# VI. Appendix

## *VI.A. Survey Methodology*

The household survey was conducted in face-to-face format between October 2013 and April 2014. It consisted of interviews of 4200 residents (18 year old and older) of the Gran Area Metropolitana (GAM), the principal urban center in Costa Rica. The GAM, which includes cantons in the provinces of Alajuela, Cartago, Heredia, and San José, contains approximately 2.6 residents, and accounts for 60% of the country entire population. The survey was administered by Borge y Asociados, the most prominent survey research firm in Central America. On average, interviews lasted 25 minutes.

A two-stage clustered random sample based on the 2000 national census was generated (with fixed proportions defined for age and gender groupings). Three hundred and fifty primary sampling units (PSUs), the smallest geographic unit in the census, were selected from the total contained within the GAM, with twelve interviews conducted in each PSU. Interviewers began from the northernmost point of the PSU and proceed in a clockwise direction. Within each household, individuals were selected based on quotas by age and gender, so that half of the surveys are obtained from each gender, and one third fall into each of the categories of 18-29 years old, 30-45 years old, and 45 or more years old. In cases of refusals or when no one responded, the household was replaced with the adjacent household.

The survey was preceded by a small pilot consisting of 48 cases, administered in October 15 and 16, 2013. The goal of the pilot was for enumerators to familiarize themselves with the questionnaire on the field and to test their skills in administering the questionnaire. All survey enumerators utilized PDAs (personal digital assistants) to conduct the survey.

Survey enumerators were recruited by *Borge y Asociados* and were mostly experienced with the administration of surveys. They went through extensive training on the administration of

the survey instrument. For the purpose of survey verification, enumerators recorded the first name only and phone number of each respondent. Verification was conducted on a randomly selected 30% percent of the sample by phone, after which this information was destroyed. Team leaders also conducted verification on the field by randomly selecting households for verification the same day that the interview was conducted. If mistakes were found by either method, interviews were replaced by new ones. The contact rate for the survey was 87 percent, the response rate was 29 percent, the cooperation rate 39 percent, and the refusal rate 44 percent.[16]

**Focus Groups**. Focus groups were conducted in San José with residents of varied backgrounds in August 6, 7, and 8, 2013, prior to fielding the household survey. The goal of these focus groups in relation to this paper was to get a general sense of individuals' perceptions of the main topics of the survey: corruption and inefficiency of different areas of the government, crime and issues of citizen security, and reporting of crime.

**Phone Survey**. Before the household survey and the focus groups sessions, a nationally representative telephone survey of 1200 Costa Rica residents (older than 18) was conducted by *Borge y Asociados* between July 15 and July 20, 2013. The goal of this survey in relation to this paper was to test the questions, the questions' wording, and the order for the household survey.

16. Rates calculated according to the American Association of Public Opinion Research.

Table A1: Survey Sample Representativeness

| Variable | 2013 Survey | 2000/2011 Census |
|---|---|---|
| | **Gender** | |
| Male | 50.0 | 47.8 |
| Female | 50.0 | 52.2 |
| | **Age** | |
| 18-29 | 36.8 | 31.8 |
| 30-45 | 35.8 | 32.1 |
| $\geq$46 | 27.4 | 36.1 |
| | **Population** ($\geq$20 yrs) | |
| San José | 56.7 | 56.9 |
| Alajuela | 11.7 | 11.6 |
| Cartago | 15.6 | 15.5 |
| Heredia | 16.0 | 16.0 |

*Note*: Age and gender correspond to the 2011 Costa Rica Census (www.inec.go.cr). Data by province correspond to the 2000 Census and was only available for ages 20 and older (recall that the survey included residents 18 years old and older).

