

Gender and Policing Norms: Evidence from Framing Experiments among Police Officers in Uganda*

Natascha Wagner^{a,b}, Matthias Rieger^a, Arjun Bedi^a, Wil Hout^a

^aInternational Institute of Social Studies of Erasmus University Rotterdam, Kortenaerkade 12, 2518 AX Den Haag, Netherlands

^bCorresponding author: wagner@iss.nl

Abstract – Can the feminization of public services improve service quality and lower corruption? The underlying logic of such efforts is the belief that women have higher ethical standards than men. To answer this question, we examine the link between gender and policing norms using data from twelve vignette cases assessed by 600 Ugandan police officers. Our empirical strategy is based on a survey experiment that randomly varies the gender of police officers and victims/perpetrators presented in twelve hypothetical cases. Pooling the twelve cases we obtain a sample of 7,200 observations that allows us to control for individual officer characteristics, case and district fixed effects. We find that the gender of the police officer and the gender of the victim depicted in the cases are not related to the judgment of police malpractice, nor to suggested disciplinary measures. We also assess the association between respondent gender and policing norms. Male respondents tend to be stricter when assessing the hypothetical cases. However, male officers are slightly more lenient when there is a male rogue officer depicted in the vignette cases. The findings point to and complement descriptive evidence on power structures within the police, where men in leading positions are aware of the official rules and know whom to turn to for reports of inappropriate behavior. But gender perceptions do not differ between male and female police officers. Overall, the results indicate that simply feminizing the police force is unlikely to enhance service quality.

Keywords: Gender, Discrimination, Stereotyping, Police, Survey Experiments, Uganda

JEL Classification Codes: C90, J16, O12

1. Introduction

This paper studies gender, corruption and ethical judgments among police officers in Uganda using randomized survey experiments and vignette cases. We contribute novel evidence that sheds doubt on the assumption that *feminizing* public services, in itself, will improve the quality of services and lower corruption. Uganda is a particularly interesting case study since the majority of citizens consider the police the most corrupt institution in the country (Transparency International-Kenya, 2011; Flanary and Watt, 1999; Watt et al., 1999; Biddle et al., 1998). This has been consistently documented in surveys carried out in 1998, 2003 and 2005 (Commonwealth Human Rights Initiative, 2006b). Such high levels of corruption, in turn, raise the question of what factors might influence the quality of policing practice. This paper focuses on one such factor, namely the role of gender, and more specifically the gender of fellow police officers and the gender of victims and perpetrators.

Are gender and policing practice related? Indeed, the role of women in the police is often seen as a silver-bullet in the fight against corruption and crime (Gutierrez, 2003). The line of reasoning is that women are more trustworthy, competent, pro-social and respectful (Gray, 2013; Kahn, 2013; Sichel, 1978). Based on such reasoning, a number of initiatives have been implemented to induct women into law-enforcement activities. One such program that has received considerable international attention is the ‘Black Mambas Anti-Poaching Unit’.¹ This group of young South African women, hunting poachers in Kruger National Park, has received the UN's 2015 Champions of the Earth Award for Inspiration and Action (Wills, 2015). The United Nations Environment Programme (2015) claims that since the Black Mamba Anti-Poaching Unit started its work in 2013, “the number of rhinos lost to poaching has plummeted, snaring and illegal bush-meat incidents have been reduced by 75 per cent, and nine poacher incursions have been detected, leading to the arrests of the offenders.” For the case of Nigeria, where female police officers are also very common, Adebayo (2005) claims that female and older officers are more ‘ethical’. There is similar, anecdotal evidence on the feminization of the police in South America. The World Bank, in its 1994-2000 Transport Rehabilitation Project, championed Peru’s efforts to recruit female police officers to restore the image of the traffic police (Gutierrez, 2003). A decade after the launch of this feminization campaign, female police officers believe that they have contributed to a reduction in low-level corruption, although, they continue to identify corruption as the biggest problem of the Peruvian police (Karim, 2011). In a similar vein, in 1999, the police chief of Mexico City initiated a major anti-

¹ See the website of the group: <http://www.blackmambas.org/> [Last accessed January 10, 2017]

corruption initiative by setting up an all-female traffic police force (Moore, 1999).² While there is no systematic study of this initiative, casual evidence suggests that female police officers also ask for bribes (Gray, 2013; Kahn, 2013).

Existing evidence on the feminization of the police is intriguing, but purely anecdotal. Rigorous causal evidence on such feminization efforts is scant and it is unclear how much of the ‘success’ simply stems from the publicity and scrutiny that comes with such programs. In addition, much of the evidence stems from male-dominated societies and institutions, where women even if they enter male domains, tend to fill positions of lower rank and have less room to engage in fraudulent behavior (Prenzler and Sinclair, 2013).³

In this paper, we contribute novel evidence that sheds doubt on the assumption that feminizing public services, in itself, will improve the quality of policing and lower corruption. We contribute to the debate on gender and policing by examining the underlying logic of feminization campaigns that have been implemented in a number of developing countries, namely potentially gendered judgments and norms. The paper adds novel quantitative evidence based on an original survey among 600 police officers.

Surveying police officers, especially on gender issues is not straightforward. Ideally we would want to observe actual behavior of police officers in the field by gender of officer, victim and perpetrator. This poses several problems: First, it is not easy to observe and measure illegal policing behavior. Second, quantitative analysis requires many, relatively uniform cases. Third, the gender of the police officer, victim and perpetrator is likely to be endogenous to the case and circumstances. For instance, it is possible that male officers or victims are more likely to be involved in violent crime (Rowe et al., 1995) or female police officers are more likely to be deployed in cases of domestic violence and sexual assault. Alternatively, police officers could be directly asked about the role of gender in their daily work practice. This, however, raises concerns about over-reporting of socially expected behavior or under-reporting of bad practices in relation to gender.

To address these empirical concerns and add plausibly causal evidence to the literature, we make use of a survey experiment. Police officers were confronted with twelve hypothetical vignette cases carefully tailored to the local context. The cases ranged from traffic offences to

² Similar efforts have been undertaken in Mexico State, but the female police officers were not allowed to issue tickets, as their unit was not officially approved by the state (Kahn, 2013).

³ In Africa, female police officers are not uncommon. As early as in 1995, women constituted 18 percent of the total South African police which rose to 20 percent in 1998 (Rauch, 2001). In our representative survey from Uganda, 23 percent of the respondents are female. In 2011, in Uganda, 33 percent of the positions in public service were held by women (UNDP, 2012).

robbery and murder. In these cases we randomly changed the gender of the police officer depicted in the case and the gender of the victim/perpetrator.⁴ This exogenous variation allows us to estimate the impact of gender ‘framing’ in these cases. It is important to note that a number of measures were taken to ensure anonymity and reduce the possibility of providing socially desirable responses. For instance, police officers were unaware of the gender angle or framing of the study and the survey was self-administered by the officers. An additional novelty of the paper is that it is not a laboratory experiment with students, but that our experiment involves actual police officers.

Overall, we find that the cases are judged independently of the gender of the victim/perpetrator and the gender of the police agents depicted in the case. The estimated effect sizes are on average very small. Even if one is willing to assume that there is some sort of gender bias in the perception of female police officers and female victims, the average effect explains less than 5% of a standard deviation independent of the model we apply. We also check the robustness of our results with respect to social desirability bias using respondent fixed effects, as well as internal and external proxies for honest reporting. While we cannot fully rule out social desirability bias, it does not seem to drive our (null) findings.

In addition to the random gender framing depicted in the vignette cases, we also assess the role of respondent gender in assessing the cases. These relationships are interpreted as associations since respondent gender might be driven by gendered (self-) selection into the police. We find that male respondents are more critical than their female colleagues. For instance, they indicate a greater willingness to report fellow officers who engage in errant behavior and also appear to be more familiar with the official rules and regulations. We further interacted respondent gender with the gender framing treatments. We find that male officers are more lenient towards male, rogue colleagues depicted in the cases, potentially suggesting double standards within the police with female officers having to live up to higher expectations. We do not find similar patterns with regard to treatment of male victims by male respondents.

Our findings point to power structures within the police, where men in leading positions are aware of the official rules and know whom to turn to for reports of inappropriate behavior. To put our findings in context it is important to know that at the level of the Ugandan central police station, 79% of the administrative positions are held by men (Mutagubya, 2013). Because of their leading positions, male police officers are perhaps more aware of

⁴ We used common Ugandan male and female surnames such as Mary, Charlotte, Joseph, and James.

inappropriate behavior according to the official code of conduct. Women tend to hold lower positions, are less aware and more restricted in their responses when observing misbehavior among their colleagues. Overall, our results suggest that the effect of feminization campaigns on service quality is likely to be limited.

The rest of the paper is organized as follows. In section 2 we discuss existing evidence on gender differences in ethical judgments and perceptions. Section 3 describes the survey instruments and data. Section 4 outlines the framing of the survey experiment and the empirical strategy. Results are discussed in Section 5 while section 6 concludes the paper.

