Forthcoming in *Review of Development Economics*

Improving Police Integrity in Uganda:
Impact Assessment of the Police Accountability and Reform Project[*]

Natascha Wagner[a,c], Wil Hout[a], Rose Namara[b]

[a] International Institute of Social Studies of Erasmus University Rotterdam, Kortenaerkade 12, 2518 AX Den Haag, Netherlands

[b] Uganda Management Institute, Kampala, Uganda

[c] Corresponding author: wagner@iss.nl, phone: + 31 70 426 05 74

Abstract

Uganda and in particular the Ugandan police are perceived as highly corrupt. To address the integrity of police officers, an intervention called the Police Accountability and Reform Project (PARP) was implemented in selected police districts between 2010 and early 2013. This paper studies the impact of PARP for a sample of 600 police officers that were interviewed about police integrity by means of twelve hypothetical vignette cases depicting context-specific, undesirable behaviour of varying degrees of severity. The assessments of the cases by the police officers are analysed using propensity score matching, inverse probability weighting and seemingly unrelated regression techniques. We show that self-selection of police officers into the program is unlikely to drive the results. The results suggest that officers participating in PARP activities (i) judge the presented cases of misconduct more severely, (ii) are more inclined to report misconduct and (iii) also expect their colleagues to judge misbehaviour at the level of the police more critically although the latter two coefficient estimates are smaller in size. This suggests that PARP activities have affected the perception of police officers but only encouraged them moderately to actually take action against bad practices.

Keywords: Integrity, Corruption, Police, Human Rights Training, Propensity Score Matching, Police Accountability and Reform Project, Uganda

JEL Classification Codes: D73, H40, O12

---

1. Introduction

Integrity and accountability of public servants are key to discussions on 'good governance' or 'good government' (Holmberg and Rothstein, 2012). Integrity is normally understood as 'the quality of acting in accordance with relevant moral values, norms, and rules' and can be a quality of individuals and organisations (Lasthuizen et al., 2011: 387). Corruption, generically defined as 'the abuse of entrusted power for private gain' (Transparency International, 2016a), is a prominent aspect of public integrity, because there is almost universal agreement that this signals a serious governance quality deficit. The fight against corruption is not only seen as important for public sector performance, but also for development more generally and is therefore included among the Sustainable Development Goals (Rose-Ackerman and Palifka, 2016: 5).

Traditionally, the police has received much attention in discussions about integrity because of its central role among the instruments of the state, in particular its 'strong arm'. The functioning of the police, and more specifically the curbing of malpractice of its officers, is a measure of governance quality because of the centrality of the rule of law and law enforcement (Rose-Ackerman and Palifka, 2016: 81-82).

Previous research on police integrity has focused on (un)acceptable behaviour (Klockars et al., 2000, 2006; Kutnjak Ivković, 2005a, 2005b), as well as on the impact of organisational characteristics and the external environment on police officers' attitudes (Chappell and Piquero, 2004; Kutnjak Ivković and O'Connor Shelley, 2010; Kutnjak Ivković and Sauerman, 2013; Kutnjak Ivković and Haberfeld, 2015). Recent literature on the topic has also zoomed in on the extent to which policing reforms may reduce police misconduct (Collins et al., 2016) next to focusing on how corrupt officials respond to accountability mechanisms that increase the moral costs of misconduct (Olken and Pande, 2012).

The current paper contributes to the literature on public integrity and policing by analysing a specific intervention, the Police Accountability and Reform Project (PARP), implemented in Uganda between 2010 and 2013. PARP provided training to police officers and expanded contact with social stakeholders, thus aiming to influence the attitudes of Ugandan police officers on corruption and associated integrity violations.

A main challenge of the crime and justice literature is to identify how integrity-enhancing measures impact on the behaviour of public servants. Our attempt to identify causal effects rests on a quasi-experimental design. In line with earlier research on police integrity, and building on the approach introduced by Klockars et al. (2000), we do not ask police officers about their *own* policing practices and those of others but present them with hypothetical cases that they are asked to assess. Propensity score matching is applied to assess the impact of PARP. We find systematic observable differences in police integrity between PARP participants and non-participants, which suggest that participants (i) judge the presented cases of misconduct more severely, (ii) are more inclined to report misconduct and (iii) expect their colleagues to judge misbehaviour at the level of the police more critically. The results are most pronounced for normative judgements of case severity; they are reduced when it comes to actual reporting of misconduct. Thus, PARP was successful in diffusing knowledge about proper policing and human rights, which is demonstrated by the finding that more severe cases of integrity violations are judged more rigorously.

The remainder of this paper is organized as follows. In section 2, we situate our study in the literature on police integrity. Section 3 gives a brief introduction to the Ugandan national police, while PARP is introduced in section 4. Sampling and survey are described in section 5 and descriptive statistics presented in section 6. The empirical model is outlined in section 7. The results are discussed in section 8. Section 9 concludes.

2. Situating the study

Research on the quality of policing has focused on (degrees of) integrity to get around the problem of studying actual acts of corruption and breaches of integrity. Building on the work of scholars such as Klockars et al. (2000; 2006) and Kutnjak Ivković (2005a), our study includes forms of corruption such as bribery but further extends to other integrity issues such as the maltreatment of suspects. We use Kutnjak Ivković's (2005a: 16) definition of police corruption as 'an action or omission, a promise of action or omission, or an attempted action or omission, committed by a police officer or a group of police officers, characterized by the police officer's misuse of the official position, motivated in significant part by the achievement of personal gain'. Police integrity is understood as 'the normative inclination among police to resist temptations to abuse the rights and privileges of their occupation' (Klockars et al., 2006: 1).

Klockars et al. (2000; 2006) have argued that research should focus on police integrity rather than corruption, because the so-called 'administrative/individual approach', which aims at measuring the level of corrupt *behaviour*, encounters 'enormous…obstacles' (Klockars et al., 2000: 3). The problems are illustrated in research that has tried to measure corruption by recording experiences in a survey (Tankebe, 2010), assessing activities through analysing written records (McMillan and Zoido, 2004) or accompanying corrupt individuals and observing the payment of bribes (Olken and Barron, 2009). An 'organizational/occupational approach' lends itself to asking 'questions of *fact* and *opinion* that can be explored directly, without arousing the resistance that direct inquiries about corrupt *behaviour* are likely to provoke' (Klockars et al., 2000: 3).

While most earlier studies have mainly described the *level* of police integrity and misconduct, contemporary studies seek explanations of differences among groups of officers at the meso- and macro-level related to characteristics of police precincts, gender and race differences, and attitudes (Gottschalk, 2010; Hickman et al., 2016a; 2016b). Studies on non-

Western countries have mainly focused on measuring police integrity within specific types of countries or regions, specific regimes and institutional cultures (Kutnjak Ivković, 2015: 21-27). These studies aimed at establishing a relationship of a variety of factors with police integrity. Such factors include differences between supervisors and line officers in Bosnia and Herzegovina, Croatia, the Czech Republic, Hungary, Slovenia and South Africa (Kutnjak Ivković and Sauerman, 2013), individual characteristics of officers in South Korea and Turkey (Cetinkaya 2010), gender differences in Romania (Andreescu et al., 2012), characteristics of organisational culture of police agencies in Turkey (Kucukuysal, 2008) and differences between urban and rural areas in Armenia (Kutnjak Ivković and Khechumyan, 2014).

Kutnjak Ivković (2015: 18-27) has presented a comprehensive overview of the variety of research designs applied in the studies on police integrity in the tradition of Klockars et al. (2000; 2006); her overview has not, however, identified a single study that assesses the *impact of an intervention* that aims at enhancing police integrity. Our study focuses on a police reform project in Uganda. Contrary to earlier studies we compare participants benefitting from the intervention with non-participants, making use of information about individual police officers, their police station and district characteristics to control for confounding factors stemming from these levels.