## *VI.B. Supplementary Tables*

Table A2: Covariate Means Before and After Matching

| | | Before matching | | After matching | |
|---|---|---|---|---|---|
| Covariate | | treatment | control | treatment | control |
| male | | 0.63 | 0.49 | 0.63 | 0.65 |
| age | | 32.4 | 39.1 | 32.4 | 32.8 |
| education: | primary or less | 0.24 | 0.28 | 0.24 | 0.22 |
| | secondary incomplete | 0.34 | 0.27 | 0.34 | 0.34 |
| | secondary complete | 0.27 | 0.23 | 0.27 | 0.27 |
| | some education | 0.03 | 0.04 | 0.03 | 0.02 |
| Costa Rican national | | 0.93 | 0.90 | 0.93 | 0.94 |
| head of household | | 0.49 | 0.48 | 0.49 | 0.49 |
| material wealth: | laptop | 0.44 | 0.41 | 0.44 | 0.43 |
| | tablet | 0.31 | 0.24 | 0.31 | 0.31 |
| | car | 0.37 | 0.38 | 0.37 | 0.39 |
| | internet | 0.57 | 0.55 | 0.57 | 0.57 |
| knows police officer | | 0.54 | 0.40 | 0.54 | 0.54 |
| knows someone prosecuted | | 0.62 | 0.39 | 0.62 | 0.62 |
| direct contact with police | | 0.45 | 0.23 | 0.45 | 0.47 |
| crime victim | | 0.45 | 0.32 | 0.45 | 0.46 |
| district characteristics: | unmet basic necessities | 0.20 | 0.19 | 0.20 | 0.19 |
| | population density | 489 | 473 | 489 | 494 |
| | higher education | 0.29 | 0.30 | 0.29 | 0.30 |
| | laptop | 0.34 | 0.36 | 0.34 | 0.35 |
| | car | 0.45 | 0.48 | 0.45 | 0.46 |
| | internet connection | 0.46 | 0.47 | 0.46 | 0.47 |
| | homes in poor condition | 0.06 | 0.05 | 0.06 | 0.05 |

Table A3: Propensity Score Estimating Equation - Logistic Regression of Witnessing Police Violence on Covariates

| variable | | | coefficient | s.e. | p.value |
|---|---|---|---|---|---|
| constant | | | -3.47 | 0.73 | <0.001 |
| male | | | 0.37 | 0.11 | <0.001 |
| age | | | -0.03 | 0.005 | <0.001 |
| education: | primary or less | | 0.67 | 0.20 | <0.001 |
| | secondary incomplete | | 0.65 | 0.18 | <0.001 |
| | secondary complete | | 0.50 | 0.18 | 0.005 |
| | some technical | | 0.13 | 0.34 | 0.71 |
| Costa Rican | | | 0.25 | 0.20 | 0.22 |
| head of household | | | 0.12 | 0.12 | 0.31 |
| wealth: | laptop | | -0.01 | 0.13 | 0.97 |
| | tablet | | 0.27 | 0.13 | 0.03 |
| | car | | -0.14 | 0.12 | 0.24 |
| | internet | | -0.02 | 0.12 | 0.90 |
| knows police officer | | | 0.24 | 0.11 | 0.02 |
| knows someone prosecuted | | | 0.62 | 0.11 | <0.001 |
| contact with police | | | 0.74 | 0.11 | <0.001 |
| crime victim | | | 0.31 | 0.11 | 0.004 |
| district characteristics: | | unmet necessities | 2.27 | 2.09 | 0.28 |
| | | population density | 0.00002 | 0.00002 | 0.27 |
| | | higher education | 1.01 | 2.39 | 0.67 |
| | | laptop | -5.62 | 3.71 | 0.13 |
| | | car | 0.39 | 1.40 | 0.78 |
| | | internet | 4.02 | 2.20 | 0.07 |
| | | homes in poor condition | -3.59 | 4.96 | 0.47 |
| AIC=2624 | | | | N=4,062 | |