2. Existing Evidence on Gender Differences

While the body of work on gender and the quality of police services and corruption in the police is limited, there are a number of non-experimental and experimental general studies which have shown that women are the ‘better’ sex, i.e. women have been found to be less corrupt (Dollar et al., 2001; Swamy et al., 2001), more altruistic (Simmons and Emanuele, 2007; Eckel and Grossman, 1998), more trustworthy (Buchan et al., 2008), more peaceful (Gizelis, 2009; Caprioli, 2000) more egalitarian (Andreoni and Vesterlund, 2001), more likely to reciprocate (Croson and Buchan, 1999) and more likely to favor welfare spending (Gidengil, 1995; Welch and Hibbing, 1992).

However, the entire body of work on gendered attitudes and behavior does not unequivocally support the claim that women are inherently more pro-social and ethical. Women have been found to be neither more nor less socially oriented, but simply more sensitive in accounting for social conditions compared to men (Croson and Gneezy, 2009). When confronted with ethical dilemmas in the form of vignettes, women do not appear to have stronger ethical beliefs (Loo, 2003). With regard to gender and corruption, laboratory evidence suggests that women are not intrinsically more honest, but more opportunistic when they have a chance to break an implicitly corrupt contract. This results in lower corruption in mixed gender teams (Frank et al., 2011). Moreover, Alatas et al. (2009) conducted corruption experiments with students from four countries (Australia, India, Indonesia, and Singapore) and documented gender-differences in Australia, but not for the three Asian countries. The authors concluded that gender differences in attitudes towards corruption are not universal and thus, not only biological but also culturally formed. As opposed to lab experiments, Armantier and Boly (2008) compared gender differences in corruption amongst teachers in Burkina Faso. Experimental subjects had to grade 20 exams of which one exam paper came with a bribe. They

found that females and males were equally likely to accept a bribe, but women were more responsive to monitoring and punishment.

The accumulated evidence on women's willingness to engage in corrupt behavior suggests that contextual factors matter rather than gender *per se* (Esarey, and Chirillo, 2013; Alatas et al., 2009; Alhassan-Alolo, 2007; Schulze and Frank, 2003; Sung, 2003). For instance, Goetz (2007) argues that most of the findings associated with reduction in corruption and the representation of women are confined to liberal democracies, which are less corrupt in the first place. Additionally, it is a matter of opportunities to engage in corrupt practices and due to the gendered nature of access to politics and public office women have fewer opportunities to engage in corruption (Goetz, 2007).

3. Survey Instruments and Data

This paper uses data from a survey among 600 police officers in 10 districts of Uganda.⁵ The survey took place in April 2015 and was carried out as part of an impact evaluation of a police-civil society project known as Police Accountability and Reform Project (PARP) that was implemented to foster exchanges between civil society and the police.⁶ The project was carried out by the Human Rights Network Uganda (HURINET-U). The police officially participated as counterpart.

The Ugandan police force was institutionalized in 1906 (Uganda Police Force, 2007) and in April 2010 was officially named 'Uganda National Police' (Lumu, 2014). It is divided into 20 directorates based on tasks, and in regional and district offices, police stations and posts (Uganda Police, 2015). Reliable data on the police and policing activities is scant since the statistical capacity of the administration is not very strong (Uganda Police Force, 2007). There are roughly 128 officers per 100,000 inhabitants, which is substantially lower than the average in the European Union.⁷

Sampling was informed by the administrative organization of the police. Geographically, the police force is organized into 16 police regions, which are further sub-

⁵ The districts are Bushenyi, Iganga, Jinja Kabale, Kabarole, Tororo, Luwero, Mbarara, Mityana and Soroti.

⁶ Detailed information about PARP is available in Hout et al. (2016).

⁷ The average EU country had about 350 officers per 100,000 inhabitants in 2012. The EU member with the lowest number of officers is Finland with 151 officers per 100,000 inhabitants. For more details see EuroStat (2014), http://ec.europa.eu/eurostat/statisticsexplained/index.php/Crime_statistics#Further_Eurostat_information [Accessed October 13, 2015].

divided into police districts. Districts include police stations and posts. We picked ten districts at random, but never more than one district per region to ensure regional coverage. Within each of the selected districts, 60 police officers were selected and requested to participate in a self-administered pen and paper questionnaire in a classroom setting. Individual police officers were picked in a stratified way in collaboration with HURINET-U. To capture officers across all ranks, officers in leading positions at the regional level were identified and asked to participate in the survey. Similarly, at the district level, officers in a leadership position were purposively included in the survey. Police stations within districts were then randomly sampled with half the officers in our sample coming from stations that have no more than 10 officers and 70% of the officers are from stations with up to 25 officers. The day of the survey was arbitrarily picked and police officers from the local stations participated in the survey based on availability or presence. Since local police stations consist of a small number of officers, we do not expect any systematic selection of participants into our sample. The aim of our sampling procedure was to have a stratified sample of officers that represents the full spectrum of police work, functions, positions and hierarchies.

Substantial efforts were made to minimize the possibility of obtaining socially desirable responses. The survey was self-administered and during the survey each officer was provided enough personal space so as to ensure privacy and confidentiality. To ensure anonymity, officers were not asked to provide their names or addresses and hence it is not possible to identify the individual respondent. Furthermore, gender issues were not discussed and respondents were unaware of the gender framing of the vignettes.

The survey had two parts. In the first part, officers provided information on socio-economic characteristics. In the second and core part, officers were asked to review a series of twelve vignette cases that were adapted from Klockars et al. (2000). In collaboration with HURINET-U and the Uganda Police Force, the cases were changed to suit the local context, while irrelevant cases were deleted and new context-specific cases were developed.

The vignettes reflect real dilemmas and situations faced by the Ugandan police. In the survey, the cases were presented in no particular sequence to rule out ordering effects.⁸ We may group the cases into five categories: The *first group* of cases looks at the code of conduct among police officers themselves. The *second and largest group* of cases covers undue advantage, bribery and fraud. In particular, the least severe case depicts a situation where

⁸ The original ordering of the cases was as follows: Case 3, Case 4, Case 6, Case 9, Case 1, Case 2, Case 7, Case 11, Case 5, Case 10, Case 12, Case 8.

officers accept free meals and gifts. The next case (4) clearly mentions the word bribe and sets its value at half the amount of the official fine. Case 5 describes personal enrichment of about a daily wage. The sixth case describes theft, during a burglary investigation, of about a month's pay (see Appendix for details). The *third group* of cases contains situations, in which officers refuse to register a complaint against their colleagues. In the *fourth group*, we include two cases of severe crimes against individuals, which are not followed up by the police. Finally, the *fifth group* contains situations of undue force used by the police against suspects and demonstrators. For the purpose of this study, we further ordered the cases by severity within each of the five categories. This yields the following grouping of the twelve cases:

Group 1: Code of conduct among police officers

Case 1: Police mechanic repairing supervisor's car in exchange for holidays

Case 2: Police officer driving drunk and having an accident goes unreported by colleague

Group 2: Situations of undue advantage, bribery and fraud

Case 3: Acceptance of freely offered meals and small gifts while on duty

Case 4: Speeding not reported in exchange for a bribe

Case 5: Officer taking money from a found wallet

Case 6: Police officer stealing goods while investigating a burglary

Group 3: Refusal to register complaints

Case 7: Refusal to register a complaint and humiliation of the complainant

Case 8: Refusal to register a complaint and a one-week detention for the complainant for false accusation

Group 4: Reported severe crimes against individuals not followed up upon

Case 9: Police officer refusing to register wife beating

Case 10: Reported murder not being followed up on

Group 5: Undue force used by the police

Case 11: Foot patrol torturing a thief

Case 12: Brutal strike down of a demonstration

The exact wording of the cases and the questionnaires as administered may be found in the appendix and attachment to the article. In turn, after each case, the police officers answered the following questions that are used in the analysis:

1. *How serious do YOU consider this behavior to be?*

2. *Do you think YOU would report a fellow police officer who engaged in this behavior?*
3. *How serious do MOST POLICE OFFICERS IN YOUR OFFICE consider this behavior to be?*
4. *If an officer in your agency engaged in this behavior and was discovered doing so, what if any discipline do YOU think SHOULD follow?*
5. *Would this behavior be regarded as a violation of official policy in your agency?*

The possible answer categories range on a Likert scale: 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’ [1] to ‘definitely yes’ [5]. Question 4 on disciplinary measures that should follow the discovery of such wrongdoings included the categories ‘none’ [1], ‘verbal reprimand’ [2], ‘written reprimand’ [3], ‘period of suspension without pay’ [4], ‘demotion in rank’ [5] and ‘dismissal’ [6].

Using these twelve hypothetical cases we attempt to obtain an idea of the culture of policing in Uganda. The advantage of using vignettes is that the cases are identical across officers and since we do not ask about the officers’ actual behavior but rather their assessment of theoretical cases in a private and anonymized setting such an approach should be less likely to yield socially desirable responses.