We build on the method that was pioneered by Klockars et al. (2000, 2006) and Kutnjak Ivković (2005a; 2005b). The approach is based on presenting a series of 'vignette cases' – short hypothetical descriptions of forms of police misconduct – to police officers and registering their responses on the seriousness of the behaviour that is described and their willingness to report police officers who are responsible for the misconduct. The cases range from small-scale bribery to traffic offences, and from robbery to murder. Our survey questions ask for an assessment of the cases based on officers' judgments about good policing and perceived best practices. Thus, instead of framing the survey as an assessment of police corruption it was

presented as a review of the challenges that police officers face. The advantage of the approach is the uniformity and resulting comparability of case assessments across individual officers. Since all officers assess identical scenarios, we can directly compare their judgments within each of the twelve cases.

By focusing on Uganda, this study aims to enhance our knowledge about police integrity in the non-Western world. In recent years, the attention for policing in developing countries has increased (Tankebe, 2010; Banerjee et al., 2012, 2014; Kutnjak Ivković and Haberfeld, 2015; Collins et al., 2016) but there has been limited research on Africa or lower-income countries. Kutnjak Ivković's (2015: 18-27) overview indicates that the vignette-based analysis has been used in 23 countries, including the US where the approach was developed and most studies were conducted. Of the 23 countries, only eight are outside Europe and North America, and only two are from Sub-Saharan Africa.[†] Studies on Eritrea and Pakistan have concentrated on measuring police integrity (Kutnjak Ivković, 2015: 22-23). South Africa has received more attention from researchers resulting in studies about police integrity in the Johannesburg area and at the national level; there is also a study about the code of silence in the South African police force (Kutnjak Ivković, 2015: 23-24).[‡] Given the dearth of research done in Sub-Saharan Africa, it seems relevant to obtain further evidence from countries on the continent. As Kutnjak Ivković and Haberfeld (2015: 365) have observed, 'the contours of police integrity vary across the world. What is acceptable and tolerated in one country or one police agency may not be acceptable at all in another, and may be disciplined severely.'

[†] The non-Western countries include two low- and lower-middle-income countries (Eritrea and Pakistan), five upper-middle-income countries (Armenia, Malaysia, South Africa, Turkey and Thailand) and one high-income country (South Korea).
[‡] For a more detailed discussion, the overview by Kutnjak Ivković (2015) can be consulted.

3. Background: The Uganda National Police

The 2015 Corruption Perceptions Index, which measures perceived levels of public sector corruption worldwide, included Uganda in the top quintile of most corrupt countries (Transparency International, 2016b). The Ugandan police is regarded as particularly corrupt (Wambua, 2015; Basheka, 2013; Transparency International-Kenya, 2013). Surveys of the Commonwealth Human Rights Initiative (2006a) demonstrated that a majority of Ugandan citizens perceive the police as the most corrupt institution in the country.

The Ugandan police force, which was institutionalized in 1906 (Uganda Police Force, 2007), is divided functionally into 20 directorates based on tasks and geographically into regional and district units (Uganda Police, 2015). In the early 2000s, Uganda had fewer than 15,000 police officers (Commonwealth Human Rights Initiative, 2006b). At the end of 2014, the Inspector General announced the expansion of the police to 65,000 officers (Kakamwa, 2014).

In 2013, the crime rate was 273 per 100,000 Ugandans with public sector crime investigations being on the rise. The Ugandan police reported 413 investigations in 2013, compared to 214 in 2012. The Ugandan Police (2013) mentioned 19 cases in which police officers were under investigation of suspected crimes.

4. The intervention: The Police Accountability and Reform Project

Between 2007 and early 2013, the Police Accountability and Reform Project (PARP) was implemented by the civil society organization Human Rights Network Uganda (HURINET-U), with financial support from the Dutch Embassy in Uganda.[§] The project was realized against the background that the police force in Uganda is widely perceived as a partisan force. The main concerns were brutality, lack of respect for human rights, abuse of power and corruption.

[§] The Embassy supported the project with €230,000 during the first phase (2007-2010) and €260,000 during the second phase (2010-2013). This article focuses on the activities implemented between 2010 and 2013.

The project objectives revolved around improving accountability and democratic governance within the police, in close cooperation with civil society organizations (CSOs). The assumption was that police integrity would be enhanced when external accountability mechanisms get established, as they strengthen local democratic control, citizen and media involvement (Newburn, 2015). The project brought together the police and civil society to foster exchange and implement external control. PARP objectives were to: (i) create stronger civilian oversight of the police, (ii) establish public safety and security networks based on the premise of a shared responsibility between the police and the public, (iii) enhance civil society's contribution to the police review process, and (iv) contribute to a public order management system that protects the rights and freedoms of Ugandans to assembly (HURINET-U, 2013).

The Dutch Embassy funded PARP because of HURINET-U's longstanding relationship with the Ugandan police force. PARP was delivered by HURINET-U in collaboration with seven other civil society organizations in the form of advocacy work, workshops involving civil society representatives, the media and the police, field visits, information campaigns and radio broadcasts (HURINET-U, 2013). Next to disseminating the findings of the government's police review process and publishing an analysis of the Public Management Order Bill 2010, HURINET-U organized various targeted activities during the second phase of the PARP project. Activities included five one-day CSO-police, three media-police and two student-police dialogues, around 40 work sessions involving the police and the project team, and field missions to document the role of the army and police during elections, in particular the heavily contested general elections of 2011 (Perrot, 2014). The aim of the dialogues was to discuss the abuse of power and brutality by the police and find ways to overcome misbehaviour by giving the public a role in general oversight. Further, HURINET-U distributed 700 copies of the police accountability newsletter 'Police Watch', organized visits of more than 850 citizens to police stations in four districts under the motto 'Taking the police to the people: Enhancing

accountability' and arranged for 15 radio talk shows in selected districts, usually in the local language and limited in ambit. The police accountability newsletter and the station visits aimed to increase transparency and reduce the potential for corruption. HURINET-U placed particular emphasis on human rights: after several meetings with representatives of the Ugandan police in early August 2012, the NGO distributed 10,000 copies of a newly introduced complaint form and 5,500 copies of a complaints handling manual. The form allows the filing of complaints against police officers who violate human rights and act unprofessionally. In particular, the latter activities aimed at putting human rights at the centre of the police-general public interactions.[**]

The overview illustrates that PARP activities were heterogeneous and that each activity was limited in scope. The good and sustained relationship between HURINET-U and the Ugandan police was an important necessary condition for the implementation of PARP. HURINET-U has been present on the ground in Uganda since 1993, and it has actively and promptly followed up on human rights violations and has been in constant dialogue and exchange with the police.

5. Sampling and survey

The implementation of PARP was limited to eleven Ugandan police districts.[††] The restricted geographical ambit of PARP is used as a cornerstone of the empirical assessment. We sampled five districts where PARP activities took place and five comparable districts that were not included in the project.

We conducted a survey among 600 police officers in the 10 districts, sampling 60 officers within each district.[‡‡] The survey took place in April 2015. Since the survey was

---

[**] Further details about PARP can be found in Hout et al. (2016).
[††] The eleven police districts in which HURINET-U mainly worked are Arua, Bushenyi, Gulu, Kabale, Kabarole, Kampala, Lira, Masaka, Mbarara, Moroto and Soroti.
[‡‡] The survey districts are Bushenyi, Iganga, Jinja Kabale, Kabarole, Tororo, Luwero, Mbarara, Mityana and Soroti.

conducted roughly two years after the end of the intervention, we can only identify effects that have 'survived'. We consider this a strength of the analysis since assessments done right after interventions that aim to increase knowledge and change behaviour are mainly registering immediate effects.

Individual officers were selected in a stratified way to capture officers across all ranks. The data collection was carried out by our local university partner, the Uganda Management Institute, in consultation with the police. Police officers were approached after authorization from the police headquarters and the regional police. Importantly, HURINET-U did not participate in the selection of respondents and/or data collection.

