overall much more satisfied with the police.

Finally, the three interventions that reduced middle manager’s autonomy were less successful. Only the freeze on transfers may have had any positive effect, while the duty rotation/day off seems to have had a negative effect on some outcomes. The community observer had no significant effect whatsoever.

³The freeze on transfers, guaranteed time off, and training interventions came directly from the Padmanabhaiah Committee on Police Reforms, constituted largely of active and retired senior police officials, while the community observer program was developed in response to the Committee’s recommendations on community policing.

Taken together, these results argue against the view that managers can readily identify the interventions that work best in their context. The broad management principles that apply in private firms or in hospitals may very well work in the police context too, in spite of the obvious differences in culture, but the management needed help to identify the specific interventions that would help translate these into concrete, implementable programs. Of course, we have only six interventions that we are comparing; it could be the case that some unidentified force caused the interventions we suggested to work, while three of the four chosen by the police did not.

However, the evidence suggests that a main reason the three interventions that reduced management autonomy had very little effect (either good or harm) is that they were not implemented properly, despite the support of they enjoyed from the top brass. (In contrast, the three interventions that clearly had a positive impact were also the ones that were best implemented.) This is of course not entirely surprising since the programs reduced managerial discretion at the station level and therefore the heads of the police stations had every reason to try to undermine them. Similarly, in a private sector context, a productivity-enhancing innovation was not adopted in firms in Pakistan because it was successfully blocked by cutters and printers who stood to lose from its adoption (Atkins, Chaudhry, Chaudhry, Khandelwal, and Verhoogen 2014).⁴ However, it is notable that the senior management did not anticipate the ability of this layer of management to undermine these reforms or to reinforce their implementation when, over the course of the project, signs of poor implementation became evident. The basic instinct of the top brass (to rely on discipline over autonomy, control favoritism through regimentation, and value intrinsic motivation rather than incentives, explicit or implicit) was a hindrance in selecting implementable interventions. Perhaps the hierarchical nature of the police does not affect the effectiveness of standard management principles, but reduces the chance that the management can identify on their own reforms that go in the right direction.

2 The Rajasthan Police Experiments: Design and Objectives

2.1 Policing and Police Reform in India

As in many developing countries (e.g. parts of Africa (Opolot 1992), the Philippines (Varona 2010)), the Indian police largely retained the organization and rules established by the colonial power—indeed, the Police Act of 1861 is still in effect 67 years after independence and 64 years after the first democratic elections. As might be expected of a colonial law, the emphasis of the Police Act is on keeping the populace quiescent, in part through intimidation, rather than helping them with their problems (Kolsky 2010).⁵ Moreover the military-style hierarchical organization, with its rigid pay scale and promotion schedule and no guaranteed leave, allows essentially no scope for formal incentives or disincentives. Since 1977, a succession of police reform commissions have

⁴Banerjee, Duflo, and Glennerster (2008) and Hanna and Dhaliwal (2013) are two other examples of how middle-level managers in government bureaucracies undermine programs that limit their discretionary authority.

⁵Especially because the force is extremely overstretched. According to 2011 UNODC data, India has 129 police per capita, versus 227 in the USA. <http://www.unodc.org/unodc/en/data-and-analysis/statistics/data.html>.

recommended wide-ranging changes, but they have very rarely been implemented.

The recent history and current position of the Rajasthan State Police are broadly consistent with that of police across India. In 2007 the total force consisted of 70,767 personnel, divided among 711 police stations with 76% constables—the lowest rank—and only 0.2% members of the elite IPS cadre that supplies most of the top brass. Crime rates were comparable with other Indian states: Rajasthan ranked 4th out of 28 states in total crimes serious enough to be investigated without court order, but 25th in murder rates.

This long history of estrangement from the public was apparent in the data collected in the baseline of our study. The representative household survey data that we collected (it turns out, for the first time in India) showed that only 29% of crime victims had bothered trying to register the crime, often because they felt that the police would not make an effort to assist them. They were probably correct. Field surveyors posing as regular citizens trying to report incidents were sent away without being able to file a report 52% of the time. The lack of trust is further demonstrated by the 53% of baseline survey respondents who claimed that law-abiding citizens fear the police.

2.2 First Reform Experiment

This alienation of the public from the police was becoming increasingly costly in a time of rising terrorism, when it was important to get citizens to report anything that they felt might be suspicious. Against this backdrop, in 2005 the Director General of the police in Rajasthan, A.S. Gill took the unusual step to request our team (made up of Nina Singh, then a senior officer in the Rajasthan Police and an author of this article, and researchers from J-PAL) help the department identify a set of reforms that could both improve the public perception of the police and their effectiveness through randomized control trials.

The main interventions for this experiment were chosen based by the police top brass after discussion with the research team, and were strongly influenced by the reports of the various police commissions, which were co-written by retired senior police officers. Within that set of options, choice was dictated in part by a preference for ease of implementation and our requirement that the unit of implementation be a police station (rather than something larger) to ensure that we had enough independent units to randomize across.⁶ With a view towards potentially scaling up the interventions throughout Rajasthan and elsewhere, the police leadership ruled out any reforms that would require a large increase in budget outlay or manpower, or require changes in the fundamental processes used by the police (changes in the pay scale, performance bonuses, etc.). Interventions that explicitly targeted corruption were also ruled out as politically infeasible as a first step.

Partly to ensure statistical power and partly to make sure that the experiments took place at scale, the interventions covered large parts of the state—the first round in 11 districts and the second round in 10, each covering populations larger than the nation of Denmark. With the exception of the decoy program, which was run by the research team, the primary implementation of the reforms

⁶The emphasis on not getting into too many fights right at the outset meant that the project did not target corruption, bribe taking, or ability to handle complicated or violent law-and-order situations.

was performed through the police hierarchy, with general supervision done at the level of state police headquarters. The district superintendents of police supervised the overall implementation in their respective districts and, except when specified otherwise by the experiment, police station chiefs were responsible for carrying out the interventions in their stations.

The four main interventions were

1. **Freezing of transfers:** In the Indian police, transfers are one of the very few incentives instruments in the hands of supervisors. Partly as a result, transfers are very frequent: one-third of all policemen were transferred in our control stations over a period of eighteen months. However, these transfers are not only controlled by the police management. While nominally under the control of the police, they are frequently imposed by politicians (who are able to apply pressure at the relevant level), often allegedly for partisan or corrupt motives (National Police Commission 1980; Centre for Media Studies 2005; Wade 1985). Unchecked transfers were seen as both bad for the morale of the police and a disincentive for them to invest in acquiring local knowledge, which in turn might affect their police work and relationship with the population. The intervention specified *that for personnel posted to all police stations selected for this intervention (except the control group) all administrative transfers would be frozen for period of at least two years. Exceptions could be made for well-documented cases of police misconduct that required a transfer.*
2. **Weekly day off and duty rotation system:** Indian police procedural regulations provide little guidance on management of human resources at the police station level. In theory, police staff remain on duty for 24 hours a day, seven days a week, and all assignment of leave and duties (such as clerk constable, night patrolling, etc.) is at the discretion of the police station chief. Such an ad-hoc system is susceptible to accusations of favoritism in assigning desirable duties, and many officers report feeling overworked and burnt out (Sidhu 2003). Replacing this with a transparent and fair system of work allocation would lead to less jockeying for positions, more predictability, better informed officers, and higher overall productivity. Note, however, that this intervention reduced the ability of the station head to allocate staff optimally, and may have blunted incentives. The intervention specified *that police station-level duties were to be allocated to staff on a rotating basis according to a written schedule. The entire staff in selected police stations (except the station chief) received one day off every eight days. In smaller police stations, where the shortage of manpower might be more acute, the station chief had the option of extending the work period up to fifteen days.*
3. **Community observers:** Recognizing the lack of police accountability vis-à-vis the public, and limited communication between the police and the public, several of the Indian police reform commissions have recommended the creation of civilian oversight institutions for the police. They have also emphasized the need to communicate to the public a “clear understanding of the limitations and constraints within which the police has to function” (NPC 61.46). Such experiments have been tried in several developed countries, including the ex-

Soviet States (Caparini and Marenin 2005) and South Africa (Bruce and Neild 2005) but this was the first time in Rajasthan. The intervention specified *that two volunteers drawn from a (long) list of potential volunteers would spend about three hours in the police station during peak operating hours, on rotating basis. The observers' sole task was to watch the activities within the police station and become familiar with the duties, procedures and challenges faced by the police.* This program had two main goals: first, to give a group of citizens firsthand experience with the police in a positive setting and encourage them to share their experience with others; and second, to provide community oversight in the police station.

4. **In-Service training program:** A lack of skills, both professional (investigation methods, etc.) and “soft skills” (mediation, communication) have frequently been recognized as significant barriers to effective policing (*The Hindu*, 2011),. Constables are recruited mainly on physical characteristics (height, chest measurements, etc.), and only education through class 10 (less than high school) is required. After the initial training of 12 months for constables or two years for Sub-Inspectors, refresher courses are scarce. Under this intervention *randomly selected police personnel were given training in at least one of two modules:*

- **Professional/Investigation skills:** *292 investigating officers (inspectors, sub-inspectors and assistant sub-inspectors) were trained for six days at the Rajasthan Police Academy with inputs on improving investigation procedures, such as field techniques and documentation, with emphasis on scientific techniques.*
- **Soft skills:** *1541 police personnel of all ranks were trained for three days on improving attitude with the public with inputs on “soft skills” such as communication, mediation, stress management, motivation, team building, leadership, attitudinal, change, etc.*

This training was rolled out across the state of Rajasthan in groups of two to three districts in the months after the other reform interventions had begun.

As mentioned earlier, one important public complaint was that police would often refuse to register cases. To check whether our interventions improved police behavior along this dimension field officers posing as regular citizens attempted to register complaints at the police station. The crimes chosen were not representative of the entire range of possible crimes; rather, we chose scenarios where the crime was relatively petty,⁷ which is where the tendency to under-register may be the greatest. These visits were unannounced and the police learned only ex-post that they had been assigned to a decoy visit, either when the field officer revealed himself in order to prevent the registration of a false case, or (if the police refused to register the case) when a supervisor called shortly after the visit.⁸

Early in the project, our team observed that the police were animatedly discussing decoy visits after they happened. It became clear that, although police had already been told that such visits

⁷Stolen mobile phone (29%), stolen suitcase (25%), stolen bicycle (15%), stolen railway ticket (10%), harassment of female relative (9%), victim of dangerous driving (8%), domestic violence (2%), burglary (1%).