Descriptive statistics of the respondent’s socio-demographic characteristics, as well as all cases and related responses are presented in Tables 1 and 2. Our sample of 600 police officers is representative of the gender-mix of the Ugandan Police Force as 77% of the respondents are male. The average age is almost 42 years, indicating that we interviewed experienced police officers. Concerning the household situation of the police officers, 84% are married, an equal share identifies themselves as household heads and they live in households with an average of between six and seven people; 75% of the officers have at least a secondary education. The large majority of the officers, 60%, indicate that they earn between 300,000 and 500,000 Ugandan Shilling corresponding to 81-136 US\$.⁹ Concerning their work characteristics, 46% of the officers have general duties and 26% are part of the investigative police, the remaining officers work for the traffic or intelligence unit or have other functions. According to the officers’ self-declarations, the majority of the officers namely 62% are of low rank, 32% are of medium rank and the remaining 6% are of high rank. The average years of service is 18.8 years. However, there is considerable variation in that variable as indicated by

⁹ We apply the exchange rate of 0.00027 US\$ per Ugandan Shilling.

a standard deviation of 10.6 years. Almost all the respondents (95%) indicated that they responded honestly to the questions

Results of the case assessments are presented in Table 2. The mean value as well as the share of respondents picking the two lowest and two highest scores are presented. None of the offences are seen as trivial. However, we do observe the expected patterns in terms of judging the severity of the cases. The first group of cases pertaining to the code of conduct among the police tends to be judged mildly. One fourth of the respondents are not very critical (answer categories 1 and 2) when a police mechanic repairs a supervisor's car in exchange for holidays or when an inebriated police officer has an accident and is not reported by a colleague. Similarly, the receipt of free meals and small gifts is perceived as a minor offence as indicated by the average score of 3.53. Acceptance of a bribe is considered a serious offence by 85% of the officers and the average score is 4.37. Taking a small bribe is judged more severely as compared to stealing the equivalent of a day's pay (score of 4.37 versus 4.04) from a lost wallet. This judgment seems to be linked to opportunities since it is far more likely for an officer to be tempted by a speeding bribe. Stealing goods, worth about a month's pay, after a burglary, receives the most critical assessment with 87% of the respondents judging this as a very serious offence.

Interestingly, even if officers do not judge an offence too harshly they appear to be more inclined to report police officers who engage in an offence. For instance, despite a relatively mild average judgment of 3.72 for a case where a police officer repairs a supervisor's car in exchange for holidays, the average score for reporting such an offence is 4.06. In all except two cases a police officer's own assessment of the severity of offences is higher than the reaction expected by fellow police officers. This suggests overconfidence in the adequacy of one's own behavior. The overestimation of own ability relative to others is a feature well documented in the literature (Dunning et al., 2004; Camerer and Lovallo, 1999; Svenson, 1981). Individuals often think that they are 'better' or above average.

Officers are aware that official police policy is likely to be more stringent than their own beliefs. In almost all cases the officers judge that their agency would take a more rigorous view of violations as compared to their own views. For instance, the average difference between own judgment and perceived agency policy is more than 0.6 for a police officer who

refuses to register a complaint by a former arrestee and reported murder that is not being followed up on.¹⁰

4. The Framing of the Survey Experiment and the Empirical Strategy

Our aim is to examine the role of gender in Ugandan policing practice using the hypothetical cases discussed in the previous section. Policing in general, but even more so in combination with gender, is a sensitive issue and asking explicit questions is unlikely to yield credible responses. To elicit responses which permit a credible analysis of gender differences, we designed a survey experiment where we randomly administered four versions of the questionnaires to the police officers. The four versions are identical except that in the vignette cases we varied the gender of the depicted police officers, who we will refer to as *rogue officers* in the rest of the paper, and we also varied the gender of the victim/perpetrator. Note that all officers faced male and female-framed cases. In other words, there is within question variation in framing. The randomized gender framing allows us to examine whether the assessment of the severity of the cases and potential ensuing actions depend on the gender of the rogue officer and the victim. Most importantly, the police officers themselves were not aware of the randomized gender framing. In other words, we can identify the causal effect (if any) of gender framing on the survey responses. Descriptive statistics assessing treatment balance, i.e. the balance of individual characteristics for the experimental gender framing, indicate that the randomization is well balanced across pre-determined observable characteristics of the respondents (Table 1, Columns 5 and 6). Balancing with respect to the rogue officer cannot be rejected at the 5% level for all except one pair-wise comparison. Balancing for the victim is rejected in two pair-wise comparisons. The remaining columns of Table 1 examine balance in terms of respondent gender, which we will discuss in more detail below.

To examine whether, on average, assessments of the severity of the cases and potential actions depend on the gender of the rogue officer or the gender of the victim, we pool all twelve cases. We estimate the average effect of the framings for each of the five questions pooled

¹⁰ Examining the raw data, we observed that police officers tend to judge misbehavior where colleagues gain undue advantages in a harsher manner than neglect of duty (“reported murder not being followed up”, see Table 2). For the latter case the focus does not seem to be on the explicit deed to be investigated, the murder, but rather on the general neglect of duty. In turn, the respondents indicated that they considered it as a personal disadvantage when colleagues gained considerable advantages from cheating the system (compare Table 2). We further observed in the raw data a tendency to distinguish between the inner circle of the police, i.e. the police officers themselves, and outsiders, i.e. criminals and victims. In order not to give too much weight to a single case we now focus on the results based on the pooled data including case fixed effects (Tables 3 and 4).

across the twelve cases, as well as derive an average effect across all five questions. The latter treatment effect indicates if the overall case assessments are influenced by such gender framings without cherry-picking possible effects on the five individual questions.

We employ a seemingly unrelated regression model where the outcome variable Y_{idcq} is the response provided by individual police officer i in district d to case c for one of the five questions $q=1, 2, \dots, 5$ which follow every vignette. The model then looks as follows:

$$Y_{idcq} = \beta_{1q} \text{rogue_officer}_{idcq} + \beta_{2q} \text{victim}_{idcq} + \beta_{3q} \text{rogue_officer}_{idcq} \times \text{victim}_{idcq} + \lambda_d + \delta_c + \varepsilon_{idcq}. \quad (1)$$

The gender of the rogue officer which randomly varies across vignettes is represented by $\text{rogue_officer}_{idcq}$ and the gender of the victim, which also randomly varies, is given by victim_{idcq} . We further included an interaction term between the gender of the rogue officer and the victim, $\text{rogue_officer}_{idcq} \times \text{victim}_{idcq}$. We coded our gender variables such that 1 identifies males and 0 females. We further include district λ_d and case δ_c fixed effects. The latter allows us to account for case-specific heterogeneity. Standard errors are clustered at the level of the individual respondent. The subscript q associated with the β coefficients indicates that we estimate effects associated with each of the five outcome variables. Following Clingingsmith, Khwaja and Kremer (2009), the five response variables were standardized to make them comparable in magnitudes. More importantly, we also calculate an average effect across all five questions to pre-empt concerns about multiple hypothesis testing. This average effect is an index based on the five coefficient estimates associated with the rogue officer, victim and the interaction term, respectively (compare Clingingsmith, Khwaja and Kremer, 2009).¹¹ The standard error of the average effect size depends on both the variances of each individual coefficient associated with the randomized police or victim gender as well as any covariances across the five survey questions. Therefore, a seemingly unrelated regressions framework is needed to test the cross-equation hypothesis that the average effect size consisting of the five coefficient estimates equals zero (Casey, Glennerster and Miguel, 2012). We opted for this approach since results associated with any question could potentially be due to chance. This is less likely when one simultaneously considers several related questions in an index.¹²

¹¹ We use the “avg_effect” command in STATA to estimate the model.

¹² Note that an earlier draft of the paper examined all 600 cases and outcomes separately using ordered logit models. We thank an anonymous referee for the helpful suggestion to pool cases. We originally calculated sample sizes to be large enough to detect odds ratios slightly larger than a ‘small’ effect of 1.5. We used the power

In a second specification we augment Equation (1) with individual level control variables Ind_{id} . The matrix Ind_{id} contains socio-economic characteristics such as age, marital status, education level, household size, whether the respondent is the head of the household and a set of dummy variables indicating the income band in which a respondent lies. In addition, we control for work related characteristics such as years of service, rank and the police section in which the respondent works (see list of covariates in Table 1).

In a third and final specification, we exploit the fact that we have multiple answers per individual. We can thus replace the individual level controls Ind_{id} and the district fixed effects λ_d with respondent fixed effects ζ_{id} .

4.1 Correlational Analysis with respect to Respondent Gender

The analysis of the survey experiment does not examine the effect of respondent gender on the assessment of the vignettes. Due to the possibility of gender specific self-selection into the police force the effect of respondent gender should not be imbued with a causal interpretation. Therefore, we analyze the effect of respondent gender in a separate, correlational analysis and also carry out a heterogeneity analysis by interacting respondent gender with the experimentally varied gender of the rogue officer and the victim.