Regional-level officers were purposively chosen to participate in the survey because of their leading position. Similarly, the leading police officers of the district headquarters were purposively included. Police stations within districts were randomly sampled, with half the officers in our sample coming from small stations (with up to 10 officers) and an additional 20% from medium-sized agencies (with 11 to 25 officers), resulting in a total of 70% of our sampled police officers being employed in agencies of up to 25 officers. The day of the survey was picked randomly and police officers from the districts participated in the survey based on availability or presence. Since local police stations only have few officers, we do not expect any systematic selection of participants into our sample. We applied this sampling procedure to have a stratified sample of officers that represents the full spectrum of police work, functions, positions and hierarchies.

The survey consisted of a self-administered pen and paper questionnaire in a classroom setting. During the survey, each officer was provided enough personal space to ensure privacy and confidentiality. To protect anonymity, we did not ask the officers to provide their names or addresses. The survey had two parts. In the first part, officers reported their basic socio-economic characteristics. In the second and core part, officers were asked to review twelve

vignette cases that were formulated following the example of Klockars et al. (2000, 2006) and Kutnjak Ivković (2005a, 2005b). In collaboration with HURINET-U and the Uganda Police Force, the cases were adapted to the local context to ensure that they are relevant, i.e., that the vignette cases reflect dilemmas faced by the Ugandan police. For example, we replaced the original scenario 4 (Klockars et al., 2006): 'A police officer is widely liked in the community, and on holidays local merchants and restaurant and bar owners show their appreciation for his attention by giving him gifts of food and liquor.' This scenario had to be modified since this type of behaviour is not perceived as bribery in the context of Uganda. Gifts around Christmas time are considered acceptable. Similarly, since jewellery shops are not common in Uganda, we changed the original scenario 5 and introduced a burglary in a general merchandise shop. Further, we introduced the police complaint form and the treatment of demonstrators in other scenarios since these are important issues in the Ugandan context. The modified cases were pre-tested for their relevance in the field. Thus, we ensured that the changed case scenarios have cultural resonance. We feel that the context-specific adaptation has enhanced the quality of our study since it is based on an in-depth analysis of the local context and conditions prior to the use of the scenarios.

In the survey, the cases were presented randomly to avoid an order by severity. For the sake of clarity we have grouped the cases into six categories of two cases each in the paper: the first group focuses on the code of conduct among police officers, the second on bribery, the third on fraud, the fourth on the refusal to register a complaint against the police, the fifth on severe crimes against individuals that are not followed up by the police, and the sixth on undue force used by the police against suspects and demonstrators. A detailed grouping of the twelve cases is presented in table A1 in appendix A1, while the exact wording can be found in appendix A2. Our survey also assessed gender dynamics, the findings of which are presented in a separate article (Wagner et al., 2017).

In line with Klockars et al.'s (2000; 2006) organisational/occupational approach that was addressed in section 2 above, our survey did not directly ask police officers about their behaviour in order to avoid biased responses. Instead, the police officers answered the following normative questions for each case:

1. How serious do YOU consider this behaviour to be?

2. Do you think YOU would report a fellow police officer who engaged in this behaviour?

3. How serious do MOST POLICE OFFICERS IN YOUR OFFICE consider this behaviour to be?

4. If an officer in your agency engaged in this behaviour and was discovered doing so, what if any disciplinary measure do YOU think SHOULD follow?

5. Would this behaviour be regarded as a violation of official policy in your agency?

The possible answer categories range on a Likert scale from 1 to 5. Questions 1 and 3 could be answered on a categorical scale from 1 (not at all serious) to 5 (very serious). Responses to questions 2 and 5 ranged from 'definitely not' to 'definitely yes'. Question 4 on disciplinary measures could be answered with 'none' [1], 'verbal reprimand' [2], 'written reprimand' [3], 'period of suspension without pay' [4], 'demotion in rank' [5] and 'dismissal' [6].

The advantage of using vignettes is that all officers are presented with the same cases; the disadvantage is that we do not observe actual behaviour. Clearly, we cannot determine whether police officers are indeed honest or corrupt. But the vignette approach has received strong validation in public health research, which documented consistency between hypothetical cases and actual behaviour (Peabody et al., 2000; Van der Meer and Mackenbach, 1998).

6. Descriptive statistics

6.1 Socio-demographic characteristics of the police officers

Descriptive statistics of the respondents' socio-demographic characteristics are presented in Table 1. The average officer in the sample is almost 42 years old. Slightly less than 25% of the respondents are female and most of the officers are married (84%). On average, they live in a household with almost 7 people and the majority of the interviewees are household heads (84%). Almost half the officers have secondary education, 27% completed advanced secondary education and 25% have a higher education degree. The remainder (less than 3%) have only primary education.

As to economic well-being, around 60% of the respondents earn between UGX 300,000 and 500,000 on a monthly basis.[§§] The average respondent owns 1.34 mobile phones and has almost 2 habitable rooms at home. Membership in clubs or community organizations is reported by almost half the respondents, and sports activities by 53%. These latter two variables serve as controls for the activity levels of the respondents and their readiness to engage in extra activities.

A comparison of PARP participants and non-participants shows very few differences. All but three characteristics are statistically identical. Significant differences, which are controlled for in the multivariate analyses, relate to income (with PARP participants earning less than non-participants) and housing.

The specific police work of the respondents is part of the second set of control variables related to the duration of their work as police officer, their rank, the section of the police force they work in, and the available infrastructure at their station (Table 1).

The average length of service is 18.8 years, with no differences between PARP participants and non-participants. Rank, however, does seem to matter. Of all respondents, 6% are of higher ranks; PARP participants are more likely to hold a higher rank (8% among PARP

[§§] This roughly corresponds to a range of USD 80 and 140 (UGX/USD exchange rate of 0.00028 on July 28, 2017).

participants versus 5% among non-participants) since the intervention targeted high-ranking officials. Roughly one third of the officers are of middle rank; among these more officers work in non-PARP districts. The majority of the officers are of low rank; here there are no differences between our two groups of respondents. The data further indicate that police officers with general duties are overrepresented among PARP participants, while significantly fewer participants work in the investigation section. Finally, in comparison to non-participants, PARP participants tend to come from smaller stations with fewer cars. Thus, policing conditions differ to some extent between the two groups and we therefore control for work- and infrastructure-related variables in the analysis.

Data on the geographical distribution of the respondents form the last group of control variables (Table 1). As the descriptive statistics show, the control districts were well chosen since the average population size and growth as well as the average poverty level are identical across PARP and non-PARP districts. Differences between PARP and non-PARP districts show up in relation to inequality and the number of police officers per 100,000 inhabitants, but these are significant only at the 10% level. Average crime and homicide rates across districts are identical.

To give an overview of the composition of our sample, we provide the distribution of PARP participants and non-participants across districts in the Appendix, Table A2. We show that most PARP participants are still located in the five districts where training activities took place, but roughly 13% of participating officers reside in non-targeted districts. The non-PARP district with the largest share of PARP participants is Iganga (6.8%). Similarly, the control sample of non-participants is mainly drawn from the five control districts. They make up for 76% of the control sample. The remaining 24% of the control sample resides now in former intervention districts. Since this can result in potential spillover effects, we control for it in the robustness tests of our multivariate analysis. Furthermore, Table A2 shows that we reached the

target sample of 60 participants in all but two districts where we only sampled 59 participants. The missing two participants were sampled from two other districts. Thus, the extent to which police officers were affected by PARP activities results predominantly from work-related characteristics and community features. We control for these two sets of confounding factors along with the individual characteristics in the multivariate analyses.

6.2 Descriptive statistics of the case assessments

Detailed descriptive statistics of the five outcome variables that we collected for the twelve cases are presented in Table 2. We show the simple averages resulting from the Likert scale answers.