Table A4: Graded Response Model of Propensity against Reporting Crime

| indicator | discrimination | s.e. | cutpoint.1 | cutpoint.2 | cutpoint.3 |
|---|---|---|---|---|---|
| *proclivity against reporting...* | | | | | |
| robbery w/o weapon | 1.05 | 0.04 | -1.24 | 0.23 | 1.57 |
| robbery w/ weapon | 1.28 | 0.05 | -0.54 | 0.70 | 1.87 |
| robbery of home | 1.20 | 0.05 | 0.21 | 1.75 | 2.83 |
| suspicious activity | 2.72 | 0.09 | -0.45 | 0.74 | 1.56 |
| drug sales | 2.58 | 0.09 | -0.64 | 0.41 | 1.16 |
| gang activity | 2.59 | 0.09 | -0.55 | 0.66 | 1.39 |
| gun shots | 1.96 | 0.06 | -0.59 | 0.58 | 1.33 |

*Note*: The R package ltm was utilized to estimate the Graded Response Model (Rizopoulos 2013).

Table A5: Nearest Neighbor Matching Analysis of Impact of Police Violence on Willingness to Report Crime - Outcome: Latent Propensity against Reporting Crime

| | |
|---|---|
| ATT | 0.25 |
| s.e. | 0.06 |
| p. value | <0.001 |
| *N* | 4062 |
| *N.treated* | 458 |
| *N.matched* | 458 |

*Note*: Latent propensity against reporting crime was estimated using the factor scores obtained from applying the Graded Response Model to the seven ordinal reporting outcomes. Factor scores were estimated using the Empirical Bayes method.

Table A6: OLS Analysis of Impact of Police Violence on Willingness to Report Crime - Outcome: Latent Propensity against Reporting Crime (N=4062, $R^2$=0.08)

| variable | | coefficient | s.e. | p.value |
|---|---|---|---|---|
| constant | | −0.77 | 0.18 | <0.001 |
| **police violence** | | **0.23** | **0.04** | **<0.001** |
| male | | 0.01 | 0.03 | 0.69 |
| age | | 0.003 | 0.001 | 0.01 |
| education: | primary or less | 0.35 | 0.05 | <0.001 |
| | secondary incomplete | 0.27 | 0.04 | <0.001 |
| | secondary complete | 0.23 | 0.04 | <0.001 |
| | some technical | 0.06 | 0.08 | 0.42 |
| Costa Rican national | | −0.08 | 0.05 | 0.10 |
| head of household | | −0.05 | 0.03 | 0.07 |
| wealth: | laptop | −0.06 | 0.03 | 0.07 |
| | tablet | −0.02 | 0.03 | 0.47 |
| | car | −0.07 | 0.03 | 0.03 |
| | internet | −0.05 | 0.03 | 0.15 |
| knows police officer | | 0.01 | 0.03 | 0.81 |
| knows someone prosecuted | | 0.05 | 0.03 | 0.07 |
| direct contact w/police | | −0.13 | 0.03 | <0.001 |
| crime victim | | 0.02 | 0.03 | 0.47 |
| district characteristics: | unmet necessities | 0.12 | 0.53 | 0.81 |
| | population density | 0.00002 | 0.000006 | 0.01 |
| | higher education | −1.95 | 0.60 | 0.001 |
| | laptop | 0.21 | 0.93 | 0.82 |
| | car | −0.44 | 0.35 | 0.20 |
| | internet | 2.22 | 0.57 | < 0.001 |
| | homes in poor condition | 3.17 | 1.23 | 0.01 |

*Note*: Latent propensity against reporting crime was estimated using the factor scores obtained from applying the Graded Response Model to the seven ordinal reporting outcomes. Factor scores were estimated using the Empirical Bayes method.