To inform this correlational analysis and to examine the possibility of self-selection, we examine co-variate balance by respondent gender in Table 1, Columns 7 to 9. These balance tests show that covariates need to be included when analyzing associations with respect to respondent gender. Female officers tend to be younger and are considerably less likely to be married, they are also less likely to be household heads and tend to earn less than their male peers. All these differences are statistically significant at the 5% level suggesting that female officers are systematically different from their male counterparts. These differences are perhaps not surprising as married women are more likely to drop out of the police force and women are also less likely to be household heads. While there are differences between male and female police officers there are similarities as well. Women and men in the survey have similar education levels and come from similar types of households. By and large female officers also have similar duties within the police, with the exception that fewer women are on general duties (significant at 5%). But in line with their younger age female officers have gained less work

calculations proposed by Whitehead (1993) for proportional odds ordinal logistic model. As usual, we set power to 0.8 and type I error 0.05. When we assume equal proportions across the five answer categories, as well as equal group size, sample size is 597. Therefore, we opted for an estimation sample size of 600. We included co-variates to further enhance power. Since we are now pooling responses across the twelve cases, we have sufficient power to detect even smaller gender differences.

experience in the force. As a result of these differences, the interaction between respondent gender and the experimental framings is not always well-balanced (Table 1, Columns 10 and 11). This additional caveat has to be kept in mind when interpreting the heterogeneous effects of the gender framing as a function of respondent gender.

5. Results

5.1 Main results

The main results are reported in Table 3. The table features question-specific effects as well as mean effect sizes across the five questions (Kling and Liebman, 2004). All effects may be interpreted in terms of standard deviations. For instance, a coefficient of 0.1 implies that the effect size is one-tenth of a standard deviation for male versus female gendered frames. In what follows, we will first discuss findings for each question and then examine mean effects across questions.

We find that male rogue officers tend to get less strictly judged compared to female rogue officers (Table 3, Panel A, Column 1). However, the mean effect size of -0.072 suggests that less than one-tenth of the standard deviation is explained by the framing. Not only is the coefficient estimate small (in absolute) magnitude it is also statistically insignificant. If we control for individual characteristics along with case and district fixed effects we get a similar result with the coefficient estimate dropping in (absolute) magnitude to -0.067 (Table 3, Panel B, Column 1). In Panel C we control for individual fixed effects, which leads to a further reduction in (absolute) magnitude (-0.016). Taken together, while the coefficient estimate is negative, it is small in magnitude and statistically insignificant. In other words, the judgment of police officers is not influenced by the gender of the rogue officer depicted in the cases.

We observe the same pattern for randomly varying the gender of the victim depicted in the cases. The severity of cases depicting male victims is judged less critically with the effect size varying between -0.100 and -0.034 depending on the type of model we employ (Table 3, Column 1). Again, the magnitude of the estimated effect is small and none of the coefficient estimates is statistically significant.

We also assessed the interaction between officer gender and victim gender. The estimated effect is even smaller and unstable across models as the positive coefficient estimate found in Panels A and B turns negative in Panel C (Table 3, Panel A, Column 1).

Police officers were further asked to indicate how likely it is that they would report a colleague who engages in any of the behaviors presented in the cases (Table 3, Column 2).

Concerning the gender-sensitivity of the responses we find a similar pattern as for the evaluation of the severity of the cases. Across empirical models we find that respondents indicate that they are less likely to report male colleagues if engaged in any such behavior. However, the coefficient estimates are small with effect sizes ranging between -0.049 and -0.029. Again they are all statistically insignificant suggesting that neither male nor female police officers are more likely to get reported by their colleagues for the depicted misbehaviors. The coefficient associated with being a male victim is also negative and statistically insignificant across models. However, the coefficient associated with the interaction term (Male Rogue Officer \times Male Victim) is positive, moderately sized and statistically significant with a mean effect size of between 12.5 and 15% of the standard deviation in our outcome variable.

Next we turn to the officers' perception of the assessment of the severity of the cases by their fellow colleagues. We had asked "How serious do MOST POLICE OFFICERS IN YOUR OFFICE consider this behavior to be?" Again we find a negative, albeit small and statistically insignificant coefficient estimate associated with being a male rogue officer (Table 3, Column 3). The coefficient estimate becomes significant at the 10% level once we apply individual fixed effects. However, the effect size is small, namely -0.075 indicating that less than one-tenth of a standard deviation is explained. The picture looks identical for male victims with the model employing individual fixed effects being the only one that yields a statistically significant result at the 10% level. The interaction term between the male rogue officer and the male victim is again positive but insignificant across models.

With regard to disciplinary measures they would deem appropriate if an officer in their agency engaged in any of the depicted behaviors (Table 3, Column 4), we find similar patterns. That is, the coefficient estimates are small and none of them are statistically significant.

Finally the officers were asked whether the depicted behavior would be regarded as a violation of official policy in their agency (Table 3, Column 5). There is some moderate indication that it is less likely to be considered as violation of official policy if the rogue officer is male. However, this finding is only statistically significant with a meaningful magnitude when we employ individual fixed effects (-0.078). The found evidence is inconclusive for male victims as we see a sign reversal across models. It is positive, very small and statistically insignificant for the interaction term.

The last column of Table 3 presents a summary measure of the effect sizes across the five questions. Except for one of the coefficients, which is statistically significant at the 10% level, all other coefficient estimates are statistically insignificant at conventional levels. The

average effect size is small –even smaller than most of the question-specific effect sizes– and the overall impression conveyed by these estimates is that the gender of the rogue officer and the gender of the victim do not influence the assessment of the cases. Even if one is willing to assume that there is some sort of gender bias in the perception of female police officers and female victims, the found effect explains less than 5% of a standard deviation in the index consisting of all five questions independent of the model we apply.

5.2 Robustness checks

In the context of this study social desirability bias is a major concern. It is well-known that in self-reports on sensitive topics, there is a tendency to present oneself in the best possible way (Tourangeau and Yan, 2007). In other words, police officers might give “socially” expected answers. Potential underlying reasons are self-deception, impression management and other-deception. While we introduce the gender framing in a random fashion, social desirability can nevertheless affect the validity of experimental findings (Nederhof, 1985).

We took a number of steps to minimize the possibility of obtaining socially desirable responses. First, the survey was entirely self-administered by the police officers. We intentionally did not use interviewers so as to eliminate bias stemming from a desire to please an enumerator. Second, the questionnaires did not contain any information that would allow us to identify the respondents and the respondents were informed of this prior to responding to the questionnaire. Third, and perhaps most importantly, respondents were not sensitized towards gender issues and completely unaware of the gender framing of the vignettes. If we expect respondents to be dishonest independent of the gender framing we would simply obtain inflated standard errors. Since we are estimating average effects using a large sample of 7,200 respondent-case observations, we do not consider inflated standard errors a major problem. Moreover, estimated effect sizes (some of which are also statistically significant) tend to be “economically” small.

In addition to the survey-based measures we attempt to deal with social desirability in three ways. *First*, as shown in Table 3, Panel C, we present estimates with respondent specific fixed effects. We expect that respondent specific social desirability bias is at least to some extent absorbed by the fixed effects. Results are indeed stable and qualitatively similar (compare Panels A/B to C in Table 3). *Second*, we make use of information gathered from the survey. At the end of the questionnaire, police officers were asked the following two questions: 1) “Do you think most police officers give their honest opinion in assessing these hypothetical cases?” followed by 2) “Did you give your honest opinion?” As a first robustness check we

excluded those individuals from the analysis who indicated that they did not honestly report their opinion. Estimates for the average effect size across the five questions are presented in Table 4 while detailed question-specific results are found in the appendix (Table A1). Comparing the results of the full sample (Table 4, Column 1) with the ones from the robustness check (Table 4, Column 2) we draw qualitatively similar conclusions with regard to the size and significance of the coefficient estimates.

But of course, respondents might lie about their honesty. So in a second robustness check we make use of the perceived honesty of the other respondents. For every respondent, we calculate the perceived average honesty of the other officers in the district (excluding the respective respondent). This allows us to get a measure of the average honesty in the district. We then split our sample in districts with above and below median honesty and carry out the analysis on the subsample of respondents who work in districts with above median honesty. If there is social desirability bias, we expect to get more honest responses from districts that have a more honest policing environment. Again we assess whether the results from the full sample differ systematically compared to the results from this sub-sample analysis. Results are qualitatively similar and stable (see Table 4, compare columns 1 and 3). These patterns suggest that honesty is balanced across gender framings.

Third, we obtained information on district level crime rates over the period 2007-2013. We calculate the average crime level per district over that period and split our sample of police officers at the median crime level. We hypothesize that officers in high crime environments might be more corruptible and therefore restrict our sample to districts with below average crime levels. Again our sub-sample estimates are qualitatively similar to the full sample estimates indicating that results are not driven by officers from low crime environments (see Table 4, compare columns 1 and 4).

While the survey set-up and the empirical findings from the sub-sample analyses include attempts to tackle social desirability bias, we cannot objectively measure honesty and therefore we cannot fully rule out the effects of social desirability. This caveat needs to be kept in mind when interpreting our results.

5.3 Correlations with respondent gender and descriptive heterogeneity analysis

This section provides estimates of the link between the gender of the respondent (see Table 5) and the responses to the five post-vignette questions. Panel A presents baseline results controlling for respondent characteristics, gender framings and their interactions, as well as

case and district fixed effects. Panel B adds interactions of respondent gender with the experimental gender framings.