Across the twelve cases and five assessment criteria PARP participants tend to be more critical compared to non-participants. The first two vignette cases, on police code of conduct, are judged rather mildly. Receiving holidays in exchange for repairing a supervisor's car is assessed moderately negatively (average score 3.72), although officers tend to be generally aware that such behaviour violates official policy (average score 4.24), which they would report (average score 4.06). PARP participants feel more strongly that disciplinary measures should follow (PARP: 4.016 versus non-PARP: 3.651). The misbehaviour described in the second case, related to covering a drunk colleague who caused an accident, is by and large seen as a light offence. Overall, PARP participants and non-participants tend to differ on the need to report the behaviour and on their judgment of the behaviour, with the former group taking, on average, a stricter position than the latter.

The second group of cases depicts situations of bribery. Case 3 on accepting gifts while on duty has the lowest Likert score of all cases. All respondents are close to neutral (value of 3) when it comes to reporting a colleague, although PARP participants tend to be slightly more critical than non-participants. Case 4, related to the acceptance of a bribe after observed

speeding, is evaluated very critically, as is reflected in the score of 4.58 among PARP participants and 4.22 among non-participants. Fraud cases are shown in the third group: case 5 (the misappropriation of money from a found wallet) and case 6 (illegal enrichment when investigating a burglary) are judged harshly and a great majority of respondents among PARP participants and non-participants indicate they would report a colleague who shows this type of behaviour.

Overall, the first six cases suggest that police officers have a clear idea about acceptable and non-acceptable behaviour: the acceptance of bribes and misappropriation are evaluated more critically than violations of the police code of conduct. In most cases, officers see themselves as more critical of misbehaviour than their colleagues. Responses to the question whether forms of behaviour violate official policy indicate the existence of a gap between formal rules and actual practices. Overall, PARP participants tend to evaluate the vignette cases more critically than non-participants.

The next two cases depict situations of how police officers deal with complaints. The refusal to register a complaint and the humiliation of the complainant (case 7) is judged rather mildly, but the majority of officers considers the arrest of a complainant on false grounds (case 8) to be unacceptable. This is in line with our expectations and shows the internal coherence of the vignette cases. The differences across respondents indicate that the accountability project may have left an impact: PARP participants rate the severity of case 7 with 3.988, while non-participants rate it with 3.360. The more severe case 8 is considered as an example of serious misbehaviour as indicated by the rating of 4.332 among PARP participants compared to 3.780 among non-participants.

Lastly, cases of reported severe crimes against individuals without adequate follow-up by the police (cases 9 and 10), and of the use of undue force (cases 11 and 12) are assessed very critically. Again, PARP participants tend to be much more critical than non-participants, which

suggests that officers who took part in the project apply a more careful judgment when it comes to abuse of power and human rights violations.

7. Empirical model

In our identification strategy, we rely on propensity score matching (PSM). We opted for this approach since project locations were not selected randomly and data was collected in only one round after the implementation of PARP. We pool the responses for all 12 cases so that we obtain an estimation sample of 7,200 observations.

For PSM to be valid we need to impose an assumption about conditional independence, which states that given a set of observable covariates, which are not affected by the project (i.e., exogenous to PARP), potential outcomes are independent of project assignment (Lechner, 1999). It implies the strong assumption that selection into PARP is solely based on observable characteristics for which we can control in the analysis. The discussion above indicated that selection into PARP appeared to be mainly based on work-related characteristics and community features and not on the personal characteristics of police officers. We are therefore confident that we can properly capture participation with control variables related to individual and work-related characteristics as well as community features.

By employing a logistic regression of project allocation on the observable covariates, we determine the probability of participation in PARP for every police officer i based on observable characteristics:

$$P(D_i = 1|X_i, I_i, DC_i, C_i) = l[\alpha_0 + X_i\, \alpha_X + I_i\, \alpha_I + DC_i\, \alpha_{DC} +\, C_i\, \alpha_C +\, \varepsilon_i],$$

where D_i is a dummy variable coding for participation in PARP, and $l[.]$ is the logistic function. All observable characteristics are collected in X_i, I_i, DC_i and C_i. The unobserved error term is denoted by ε_i. The individual characteristics are denoted by X_i, police station infrastructure by I_i and district characteristics by DC_i. With these control variables we account for the nested

design of the study since police officers are integrated in police stations and police stations are organized within districts. The individual level characteristics X_i are captured by age, gender, marital status, heading the household, level of education, level of income, number of habitable rooms in the house, household size, number of mobile phones owned, engagement in sport activities and membership in an organization. In addition, we control for work-experience variables. These are the years of service, rank and unit of operation. The infrastructure controls I_i are the number of rooms, police cars, motorcycles and bicycles at the police station. The district characteristics DC_i are population size (in log), population growth rate, headcount poverty rate, inequality (measured by the Gini coefficient), share of the population belonging to the largest ethnic group, number of police officers per 100,000 inhabitants, and the crime and homicide rates. Finally, we control for case-specific effects (C_i) to account for the differences in the severity of the presented vignette cases. By deriving the probability of participation from the logistic regression, we ensure that persons with the same observable characteristics as denoted by *X, I* and *DC* have a positive probability of being both participants and non-participants, which generates the common support (Heckman et al., 1999). It allows us to form matches of individuals with similar characteristics observed for PARP participants and non-participants.

We derive the PSM estimator of the impact of PARP as the mean difference in outcome variables over the common support, appropriately weighted by the propensity score distribution of participants:

$$\tau^{PSM} = E\ \{\ E[Y(1)|\ D=1, P(X,I,DC,C)] - E[Y(0)|\ D=0, P(X,I,DC,C)]\ \},$$

where *Y* is the outcome under study, *1* represents PARP participants and *0* non-participants. We apply nearest-neighbour matching by matching each individual from the group of PARP participants with suitable individuals from the control group. We employ the revised PSM

procedure that was proposed by Abadie and Imbens (2006, 2008) to derive consistent standard errors.

In addition, we compare our results to an estimator that makes use of the PSM weights in a regression framework, the so-called inverse probability weighting (IPW) (Wooldridge, 2007). In this model, the observations of the non-participants are weighted by their respective propensity scores. Those non-participants that share characteristics with participants, and thus have larger propensity scores, receive larger weights in the regression model. We derive this second estimator to gauge the robustness of the PSM results.

Finally, in order to avoid selectivity in highlighting possible effects on the five individual outcomes, we also employ a seemingly-unrelated regression model that accounts for correlation of the error term across specifications, i.e. multiple hypothesis testing. Moreover, the approach allows us to derive one single average effect of the intervention across all five outcome categories (Casey et al., 2012; Clingingsmith et al., 2009).

8. Results

8.1 Determinants of participation in PARP

Before assessing the impact of the intervention, we identify the observable characteristics that determine PARP participation with a logistic regression model. Table A3 (Appendix) shows that the individual socio-demographic characteristics of police officers are unlikely determinants of participation in PARP. This applies to all individual characteristics except for being member of a club or community organization. Members of such organizations are more likely to participate in PARP. Concerning work experience related variables, membership of particular sections within the police force – in particular, general service – and holding a higher rank are positively related to PARP participation. These findings are not surprising since they reflect the outreach strategy of PARP.

District-level covariates are the most important determinants of participation in PARP, which suggests that HURINET-U focused on less populated and less ethnically homogeneous districts with higher poverty but less inequality. PARP districts appear to have more police officers and higher crime rates but lower homicide rates. All district-level covariates are statistically significant at the 1%-level and in practical terms highly relevant. The relationship with PARP activities suggests that HURINET-U selected the intervention districts mainly on the basis of perceptions of community characteristics and the crime environment. Concomitantly, we can consider PARP activities as an exogenous event for the individual police officers concerned. Self-selection bias, resulting from *unobservable* individual characteristics, is unlikely. Therefore, in assessing police integrity we rely on the observed differences for the matched sample of PARP participants and non-participants controlling for the above-mentioned individual, work experience, police infrastructure and district level confounders.