⁸These calls were made to insure that the information set did not depend on behavior during the decoy.

could happen, they had not taken it to heart. This suggested that each actual decoy visit would reinforce the message that these visits were happening, and that knowledge would potentially have an effect on behavior (even though police staff had been assured that the monitoring was not linked to a specific punishment). We therefore decided to transform the decoy visits into an intervention. We randomized the order in which the decoy visits were carried out and used police staff behavior at subsequent visits as the outcome.

2.3 Second Reform Experiment

Soon after the conclusion of this first round of reforms, the Rajasthan Police began preparations to acquire a large number of breathalyzers and for the first time begin strict enforcement of anti-drunken driving laws across the state.⁹ Anti-drunken driving enforcement in India is carried out through sobriety checkpoints, where passing vehicles whose drivers are suspected of being drunk are given a breathalyzer test. Those who are found to be driving drunk have their vehicle confiscated, and the breathalyzer record is produced in court against them.

Police management initially planned to implement the crackdown through local police station staff. However, in the first month of piloting the same implementation failures rapidly became evident: only 60% of checkpoints were actually performed. The challenge was therefore to provide the teams with incentives to do this job well, while financial incentives continued to be ruled out. We focused on identifying specific space in the system where incentives could be provided. It was determined to provide incentives through the implicit promise of transfers based on performance, a common informal practice in the police system.

The teams who could be provided with strong incentives were drawn from the “police lines,” a reserve force of police kept at the district headquarters barracks who normally perform riot control, event policing, and VIP security duties. The police lines are generally considered a punishment posting in the Indian police system, since assignment there effectively removes an officer from contact with the public. As a result, the promise of a potential transfer out of the police lines could be used for incentives. Note that there is no a priori contradiction between the freeze of transfer in the first set of experiments and the use of a promised transfer in this case: the majority of the police personnel do not want to be transferred in a frequent and unpredictable fashion, but conditional on assignment to barracks the majority of personnel want to return to a “field” posting.

To base this incentive on an objective measure of performance, police leadership installed GPS tracking devices in the police vehicles allocated to the police lines staff to travel from their barracks at the district headquarters to the checkpoint locations. These devices provided a real-time display of the vehicle’s location to the district-level supervisory officers via a simple internet interface, and stored the vehicle’s travel routes for future examination. Senior officers could instantly observe, on their desktop, whether the police lines teams were at the checkpoint site, but could not observe whether they were actually stopping vehicles and testing drivers. Police lines teams were told of the

⁹In a companion paper we evaluate the effects of this crackdown on driver learning and ultimately accidents and road deaths.

installation of the GPS devices, and were informed that good performance on this assignment might improve their chances for transfer out of the police lines. This was not set up as a competition between teams: evidence of solid work on this assignment was going to play a role in the next transfer decision.

We assigned the police lines teams to man checkpoints in a randomly chosen subset of police stations; the rest the checkpoints became part of the regular duties of the staff of the police stations under whose jurisdiction they fell. This is the status quo for special crackdowns of this type in the Rajasthan Police. Henceforth we refer to these as the “station teams,” to distinguish them from the “lines teams.” Both types of teams’ ultimate results in terms of drunken drivers ticketed were reported to senior officers at regular intervals. However, since none of the senior officers had previously conducted anti-drunken driving crackdowns in these police stations, they might have had trouble distinguishing underperforming police from regional variation in propensity to drunken drive. These teams therefore had relatively weak incentives.

This comparison is one of the first pieces of evidence of the strength of the promise of transfer to desirable location as a source of incentives in a government setting. It is far from being a perfect test, however, since there were other differences between the police lines and police station teams: first, the personnel is differentially selected, though since the police lines are a punishment posting, the selection almost certainly works against finding a positive impact. Second, and probably more important in this case, the police line teams and the police station teams had different activities during the rest of the day. Station teams may have been tired by the time they had to start their checkpoints, for example. Nevertheless, since transfers are ubiquitous as an implicit or explicit way to enforce compliance, this is a useful data point, at least as motivation for further tests.

3 First Reform Experiment: Experimental design, data collection and results

3.1 Evaluation Strategy

For the first round of interventions, 162 police stations across 11 districts of Rajasthan were included in the sample, with districts chosen to be geographically representative of the state of Rajasthan. The sample represents a substantial fraction of the state: over 20% of the police stations in the state and a population of almost 8 million people. Not only was this scale necessary to get adequate power, it also gives meaning to success (or failure) in implementation; this was not a “gold-plated,” small-scale experiment, but a full-scale rehearsal of what the reforms would look like if they were to be broadly implemented.

Police stations were randomly assigned (by the research team, using a computerized randomization) to different reform strategies, using stratification to ensure a balance on geographical area (district), criminality (number of cases registered in 2005), and urban/rural environment. These variables are included as controls in all regressions performed in this study. Within each police station, staff members were randomly selected for the training intervention with the total percentage

of staff trained in each station randomized at 0%, 25%, 50%, 75% or 100% of station personnel. This allows us to detect whether the effects of training are different when few or almost all the officers are trained: is it sufficient to train a few police officers, who will then train their colleagues? Or is a critical mass necessary to get any impact, because trained officers get “contaminated” by those who have retained their old methods? Or is the effect broadly linear?

Surveyors visited sample police stations to collect data on a relatively regular basis. To assess a possible Hawthorne effect—the possibility that the police stations participating in the program might change their responses or behavior simply because they are being observed by outside evaluators—an additional group of police stations was also randomly selected to serve as the “pure control.” These police stations were not informed of the project, had no staff trained, and were never visited by investigators until the final endline survey, at which point they were surveyed like all other stations. The Hawthorne effect can be estimated by comparing the control and “pure control” groups.

Table 1 describes how the police stations were allocated to each evaluation group, with the number of police stations in each group in column 1, and the program elements implemented in that group in columns 1–4. As shown in column 4, wherever the community observer or weekly day off/duty rotation interventions were attempted, transfers were also frozen. This was done for two reasons: first, because long tenure of staff is considered necessary to get acquainted with the area and people of the police station jurisdiction and see if this has any effect on their performance; and second, due to concerns about attrition during the study. This implies that the impact of duty rotation and community observers (or the combination of the two in the all interventions stations) will be seen on top of the impact of the freezing of transfers.

Finally, to evaluate the impact of the decoy intervention, we randomly selected the order in which each police station would be visited by a surveyor pretending to register a crime. Decoy visits were rolled out in overlapping rounds, so that, for example, the final police stations of the third round were receiving their decoy visits during the same period as the first stations of the fourth round. This implies that, conditional on calendar time, the number of past decoys performed is randomly assigned for any given police station. The scripts used to attempt to register cases, as well as the identity of the surveyors, were also randomized with surveyors only attempting to register a decoy case in police stations they had never previously visited.

3.2 Data

A series of surveys and interviews were conducted to establish baseline levels of police performance and baseline public perception of the police, as well as to estimate the effectiveness of the police implementation at the endline. These included traditional household surveys, surveys of police personnel, as well as random, unannounced visits to all police stations to collect data on program implementation. These rich data sets are one of the outcomes of the project: they are available for public use at <http://dvn.iq.harvard.edu/dvn/dv/jpal>.

Visit of Decoy Victim Surveyors

As mentioned above, the “decoy” intervention was originally conceived as a source of objective data on officer behavior. Surveyors posing as victims of various types of petty crimes visited police stations in the treatment and control groups and attempted to register cases. Immediately after the visit to the police station, the surveyor completed a short form recording his success or failure in registration, and the attitudes and actions of the police. The field officer also recorded their impression on whether or not they thought the police officer may have suspected them to be a decoy, which seems to have happened very rarely.

Public Surveys

The other main source of data was a household survey that recorded both general perceptions of the police and the experiences of those individuals who happened to be victims of crimes in the previous year. In total 5,895 households were interviewed: in the spring of 2007, a baseline was conducted with 2,454 households randomly chosen from the 2004 voters list. In the autumn of 2008, the households that could be located again were re-interviewed along with 3,441 newly chosen households.¹⁰

Within each household, three modules were administered (at baseline and endline). First, a general crime-screening questionnaire administered to the head of the household, inquired whether any member of the household had been a victim of a crime in the previous year. If any victims were reported, a crime victim survey, recorded much greater detail about the crime. Finally, an opinion survey, administered to a randomly chosen adult in the household, asked about his or her contact with the police, and perceptions of police performance and integrity.

The crime-related surveys were broadly based upon the International Crime and Victimization Survey from the UN’s Interregional Crime and Justice Research Institute (UNICRI), customized to the Indian context in order to make their results comparable with the crime categories of the Indian Penal Code. Like any crime victimization survey, these questionnaires do not capture “victimless crimes” such as gambling or drug use. Also, no attempt was made to measure the incidence of domestic violence because of concerns about the accuracy of reporting and the possibility of decreased respondent cooperation.

Case-Investigation Grading

To measure the impact of the training (and potentially other interventions) on the quality of the actual police work, 982 case files were randomly selected from the project police stations and sent to a group of retired senior police officers for grading in two rounds, half before the training and half post-training. The retired officers filled out a detailed report about each case in which they

¹⁰In order to gather more data on crime victims, these newly chosen households were selected disproportionately from police stations that had proven to have high crime victimization rates in the baseline survey. To control for any imbalance this may introduce into the sample, all regression specifications without police station fixed effects are re-weighted to be representative of the population of police stations included in the survey.

graded the performance of the officer on his actions at the crime scene, whether scientific techniques were used, and the care with which the evidence was collected and documents in the case file were prepared. Due to high turnover of police staff—despite the freeze of transfers that was supposedly in operation—the sampling of the case investigation grading is not representative of the sample of officers being trained. Rather, we sampled at endline to ensure representativeness of officers working in the police stations at that point, and we will compare the quality of the investigation at the police station level. This comparison will therefore not give us the total effect of the training (since a trained investigator may now be working in a control training station), but it will shed light on the channel through which police training improves victim and public satisfaction.

3.3 Crime and Policing in Rajasthan: Baseline Status

The baseline data collected for the project provides further evidence on the context and background for the interventions and the study. In particular, it confirms many of the problems discussed in section 1, which motivated our choices of interventions. Table 2 presents some descriptive statistics with highlights from the surveys at baseline.

Crime

The detailed crime survey provides the first large-scale representative estimates of the crime rate in Rajasthan, since India does not have a standard crime victimization survey. We find that 1.7% of individuals were victims of a crime in the past year, and that 5.9% of households had at least one member who was a victim of a crime. This estimate excludes crimes of domestic violence, which were not surveyed because of accuracy concerns. Theft was by far the most common, constituting 37.9% of all reported crimes, followed by burglary at 16.6% and assault at 12%.