In the base models, the coefficients associated with respondent gender lie between 13.8 and 22.4% of the standard deviation of the outcome variable (Panel A). Estimates are statistically significant for four of the five questions (at the 10% level). The average estimate across questions amounts to 12% and is significant at the 5% level. Across the five questions we find considerably larger (standardized) coefficient estimates as compared to the coefficients associated with the randomly framed gender depicted in the cases. The results indicate that male police officers judge the behavior depicted in the cases more critically than their female counterparts (column 1), and they are also more likely to indicate that they would report a fellow colleague who engages in any of the depicted misbehaviors. Male officers indicate that stricter disciplinary measures should follow and they are more likely to consider the behavior depicted in the cases a violation of official policy in their agency. The only question that does not yield statistically different responses between male and female police officers concerns the judgment of other police officers (column 3). However the coefficient is also positive but smaller in magnitude compared to the other outcomes.

In a next step, we interact respondent gender with the gender framings depicted in the cases. Results are presented in Panel B of Table 5. Similar to the estimates in Panel A we find a positive link between respondent gender and the five post-vignette questions. In addition, across all the questions, although not always statistically significant, we find a negative link between respondent gender and the gender of the rogue officers depicted in the cases. This pattern suggests that while male officers are more critical and have higher ethical standards they tend to be more lenient towards their male colleagues. The average size of the coefficient across the five questions is -0.085 of a standard deviation (significant at the 5% level; see Column 6, Panel B, Table 5). Thus, on average male officers seem to be more willing to cover male colleagues. In contrast, we do not detect a coherent correlation of case assessments with the interaction term between respondent gender and victim gender. Across questions the coefficient associated with this interaction term changes signs and it is not statistically significant. This result indicates that victims are treated independently of their gender.

6. Discussion and Conclusion

This research contributes new evidence to the literature on gender perceptions and stereotyping. In particular, we add evidence to the debate on the quality of public service provision and

gender in the sensitive work environment of the police. Earlier studies about gender bias tended to treat gender as a residual category or black box, without disentangling the underlying institutional and cultural factors that shape gendered realities (Dollar et al., 2001; Swamy et al., 2001). In the meantime many gender myths have been repudiated. Even the widely held belief that women are more emotional has been challenged (Feldman Barrett et al., 1998).

We examined gender perceptions in the context of the Ugandan police force with a survey among 600 Ugandan police officers. The officers were exposed to hypothetical cases of policing behavior within a survey experiment that randomly varied the gender of the officers and the victims depicted in the cases. We found that gender framing does not influence responses and conclude that misbehavior by female police officers is perceived in a similar way as misbehavior by male police officers. Similarly, there is no difference in the treatment of male and female victims.

In contrast, correlations between actual respondent gender and our five post-vignette questions suggests that male police officers are more critical and that they are more likely to indicate that they would report a fellow colleague who misbehaves. Furthermore, their responses indicate that stricter disciplinary measures should follow and they are more likely to state that the behavior depicted in the cases is a violation of official policy in their agency. At the same time the evidence suggests that male respondents are more lenient to the male rogue officers depicted in the cases. However these differences may, *in part*, be driven by gendered self-selection into the police force since we do not observe male and female police officers with similar individual (observable) characteristics.

It is also possible that the relative position of men and women in the Ugandan police hierarchy is a plausible explanation of these empirical patterns with respect to respondent gender. According to a report put together by the Centre for Women in Governance (2010) only two of the top 14 positions in the Ugandan police force were held by women. In our sample, only 8.3% of high-ranked posts are held by women. Given the Ugandan context it is perhaps not surprising that female police officers, who are in the minority, are less likely to report misbehavior as it essentially means filing complaints about male colleagues to a male-dominated leadership. Needless to say, female police officers might also be less familiar with the official rules and regulations of the agency. Nevertheless, our research supports the idea that it is naïve to think that women are inherently more ethical or will be more reliable police officers.

How much do such case assessments by police officers tell us about actual policing practices? While this question has not been answered in the context of police work, the medical

literature makes considerable use of vignette techniques and has documented the consistency between hypothetical cases and actual behavior for both doctors' diagnostic skills and patients' tendency to consult a doctor (Peabody et al., 2000; Van der Meer and Mackenbach, 1998). Vignette cases and virtual patients are seen as a way of training and assessing medical skills, which ultimately have a bearing on the life and death of patients (Triola et al., 2006).

Future work on gender and police could perhaps employ both vignettes and direct observation to assess consistency between the two approaches or a combination of paper vignettes with a more behavioral approach in the form of hypothetical plays involving actors who engage in the kinds of behavior depicted in the vignettes as is extensively used in medical research (Das et al., 2008; Gibbons et al., 2002; Barrows, 1993). Finally it is worth pointing out that the methods in this paper lend themselves to study behavioral responses to other sensitive topics such as discrimination on grounds of sexual orientation.¹³

References

- Adebayo, Dada O. (2005). Gender and Attitudes toward Professional Ethics. *African Security Review* 14(2): 93-100.
- Andreoni, James, and Lise Vesterlund. (2001). Which is the fair sex? Gender differences in altruism. *Quarterly Journal of Economics* 116(1): 293–312.
- Alatas, Vivi, Lisa Cameron, and Ananish Chaudhuri. (2009). Gender, Culture, and Corruption: Insights from an Experimental Analysis. *Southern Economic Journal* 75(3): 663–80.
- Alhassan-Alolo, Namawu. (2007). Gender and Corruption: Testing the New Consensus. *Public Administration and Development* 237(3): 227–37.
- Armantier, Olivier, and Amadou Boly. (2008). Can corruption be studied in the lab? Comparing a field and a lab experiment. CIRANO-Scientific Publications 2008s-26.
- Biddle, K., C. Male and S. Neema. (1998). Uganda police project evaluation. Evaluation Department, Department for International Development, UK.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. Chichester, West Sussex, UK: Wiley.
- Buchan, Nancy R., Rachel T.A. Croson and Sara Solnick. (2008). Trust and gender: An examination of behavior and beliefs in the Investment Game. *Journal of Economic Behavior & Organization* 68(3–4): 466–476.
- Camerer, C.F. and D. Lovallo. (1999). Overconfidence and excess entry: An experimental approach. *American Economic Review* 89 (): 306–318.
- Caprioli, Mary. (2000). Gendered Conflict. *Journal of Peace Research* 37: 51-68.

¹³ The established and widely used list questions inferring discriminating behavior against ethnical or sexual minorities may be adapted as well (Rayburn et al., 2003; Kuklinski et al., 1997a; Kuklinski et al., 1997b).

- Casey, Katherine, Rachel Glennerster, and Edward Miguel. (2012). Reshaping Institutions: Evidence on Aid Impacts Using a Preanalysis Plan. *The Quarterly Journal of Economics* 127 (4): 1755-1812.
- Centre for Women in Governance. (2010). Uganda UNSCR 1325 Monitoring Report.
- Clingingsmith, David, Asim Ijaz Khwaja and Michael Kremer. (2009). Estimating the Impact of The Hajj: Religion and Tolerance in Islam's Global Gathering. *The Quarterly Journal of Economics* 124 (3): 1133-1170.
- Commonwealth Human Rights Initiative. (2006a). A Review of the Uganda Police Force Budget and its Effects on Crime Management. Report.
- Commonwealth Human Rights Initiative. (2006b). The police, the people, the politics: Police accountability in Uganda. Report.
- Croson, Rachel and Nancy Buchan. (1999). Gender and Culture: International Experimental Evidence from Trust Games. *American Economic Review* 89(2): 386-391.
- Croson, Rachel and Uri Gneezy. (2009). Gender Differences in Preferences. *Journal of Economic Literature* 47(2): 448-74.
- Das, Jishnu, Jeffrey Hammer, and Kenneth Leonard. (2008). The Quality of Medical Advice in Low-Income Countries. *Journal of Economic Perspectives* 22(2): 93-111.
- Dollar, David, Raymond Fisman, and Roberta Gatti. (2001). Are Women Really the 'Fairer' Sex? Corruption and Women in Government. *Journal of Economic Behavior & Organization* 46 (4): 423-29.
- Dunning, David, Chip Heath and Jerry M. Suls. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychological Science in the Public Interest* 5(3): 69-106.
- Eckel, C.C. and P.J. Grossman. (1998). Are women less selfish than men? Evidence from dictator experiments. *Economic Journal* 108: 726-735.
- Esarey, Justin and Gina Chirillo. (2013). 'Fairer Sex' or Purity Myth? Corruption, Gender, and Institutional Context. *Politics & Gender* 9: 361-389.
- Feldman Barrett, Lisa, Lucy Robin, Paula R. Pietromonaco and Kristen M. Eyssell. (1998). Are Women the 'More Emotional' Sex? Evidence From Emotional Experiences in Social Context. *Cognition and Emotion* 12(4): 555-578.
- Flanary, Rachel and David Watt. (1999). The state of corruption: A case study of Uganda. *Third World Quarterly* 20(3): 515-536.
- Frank, Björn, Johann Graf Lambsdorff and Frédéric Boehm. (2011). Gender and Corruption: Lessons from Laboratory Corruption Experiments. *European Journal of Development Research* 23: 59-71.
- Gibbons, Susanne W., Graceanne Adamo, Diane Padden, Richard Ricciardi, Marjorie Graziano, Eugene Levine and Richard Hawkins. (2002). Clinical evaluation in advanced practice nursing education: Using standardized patients in health assessment. *Journal of Nursing Education* 41(5): 215-221.
- Gidengil, Elisabeth. (1995). Economic Man, Social Woman: The Case of the Gender Gap in Support of the Canada-United States Free-Trade Agreement. *Comparative Political Studies* 23(3): 384-408.