8.2 Impact of PARP

Results of the comparisons of the outcome variables between PARP participants and non-participants are presented in Table 3. We consider the impact of the PARP intervention across the five normative questions. Panel A of Table 3 presents the simple *comparison of means* without accounting for confounding actors.

Results show that across all five questions PARP participants tend to give a more critical rating, with the difference being statistically significant at the 1%-level. In line with the case-specific descriptive statistics (Table 2), PARP participants score higher on average, which indicates that they assess the depicted behaviour more critically: Among PARP participants, the assessment of the perceived seriousness of the cases is most critical (0.49); the least critical rating is given for the action that should follow according to the police officers (0.22).

Next, we present the impact estimates resulting from the propensity score model (PSM, Panel B), those from inverse probability weighting (IPW, Panel C) and the ones from the seemingly unrelated regression model (SUR, Panel D). The coefficient estimates are similar to the raw comparison of means. In the PSM and IPW models, the judgment of case severity differs by over half a point between PARP participants and non-participants and is highly statistically significant. This supports the conclusion that PARP had a positive impact on normative judgments, importantly including human rights as one of its main targets. The coefficient estimate that accounts for the covariance in the error term across the five questions is smallest in magnitude but again it leaves no doubt that PARP participants judge case severity more critically.

With regard to reporting of misbehaviour, PARP participants are on average only 0.41 points (PSM) more likely to report a colleague's misbehaviour. While the coefficient on reporting is smaller in size than the one on case severity, it is still highly statistically significant. This suggests that PARP participants are not only more critical about inappropriate behaviour but also more inclined to report it. The IPW and SUR models fully support the findings.

PARP activities also seem to have impacted on the way the judgments of fellow police officers are perceived. Yet, the average effect is 0.21 (=0.50-0.29) points smaller compared to the estimate on case severity (PSM). Police officers have the impression that colleagues take misbehaviour less seriously than they do themselves. PARP participants have more confidence in their own judgment than that of their colleagues. Again, the IPW and SUR models confirm the PSM findings with even slightly higher coefficients (0.39 and 0.35, respectively, compared to 0.29).

Differences among respondents are less pronounced when it comes to the disciplinary measures they consider appropriate in case of misdemeanours (question 4). According to the PSM model, the estimated difference in average scores is 0.21. This is statistically significant,

but not large enough to ascribe considerable impacts to PARP. Responses to question 5 indicate that PARP participants appear to know more about their agency's official rules of conduct. The impact estimate of 0.27 indicates that PARP participants are more ready to define misbehaviour as a violation of official policy. The IPW and SUR models produce lower coefficient estimates. Nevertheless, all differences are statistically significant at least at the 5% level.

Finally, we calculated the regression-based average difference across the five questions. Calculations on the basis of the PSM, IPW and SUR model result in global average effects ranging between 0.33 and 0.34, showing that across models we coherently identify a positive and practically meaningful impact of PARP. Statistical significance can only be assessed with SUR indicating that the effect is significant at the 1%-level.

To further assess the robustness of our findings, we employed five additional models. First, we replaced the police-infrastructure and district-level covariates with district-level dummies (Panel E). We identify similar but considerably bigger effects. The findings show that the estimates presented so far can be seen as conservative impact estimates. From a methodological point of view, the results suggests that police infrastructure and the situation in the district are related to police integrity, i.e. if the police officers are already unsatisfied with the provided infrastructure, they are less likely to take their work seriously. Thus, the quality of the provided infrastructure is very likely to be reflected when the officers are asked about their work attitude and integrity.

Second, we further challenge the role of district level control variables by estimating a specification that only includes individual level control variables. Results are presented in Table 3, Panel F. The coefficient estimates tend to be even larger, suggesting that we overestimate the impact of PARP if we fail to control for infrastructure and district variables.

Third, we addressed possible spillovers from PARP participants who relocated to non-PARP districts and vice versa by excluding PARP participants who reside in non-PARP

districts and non-participants that reside in PARP districts (Panel F). This results in a smaller sample, as sample size drops to 5,820 observations. The effects we identify are slightly smaller when compared to the PSM model with all observations and the full set of controls (Panel B). Yet, in practical terms the impacts are still meaningful and all findings are highly statistically significant. Moreover, the identified global average effect is bigger than the one identified with the SUR model. We prefer the original specification with the 'movers' as it allows us to address structural effects at the district level. The spillover specification does not allow us to control for district-related variables as they are perfectly collinear.[***] Furthermore, if there are indeed spillovers, it makes it more difficult for us to find an effect in general since the control group will also show higher support for police integrity.

Fourth, we employed an empirical specification that excludes individuals in leading positions, i.e. those with high rank. Since the sub-sample of PARP participants has more police officers of high rank, the results could have been driven by these individuals. Results of the model without police officers in leading positions are presented in Panel H. Except for some numerical differences our results are well aligned with the full model (Panel B). It is therefore unlikely that the results are driven by officers of high rank, who account for only 6% of the overall sample.

Fifth, in a last robustness test we excluded all 'spillover' police officers along with those

---

[***] Concerning spillovers, we note that since HURINET-U is an NGO with limited funds, it had to restrict its work ambit in particular in the second phase of the program, the phase that we are evaluating. Yet, police officers in leading positions also meet at the national level and exchange about the activities in their districts. At the same time, we are not aware of any attempt from non-PARP districts to be part of the intervention and merely hearing about PARP is not likely to change operations. We consider it unlikely that information about the intervention is spread very vocally since the intervention aims at an a priori unpopular change and implies outside involvement in operations that are traditionally considered as being exclusively controlled by the police. Based on our knowledge of the Ugandan context and the fact that the police is considered to be the most corrupt institution by the Ugandan citizens (Commonwealth Human Rights Initiative, 2006a), we do not expect police officers to full-heartedly fight for the implementation of the complaints form, a stronger focus on human rights (of suspects and arrestees) and more respect for demonstrators. Lastly, the Ugandan situation is such that the regime, including its agents such as police officers, tries to control society (Anderson and Fisher, 2016). Therefore, it is a valid assumption that PARP-participants who moved to non-PARP districts are likely to have little influence on the behaviour of their colleagues.

of high rank (Panel I). As for the previous specifications, the results are virtually identical with those of the full model (Panel B).

Further, we note that our analysis identified only the effects that have 'survived', since the survey was conducted roughly two years after the end of the PARP intervention. In research that is done shortly after an intervention, the analysis may pick up knowledge about best practices that is still fresh but that may dwindle after some time. The time lag of two years adds strength to our analysis as it permits a focus on long-lasting impacts rather than short-term effects.

To sum up, the analyses reported above indicate that PARP seems to have had an impact mainly through *normative perceptions* about the severity of cases. The replies of the police officers to the vignette cases indicate that participants in PARP activities score higher on average across all cases and all questions, thus indicating that PARP has been successful in creating heightened awareness of what is right and wrong police behaviour. Most importantly, PARP was successful in diffusing knowledge about proper policing in relation to cases of more severe police misbehaviour. The most noticeable differences result from the treatment of 'clients' (former arrestees and suspects, thieves, persons complaining), which indicates that the human rights agenda of PARP has been translated into better knowledge of the police officers who participated in the intervention.

In addition, our findings indicate that there seems to be a disparity between officers' *own* assessment of the severity of the cases and their perception of how violations should be treated. Thus, although police officers know the rules about good policing, they do not fully comply with them in their daily practice. This may imply that official standards are only partially enforced, and that individual officers have room to interpret the rules to their

advantage. Consequently, the change in *normative* views about acceptable and non-acceptable behaviour may not have produced a *behavioural* change of the police officers themselves.[†††]

9. Conclusion

The findings of our research on PARP, which was implemented in Uganda between 2010 and 2013, indicate that the intervention seems to have contributed to greater awareness among police officers about the need for on-the-job integrity and proper behaviour vis-à-vis the Ugandan citizens. Comparing police officers who took part in PARP activities with non-participants, the attitudinal difference between the two groups on a variety of vignette cases suggests that the project has had lasting positive results. We conclude from our findings that police accountability may be enhanced by targeted attention to unacceptable police behaviour, breaches of integrity and corruption. Yet, activities on good and accountable policing are not very likely to assume their full potential when used as stand-alone instruments; they need to be combined with credible internal enforcement mechanisms.