However, most crime victims never report their incidents to the police. Only 29% of the crime victims we surveyed stated that they had visited a police station to report the crime.¹¹ Among those who attempted to report their crimes, 17% did not succeed in registering a case. The comparison of crime surveys and official police records confirms that per-capita crime rates are substantially higher from survey data, especially for property crimes: survey theft rates are 9.24 times registered rates and robbery rates are 11.35 times higher than in police records. The differential decreases for violent crimes: rape or molestation surveyed rates are 43% rates reported to the police, and assault is only 11% higher in the survey (neither difference is statistically significant). While the biggest reason for not reporting a crime was lack of importance perceived by the victim (28%), substantial numbers of crime victims did not go to the police because they thought that the police were incapable of helping (20%) or unwilling to help (17%).¹²

¹¹The comparable number from the United States is 42%. Source: National Crime Victimization Survey, 2006–2010.

¹²An interesting analysis of the discrepancy between reporting and actual crime is provided in Iyer et al. (2012). They show that reporting of crime against women increases when political positions are reserved for women at the local level, both in our data and in nationally representative data. However, they also find that in our data there is no actual increase in the incidence of crime.

Police attitude with crime victims

The poor behavior of police officers with victims is clearly demonstrated by the results of the decoy survey. Overall, in the first decoy visits conducted by our investigators, police were willing to register a First Information Report (known as an FIR, the necessary first step to a criminal investigation) from the decoys only 54% of the time. More serious and verifiable cases are registered significantly more frequently: house break-ins led to 92% FIR registration while stolen mobile phones had only 40% success in registration. Crime victims who attempted to file an FIR had better luck on average (83%), likely due to the high threshold before deciding to actually bother to register: they were less likely to attempt to report the sort of minor crime that were reported by our decoys. For example, only 7.4% of theft victims who lost less than 1,500 rupees (roughly the value of a mobile phone) even attempted to register their case with the police.¹³

Decoy visits also allow us to assess the politeness of the police officer. They found that 31% of the time the police were not polite, and only 11% of the time they were “extremely polite.”

Public Perception of Police and Victim Satisfaction

Few respondents have experience with the police or ever interacted with them. Only 9% of those interviewed have ever had an interaction with the police in their lives, and only 3% of women have spoken with a police officer. Even in urban areas, only 20% of male respondents ever reported interacting with the police. In the absence of personal experience, 72% of those surveyed claimed that they had formed their opinions based on word of mouth from other citizens, with only 14% and 16% stating that they based their opinions on print and television news sources, respectively.¹⁴ This pattern of information dispersal—opinions formed by hearsay and few real experiences—would suggest that an intervention like the community observer program, which gives a few citizens positive experiences with the police, does have potential to improve perceptions of the police.

Unlike the average citizen, crime victims frequently have personal experience with the police. 39% of victims were either completely satisfied or satisfied, but 35% of victims reported being completely unsatisfied, and 27% mostly unsatisfied. 82% of the unsatisfied victims stated that they were unhappy because the police “didn’t do anything special” or “didn’t take interest” in their case. In contrast, only 15% complained of police corruption, and another 20% complained of discourtesy when attempting to report a crime.

Case review performance

The reviews of case files by retired officers revealed substantial heterogeneity in the quality of investigations of different types of cases. The retired officers gave 6.18 out of 10 on average for overall field investigation quality, and 5.91 out of 10 for quality of case documentation. However, these grades were very dispersed: standard errors were 1.95 for field investigation and 2.42 for

¹³We took care that decoys would look and act like typical young men in the area, although if they had been identified as decoys, the rate of registration would presumably have been greater, not lower.

¹⁴Respondents were permitted to list as many sources of information as they wished.

documentation. Only 19% of cases investigated during 2006–2007 and reviewed by our reviewers made any use of scientific investigation, and most of these were accident cases.

3.4 Reform Implementation

In the previous literature, a major obstacle in the implementation of bureaucratic reform at scale has been the resistance to sustained implementation on the part of the middle management, on whom the reform where often imposed. We document similar difficulties in this setting for all the programs that relied on implementation by the station chiefs. The reasons may be manifold: few direct benefits to station chiefs from the intervention, lack of manpower and other resources, or competing priorities such as a violent statewide ethnic agitation and a series of terror blasts in the state capital that occurred during this period. A key reason, in our view, is that all these interventions essentially took autonomy away from police station chiefs. In this section, we present evidence on the strength of implementation.

No Transfer The extent to which transfers were actually frozen is measured through administrative data from personnel records. We matched names of officers from police station staff lists at the beginning of the program with those from the end of the program, and counted any officers not found as having been transferred between the beginning and the end of the project. Table 3 shows that there was indeed a significant reduction in the number of transfers out of police stations in the ranks that were targeted: 13 percentage points for constables (or a 44% reduction from the 30% mean in the control group), and 15 percentage points for inspectors (a 23% reduction from the 63.6% rate of transfer in the control group). Still, transfers during the course of the project remain quite frequent (21% on average across all ranks, in the treatment group), particularly at higher ranks.¹⁵

While far from a complete freeze, the ban thus resulted in significantly longer posting duration. Transfers require approval from a senior officer at headquarters, and therefore the ultimate control resides at the top, which insured some amount of adherence to the policy. However there were pressures to transfer particular staff members both from below and from above (i.e. from politicians), and some instances were more successfully resisted than others.

Training The training was the easiest intervention to systematically implement, and there is administrative evidence that it took place largely as planned. Rajasthan Police Academy records show that 88% of officers selected for training reported to the academy for training. Like the ban on transfers, senior officers have tight control over implementation and can instruct specific constables to report to training.

¹⁵These figures are lower bounds, since some personnel may have been transferred into then back out of the sample police stations during this period and thus included among the transferred staff.

Community Observers To measure the strength of implementation of the remaining two interventions (community observers and the combination of weekly day off and duty rotation), surveyors visited the police station several times during the course of the project. Each police station in the program designated a specific time during which the observer was available on each day, but in actuality observers were present only 10.29% of the time during designated hours—and were about as likely to be at the station at any other time. A villager wishing to visit the police station when a community observer was present would be very hard pressed to decide when the right time would be.

Overall, it appears that the community observer intervention was not properly implemented on a sustained basis. Unlike the previous two programs, this one was under the direct control of the station chiefs and their imperfect commitment was exacerbated by the difficulty of sustaining community participation. The observers had little to gain from spending three hours in the police station, and goals of better public relations for the police department may have had little relevance to their own interests. These results echo Olken (2007) and Banerjee, Banerji, Duflo, Glennerster, and Khemani (2010), who provide further evidence of the difficulty of sustaining collective monitoring to improve governance.

Weekly Day Off and Duty Rotation

On paper, the implementation of the weekly day off program was good: 84% of treatment police stations reported executing some type of rotating day off, whereas only 34% of control police stations reported having any scheduled day off.

In order to verify that the police station chief was in fact giving days off, surveyors randomly selected two constables from the staff and interviewed them separately at each of their visits, asking them to recall the last time that they had had a day off for any reason, the weekly day off program or otherwise. The results, presented in Table 4, panel A, show that early on, the weekly day off program did succeed in shortening the time since the last day off, going from an average of 30 days in control stations to about 21 days in treatment stations. However, this difference shrinks over the course of the program, eventually disappearing by round five, our last visit.

The implementation of the duty rotation was also close to complete on paper, with 91% of station house officers of treatment police stations able to produce a duty roster when asked. 58% of program police stations posted the duty roster on the wall as stipulated by the implementation guidelines, as opposed to 39% of control police stations, which had independently created duty rosters. Table 4, panel C shows the responses over time to questions about the rapidity of changing duties. It turns out that, even in control stations, the rotation from duty to duty is rapid, if haphazard. Overall there is no significant impact on time spent on specific duties.¹⁶

¹⁶One objective of the rotation was to create some predictability in the assignment to limit the jockeying for positions. Constables in police stations with duty rotation did have a significantly higher probability of reporting that they knew what their next duty would be, with an overall average of 50% in treatment police stations and 41% in controls. Unlike the weekly off program, this effect appears to have grown stronger over time, with the greatest treatment/control differences in the final periods.

3.5 Were the reforms effective?

We now turn to the impact of the interventions on police behavior. We start from the most direct measure of impact, namely the behavior on the decoy surveys. We then move to crime victim satisfaction measured from our victimization survey, and finally to any possible broader impact on the population.

3.5.1 Using the decoy survey outcomes

Table 5 presents regressions of the following form for the two outcomes of interest (whether the case was registered, and whether the police were very polite):

$$y_{dj} = \beta_0 + \beta_T T_j + \beta_D D_{tj} + \beta_P (\%trained)_j + \beta_X X_{dj} + \epsilon_{dj}$$

where d indexes the decoy visit, j indexes the police station, T_j is a vector of binary variables describing the treatment conditions, D_{tj} is the number of past decoys performed in this police station, $(\%trained)$ is the percentage of the police station staff that were trained, and X_{dj} is a vector of control variables. In particular, all regressions include month fixed effects, crime script used, district fixed effects (districts include multiple police stations but have a single chief of police), controls for urban/rural location, and the 2005 crime rate to account for the stratified sampling design. Since order of decoy visits was randomly assigned, conditional on the calendar date, the number of past decoy visits is randomly assigned and indicates impact on police behavior of more past decoy exposure.

Standard errors in all regressions are clustered at the police station level to accommodate arbitrary serial correlation across decoy visits within police station.

Given the combinations of treatment, the treatment variables could have been parametrized in different ways: this, of course, does not affect the results, but it does affect how the tables must be read and interpreted. The list below summarizes our parametrization choices, and will be useful reference for the analysis below.

Summary of definition of treatment variables:

No transfers: A dummy equal to 1 if transfers were intended to be frozen in the police stations.

Recall that all treatment stations had a no transfer policy, so this dummy is 1 for every police station that received any treatment (except for the training program which was not necessarily paired with the no transfer policy).

Weekly off and duty rotation: A dummy equal to 1 if transfers were intended to be frozen and the station was assigned the weekly day off/duty rotation intervention but not the community observer treatment.

Community observers: A dummy equal to 1 if the station got the no transfer program and the community observer program but no weekly day off/duty rotation.

All interventions: A dummy equal to 1 if the station was assigned no transfers, community observers, and weekly day off/duty rotation.

Percentage trained: Percentage of police staff in the station that were trained.

Number decoy visits: Number of decoy visits that had previously taken place in this police station. Since order of decoy visits were randomly assigned, conditional on the calendar data, the number of past decoy visits is randomly assigned and indicates impact on police behavior of more past decoy exposure.

In Study: A dummy equal to 1 if the police station is not in the pure control group (to identify any possible Hawthorne effect). Since pure control stations were only surveyed in the endline, this effect cannot be identified in regressions with police station fixed effects.