- Gizelis, Theodora-Ismene. (2009). Gender Empowerment and United Nations Peacebuilding. *Journal of Peace Research* July 46: 505-523.
- Goetz, A. M. (2007). Political Cleaners: Women as the New Anti-Corruption Force. *Development and Change* 38(1): 87-105.
- Gray, Madison. (2013). How to Curb Police Corruption? Hire More Female Cops. *Time*. September 30, 2013.
- Gutierrez, Maria. (2003). Integrating Gender into World Bank Financed Transport Programs – Case Study. Report 34495, World Bank.
- Hout, Wil, Ria Brouwers, Jonathan Fisher, Rose Namara, Lydeke Schakel and Natascha Wagner. (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.
- Kahn, Carrie. (2013). Mexican State's Anti-Corruption Plan: Hire Female Traffic Cops. *National Public Radio (NPR)*. September 28, 2013.
- Kakamwa, Charles (2014). Uganda to Increase Size of Police Force. *New Vision*, December 26, 2014.
- Karim, Sabrina. (2011). Madame Officer. *Americas Quarterly* 5(3).
- Kling, Jeffrey R., and Jeffrey B. Liebman, (2004). Experimental Analysis of Neighborhood Effects on Youth. NBER Working Paper.
- Klockars, C.B., S.K. Ivkovich, W.E. Harver, and M.R. Haberfeld. (2000). The Measurement of Police Integrity, Research in Brief, NCJ 181465, Rockville: National Institute of Justice.
- Kuklinski, J. H., M. D.Cobb, M. and Gilens. (1997a). Racial attitudes and the New South. *Journal of Politics* 59(2): 323-349.
- Kuklinski, J. H., P. M. Sniderman, K. Knight, T. Piazza, P. E. Tetlock, G. R. Lawrence B. and Mellers. (1997b). Racial prejudice and attitudes toward affirmative action. *American Journal of Political Science* 41(2): 402-419.
- Loo, Robert. (2003). Are women more ethical than men? Findings from three independent studies. *Women in Management Review* 18(4): 169–181.
- Lumu, David. (2014). Uganda Police changes name. *New Vision*, Kampala. <http://www.newvision.co.ug/news/654457-uganda-police-changes-name.html> (Accessed September 30, 2015).
- Moore, Molly. (1999). Mexico City's Stop Sign to Bribery; To Halt Corruption, WomenTraffic Cops Replace Men. *The Washington Post*. July 31, 1999. <http://www.highbeam.com/doc/1P2-605613.html> (Accessed September 9, 2015).
- Mutagubya, Nelson. (2013). A report on Field attachment at Central Police Station, Kampala. BA thesis. College of Humanities and Social Science, Makerere University,
- Nederhof, A. J. (1985). Methods of coping with social desirability bias: a review. *European Journal of Social Psychology*, 15: 263-280.
- Peabody, John W., Jeff Luck, Peter Glassman, Timothy R. Dresselhaus and Martin Lee. (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.
- Prenzler, Tim and Georgina Sinclair. (2013). The status of women police officers: an international review. *International Journal of Law, Crime and Justice* 41(2): 115–131.
- Rauch, Janine. (2001). Police Reform and South Africa’s Transition. Proceedings of an international conference [30 August-September 1, 2000] on Crime and policing in transitional societies.
- Rayburn, N. R., M. Earleywine, and G. C. Davison. (2003). An investigation of base rates of anti-gay hate crimes using the unmatched-count technique. *Journal of Aggression, Maltreatment and Trauma* 6: 137-152.
- Rowe, David, Alexander Vazsonyi, Daniel Flannery. (1995). Sex Differences in Crime: Do Means and Within-Sex Variation Have Similar Causes? *Journal of Research in Crime and Delinquency* 32: 84–100.
- Schulze, Günther G. and Björn Frank. (2003). Deterrence versus Intrinsic Motivation: Experimental Evidence on the Determinants of Corruptibility. *Economics of Governance* 4(2): 143–60.
- Sichel, Joyce L. (1978). Women on Patrol: A Pilot Study of Police Performance in New York City.
- Simmons, Walter O. and Rosemarie Emanuele. (2007). Male-female giving differentials: are women more altruistic?. *Journal of Economic Studies* 34(6): 534–550.
- Sung, Hung-En. (2003). Fairer Sex or Fairer System? Gender and Corruption Revisited *Social Forces* 82(2): 703-723.
- Svenson, Ola. (1981). Are We All Less Risky and More Skillful Than Our Fellow Drivers?. *Acta Psychologica* 47(2): 143-148.
- Swamy, Anand, Stephen Knack, Young Lee and Omar Azfar. (2001). Gender and corruption. *Journal of Development Economics* 64(1): 25–55.
- Transparency International-Kenya. (2011). The East African Bribery Index 2011.
- Triola, M., H. Feldman, A.L. Kalet, S. Zabar, E.K. Kachur, C. Gillespie, M. Anderson, C. Griesser and M. Lipkin. (2006). A randomized trial of teaching clinical skills using virtual and live standardized patients. *Journal of general internal medicine* 21(5): 424–429.
- Tourangeau, Roger and Ting Yan. (2007). Sensitive questions in surveys. *Psychological Bulletin* 133(5): 859–883.
- Uganda Police. (2015). Directorates.
- Uganda Police. (2013). Annual Crime and Traffic/Road Safety Report.
- Uganda Police Force. (2007). Strengthening Statistics for Development Planning. Sector Strategic Plan for Statistics (2006/07 - 2010/11). Report.
- United Nations Development Programme (UNDP). (2012). Gender Equality and Women’s Empowerment in Public Administration – Uganda Case Study.

- United Nations Environment Programme (UNEP). (2015). Black Mamba APU – Anti-Poaching Unit, Balule Nature Reserve, South Africa. 2015 Champion of the Earth Inspiration and Action. <http://web.unep.org/champions/laureates/2015/black-mamba-apu> (Accessed September 29, 2015).
- Van der Meer, J. and J. Mackenbach J. (1998). Low education, high GP consultation rates: the effect of psychosocial factors. *Journal of Psychosomatic Research* 44: 587–597.
- Watt, David, Rachel Flanary and Robin Theobald. (1999). Democratisation or the democratisation of corruption? The case of Uganda. *Journal of Commonwealth & Comparative Politics* 37(3): 37-64.
- Welch, Susan, and John Hibbing. (1992). Financial Conditions, Gender, and Voting in American National Elections. *Journal of Politics* 54: 197–213.
- Whitehead J. (1993). Sample size calculations for ordered categorical data. *Stat in Med* 12: 2257–2271.
- Wills, Kate. (2015). Kruger National Park: Meet the women hunting South Africa's poachers – A new anti-poaching unit comprised almost entirely of women is proving remarkably adept at stalking the stalkers. *The Independent*. September 28, 2015. <http://www.independent.co.uk/news/world/africa/kruger-national-park-meet-the-women-hunting-south-africas-poachers-10514587.html> (Accessed September 29, 2015).
- Xinhua News Agency. (2007). Ugandan Police Force Undergoes Massive Overhaul Ahead of Commonwealth Summit. , July 1, 2007.

Tables

Table 1: Descriptive statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
					Gender framing balance		Gender framing balance by respondent gender			Interaction: Respondent gender and framing	
	Mean	SD	Min	Max	Rogue Officer <i>p</i> -value Δ	Victim <i>p</i> -value Δ	Male Mean	Female Mean	<i>p</i> -value Δ	Rogue Officer \times Male Respondent <i>p</i> -value Δ	Victim \times Respondent <i>p</i> -value Δ
<i>Socio-economic characteristics</i>											
Male respondent	0.77		0	1	0.68	0.87					
Age in years	41.79	9.42	21	60	0.98	0.87	42.71	38.66	0.00	0.00	0.00
Married	0.84		0	1	0.18	0.61	0.91	0.64	0.00	0.00	0.00
Household size	6.67	3.99	1	25	0.74	0.55	6.75	6.41	0.30	0.44	0.42
Secondary education or more	0.75		0	1	0.98	0.54	0.75	0.75	0.83	0.68	0.99
Household head	0.84		0	1	0.85	0.86	0.94	0.53	0.00	0.00	0.00
Income group 1: Less than 100,000 UGX	0.03		0	1	0.33	0.65	0.04	0.01	0.02	0.03	0.08
Income group 2: 100,000-200,000 UGX	0.02		0	1	0.12	0.13	0.02	0.03	0.36	0.15	0.18
Income group 3: 200,000-300,000 UGX	0.12		0	1	0.02	0.02	0.10	0.17	0.05	0.01	0.01
Income group 4: 300,000-500,000 UGX	0.60		0	1	0.14	0.26	0.59	0.65	0.20	0.70	0.48
Income group 5: 500,000-700,000 UGX	0.14		0	1	1.00	0.59	0.15	0.12	0.53	0.74	0.48
Excluded category: Income group 6: More than 700,000 UGX											
<i>Function in the police</i>											
Traffic police	0.04		0	1	0.52	0.53	0.04	0.04	0.98	0.98	0.98
Investigative police	0.26		0	1	0.07	0.01	0.25	0.31	0.13	0.01	0.00
Intelligence	0.06		0	1	0.31	0.91	0.07	0.04	0.08	0.06	0.15
General duties	0.46		0	1	0.55	0.58	0.49	0.39	0.04	0.03	0.03
Excluded category: Other function											
<i>Rank and experience</i>											
Low rank	0.62		0	1	0.48	0.52	0.61	0.64	0.51	0.91	0.80
Medium rank	0.32		0	1	0.41	0.41	0.32	0.34	0.69	0.41	0.35
Excluded category: High rank											
Years of service	18.80	10.55	0	48	0.87	0.65	19.30	17.12	0.03	0.15	0.25
Honest reporting	0.95		0	1	0.19	0.06	0.95	0.94	0.77	0.48	0.41