Our research suffers from two non-negligible limitations: First, we had to resort to a quasi-experimental evaluation design. Second, we cannot fully rule out spillover effects. Future work on the impact of police integrity trainings should resort to more rigorous evaluation designs to gauge whether the current findings can be substantiated.

While PARP activities were scattered, heterogeneous interventions, the project seems to have impacted on police integrity by altering the *perceptions* and *attitudes* of participating police officers. We cannot, however, be sure that the changes in perceptions and attitudes have translated into improved practices since our survey tool does not allow us to observe behavioural outcomes. Overall, we conclude that the measurement and systematic analysis of

---

[†††] A qualitative analysis of in-depth interviews confirms the quantitative findings. Results of the qualitative analysis can be found in Hout et al. (forthcoming).

(changes in) perceptions and attitudes remains a challenge. The findings highlight the need for future research on behavioural changes and on measurement of perceptions and attitudes, particularly from a comparative perspective.

References

Abadie, A. and Imbens, G.W. (2006) Large sample properties of matching estimators for average treatment effects, *Econometrica,* 74:235-267.

Abadie, A. and Imbens, G.W. (2008) On the failure of the bootstrap for matching estimators, *Econometrica,* 76:1537-1557.

Anderson, D.M. and Fisher, J. (2016) Authoritarianism and the Securitization of Development: The Case of Uganda, in Hagmann, T. and Reyntjens, F., eds, *Development without Democracy? Foreign Aid and Authoritarian Regimes in Africa*, London: Zed Books, pp. 67-90.

Andreescu, V., Keeling, D. G., Voinic, M. C., and Tonea, B. N. (2012b) Future Romanian law enforcement: Gender differences in perceptions of police misconduct, *Journal of Social Research & Policy*, 3(1): 97-113.

Banerjee, A., Chattopadhyay, R., Duflo, E., Keniston, D. and Singh, N. (2012) *Can Institutions be Reformed from Within? Evidence from a Randomized Experiment with the Rajasthan Police*. CEPR Discussion Paper No. 8869. London: Centre for Economic Policy Research.

Banerjee, A., Chattopadhyay, R., Duflo, E., Keniston, D. and Singh, N. (2014) *Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy and Training*. NBER Working Paper No. 17912.

Basheka, B.C. (2013) Public Administration and Corruption in Uganda, in Vyas-Doorgapersad, S., Lukamba-Muhiya, T. and Peprah Ababio, E., eds, *Public Administration in Africa: Performance and Challenges*, Boca Raton: CRC Press, pp. 45-81.

Casey, K., Glennerster, R. and Miguel, E. (2012) Reshaping Institutions: Evidence on Aid Impacts using a Preanalysis Plan, *The Quarterly Journal of Economics*, 127(4):1755–1812.

Cetinkaya, N. (2010) *Perceptions of police corruption among Turkish police cadets*, PhD dissertation, East Lansing: Michigan State University.

Chappell, A.T. and Piquero, A.R. (2004) Applying Social Learning Theory to Police Misconduct, *Deviant Behavior,* 25(2):89-108.

Clingingsmith D., Khwaja A. I. and Kremer M. (2009) Estimating the Impact of The Hajj: Religion and Tolerance in Islam's Global Gathering, *The Quarterly Journal of Economics,* 124 (3):1133–70.

Collins, P., Marenin, O. and Chu., M.C., eds (2016) The Politics and Management of Policing Reforms in Newly Industrialized, Industrializing and Developmental States, Special Issue, *Public Administration and Development,* 36(2):71-168.

Commonwealth Human Rights Initiative (2006a) *The Police, the People, the Politics: Police Accountability in Uganda*. Report, http://www.humanrightsinitiative.org/publications/police/uganda_country_report_2006.pdf (Accessed 29 July 2017).

Commonwealth Human Rights Initiative (2006b) *A Review of the Uganda Police Force Budget and its Effects on Crime Management.* Report, http://www.humanrightsinitiative.org/publications/police/uganda_report.pdf (Accessed 29 July 2017).

Gottschalk, P. (2010) Crime-Based Survey Instrument for Police Integrity Measurement, *Policing: An International Journal of Police Strategies & Management,* 33(1):52-68.

Heckman, J., LaLonde, R. and Smith, J. (1999) The economics and econometrics of active labor market programs, in Ashenfelter, O. and Card, D., eds, *Handbook of Labor Economics*, Vol. III, Amsterdam: Elsevier, pp. 1865–2097.

Hickman, M.J., Piquero, A.R., Powell, Z.A. and Greene, J. (2016a) Expanding the Measurement of Police Integrity*, Policing: An International Journal of Police Strategies & Managemen,t 39(2):246-267.*

Hickman, M.J., Powell, Z.A., Piquero, A.R. and Greene, J. (2016b) Exploring the Viability of an Attitudes Toward Ethical Behavior Scale in Understanding Police Integrity Outcomes*, Policing: An International Journal of Police Strategies & Management,* 39(2):319-337.

Holmberg, S. and Rothstein, B. (2012) *Good Government and the Relevance of Political Science*, Cheltenham: Edward Elgar.

Hout, W., Brouwers, R., Fisher, J., Namara, R., Schakel, L. and Wagner, N. (2016) *Policy Review Good Governance: Democratisation, Promotion of Rule of Law and Control of Corruption – Uganda Field Study*, Final Report for the Policy and Operations Evaluation Department (IOB), Ministry of Foreign Affairs, The Netherlands.

Hout, W., Wagner, N. & Namara, R. (forthcoming) Holding Ugandan Police to Account: Case study of the Police Accountability and Reform Project, in Bergh, S., Pellissery, S. and Sathyamala, C., eds, *The State of Accountability in the Global South: Challenges and Responses*, Cheltenham: Edward Elgar.

HURINET-U (2013) *Police Accountability and Reform Project: Narrative Project Report (November, 2010-January, 2013)*, Kampala: Human Rights Network Uganda.

Kakamwa, C. (2014) Uganda to Increase Size of Police Force, *New Vision*, December 26, 2014.

Klockars, C.B., Kutnjak Ivković, S., Harver, W.E. and Haberfeld, M.R. (2000) *The Measurement of Police Integrity*, Research in Brief, NCJ 181465, Rockville: National Institute of Justice.

Klockars, C.B., Kutnjak Ivković, S. and Haberfeld, M.R. (2006) *Enhancing Police Integrity.* Dordrecht: Springer.

Kucukuysal, B. (2008) *Determinants of Turkish police officers' perception of integrity: Impact of organizational culture*, PhD dissertation, Orlando: University of Central Florida.

Kutnjak Ivković, S. (2005a) *Fallen Blue Knights: Controlling Police Corruption*, Oxford: Oxford University Press.

Kutnjak Ivković, S. (2005b) Police (Mis) Behavior: A Cross-Cultural Study of Corruption Seriousness*, Policing: An International Journal of Police Strategies & Management,* 28(3):546-566.

Kutnjak Ivković, S. (2015) 'Studying Police Integrity', in Kutnjak Ivković, S. and Haberfeld, M.R., eds, *Measuring Police Integrity Across the World: Studies from Established Democracies and Countries in Transition*, New York: Springer, pp. 1-36.

Kutnjak Ivković, S. and Haberfeld, M.R., eds (2015) *Measuring Police Integrity Across the World: Studies from Established Democracies and Countries in Transition*, New York: Springer.