The treatment police stations are compared to the regular control. Further, the “community observer,” “weekly off,” and “all interventions” dummies indicate the additional effect of being assigned those additional interventions compared to being assigned no transfers only. So, for example, to compare a control police station to a police station that received the no transfers and community observers interventions, we need to add the two coefficients. To compare the stations that were assigned no transfers, weekly day off/duty rotation, and community observers to the control group, we need to add the “no transfer” coefficient and the “all three interventions” coefficients.

Table 5 presents the results. The baseline FIR registration rate is very low: only 48% of crime reports that the decoys tried to get registered in the first visit were in fact filed. Columns 1 and 4 are exactly as above, and column 2 and 5 include in addition an indicators for whether the surveyor thought the police might have suspected that he was a decoy.¹⁷ In columns 3 and 6 we include police station-level fixed effects to control for any random sampling imbalance across police stations.

The decoys themselves appears to have a substantial impact on FIR registration. Each previous decoy visit increases the probability of FIR registration at the subsequent visit by 16–17%. Figure 1 shows these effects graphically, plotting the coefficients on dummy variables indicating the number of preceding decoy visits from a regression otherwise the same as column 1 in Table 5. This figure implies that by five previous decoys, the registration rate is close to 100%. We do not believe that this is due to a special treatment of the decoy visitor. Decoys were local people, dressed normally, and reporting very common crimes. We asked the decoys to record whether they suspected that the police officer thought they were decoys. This was reported to be likely in 3% of cases, and possible in 15% more. Results in columns 2 and 4 show that controlling for cases in which the surveyor suspected that the police were aware he was a decoy does not attenuate the coefficient on the number of previous decoy visits. In addition, the staff training shows a positive effect on

¹⁷We have also estimated a specification where the effects of training on FIR registration was nonlinear. There was no effect on that specification either (results available upon request).

registration, becoming much larger and significant when fixed effects are included in column 3. The other interventions did not have any impact.

Turning to the effect on politeness, both the number of past decoy visits and police officer training had a significant impact on the politeness with which the “victim” was dealt with, with $p \approx .06$ in columns 4 and 5. Recall that the training for constables (who are in charge of the FIR registration) focused on “soft skills,” and in particular on basic notions of how to interact with victims and regular citizens. This seems to have been effective in changing behavior. Again the size and significance of the coefficient increases with police station fixed effects in column 6.

Thus, the decoy survey program had a substantial effect on both FIR registration and behavior of the police. This, despite the fact that the program was quite explicitly unlinked to any possible sanction: the police officers were warned at the outset that decoy visits could take place, but that the results would not be communicated to their superiors. Yet, they clearly believed that the information might be used in some form (in a more permanent intervention not linked to a research project, this would be made explicit). This result is consistent with other results in the literature showing the potential of the threat of “top-down” monitoring to improve rule compliance, even within generally corrupt environments (Olken 2007).

3.5.2 Victim Satisfaction with Police

After the behavior most directly measured by decoys, the second most direct effects of the reform intervention would be found on those citizens who have the most at stake in their interactions with the police: crime victims. We analyze data from the household survey, using the detailed crime reports to infer whether a crime was committed after the intervention and/or training had been conducted in a district, and comparing treatment and control police stations while controlling for district and time effects.¹⁸ In each case in which the victim affirmed that the police had become aware of the crime we collected data on his or her satisfaction with the police conduct on the case. The interventions had no effect on either incidence of crime (Table A2, not surprising, since it was not a target of the operation), or the probability of police becoming aware of the case (Table A3). Thus this selection rule did not induce differential selection in the different treatment police stations.

The main specifications in Table 6 shows the results of the project on the probability that the respondent reports being “satisfied” or “completely satisfied” with the police handling of his or her case. We estimate two specifications. First, a regression similar to the decoy regression above and with a slightly different set of control variables.¹⁹

$$y_{ij} = \beta_0 + \beta_T T_j + \beta_P (\%trained)_j + \beta X_{ij} + \epsilon_{ij}$$

¹⁸All endline satisfaction reports for which victims could not recall the date of the crime were tagged as post-training (24 cases).

¹⁹The control variables we include are the variables on which the stratification was conducted (urban/rural area, past case registration, and district fixed effects) as well as household-level characteristics (indicator variables for occupation, religion, sex and age of victim, and motorcycle ownership) and controls for the type of crime.

In this regression y_{ij} is outcome y reported by citizen i living in the area covered by police station j . In this specification we use only data collected from crimes that occurred after the training and intervention had both been carried out in police station j . Because the training was rolled out district by district, the relevant control observations for the staff training intervention are those crimes which occurred in the jurisdiction of a district’s control police stations after the staff in that district’s treatment police stations had received the training.

Our second approach takes advantage of the police station panel to include fixed effects at the police station level. Here we estimate difference-in-difference type regressions of the form,

$$y_{ijt} = \beta_0 + \beta_{TFE} (T_j^{FE} \times POST_{1t}) + \beta_{PFE} (\%trained_j \times POST_{2ijt}) + \beta_X X_{ijt} + \gamma_1 POST_{1t} + \gamma_2 POST_{2ijt} + \eta_j + \epsilon_{ijt}$$

where t now indicates the time at which citizen i was a victim of a crime. $POST_{1t}$ is an indicator for the endline survey, and $POST_{2ijt}$ is an indicator for whether the personnel training had been completed in the district where station j is located by the time crime i was committed.²⁰

The coefficients of interest (reported Table 6) are those on the interactions with the treatment vector, β_{TFE} and β_{PFE} . X_{ijt} is a vector of the controls listed above, and η_j denotes the police station-level fixed effects. To conserve space we do not explicitly write the full description of the interaction variable into the row labels of the tables when reporting results from regressions with station-level fixed effects. Standard errors are clustered at the police station level.

Consistent with the decoy results on politeness, the training program shows a large impact on the satisfaction of crime victims, both in the fixed effect and the endline-only specifications. The effect of going from 0% trained officers to 100% is to raise the probability that victims are satisfied with police investigation by between 15 and 29 percentage points, depending on the specification. Since on average only 30% of victims report being entirely or partially satisfied, these changes represent a greater than 50% increase in satisfaction.

Unfortunately the crime victim surveys are not very well suited to estimating the effects of the decoy survey program. Only 149 crime victims were, like the decoy surveyors, males in the age range 20 to 50 who personally reported the crime to police, and of these only 51 reported crimes types that decoys reported (burglary, theft, and sexual harassment). Nonetheless, the analysis performed in this very limited sample is quite suggestive: in the fixed effect specification, each extra decoy visit performed before the crime reportedly happened increased crime satisfaction by 0.17, ($p > 10.5$) in the specification without control for types of crime. This increases to 0.26, ($p > 17.5$) in the specification with control for type of crime—more than doubling baseline satisfaction rates (Table A4)

The other interventions do not appear to have been as robustly effective. There is a 14% to 28% increase in satisfaction attributable to the freezing of transfers (which took place in all the police stations with any other intervention). However, this average effect is largely or completely

²⁰The variation in $POST_{2ijt}$ comes both from the roll out of the training by district and from the date at which the crime was committed. Note that it is defined similarly in treatment and control police station since we use only district-level variation in the timing of training to construct it.

undone by combining it with any of the other two interventions (community observer and weekly day off/duty rotation) or with both of them together.

The effect of the training might operate through two different mechanisms: first, it may affect public perception by changing what the police do, for instance by improving their investigation skills, allowing for faster resolution of the case. Alternatively, the effect may be due to the soft-skills training, which would manifest itself through the way in which the police treat victims and not their investigation actions. We examine these two alternative hypotheses in Table 7, Table 8, and column 3 of Table 6.

Table 7 presents the effects of the training on other police activities, such as registering cases, asking for bribes, making arrests, or re-contacting the victim. We also add in column 7 variables collected from the official police statistics about the number of cases registered by month. The results seem to suggest that most of the effects of the training operated through the behavior of the police with crime victims rather than their increased investigative success. The only significant result relates to property recovery, for which the probability increases by 4%. Consistent with the evidence from the decoy survey, we find suggestive evidence that police officers are more likely to register crime: a 18% increase with $p = .108$. The estimated impact of the training on the number of cases registered in the administrative records (column 7) is also significant at $p < 10\%$, and suggests a 9.4% increase in case registration for stations with 100% of staff trained.

However, crime victims may not be well equipped to judge police competence or effort. Table 8 reports the effects of training on the grades given by retired police officers to the case files of the investigating officers. We find no impact of either an officer's own training or the training of his fellow officers on the quality of case paperwork, the reviewer's evaluation of the officer's field investigation, or whether scientific investigation was used in the case.

Controlling for the police action and success used as outcomes in Table 7 in the satisfaction regression in column 3 of Table 6 only slightly attenuates the coefficient on training, suggesting that other, more subtle, changes in attitude affect victim satisfaction. While for the lowest levels in the police hierarchy (the constables), the training was focused on soft skills, it also included a week of refresher training on investigations for the more senior officers. These results support the conclusion that it is the soft skill training that is likely to have had the most impact.

These regressions impose the restriction that the effect of the fraction of police officers who are trained is linear in the fraction trained, i.e. that when training 50% of officers, the effect is twice as large as when training 25%. In reality, the effect could have been non-linear. Our design allows us to answer this question, since the fraction of officers trained varied randomly from station to station. Figure 1 displays the coefficients on percentage of personnel trained from a regression like that in column 2 of Table 6, with the fraction of staff trained broken down by each randomization category. The effect of training 25% of the station staff is close to zero and insignificant. Only when 50% of the staff are trained do we see a 20% increase in victim satisfaction, after which the effects are roughly constant or even decreasing. These results suggest that police training, or at least police training intended to change attitudes and treatment of the public, must be given to at

least half the staff in order to take root.

3.5.3 Public Perception of Police

Finally, we investigate whether the interventions had an impact on the more general public perception of the police. Given that very few people actually interact with the police in any capacity, it appears ex-ante unlikely that we would find results on this dimension, but Table 9 directly investigates this question.

Since the public opinion survey covered a broad range of qualitative indicators, the 14 major questions on police performance are aggregated into four broad categories in order to simplify the interpretation, make sure all questions asked are covered, and guard against false positive results. The categories were: first, responsiveness of police to citizens, second, fear of police, third, corruption, and fourth, adequacy of police resources²¹

The group outcomes are then analyzed following the approach outlined in Kling, Liebman, and Katz (2007). We first normalize the values of each outcome by the variance of the control group, and jointly estimate all regressions in the group. The group outcome is, then, the mean of the normalized individual outcome coefficients, with the standard errors of the mean coefficient incorporating the variances and covariances of the individual coefficients. The specification is identical to the specification in Table 6.