Note: Summary statistics of pooled analysis sample (600 respondents \times 12 vignette cases). Standard errors are clustered at the respondent level.

Table 2: Case summary statistics – Means and proportions of responses

	How serious do YOU consider this behavior to be?			Do you think YOU would report a fellow police officer who engaged in this behavior?			How serious do MOST POLICE OFFICERS IN YOUR OFFICE consider this behavior to be?			If an officer in your agency engaged in this behavior and was discovered doing so, what if any discipline do YOU think SHOULD follow?			Would this behavior be regarded as a violation of official policy in your agency?		
	Mean	Bottom of scale	Top of scale	Mean	Bottom of scale	Top of scale	Mean	Bottom of scale	Top of scale	Mean	Bottom of scale	Top of Scale	Mean	Bottom of scale	Top of scale
<i>Group 1: Code of conduct among the police officers</i>															
1. Police mechanic repairing supervisor's car in exchange for holidays	3.72	0.25	0.67	4.06	0.26	0.54	3.69	0.22	0.63	3.80	0.17	0.54	4.24	0.14	0.81
2. Police officer driving drunk and having an accident goes unreported by colleague	3.71	0.26	0.65	3.92	0.32	0.50	3.60	0.26	0.58	3.65	0.23	0.52	4.11	0.17	0.77
<i>Group 2: Situations of undue advantage, bribery and fraud</i>															
3. Acceptance of freely offered meals and small gifts while on duty	3.53	0.28	0.59	3.17	0.38	0.45	3.45	0.30	0.55	3.15	0.37	0.33	3.96	0.20	0.72
4. Speeding not reported in exchange for a bribe	4.37	0.13	0.85	4.25	0.26	0.54	3.92	0.20	0.70	4.31	0.06	0.70	4.54	0.09	0.89
5. Officer taking money from a found wallet	4.04	0.20	0.75	4.27	0.25	0.56	3.80	0.21	0.65	4.11	0.13	0.64	4.43	0.10	0.86
6. Police officer stealing goods while investigating a burglary	4.47	0.11	0.87	4.60	0.15	0.69	4.33	0.10	0.83	5.07	0.04	0.88	4.64	0.06	0.92
<i>Group 3: Refusal to register complaints</i>															
7. Police officer refusing to register a complaint by a former arrestee	3.62	0.28	0.65	4.05	0.26	0.57	3.54	0.26	0.57	3.48	0.23	0.42	4.31	0.12	0.83
8. Refusal to register complaint instead keeping victim in detention	4.01	0.27	0.69	4.47	0.10	0.83	3.97	0.22	0.71	4.12	0.11	0.62	4.56	0.09	0.89
<i>Group 4: Reported severe crimes against individuals not followed up upon</i>															
9. Police officer refusing to register wife beating	4.28	0.16	0.83	4.47	0.15	0.70	4.13	0.14	0.79	4.01	0.11	0.58	4.59	0.06	0.90
10. Reported murder not being followed up	3.92	0.24	0.73	4.47	0.07	0.86	4.10	0.15	0.77	2.73	0.67	0.25	4.53	0.08	0.89
<i>Group 5: Undue force used by the police</i>															
11. Foot patrol torturing a thief	3.9	0.22	0.71	4.08	0.25	0.54	3.78	0.21	0.66	3.88	0.18	0.57	4.35	0.12	0.84
12. Brutal strike down of a demonstration	4.12	0.20	0.78	4.05	0.20	0.75	4.24	0.14	0.81	3.92	0.32	0.63	4.51	0.10	0.88

Note: N=600. Bottom of scale refers to responses 1 and 2 on the 1-5 (1-6) point Likert scale while top of scale refers to responses 4 (4) and 5 (6) on the 1-5 (1-6) point Likert scale (in the case of the disciplinary measure asked for in question 4).

Table 3: Main results

	(1)	(2)	(3)	(4)	(5)	(6)
	How serious do YOU consider this behavior to be?	Do you think YOU would report a fellow police officer who engaged in this behavior?	How serious do MOST POLICE OFFICERS IN YOUR OFFICE consider this behavior to be?	If an officer in your agency engaged in this behavior and was discovered doing so, what if any discipline do YOU think SHOULD follow?	Would this behavior be regarded as a violation of official policy in your agency?	<i>Average Effect [Columns 1-5]</i>
Panel A: District and case fixed effects						
Male Rogue Officer	-0.072 (0.063)	-0.049 (0.049)	-0.056 (0.055)	-0.020 (0.050)	-0.020 (0.049)	-0.031 (0.025)
Male Victim	-0.100 (0.084)	-0.078 (0.052)	-0.082 (0.070)	-0.024 (0.060)	0.033 (0.057)	-0.034 (0.029)
Male Rogue Officer × Male Victim	0.057 (0.105)	0.157* (0.070)	0.051 (0.091)	0.040 (0.078)	0.021 (0.075)	0.048 (0.038)
Panel B: Respondent covariates, as well as district and case fixed effects						
Male Rogue Officer	-0.067 (0.061)	-0.038 (0.047)	-0.044 (0.054)	-0.010 (0.049)	-0.014 (0.049)	-0.025 (0.024)
Male Victim	-0.091 (0.081)	-0.062 (0.051)	-0.069 (0.069)	-0.010 (0.061)	0.043 (0.057)	-0.025 (0.028)
Male Rogue Officer × Male Victim	0.046 (0.102)	0.141* (0.069)	0.039 (0.090)	0.025 (0.078)	0.009 (0.074)	0.038 (0.037)
Panel C: Respondent and case fixed effects						
Male Rogue Officer	-0.016 (0.044)	-0.029 (0.039)	-0.075+ (0.043)	-0.022 (0.041)	-0.078* (0.037)	-0.033+ (0.018)
Male Victim	-0.034 (0.052)	-0.042 (0.048)	-0.099+ (0.053)	-0.037 (0.058)	-0.004 (0.041)	-0.031 (0.022)
Male Rogue Officer × Male Victim	-0.003 (0.075)	0.125+ (0.065)	0.081 (0.072)	0.057 (0.071)	0.037 (0.061)	0.044 (0.032)

Note: Sample size is 7,200 (12 vignette cases × 600 respondents). Respondent covariates are listed in Table 1. Standard errors are clustered at the respondent level. + and * denote statistical significance at the 10% and 5%, respectively. All models are estimated using the command "avg_effect" in STATA based on Kling et al. (2004) and Clingingsmith et al. (2009), calculating average standardized effect sizes within a seemingly-unrelated regression model to adjust for covariance across estimates.

Table 4: Robustness checks

	(1)	(2)	(3)	(4)
<i>Average Effects of Gender Framing on Case Evaluations</i>				
Male Rogue Officer	-0.031 (0.025)	-0.024 (0.025)	-0.036 (0.029)	-0.026 (0.037)
Male Victim	-0.034 (0.029)	-0.035 (0.029)	-0.034 (0.032)	-0.075+ (0.040)
Male Rogue Officer × Male Victim	0.048 (0.038)	0.055 (0.039)	0.038 (0.043)	0.052 (0.054)
Analysis Sample	Full (repeated from Panel A, Table 3)	Only respondents included who report to have answered the questions honestly	Average opinion on the honesty of other officers in the district > sample median (excl. the respondent)	Respondents working in districts with crime rate < sample median

Note: Sample size is 7,200 in column 1 (12 vignette cases × 600 respondents), 6,816 in column 2, and 3,600 in columns 2-4. Respondent covariates are listed in Table 1. District and case fixed effects included in all models. Standard errors are clustered at the respondent level. + and * denote statistical significance at the 10% and 5%, respectively. All models are estimated using the command "avg_effect" in STATA based on Kling et al. (2004) and Clingingsmith et al. (2009), calculating average standardized effect sizes within a seemingly-unrelated regression model to adjust for covariance across estimates.