Kutnjak Ivković, S. and O'Connor Shelley, T. (2010) The Code of Silence and Disciplinary Fairness: A Comparison of Czech Police Supervisor and Line Officer Views*, Policing: An International Journal of Police Strategies & Management,* 33(3):548-574.

Kutnjak Ivković, S. and Khechumyan, A. (2014) Measuring police integrity among urban and rural police in Armenia: From local results to global implications, *International Journal of Comparative and Applied Criminal Justice* 38(1): 39-61.

Kutnjak Ivković, S. and Sauerman, A. (2013) Curtailing the Code of Silence among the South African Police*, Policing: An International Journal of Police Strategies & Management,* 36(1): 175-198.

Lasthuizen, K., Huberts, L. and Heres, L. (2011) How to Measure Integrity Violations: Towards a validated typology of unethical behavior, *Public Management Review,* 13(3):383-408.

Lechner, M. (1999) Earnings and employment effects of continuous off-the-job training in East Germany after unification, *Journal of Business Economic Statistics,* 17(1):74–90.

McMillan, J. and Zoido, P. (2004) How to subvert democracy: Montesinos in Peru, *Journal of Economic Perspectives,* 18(4):69–92.

Newburn, T. (2015) *Literature Review: Police Integrity and Corruption,* London: Her Majesty's Inspectorate of Constabulary (HMIC), http://www.justiceinspectorates.gov.uk/hmic/wp-content/uploads/pic-literature-review.pdf (Accessed 29 July 2017).

Olken, B.A. and Barron, P. (2009) The simple economics of extortion: evidence from trucking in Aceh, *Journal of Political Economy,* 117:417–52.

Olken, B.A. and Pande, R. (2012) 'Corruption in Developing Countries', *Annual Review of Economics,* 4:479–509.

Peabody, J.W., Luck, J., Glassman, P., Dresselhaus, T.R. and Lee, M. (2000) Comparison of Vignettes, Standardized Patients, and Chart Abstraction: A Prospective Validation Study of 3 Methods for Measuring Quality, *Journal of the American Medical Association,* 283(13):1715-1722.

Perrot, S. (2014) From the February 2011 Elections to the Walk to Work Protests: Did Uganda Really Want "Another Rap"?, in Perrot, S., Makara, S., Lafargue, J. and Fouere, M.-A., eds, *Elections in a Hybrid Regime: Revisiting the 2011 Ugandan Polls*, Kampala: Fountain Publishers.

Rose-Ackerman, S. and Palifka, B.J. (2016) *Corruption and government: Causes, consequences and reform*. Second edition. Cambridge: Cambridge University Press.

Tankebe, J. (2010) Public Confidence in the Police: Testing the Effects of Public Experiences of Police Corruption in Ghana. *British Journal of Criminology,* 50(2):296-319.

Transparency International (2016a) *What is corruption?* http://www.transparency.org/what-is-corruption (accessed 29 July 2017).

Transparency International (2016b) *Corruption perceptions index 2015*, http://www.transparency.org/cpi2015 (accessed 29 July 2017).

Transparency International-Kenya (2013) *The East African Bribery Index 2013*, http://tikenya.org/wp-content/uploads/2017/06/the-east-african-bribery-index-2013.pdf (accessed 29 July 2017).

Uganda Police (2013) *Annual Crime and Traffic/Road Safety Report*, http://www.upf.go.ug/download/publications(2)/Annual_Crime_and_Traffic_Road_Safety_Report_2013(2).pdf (accessed 29 July 2017).

Uganda Police (2015) *Directorates*, http://www.upf.go.ug/directorate (accessed 29 July 2017).

Uganda Police Force (2007) *Strengthening Statistics for Development Planning. Sector Strategic Plan for Statistics (2006/07 - 2010/11)*. Report, http://www.ubos.org/onlinefiles/uploads/ubos/pdf%20documents/PNSD/UPF%20SSPS.pdf (accessed 29 July 2017).

Van der Meer, J. and Mackenbach, J. (1998) 'Low education, high GP consultation rates: the effect of psychosocial factors', *Journal of Psychosomatic Research,* 44:587–597.

Wagner, N., Rieger, M., Bedi, A. and Hout, W. (2017) Gender and Policing Norms: Evidence from Survey Experiments Among Police Officers in Uganda, *Journal of African Economies*. 26(4):492–515

Wambua, P.M. (2015) 'Police corruption in Africa undermines trust, but support for law enforcement remains strong', *Afrobarometer Dispatch* No.56.

Wooldridge, J.M. (2007) 'Inverse probability weighted estimation for general missing data problems'. *Journal of Econometrics*, 141(2):1281–1301.

Tables

Table 1: Descriptive statistics of control variables

	Overall		PARP participants	Non-participants	DiM
	Mean	Std.Dev.	Mean	Mean	*p*-value
Socio-demographic covariates					
Age	41.785	9.426	42.244	41.457	0.314
Gender: Female	0.228		0.208	0.243	0.317
Marital status (Excluded category: Not married)					
Married	0.843		0.852	0.837	0.622
Household size	6.672	3.992	6.868	6.531	0.309
Household head	0.843		0.840	0.846	0.85
Education levels (Excluded category: Primary)					
Secondary	0.455		0.476	0.44	0.383
Advanced secondary	0.268		0.248	0.283	0.343
Higher	0.248		0.252	0.246	0.861
Income level (Excluded category: Incomes<200,000UGX)					
Income 200,000-300,000UGX	0.115		0.144	0.094	0.060*
Income 300,000-500,000UGX	0.603		0.544	0.646	0.012**
Income 500,000-700,000UGX	0.140		0.148	0.134	0.634
Income>700,000UGX	0.095		0.108	0.086	0.360
# mobile phones owned	1.337	0.578	1.292	1.369	0.110
# habitable rooms	1.750	1.103	1.924	1.626	0.001***
Member of a club/organization	0.482		0.484	0.48	0.923
Does sport	0.525		0.520	0.529	0.836
Work related covariates					
Police section (Excluded category: Other sections and duties)					
Traffic	0.043		0.028	0.054	0.119
Investigation	0.262		0.216	0.294	0.032**
Intelligence	0.063		0.064	0.063	0.955
General duties	0.463		0.584	0.377	0.000***
Years of experience	18.800	10.556	19.384	18.383	0.252
Police rank (Excluded category: Low rank)					
High rank	0.060		0.080	0.046	0.082*
Medium rank	0.322		0.300	0.337	0.338
# rooms of police station	12.758	9.259	11.472	13.677	0.004***
# police cars at station	1.522	1.508	1.176	1.769	0.000***
# police motorcycles at station	6.762	12.557	6.064	7.260	0.250
# police bicycles at station	2.147	5.361	2.316	2.026	0.514
District level covariates					
Population size	427,400	92,560	394,880	459,920	0.292
Population growth rate	2.226	0.777	2.132	2.320	0.726
Poverty head count rate	22.310	8.539	23.280	21.340	0.742
Gini index	0.398	0.073	0.356	0.440	0.064*
Population share belonging to largest ethnicity	73.980	14.454	76.240	71.720	0.649
Police officers per 100,000 inhabitants	133.214	0.090	133.261	133.166	0.095*
Crimes per 100,000 inhabitants	338.232	143.510	296.802	379.662	0.393
Homicides per 100,000 inhabitants	8.629	3.910	9.629	7.628	0.451

Note: The sample consists of 600 police officers, 250 are PARP participants, 350 are non-participants. ***/**/* denotes p<0.01/0.05/0.1, respectively. Descriptive statistics of district level control variables are calculated on the basis of 10 district level observations. DiM abbreviates difference in means, the associated *p*-value is presented.