Table 9 displays the effect of each reform on each category, with results in columns 1 limited to the endline data, and those in columns 2 including the baseline data with police station-level fixed effects. The results are expressed in standardized effect size, with standard errors in parentheses. This table indicates that the perceptions of the general public were not particularly affected. Of the four categories, the project has significant positive effect across both specifications only in one, the “Fear of Police” category, and for one intervention, the freezing of transfer. The freeze of transfer seems to reduce the fear of police among ordinary citizens. However, this effect is undone in stations that also had the weekly day off/duty rotation intervention and, given the limited implementation of this intervention, it seems a little surprising that it has such a robust effect on its own.²² The most likely interpretation, assuming that this is not just a statistical accident, given the lack of

²¹The questions incorporated into the categories were as follows:

A. Responsiveness of police to citizens: “How do the police behave with normal citizens?” “Do the police help citizens when required?” “How quick is the police response to distress calls by citizens?”

B. Fear of police: “Do you think that citizens like you are afraid of the police?” “Are law-abiding citizens afraid of the police?” and when asked how the population thinks of the police, replies that they “fear them.”

C. Corruption: “Would you say that the police in your area are generally honest or generally corrupt?” “Is it necessary to pay the police some money in order to get them do their job?” and “Do policemen themselves violate the law than the average citizen?”

D. Amount of police resources: “Do the police have enough personnel to do the work required of them?” “Do the police have enough money and resources to do the work required of them?” “Should the size of the police force be increased, decreased, or stay the same?” and “Do you think that the government should spend more money on the police, even if it means spending less on things like education and roads?”

²²Appendix Table A3 unpacks the effect on fear of police by examining the effects of all interventions on the individual questions that constitute the fear of police outcome separately. While the signs of all coefficients are the same across questions, the effect seems to be driven primarily by the single question that asks respondents whether they believe that “law-abiding citizens” in general fear the police.

familiarity with police that the baseline data demonstrate, is that less frequent transfers allowed the public to become more familiar with the same police staff and hence came to trust them more and fear them less. It is also possible that once the police staff remained in a post for a longer period of time, their behavior changed with respect to the inhabitants of that area and they became less intimidating to the population and more familiar with their beat. This could also help explain why the duty rotation undid the positive effect of the no transfer policy: duty rotation made it impossible to implement the beat system, as it results in duty officers having no attachment to a particular position or neighborhood.²³

In summary, a comprehensive evaluation of the interventions chosen by the top brass from a menu of commonly suggested reforms provided several interesting lessons. First, consistent with previous experience, effective implementation is a challenge when it crucially depends on the middle management and/or community action, even in an intensely hierarchical organization like the police. Second, training turned out to be one of the most effective policies we evaluated. Third, the only other clearly effective intervention were the decoy visits which, as far we know, were not considered as a possible intervention by the police leadership, even though they seem to be easy to implement if the top hierarchy commits to them. Although they did not carry any explicit punishment, this was clearly a monitoring intervention and raised the possibility, contrary to some of the previous literature, that (well chosen and well implemented) incentives can be made to play a positive role even in a government bureaucracy.

Taken together, these results increased our motivation to identify a space in the system for a monitoring intervention combined with some incentive. In the course of the project, the central role of transfers and posting as the sole motivating tool at the disposal of the hierarchy was frequently emphasized, in particular in reaction to the duty roster and the freezing of transfer interventions. In the current system, the transfer system may often be abused, causing excessive dislocation and low morale, and as we saw, limiting transfers might have some positive effects. But the possibility of transfers to a more desirable position, used judiciously as a source of positive incentives, seemed worth exploring.

4 Second Reform Experiment: Transfers as rewards

A crackdown on drunk driving by the Rajasthan Police provided an opportunity to test the hypothesis that combining monitoring with the promise of transfer to a more desirable location could be an effective intervention. The occasion of these interventions was the purchase of breathalyzers by the police department to provide evidence of drivers' intoxication levels, without which prosecution is extremely difficult. This was intended to motivate and empower the teams carrying out sobriety checkpoints on the main routes.

²³It is much less clear why the day off would undo the positive effect of transfer. However, given that the implementation of the day off policy quickly diminished over time, the negative interaction is thus more likely to be due to the duty roster.

As described in Section 2.3, these sobriety checkpoints were performed by two different teams of police. In total, 123 police stations were assigned sobriety checkpoints (in one to three locations depending on the police stations, more details on this part of the project are available in Banerjee, Duffo, Keniston, and Singh (2014)). During checkpoints, police were to carry out breath analysis if the drivers appeared drunk. Drunken drivers' vehicles were impounded and the driver made to appear in court to pay a fine and recover the vehicle.

In the precincts of 69 randomly selected police stations the checkpoints were carried out by the regular station staff. In 54 others they were carried out by teams from the district-based "police lines," who were told that good performance on the checkpoints would increase their chances of transfer to a regular posting. Their movements were monitored by GPS trackers in their police vehicles. This incentive was not set up as a tournament, but there was a general (and credible, given the organization and the frequent use of transfer as incentives) promise that performance would matter.²⁴ The experimental design is thus straightforward, and simply involved comparing various measures of performance across police lines and police station checkpoints.

4.1 Data

We monitored police implementation of the sobriety checkpoints by sending surveyors to visit the checkpoint locations on nights when the police were scheduled to be carrying out sobriety checks. Surveyors recorded whether the police turned up at all, and if they did, noted their arrival and departure times from the checkpoint location. If the checkpoint was held, the surveyor counted the number of vehicles stopped, and the number found to be drunk. We present the summary statistics of this data, for the control (regular teams) police stations at the bottom of Table 10. The results show that program implementation, as in the first reforms, was far from complete: in police stations manned by regular teams, only 53% of the scheduled checkpoints were actually conducted, and even then the police arrived on schedule only 40% of the time.²⁵

Court records provide a secondary source of administrative data on the project outcomes. For each police station, the court provided a list of all drunken driving cases heard, including the date of the offense, level of the fine, and breathalyzer reading. These court records reflect the information received by senior officers on all teams' performances, and are net of any cases where police let drunken drivers go (in exchange for bribes or simply out of clemency). Finally, the breathalyzers used at the checkpoints provide a source of data on the actions of the police at the roadblocks. Each time that a breathalyzer is used, it records the date, time, and alcohol level measured. At the end of the project we copied the memory files off the breathalyzers and can use these to examine police conduct. Unfortunately these records are partial, since many of the breathalyzer memories were wiped when their batteries ran out. However, since the police personnel did not know of the breathalyzer memory recording feature, these erasures are as good as random, and the remaining

²⁴Several districts reported to have made these transfers in wake of the sobriety crackdown, although we do not have access to detailed personnel data during this period.

²⁵Data gathered by surveyors was not reported to police management during the crackdown, and never reported at the police station level.

data can still be used.

4.2 Results

The effects of the police line intervention on checkpoint enforcement, presented in Table 10, are clear: the police lines teams perform substantially better on all outcomes. The primary outcome, whether the checkpoint actually occurred on a night when it was supposed to have been held, is shown in column 1 using data collected by the surveyors' observations. Police lines teams were 27.6 percentage points more likely to show up for the checkpoint at all (an increase of around 50% over the control mean). Columns 2 and 3 report whether to police began checking by 7:00pm and continued until 10:00pm respectively (the scheduled times). Again, police lines teams were 29.5 percentage points (73%) more likely to arrive on time and 18.5 percentage points (27%) more likely to continue the roadblock until 10pm. While they were carrying out the checkpoints, they stopped a larger fraction of the passing vehicles (9.7 percentage points more, or 60%). Overall, they caught more than twice as many drunk drivers for any given checkpoint that was in fact conducted (since they also carried out more checkpoints, the total effect is obviously even larger).

Data from the police teams' breathalyzer units provides another insight into their performance. This information has the advantage that it reflects the behavior of the police on all nights, including those when they were not monitored by surveyors. Table 11 presents the result on breathalyzer usage during the periods when the police were assigned to carry out roadblocks for the project. Column 1 shows a simple linear probability regression of whether there was any breathalyzer use at all during the checkpoint period (a proxy whether the police actually travelled to the location), and column 2 shows the intensity of breathalyzer use conditional on any usage at all. Consistent with the results from the surveyors' reports, the breathalyzer data shows much better performance by the police lines teams on all outcomes. On the extensive margin, lines teams use breathalyzers 12 percentage points more frequently than the stations teams and, conditional on the presence of any breathalyzer data on a given night, the lines times tested more than three times as many drivers.²⁶

While breathalyzer tests provide insight into the number of drunken drivers caught, the number actually prosecuted might be substantially fewer if many drivers are let off with a warning or a bribe. Therefore we also gathered data from local courts, one per police station, on the total number of drivers who appeared in court to pay fines for drunken driving. Column 3 of Table 11 shows that over the duration of the project the police lines teams brought more than twice as many drunk drivers to court as the regular station teams. This difference persists even when controlling in column 4 for the number of breathalyzer recordings of drunken drivers, showing that not only were the lines teams more effective at apprehending drunken drivers, they also prosecuted them more effectively. Police lines teams knew that this was exactly the statistic that would be reported to the district level officer evaluating their performance. Since drivers paying fines in court by

²⁶We also tested whether police lines and police station lines acted differently on nights were the surveyors was present. We find more use of the breathalyzer when the monitor is present (though the effect is smaller than that of the police line team), but not differentially for police lines and police station teams.

definition did not pay bribes at the checkpoints, the specific monitoring mechanisms may have both increased effort and reduced corruption.

4.3 Interpretation

Compared to the mixed results of the first set of experiments, and the rather disappointing results of the use of incentives in the previous literature, these results are extremely encouraging. It is possible to build incentives into a government system that apparently has no scope for them, and it seems that these incentives have very large effects on productivity.

However, it is worth keeping in mind that the police lines teams were different from the station teams in multiple respects, not just in their incentives. For one, the teams were selected in different ways—though selection was probably working against us, since the police lines are a punishment posting. Also the police lines teams probably worked less hard during the day and did less multi-tasking (since the police lines are not a very high-intensity posting, while regular police teams still had to handle the rest of the duty at the police station). However, even if incentives per se did not cause the differences in performance, our basic claim still stands, that there are mechanisms available within the police system to substantially boost productivity but that these mechanisms are not implemented even by motivated managers seeking to accomplish fundamentally sound objectives.