Table A7: OLS Regressions of Crime Non-reporting on the Observation of Police Violence

| outcome | coef(pol. violence) | s.e. | p.value | 95% conf. |
|---|---|---|---|---|
| proclivity against reporting... | | | | |
| robbery w/o weapon | 0.26 | 0.05 | <0.001 | [0.16,0.37] |
| robbery w/ weapon | 0.18 | 0.05 | <0.001 | [0.08,0.29] |
| robbery of home | 0.13 | 0.04 | 0.003 | [0.04,0.21] |
| suspicious activity | 0.26 | 0.05 | <0.001 | [0.17,0.35] |
| drug sales | 0.21 | 0.05 | <0.001 | [0.11,0.31] |
| gang activity | 0.20 | 0.05 | <0.001 | [0.10,0.29] |
| gun shots | 0.23 | 0.05 | <0.001 | [0.12,0.33] |

*Note*: Shown are the coefficients on the observation of police violence in regressions of the proclivity to report different types of crimes. For each crime, outcomes were coded on a 1-4 scale, where 1=highly likely to report; 2=likely to report; 3=unlikely to report; 4=highly unlikely to report. Regressions included all covariates used in the matching analysis.

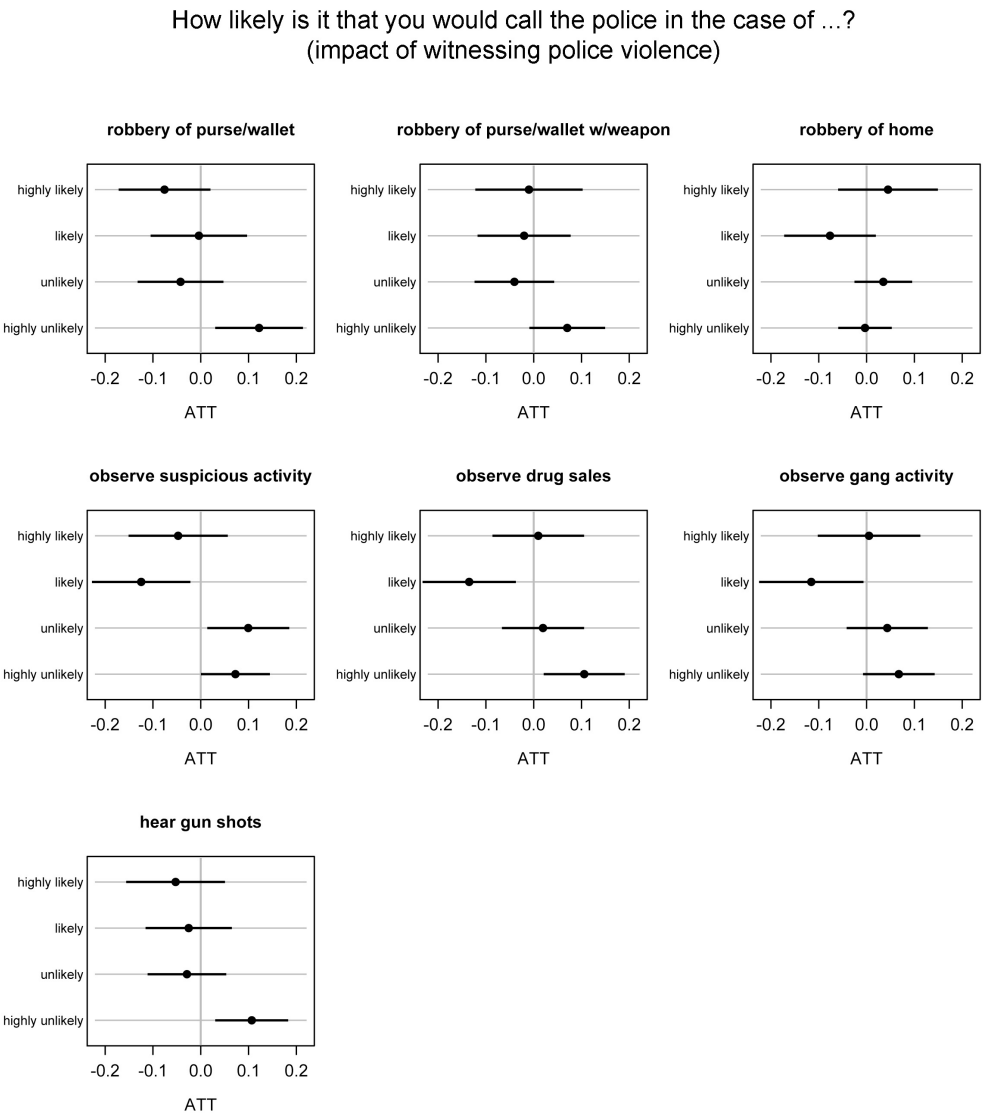
Table A8: Ordered Probit Regressions of Crime Non-reporting on the Observation of Police