Table 2: Descriptive statistics and comparison of means of the outcome variables

	Severity (own judgment)	Reporting	Severity (others)	Disciplinary measure	Official policy
Group 1: Code of conduct among the police officers					
Case 1: Police mechanic repairing supervisor's car in exchange for holidays					
Mean PARP participant	4.012***	4.408***	3.924***	4.016***	4.420***
Mean Non- participant	3.511	3.809	3.529	3.651	4.117
DiM: p-value	(0.000)	(0.000)	(0.001)	(0.001)	(0.005)
Case 2: Police officer driving drunk and having an accident goes unreported by colleague					
Mean PARP participant	3.968***	4.100***	3.708	3.800**	4.224*
Mean Non-participant	3.526	3.789	3.517	3.534	4.023
DiM: p-value	(0.001)	(0.007)	(0.115)	(0.028)	(0.076)
Group 2: Bribery					
Case 3: Acceptance of freely offered meals and small gifts while on duty					
Mean PARP participant	3.788***	3.356**	3.640***	3.312**	4.112**
Mean Non-participant	3.351	3.029	3.306	3.034	3.843
DiM: p-value	(0.001)	(0.011)	(0.007)	(0.016)	(0.029)
Case 4: Speeding not reported in exchange for a bribe					
Mean PARP participant	4.576***	4.476***	4.128***	4.404	4.684***
Mean Non-participant	4.217	4.094	3.771	4.249	4.440
DiM: p-value	(0.001)	(0.000)	(0.002)	(0.137)	(0.007)
Group 3: Fraud					
Case 5: Officer taking money of found wallet					
Mean PARP participant	4.352***	4.564***	4.084***	4.360***	4.620***
Mean Non-participant	3.820	4.063	3.589	3.931	4.300
DiM: p-value	(0.000)	(0.000)	(0.000)	(0.000)	(0.001)
Case 6: Police officer stealing goods when investigating a burglary					
Mean PARP participant	4.688***	4.752***	4.496***	5.212**	4.736**
Mean Non-participant	4.317	4.497	4.206	4.966	4.577
DiM: p-value	(0.000)	(0.001)	(0.003)	(0.013)	(0.047)
Group 4: Refusal to register complaints					
Case 7: Refusal to register a complaint and humiliation of the complainant					
Mean PARP participant	3.988***	4.244***	3.700**	3.624**	4.484***
Mean Non-participant	3.360	3.914	3.429	3.374	4.191
DiM: p-value	(0.000)	(0.001)	(0.026)	(0.020)	(0.004)
Case 8: Refusal to register a complaint and a one-week detention for the complainant for false accusation					
Mean PARP participant	4.332***	4.632***	4.160***	4.200	4.644
Mean Non-participant	3.780	4.360	3.834	4.066	4.506
DiM: p-value	(0.000)	(0.002)	(0.006)	(0.191)	(0.115)
Group 5: Reported severe crimes against individuals not followed up upon					
Case 9: Police officer refusing to register wife beating					
Mean PARP participant	4.532***	4.632***	4.312***	4.072	4.656
Mean Non-participant	4.091	4.357	3.991	3.969	4.546
DiM: p-value	(0.000)	(0.001)	(0.003)	(0.329)	(0.173)
Case 10: Reported murder not being followed up on					
Mean PARP participant	4.192***	4.648***	4.268***	2.744	4.684***
Mean Non-participant	3.729	4.337	3.980	2.717	4.420
DiM: p-value	(0.001)	(0.000)	(0.009)	(0.788)	(0.003)
Group 6: Undue force used by the police					
Case 11: Foot patrol torturing a thief					
Mean PARP participant	4.228***	4.348***	3.964***	3.976	4.504***
Mean Non-participant	3.657	3.883	3.654	3.803	4.243
DiM: p-value	(0.000)	(0.000)	(0.009)	(0.133)	(0.010)
Case 12: Brutal strike down of a demonstration					
Mean PARP participant	4.436***	4.252***	4.432***	4.064	4.664***
Mean Non-participant	3.889	3.911	4.094	3.823	4.397
DiM: p-value	(0.000)	(0.006)	(0.003)	(0.171)	(0.005)

Note: N = 600. DiM abbreviates difference in means, the associated p-value is presented. ***/**/* indicates significance at the 1/5/10% level, respectively. Column 1 "Severity (own judgment)" refers to question "How serious do YOU consider this behavior to be?", Column 2 "Reporting" refers to question "Do you think YOU would report a fellow police officer who engaged in this behavior?", Column 3 "Severity (others)" refers to question "How serious do MOST POLICE OFFICERS IN YOUR OFFICE consider this behavior to be?", Column 4 "Disciplinary measure" refers to question "If an officer in your

Table 3:Main results

	How serious do YOU consider this behaviour to be?	Do you think YOU would report a fellow police officer who engaged in this behaviour?	How serious do MOST POLICE OFFICERS IN YOUR OFFICE consider this behaviour to be?	If an officer in your agency engaged in this behaviour and was discovered doing so, what if any discipline do YOU think SHOULD follow?	Would this behaviour be regarded as a violation of official policy in your agency?	Global average across all five questions
Panel A: Simple Comparison of Means (no covariates)						
Treatment effect	0.487***	0.364***	0.326***	0.222***	0.236***	0.327
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Control group average	3.771	4.004	3.742	3.760	4.300	
Panel B: Propensity Score Matching (full set of covariates)						
Treatment effect	0.503***	0.407***	0.286***	0.213***	0.270***	0.336
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Panel C: Inverse Probability Weighting (full set of covariates)						
Treatment effect	0.578***	0.374***	0.386***	0.134**	0.246***	0.344
	(0.000)	(0.000)	(0.000)	(0.023)	(0.000)	
Panel D: Average Effect from Simultaneous Regression (full set of covariates)						
Treatment effect	0.437***	0.377***	0.350***	0.191**	0.181**	0.210***
	(0.000)	(0.000)	(0.000)	(0.019)	(0.041)	(0.000)
Robustness checks						
Panel E: Propensity Score Matching (excluding infrastructure and district variables, imposing district fixed effects)						
Treatment effect	0.676***	0.725***	0.466***	0.437***	0.381***	0.537
	(0.000)	(0.000)	(0.000)	(0.000)	(0.005)	
Panel F: Propensity Score Matching (including only individual level control variables)						
	0.832***	0.671***	0.390***	0.661***	0.519***	0.615
	(0.000)	(0.000)	(0.010)	(0.000)	(0.000)	
Panel G: Propensity Score Matching (Excluding 'spillover' police officers and district covariates)						
Treatment effect	0.460***	0.241***	0.225***	0.219***	0.131**	0.255
	(0.000)	(0.000)	(0.001)	(0.002)	(0.025)	
Panel H: Propensity Score Matching (Excluding police officers of high rank)						
Treatment effect	0.472***	0.447***	0.421***	0.255***	0.223***	0.364
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	
Panel I: Propensity Score Matching (Excluding 'spillover' police officers, police officers of high rank and district covariates)						
Treatment effect	0.506***	0.256***	0.285***	0.240***	0.308***	0.319
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	

Note: Robust *p*-values in parentheses. Only the treatment effect is presented. Covariates are included. Sample size is 7,200 for specifications in Panels A to E (12 vignette cases à 600 respondents). The specification in Panel G contains 5,820 observations since all possible 'spillover' police officers (PARP-participants in non-PARP districts and vice versa) were excluded. The specification in Panel H contains 6,768 observations since all police officers of high rank were excluded. The specification in Panel I contains only 5,436 observations since all possible 'spillover' police officers (PARP-participants in non-PARP districts and vice versa) *and* all police officers of high rank were excluded. Details about the individual and police-experience related control variables that are included in all specifications (except Panel A) can be found in Section 7. In addition, the results presented in panels B, C and D contain the police infrastructure and district level controls detailed in Section 7. Instead of containing police infrastructure and district control variables, the specification in panel E contains district fixed effects. The specification in Panel F excludes the 'relocated' police officers and district control variables. The latter need to be excluded due to perfect multicollinearity. (***/**/* indicates significance at the 1/5/10% level, respectively.