5 Conclusion

The experimental results in this paper show that it is possible to affect the behavior of the police in a relatively short period of time, using a simple and affordable set of interventions. Decoy visits were effective in getting the police to register more crimes and treat victims more politely. The promise of transfers out of a less favored assignment—police lines in this case—based on good performance more than doubled the number of drunk drivers brought to court. Training police staff in investigation techniques and public relations skills (soft skills) also increased the satisfaction of crime victims. On the other hand, some other prima facie plausible interventions were not properly implemented and turned out to be ineffective.

What does this tell us about the nature of organizational change? It is striking that two of the three most clearly successful interventions had never been advocated by the many police Reform Commissions or in the discussions we had with the police top brass despite the motivation for reform and support of the top administration. In contrast, the interventions that failed to work were all carefully selected by the police leadership, partly based on the recommendations of various Police Reform Commissions, who among them combine a huge amount of experience and expertise. These senior police officers also did not lack human capital, being selected through an extraordinarily competitive set of exams. Yet they had clearly substantially underappreciated the difficulty of implementing these interventions even with their full backing, suggesting that they did not fully realize the nature of the informal authority enjoyed by the station chiefs.

While it is not possible to rule out the possibility is that bad luck played a role, we see the

relatively poor performance of police force interventions as evidence against the view that the managers know how to translate the general principles of management into solutions that are relevant for their organization. The managers in this case were the elite of the Indian police, with decades of experience and who were motivated enough, among other things, to engage with us to launch this reform program. They had a clear and articulated understanding of the principles that motivated the interventions (“unpredictable work schedules are demoralizing,” “a chance to learn on the job can motivate an employee,” “the customer must be listened to”) and they were recognized as outstanding leaders (both the Director General A.S. Gill and our co-author Nina Singh were subsequently assigned to important positions in the central police). However, none of that could guarantee that they would automatically hone in on all the right interventions. And while they clearly shared the view that incentives, as a general principle, are important, they could not see a way to introduce them within the political and administrative constraints, until we intervened as pseudo-consultants.

What do outsiders, with less institutional knowledge and experience, provide in these situations? While we cannot directly test this question, we suggest three reasons why these insights were unlikely to have arisen within the police institution. First, the idea of introducing individual incentives for subordinate staff was an implicit acknowledgment that the traditional police hierarchy and discipline could be inadequate to obtain satisfactory program implementation. Outsiders with little stake in the existing police power structure were much more willing to consider and advocate interventions that conceded and accommodated the autonomy of middle managers.

Second, the evaluation generated evidence and information not typically available to the police leadership. The police’s own internal monitoring systems, in which subordinates report on their own performance to superiors, typically confirms that implementation is exemplary. Even in the case of the two least successful interventions, official records claimed almost complete compliance. Without external monitoring there would have been no concrete evidence that the basic problem lay not in the design of some of interventions, but rather in the incentives for their execution.

Finally, it seems that the initial underestimation of the importance of incentives and soft skills lay not so much in the lack of information by senior officers, but rather with the view that these programs would not work in that context. (This is similar to the interventions that the management in BEMMR knew about, but did not think would work for them.) In a police system based on the themes of discipline, hierarchy and duty, the idea of individual incentives, particularly financial ones, has become associated with corruption and the dilution of intrinsic motivation and discipline. Similarly, the Rajasthan Police themselves acknowledged that “soft skills” and positive public relations were not part of their institutional heritage. This explains the decision to completely outsource the soft-skills training to consultants, rather than do so through police trainers or even a training-of-trainers approach.

This interpretation may help explain the mixed results in Rasul and Rogger (2013). Perhaps the efforts to implement incentives in Nigerian government bureaucracies were sufficiently poorly designed or implemented to backfire; it does not preclude the possibility that consultants would

have been able to design them well. Meanwhile, the correlation between managers' reported organizational practices and performance in BVR may be partly due to unobserved manager type: relying on incentives is so fundamental in a private setting that admitting not to use them is the sign that there is something really wrong with the manager. The view we end up with falls somewhere in between the "best principle" and "contingencies" view of management. Organizations, both private and public, can be improved by helping managers put into practice some fundamentally sound principles: but the evidence so far suggest that this requires quite careful fine-tuning.

References

- Atkins, D., A. Chaudhry, S. Chaudhry, A. Khandelwal, and E. Verhoogen (2014). Organizational Barriers to Technology Adoption: Evidence from Soccer-Ball Producers in Pakistan. Mimeo, UCLA.
- Bandiera, O., A. Prat, and T. Valletti (2009). Active and passive waste in government spending: evidence from a policy experiment. *American Economic Review* 99(4), 1278–1308.
- Banerjee, A. V., R. Banerji, E. Duflo, R. Glennerster, and S. Khemani (2010). Pitfalls of participatory programs: evidence from a randomized evaluation in education in india. *American Economic Journal: Economic Policy* 2(1), 1–30.
- Banerjee, A. V., E. Duflo, and R. Glennerster (2008). Putting a band aid on a corpse: incentives for nurses in the Indian public health care system. *Journal of the European Economic Association* 6(2-3), 487–500.
- Banerjee, A. V., E. Duflo, D. Keniston, and N. Singh (2014). The efficient deployment of police resources: theory and new evidence from a randomized drunk driving crackdown in India. Working Paper, Massachusetts Institute of Technology.
- Bloom, N., B. Eifert, A. Mahajan, D. McKenzie, and J. Roberts (2013). Does management matter? Evidence from India. *The Quarterly Journal of Economics* 128(1), 1–51.
- Bloom, N., C. Propper, S. Seiler, and J. Van Reenen (2010). The impact of competition on management quality: evidence from public hospitals. Technical Report 16032, National Bureau of Economic Research.
- Bloom, N. and J. Van Reenen (2007). Measuring and explaining management practices across firms and countries. *The Quarterly Journal of Economics* 122(4), 1351–1408.
- Bruce, D. and R. Neild (2005). The police that we want: a handbook for oversight of police in South Africa. Center for the Study of Violence and Reconciliation, Johannesburg, South Africa.
- Caparini, M. and O. Marenin (2005). Crime, insecurity and police reform in post-socialist CEE. *The Journal of Power Institutions in Post-Soviet Societies* 2, 1–12.
- Centre for Media Studies (Ed.) (2005). *Volume 9: Corruption in the Police Department*. Transparency International, New Delhi, India.
- Di Tella, R. and E. Schargrodsy (2003). The role of wages and auditing during a crackdown on corruption in the city of Buenos Aires. *Journal of law and economics* 46(1), 269–292.
- Hanna, R. and I. Dhaliwal (2013). Making a deal with the devil: experimental evidence on bureaucratic reform in India. Working Paper, Harvard Kennedy School of Government.
- Holmstrom, B. and P. Milgrom (1991). Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization* 7, 24.

- Iyer, L., A. Mani, P. Mishra, and P. Topalova (2012). The power of political voice: women's political representation and crime in India. *American Economic Journal: Applied Economics* 4(4), 165–93.
- Kling, J. R., J. B. Liebman, and L. F. Katz (2007). Experimental analysis of neighborhood effects. *Econometrica* 75(1), 83–119.
- Kolsky, E. (2010). *Colonial Justice in British India*. Cambridge, UK: Cambridge University Press.
- National Police Commission (1980). National police commission report. Volumes 1-8.
- Olken, B. A. (2007). Monitoring corruption: evidence from a field experiment in Indonesia. *Journal of Political Economy* 115(2), 200–249.
- Opolot, J. S. (1992). The resilience of the British colonial police legacies in East Africa, Southern Africa, and West Africa. *Police Studies: International Review of Police Development* 15, 90.
- Prendergast, C. (2001, December). Selection and oversight in the public sector, with the Los Angeles Police department as an example. Working Paper 8664, National Bureau of Economic Research.
- Rasul, I. and D. Rogger (2013, October). Management of bureaucrats and public service delivery: evidence from the Nigerian Civil Service. Working Paper, University College London.
- Sidhu, H. S. (2003). Management of Reforms in Police - A Study at District Level. Master's thesis, Guru Nanak Dev University, Amritsar, Punjab.
- The Hindu*. Rebuild India's Police Forces. Sept. 9, 2011.
- Varona, G. (2010). Politics and policing in the Philippines: challenges to police reform. *Flinders Journal of History and Politics* 26, 101–125.
- Wade, R. (1985). The market for public office: why the Indian state is not better at development. *World Development* 13(4), 467–497.

Tables - First Intervention

Table 1: Program Design

| Group | Number of Police Stations | Program elements | | | |
|---|---------------------------|--------------------|------------------------------|-------------|-----------------------------|
| | | Community observer | Weekly off and duty rotation | No Transfer | Training |
| 1. All Interventions | 35 | Yes | Yes | Yes | Between 0 and 100% of staff |
| 2. Community Observer + No Transfer | 25 | Yes | No | Yes | Between 0 and 100% of staff |
| 3. Weekly Off / Duty Rotation + No Transfer | 25 | No | Yes | Yes | Between 0 and 100% of staff |
| 4. No Transfer | 40 | No | No | Yes | Between 0 and 100% of staff |
| 5. Control | 25 | No | No | No | Between 0 and 100% of staff |
| 6. Pure Control | 12 | No | No | No | None |

Table 2: Summary Statistics

| | (1) | (2) | (3) |
|------------------------------------|-------------------|-------------------|--------------------|
| | Control | Any Treatment | Difference |
| A. Opinion Survey Respondents | | | |
| Crime victim in household | 0.0547 (0.228) | 0.0610 (0.239) | 0.0063 (0.015) |
| Ever met police | 0.0844 (0.278) | 0.0915 (0.288) | 0.0071 (0.018) |
| Ever visited police station | 0.0943 (0.293) | 0.0846 (0.278) | -0.0097 (0.017) |
| Ever arrested | 0.0223 (0.148) | 0.0133 (0.115) | -0.0091 (0.010) |
| B. Decoy Surveys | | | |
| Case not filed | 0.520 (0.510) | 0.616 (0.488) | 0.0960 (0.110) |
| Police very polite to decoy | 0.200 (0.408) | 0.144 (0.353) | -0.0560 (0.087) |
| Police rude to decoy | 0.120 (0.332) | 0.192 (0.395) | 0.0720 (0.074) |
| C. Victim Satisfaction | | | |
| Victim satisfied or very satisfied | 0.348 (0.487) | 0.356 (0.481) | 0.008 (0.118) |
| Case registered with police | 0.311 (0.467) | 0.301 (0.460) | -0.010 (0.066) |
| D. Case Review | | | |
| Documentation Quality | 6.409 (1.797) | 6.133 (1.931) | -0.276 (0.300) |
| Field investigation quality | 6.537 (1.636) | 6.160 (1.823) | -0.378 (0.272) |
| Whether used scientific methods | 0.179 (0.386) | 0.202 (0.402) | 0.023 (0.053) |

Statistics reported for baseline survey, except box B which reports data from the first decoy visit. Columns 1 and 2 present means with standard errors in parentheses. Column 3 reports standard errors of the difference in parentheses, clustered at police station level for boxes A, C, and D.