| | outcome | coef(pol. violence) | s.e. | p.value | 95% conf. |
|---|---|---|---|---|---|
| | proclivity against reporting... | | | | |
| Violence | robbery w/o weapon | 0.27 | 0.06 | <0.001 | [0.16,0.38] |
| | robbery w/ weapon | 0.19 | 0.06 | <0.001 | [0.08,0.30] |
| | robbery of home | 0.16 | 0.06 | 0.008 | [0.04,0.27] |
| | suspicious activity | 0.29 | 0.06 | <0.001 | [0.10,0.32] |
| | drug sales | 0.21 | 0.05 | <0.001 | [0.11,0.31] |
| | gang activity | 0.20 | 0.06 | <0.001 | [0.09,0.31] |
| | gun shots | 0.23 | 0.06 | <0.001 | [0.12,0.3] |

*Note*: Shown are the coefficients on the observation of police violence in ordered probit regressions of the proclivity to report different types of crimes. For each crime, outcomes were coded on a 1-4 scale, where 1=highly likely to report; 2=likely to report; 3=unlikely to report; 4=highly unlikely to report. Regressions included all covariates used in the matching analysis.

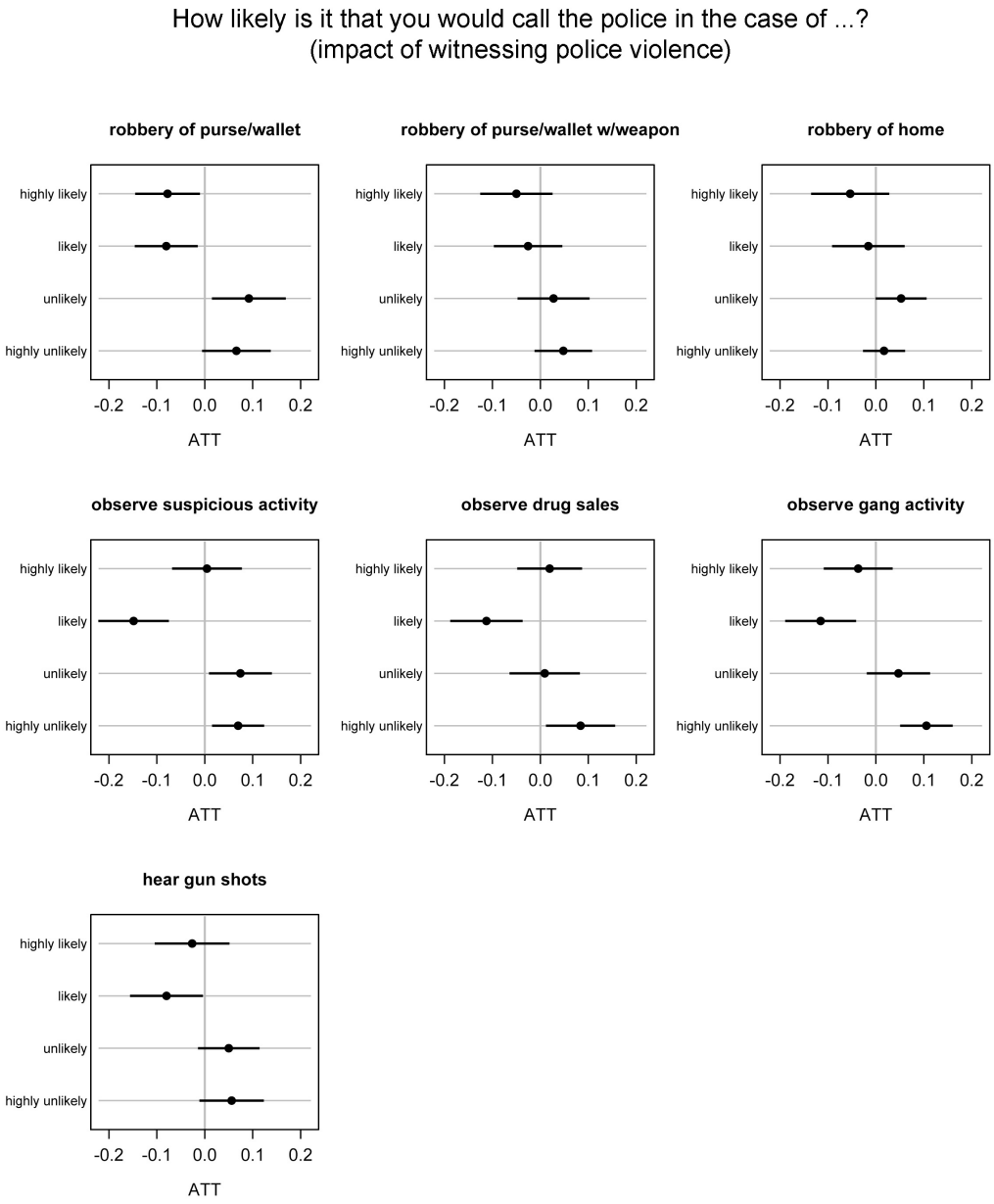
## *VI.C. Supplementary Figures*

Figure A1: Impact of Police Violence on Willingness to Report Crime - Respondents in Low Crime Districts




*Note: ATT for the observation of police violence is arrayed on x-axis. Point estimates are denoted by black dots. Ninety-five percent confidence intervals are denoted by black horizontal lines.*

Figure A2: Impact of Police Violence on Willingness to Report Crime - Respondents in High Crime Districts