Table 5: Correlations with respondent gender and descriptive heterogeneity analysis

	(1)	(2)	(3)	(4)	(5)	(6)
	How serious do YOU consider this behavior to be?	Do you think YOU would report a fellow police officer who engaged in this behavior?	How serious do MOST POLICE OFFICERS IN YOUR OFFICE consider this behavior to be?	If an officer in your agency engaged in this behavior and was discovered doing so, what discipline do YOU think SHOULD follow?	Would this behavior be regarded as a violation of official policy in your agency?	<i>Average Effect [Columns 1-5]</i>
Panel A: Respondent covariates, gender framing treatments and their interaction, as well as district and case fixed effects						
Male Respondent	0.201+ (0.117)	0.224* (0.093)	0.138 (0.101)	0.169+ (0.096)	0.171+ (0.093)	0.123* (0.043)
Panel B: Adding interactions of respondent gender and gender framing treatments to the specification in Panel A						
Male Respondent	0.225+ (0.122)	0.236* (0.095)	0.200* (0.101)	0.276* (0.100)	0.209* (0.099)	0.156* (0.045)
Male Respondent × Male Rogue Officer	-0.073 (0.090)	-0.117 (0.082)	-0.112 (0.079)	-0.219* (0.079)	-0.080 (0.072)	-0.085* (0.037)
Male Respondent × Male Victim	0.006 (0.097)	0.109 (0.081)	-0.085 (0.091)	-0.108 (0.084)	-0.036 (0.075)	-0.015 (0.036)

Note: Sample size is 7,200 (12 vignette cases × 600 respondents). Respondent covariates are listed in Table 1. Standard errors are clustered at the respondent level. + and * denote statistical significance at the 10% and 5%, respectively. All models are estimated using the command "avg_effect" in STATA based on Kling et al. (2004) and Clingingsmith et al. (2009), calculating average standardized effect sizes within a seemingly-unrelated regression model to adjust for covariance across estimates.

Appendix

The survey experiment made use of four “gendered” versions of below cases. The four questionnaires as administered are attached to this article.

The 12 cases

Group 1: Code of conduct among the police officers

Case 1: A police officer, who happens to be a very good auto mechanic, is scheduled to work during coming holidays. The supervisor Catherine offers to give him these days off, if he agrees to repair her personal car. Evaluate the supervisor’s behavior.

Case 2: At 2:00 a.m., a police officer, who is on duty, is driving his patrol car on a deserted road. He sees a vehicle that has been driven off the road and is stuck in a ditch. He approaches the vehicle and observes that the driver is not hurt but is obviously drunk. He also finds that the driver is a police officer. He transports the driver to her home. Evaluate the behavior of the police officer on duty.

Group 2: Situations of undue advantage, bribery and fraud

Case 3: A police officer routinely accepts free meals, cigarettes, and other items of small value from merchants on his duty. She does not solicit these gifts and is careful not to abuse the generosity of those who give gifts to her.

Case 4: The police officer Godfrey stops a motorist for speeding. The officer agrees to accept a personal gift of half of the amount of the fine in exchange for not taking the offending motorist to court to answer to charges for the traffic offence.

Case 5: A police officer finds a wallet in a parking lot. It contains an amount of money equivalent to a full day’s pay for that officer. She reports the wallet as lost property but keeps the money for herself.

Case 6: The police officer Glory discovers a burglary of a general merchandise shop. The display cases are smashed, and it is obvious that many items have been taken. While searching the shop, she takes 10 jerricans of cooking oil and 1 sack of posho of 100 kilograms worth about a month’s pay for that officer. She reports that the items had been stolen during the burglary.

Group 3: Refusal to register complaints

Case 7: A formerly arrested man comes to the police station and wants to fill in a complaint form. He claims that he was not treated properly during his arrest. The police officer Sarah who is in charge laughs at him and sends him away.

Case 8: Samwel goes to a police station to register a complaint over one of their officers who had beaten and tortured him. At the station he finds a friend of the officer who tortured him. The friend refuses to register his complaint and instead decides to detain him for a week over giving false information to the police. Evaluate the behavior of the friend.

Group 4: Reported severe crimes against individuals not followed up upon

Case 9: Mary goes to the police station to report a case where her husband has been beating her for the last one year. She lost one of her teeth and has a damaged eye due to the beating. The police officer on duty thinks this is a mere family dispute and not a crime for the police to handle. He refuses to register the case.

Case 10: A police officer on duty receives a woman who wants to register a case of murder of her child by a neighbor. The officer registers the case and promises to follow up and arrest the suspect in a few hours' time. Two days down the road, the suspect has not been arrested and was sending messages threatening to harm the complainant. The woman went back to the same police station to report the scenario and the suspect was arrested and detained at the police station. However, the suspect was released immediately on account that there was not enough evidence to convict him. Evaluate the behavior of the police officer who first received the woman.

Group 5: Undue force used by the police

Case 11: Two police officers on foot patrol surprise a man who is attempting to break into a shop. The man flees. They chase him for about ½ a kilometer before apprehending him by tackling him and wrestling him to the ground. After he is under control, both officers punch him a couple of times in the stomach and step on his back several times as punishment for fleeing and resisting.

Case 12: A subdistrict has a challenge of water shortage for a period of four months. The area leader together with residents decide to petition national water for the poor services and failure to deliver. However, the situation continues for two more months. The area leader and the residents opt to stage a peaceful demonstration as a way of showing their dissatisfaction. No sooner had the demonstration started than the District Police Commander deployed a team of officers with teargas and firing of live ammunitions killing 20 of the demonstrators including the area leader. Evaluate the District Police Commander's behavior.

Table A1: Robustness checks: Full details

	(1)	(2)	(3)	(4)	(5)	(6)
	How serious do YOU consider this behavior to be?	Do you think YOU would report a fellow police officer who engaged in this behavior?	How serious do MOST POLICE OFFICERS IN YOUR OFFICE consider this behavior to be?	If an officer in your agency engaged in this behavior and was discovered doing so, what if any discipline do YOU think SHOULD follow?	Would this behavior be regarded as a violation of official policy in your agency?	<i>Average Effect [Columns 1-5]</i>
Panel A: Full Sample (repeated from Panel A, Table 3)						
Male Rogue Officer	-0.072 (0.063)	-0.049 (0.049)	-0.056 (0.055)	-0.020 (0.050)	-0.020 (0.049)	-0.031 (0.025)
Male Victim	-0.100 (0.084)	-0.078 (0.052)	-0.082 (0.070)	-0.024 (0.060)	0.033 (0.057)	-0.034 (0.029)
Male Rogue Officer x Male Victim	0.057 (0.105)	0.157* (0.070)	0.051 (0.091)	0.040 (0.078)	0.021 (0.075)	0.048 (0.038)
Panel B: Honest respondents (self-reported)						
Male Rogue Officer	-0.065 (0.065)	-0.040 (0.050)	-0.041 (0.056)	0.006 (0.051)	-0.024 (0.050)	-0.024 (0.025)
Male Victim	-0.082 (0.086)	-0.102+ (0.053)	-0.058 (0.070)	-0.017 (0.062)	0.013 (0.059)	-0.035 (0.029)
Male Rogue Officer x Male Victim	0.045 (0.108)	0.205* (0.071)	0.038 (0.093)	0.038 (0.080)	0.042 (0.078)	0.055 (0.039)
Panel C: Average opinion on the honesty of other police officers in the district > sample median (calculated excl. the respondent)						
Male Rogue Officer	-0.080 (0.071)	-0.078 (0.055)	-0.020 (0.062)	-0.036 (0.057)	-0.034 (0.053)	-0.036 (0.029)
Male Victim	-0.120 (0.092)	-0.063 (0.057)	-0.064 (0.078)	-0.050 (0.065)	0.041 (0.062)	-0.034 (0.032)
Male Rogue Officer x Male Victim	0.069 (0.116)	0.140 (0.081)	-0.002 (0.101)	0.039 (0.086)	0.019 (0.084)	0.038 (0.043)
Panel D: Respondents working in districts with crime rate < sample median						
Male Rogue Officer	-0.109 (0.095)	0.032 (0.071)	-0.051 (0.076)	-0.025 (0.072)	-0.039 (0.077)	-0.026 (0.037)
Male Victim	-0.200+ (0.119)	-0.086 (0.073)	-0.158+ (0.093)	-0.103 (0.082)	0.010 (0.087)	-0.075+ (0.040)
Male Rogue Officer x Male Victim	0.109 (0.153)	0.048 (0.103)	0.065 (0.127)	0.131 (0.111)	0.023 (0.107)	0.052 (0.054)

Note: Sample size is 7,200 (12 vignette cases x 600 respondents). Respondent covariates included are listed in Table 1. District and case fixed effects included in all models. Standard errors are clustered at the respondent level. + and * denote statistical significance at the 10% and 5%, respectively. All models are estimated using the command "avg_effect" in STATA based on Kling et al. (2004) and Clingingsmith et al. (2009), calculating average standardized effect sizes within a seemingly-unrelated regression model to adjust for covariance across estimates.