Table 3: Police Staff Transfers by Rank

| | (1) Treatment PS (No Transfer) | (2) Control PS | (3) Difference |
|------------------|---------------------------------------|-------------------|--------------------|
| Inspector | 48.6% | 63.6% | 15.1% (17.16) |
| Sub. Inspector | 47.5% | 68.2% | 20.7%** (8.28) |
| Asst. Sub. Insp. | 28.0% | 29.8% | 1.8% (6.16) |
| Head Constable | 28.3% | 34.5% | 6.2% (7.00) |
| Constable | 16.6% | 30.0% | 13.4%** (4.49) |
| All Ranks | 21.1% | 32.8% | 11.7%*** (3.59) |

Standard errors in parentheses clustered by police station.

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

Table 4: Implementation of Interventions

| | Round 1 | Round 2 | Round 3 | Round 4 | Round 5 |
|---|---------|---------|---------|---------|---------|
| A. Weekly off: Days since last day off | | | | | |
| No weekly off | 29 | 32 | 28 | 31 | 30 |
| Weekly off | 21 | 22 | 25 | 26 | 32 |
| Difference: | 8 | 11 | 3 | 5 | -2 |
| B. Number of days constable has been doing current duty | | | | | |
| No duty rotation | 4.05 | 3.97 | 6.27 | 4.17 | 2.70 |
| Duty rotation | 5.58 | 4.51 | 5.33 | 4.50 | 2.78 |
| Difference: | 1.53 | 0.54 | -0.94 | 0.33 | 0.08 |
| C. Whether community observer is present | | | | | |
| Community observer station | 20% | 8% | 3% | 12% | 8% |

Surveyors did not record community observer presence in stations not assigned to that intervention.

Table 5: Decoy Survey Outcomes

| | (1) Case registered | (2) Case registered | (3) Case registered | (4) Police were very polite | (5) Police were very polite | (6) Police were very polite |
|--------------------------------------|---------------------------|---------------------------|---------------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| All interventions | 0.0285 (0.0438) | 0.0303 (0.0438) | | -0.0234 (0.0283) | -0.0216 (0.0281) | |
| No transfer | -0.0558 (0.0457) | -0.0575 (0.0447) | | -0.0341 (0.0362) | -0.0315 (0.0366) | |
| Duty rotation, weekly off | 0.0494 (0.0479) | 0.0533 (0.0483) | | 0.0297 (0.0308) | 0.0326 (0.0297) | |
| Community observer | 0.0766 (0.0579) | 0.0801 (0.0586) | | -0.00555 (0.0303) | 0.0000914 (0.0293) | |
| Percentage staff trained | 0.107 (0.0829) | 0.108 (0.0837) | 0.233** (0.116) | 0.105* (0.0543) | 0.101* (0.0539) | 0.196** (0.0869) |
| Number of decoy visits | 0.174*** (0.0577) | 0.173*** (0.0579) | 0.165** (0.0761) | 0.0700* (0.0367) | 0.0746** (0.0366) | 0.102** (0.0412) |
| Observations | 788 | 788 | 788 | 788 | 788 | 788 |
| R^2 | 0.181 | 0.185 | 0.141 | 0.096 | 0.122 | 0.112 |
| District FE | Yes | Yes | No | Yes | Yes | No |
| Station FE | No | No | Yes | No | No | Yes |
| Controls for police suspicions | No | Yes | Yes | No | Yes | Yes |
| Control mean at first decoy visit | 0.480 | 0.480 | 0.480 | 0.200 | 0.200 | 0.200 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Standard errors in parentheses clustered by police station.

Table 6: Program Effects on Victim Satisfaction

| | (1) Endline only | (2) Station FE | (3) Station FE |
|------------------------------------|------------------------|---------------------|-----------------------|
| All interventions | -0.150* (0.0777) | -0.290** (0.136) | -0.257** (0.118) |
| No transfer | 0.144 (0.0872) | 0.281** (0.123) | 0.243** (0.104) |
| Duty rotation, weekly off | -0.0614 (0.0959) | -0.0770 (0.171) | -0.0886 (0.164) |
| Community observer | 0.225* (0.116) | -0.279* (0.157) | -0.301* (0.155) |
| In study | -0.131 (0.127) | | |
| Percentage staff trained | 0.156** (0.0767) | 0.193** (0.0843) | 0.131* (0.0763) |
| Registered case | | | -0.00456 (0.0387) |
| Asked for bribes | | | -0.150*** (0.0566) |
| Made arrest | | | 0.166*** (0.0618) |
| Recovered property | | | 0.679*** (0.185) |
| Continued investigation actions | | | 0.197*** (0.0485) |
| Recontacted victim | | | 0.0877 (0.0622) |
| Observations | 269 | 571 | 571 |
| R^2 | 0.290 | 0.222 | 0.308 |
| District FE | Yes | No | No |
| Station FE | No | Yes | Yes |
| Crime and victim controls | Yes | Yes | Yes |
| Baseline/control mean | 0.246 | 0.299 | 0.299 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Standard errors in parentheses clustered by police station. Victim characteristics: age and gender of the respondent, education and dummies for the occupation of the head of household, indicators for caste or Muslim religion, and indicators for motorcycle ownership.

Table 7: Effect of Training on Police Actions

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|-----------------------------|--------------------|---------------------|-------------------|-----------------------|--|-----------------------|---|
| | Registered case | Asked for bribes | Made arrest | Recovered property | Continued investiga- tion actions | Recontacted victim | Monthly number of cases registered |
| Percentage staff trained | 0.176 (0.109) | 0.183 (0.113) | 0.137 (0.0868) | 0.0377* (0.0208) | 0.160 (0.101) | 0.0683 (0.0773) | 1.684* (0.976) |
| Observations | 571 | 571 | 571 | 571 | 571 | 571 | 3852 |
| R^2 | 0.139 | 0.107 | 0.119 | 0.137 | 0.087 | 0.089 | 0.034 |
| Station FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Crime and victim controls | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Baseline/control mean | 0.590 | 0.160 | 0.135 | 0.0123 | 0.246 | 0.135 | 17.91 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Standard errors in parentheses clustered by police station. Victim characteristics: age and gender of the respondent, education and dummies for the occupation of the head of household, indicators for caste or Muslim religion, and indicators for motorcycle ownership.

Table 8: Case Review Outcomes

| | (1) Quality of field investigation | (2) Quality of documentation | (3) Used scientific techniques |
|---|--|------------------------------------|--------------------------------------|
| Investigating officer trained | 0.152 (0.205) | 0.203 (0.308) | 0.0291 (0.0372) |
| Fraction officers trained in station | -0.0713 (0.310) | -0.251 (0.396) | -0.0000554 (0.0568) |
| Observations | 937 | 924 | 981 |
| R^2 | 0.286 | 0.327 | 0.194 |
| Station FE | Yes | Yes | Yes |
| Baseline/pure control mean | 6.199 | 6.153 | 0.195 |

Dependent variables in columns 1-2 are ratings given by retired officers to case files on a scale of 1-10. Dependent variable in column 3 is an indicator for whether the case review indicated that scientific investigation was used. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard errors in parentheses clustered by police station.

Table 9: Public Opinion

| | (1) | (2) |
|--|-----------------|-------------------|
| | Only Endline | Police Station FE |
| A. Police Responsiveness to Citizens | | |
| All interventions | 0.016 (.04) | -0.011 (.08) |
| No transfer | -0.103** (.05) | -0.014 (.08) |
| Weekly off | 0.052 (.05) | 0.133 (.08) |
| Community observer | 0.02 (.05) | 0.08 (.09) |
| Percentage staff trained | -0.03 (.04) | -0.054 (.07) |
| Not pure control | 0.112** (.05) | |
| B. Reducing Police Corruption | | |
| All interventions | 0.067 (.05) | 0.05 (.07) |
| No transfer | 0.03 (.05) | 0.02 (.08) |
| Weekly off | -0.058 (.05) | -0.048 (.07) |
| Community observer | -0.007 (.05) | 0.043 (.07) |
| Percentage staff trained | 0.01 (.05) | 0.016 (.07) |
| Not pure control | 0.078 (.07) | |
| C. Lack of Fear of Police | | |
| All interventions | 0.012 (.06) | -0.033 (.09) |
| No transfer | 0.132** (.06) | 0.242** (.10) |
| Weekly off | -0.175*** (.05) | -0.216*** (.07) |
| Community observer | -0.026 (.06) | -0.064 (.09) |
| Percentage staff trained | -0.053 (.06) | -0.003 (.08) |
| Not pure control | 0.075 (.08) | |
| D. Increase Police Resources | | |
| All interventions | 0.036 (.05) | 0.017 (.05) |
| No transfer | -0.047 (.06) | -0.034 (.06) |
| Weekly off | -0.017 (.05) | -0.037 (.05) |
| Community observer | -0.03 (.06) | -0.031 (.06) |
| Percentage staff trained | -0.027 (.05) | -0.032 (.05) |
| Not pure control | 0.13 * (.08) | |
| Police Station Fixed Effects | No | Yes |
| HH characteristics, victim/arrested, opinion source | Yes | Yes |

Notes 1. Standard errors in parentheses clustered at the police station level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ 2. Treatment variables defined as described in Section 5. 3. Control variables are as follows: HH characteristics: age and gender of the respondent, education and dummies for the occupation of the head of household, indicators for caste or Muslim religion, and indicators for motorcycle or vehicle ownership. Victim/arrested: Indicators for whether the respondent was arrested, knows someone who has been arrested, or is in a household with a crime victim. Opinion source: Indicators for the self-reported source of the victim's information on the police, including non-mutually exclusive categories such as TV news, word of mouth, personal experience, etc.

Tables - Drunken Driving Checkpoint Enforcement

Table 10: Police station vs. police lines teams

| | All Nights | Conditional on checkpoint occurring | | | |
|--------------------|---|-------------------------------------|--|--|-----------------------------------|
| | (1) Whether roadblock occurred | (2) Police arrive on time | (3) Police stay until 10:00pm | (4) Percentage vehicles stopped | (5) Drunk drivers caught |
| Police lines teams | 0.276*** (0.0392) | 0.296*** (0.0512) | 0.185*** (0.0417) | 9.718*** (2.130) | 1.897*** (0.298) |
| Observations | 1196 | 729 | 736 | 759 | 759 |
| R^2 | 0.097 | 0.097 | 0.049 | 0.098 | 0.160 |
| District FE | Yes | Yes | Yes | Yes | Yes |
| Station teams mean | 0.528 | 0.402 | 0.663 | 16.09 | 1.140 |

Standard errors in parentheses clustered by district.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Breathalyzer and Prosecution Performance

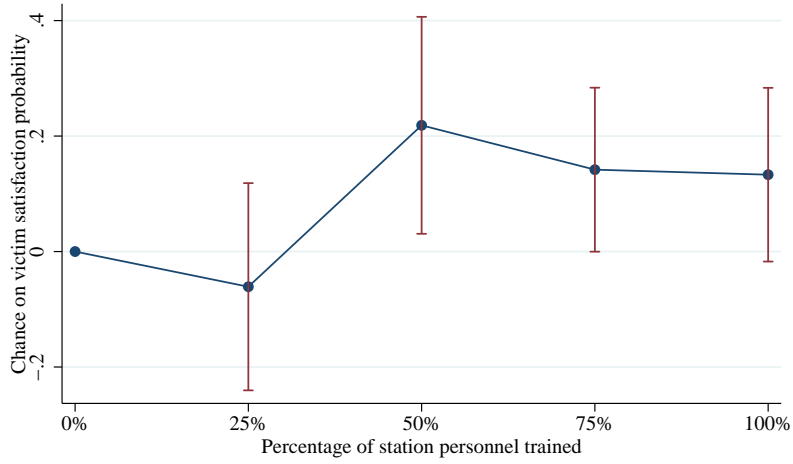
| | (1) Any breatha- lyzer data | (2) # Breath- alyzer tests if used | (3) Drunk drivers in court | (4) Drunk drivers in court |
|---------------------------------|---|---|-------------------------------------|-------------------------------------|
| Police lines teams | 0.120*** (0.0451) | 3.714*** (0.593) | 22.07*** (5.705) | 17.69*** (2.769) |
| Drunken drivers breathalyzed | | | | 0.219* (0.0986) |
| Observations | 2703 | 1023 | 123 | 123 |
| R^2 | 0.021 | 0.145 | 0.251 | 0.393 |
| District FE | Yes | Yes | Yes | Yes |
| Station teams mean | 0.599 | 1.107 | 15.33 | 15.33 |

Standard errors in parentheses clustered by district.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

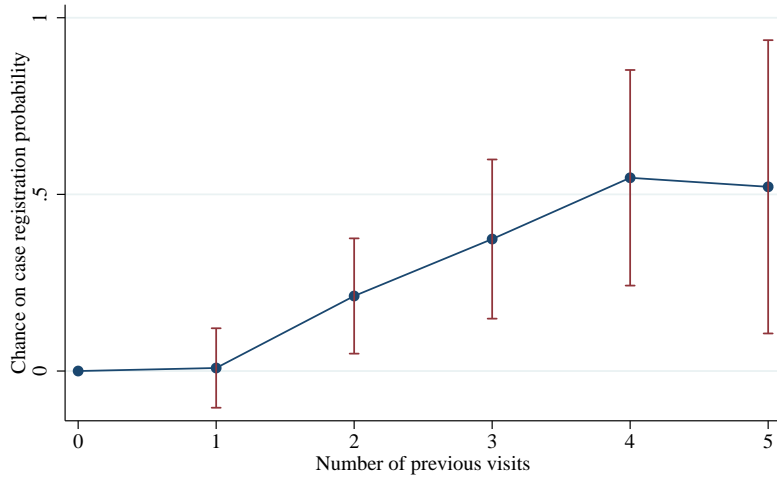
Figures

Figure 1: Effect of training on victim satisfaction



Graph points are coefficients from a regression of victim satisfaction on percentage of staff trained, including police station fixed effects and controls for crime and victim characteristics. Bars show 90% confidence interval.

Figure 2: Effect of decoy visits on case registration



Graph points are coefficients from a regression of case registered dummy on prior decoy visits, including district fixed effects and controls for case type. Bars show 90% confidence interval.

For Online Publication Appendix

Table A1: Attrition in the Police Survey

| | | |
|----------------------------|----------|---------------|
| No transfer | 0.325*** | [0.276,0.373] |
| No transfer/Weekly off | 0.385*** | [0.306,0.463] |
| No transfer/Comm. Observer | 0.278*** | [0.203,0.353] |
| All Interventions | 0.383*** | [0.330,0.437] |
| Control | 0.447*** | [0.377,0.517] |
| Observations | 1548 | |

95% confidence intervals in brackets

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2: Program Effects on Crime

| | (1) Endline only | (2) Endline only | (3) Station FE | (4) Station FE |
|-----------------------------------|------------------------|------------------------|----------------------|----------------------|
| All interventions | -0.0112 (0.00822) | -0.00848 (0.00825) | -0.00645 (0.0132) | -0.00876 (0.0131) |
| No transfer | 0.0133 (0.00849) | 0.0118 (0.00810) | 0.00776 (0.0119) | 0.00721 (0.0121) |
| Duty rotation, weekly off | 0.000763 (0.00911) | 0.00124 (0.00941) | 0.00464 (0.0151) | 0.00119 (0.0160) |
| Community observer | -0.0155* (0.00842) | -0.0138 (0.00868) | -0.00730 (0.0130) | -0.00872 (0.0135) |
| Percentage staff trained | 0.0105 (0.00847) | 0.00990 (0.00855) | 0.0156 (0.0119) | 0.0139 (0.0127) |
| In study | -0.0124 (0.00990) | -0.0119 (0.00975) | | |
| Observations | 15594 | 15550 | 22771 | 21900 |
| R^2 | 0.021 | 0.028 | 0.001 | 0.008 |
| District FE | Yes | Yes | No | No |
| Station FE | No | No | Yes | Yes |
| Victim characteristic controls | No | Yes | No | Yes |
| Baseline/control mean | 0.0824 | 0.0823 | 0.0776 | 0.0786 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Standard errors in parentheses clustered by police station.

Table A3: Program Effects on Police Awareness of Victims' Crimes

| | (1) Endline only | (2) Station FE |
|------------------------------|------------------------|---------------------|
| All interventions | 0.0266 (0.0407) | -0.0346 (0.0518) |
| No transfer | -0.0400 (0.0409) | 0.111 (0.0734) |
| Duty rotation, weekly off | 0.0553 (0.0553) | -0.0264 (0.0610) |
| Community observer | -0.0501 (0.0424) | -0.0691 (0.0730) |
| In study | 0.00937 (0.0647) | |
| Percentage staff trained | -0.0541 (0.0463) | -0.0442 (0.0449) |
| Prior decoy visits | -0.00481 (0.0171) | -0.0145 (0.0143) |
| Observations | 1567 | 2062 |
| R^2 | 0.272 | 0.245 |
| District FE | Yes | Yes |
| Station FE | No | Yes |
| Crime and victim controls | Yes | Yes |
| Date of crime controls | Yes | Yes |
| Baseline/control mean | 0.271 | 0.277 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Standard errors in parentheses clustered by police station.

Victim characteristics: age and gender of the respondent, education and dummies for the occupation of the head of household, indicators for caste or Muslim religion, and indicators for motorcycle ownership. Date of crime controls include dummies for month of crime.

Table A4: Decoy Intervention Effects on Victim Satisfaction

| | (1) | (2) | (3) | (4) |
|------------------------------|---------------------|--------------------|----------------------|---------------------|
| | District | District | District | District |
| | FE | FE | FE | FE |
| All interventions | 0.148 (0.193) | -0.0823 (0.334) | | |
| No transfer | -0.477** (0.217) | 0.272 (0.451) | | |
| Duty rotation, weekly off | 0.105 (0.116) | 0.00687 (0.398) | | |
| Community observer | -0.205 (0.247) | -0.0635 (0.390) | | |
| Percentage staff trained | 0.200 (0.188) | 0.179 (0.807) | | |
| Prior decoy visits | 0.130 (0.103) | 0.254 (0.349) | 0.174 (0.104) | 0.257 (0.185) |
| In study | -1.335** (0.567) | -2.699 (1.590) | -1.558*** (0.491) | -2.344** (0.853) |
| Observations | 51 | 51 | 51 | 51 |
| R^2 | 0.602 | 0.880 | 0.495 | 0.863 |
| District FE | Yes | Yes | Yes | Yes |
| Crime and victim controls | No | Yes | No | Yes |
| Date of crime controls | Yes | Yes | Yes | Yes |
| Baseline/control mean | 0.250 | 0.250 | 0.250 | 0.250 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Standard errors in parentheses clustered by police station.

Sample limited to males aged 20-50 who reported crimes of theft, burglary, or sexual harassment to the police themselves.

Victim characteristics: age and gender of the respondent, education and dummies for the occupation of the head of household, indicators for caste or Muslim religion, and indicators for motorcycle ownership. Date of crime controls include dummies for month of crime.

Table A5: Program Effects on Fear of Police

| | (1) Open- ended police opinion question elicits "fear" | (2) Whether citizens like respondent fear police | (3) Whether law abiding citizens fear police | (4) Open- ended police opinion question elicits "fear" | (5) Whether citizens like respondent fear police | (6) Whether law abiding citizens fear police |
|--|---|---|---|---|---|---|
| All interventions | 0.0175 (0.0427) | -0.0235 (0.0551) | -0.0121 (0.0630) | 0.00231 (0.0435) | -0.0290 (0.0570) | -0.0238 (0.0596) |
| No transfer | 0.0106 (0.0497) | 0.0699 (0.0664) | 0.208*** (0.0658) | 0.0321 (0.0516) | 0.100 (0.0657) | 0.212*** (0.0666) |
| Duty rotation, weekly off | -0.00671 (0.0480) | -0.0886 (0.0555) | -0.198*** (0.0585) | -0.0176 (0.0484) | -0.0814 (0.0549) | -0.214*** (0.0579) |
| Community observer | 0.0313 (0.0527) | -0.0989* (0.0554) | -0.00654 (0.0496) | 0.0269 (0.0530) | -0.0974* (0.0575) | -0.0383 (0.0504) |
| Percentage staff trained | 0.00716 (0.0442) | 0.0347 (0.0527) | -0.0463 (0.0534) | 0.00445 (0.0447) | 0.0565 (0.0532) | -0.0693 (0.0518) |
| Observations | 7343 | 7602 | 7397 | 6731 | 6994 | 6783 |
| R^2 | 0.002 | 0.005 | 0.014 | 0.034 | 0.029 | 0.026 |
| Station FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Ctrls for HH chars., victim/arrested, opinion source | No | No | No | Yes | Yes | Yes |
| Baseline/control mean | 0.853 | 0.467 | 0.371 | 0.844 | 0.447 | 0.363 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ Standard errors in parentheses clustered by police station.

Control variable details listed in notes of Table 6