*Note: ATT for the observation of police violence is arrayed on x-axis. Point estimates are denoted by black dots. Ninety-five percent confidence intervals are denoted by black horizontal lines.*

## *VI.D. Proofs for Game-Theoretic Model*

**Proof of proposition 1**. We begin the proof by noting that the effect of permissiveness towards violence ($\sigma$) on the probability of crime without punishment can be expressed as follows:

$$\frac{\partial \Pr(\text{CWP})}{\partial \sigma} = \frac{\partial \Pr(s=1)}{\partial \sigma} \Pr(r=0) + \frac{\partial \Pr(r=0)}{\partial \sigma} \Pr(s=1). \tag{16}$$

The expression reveals that the impact of $\sigma$ is equal to the sum of two effects: the effect of $\sigma$ on the probability of crime (holding reporting constant) and the effect of $\sigma$ on the probability of non-reporting, holding crime constant.

Consider first the non-reporting effect. This is always positive, implying that the probability of non-reporting is monotonically increasing in $\sigma$. To see this, decompose the derivative of $\Pr(r=0)$ with respect to $\sigma$ as follows:

$$\frac{\partial \Pr(r=0)}{\partial \sigma} = \frac{\partial \Pr(r=0)}{\partial u_L(r=1)} \frac{\partial u_L(r=1)}{\partial \sigma}. \tag{17}$$

The first component of the above product is negative. The second component, the change in the indirect utility of $L$ from reporting crime, is equal to

$$\frac{\partial u_L(r=1)}{\partial \sigma} = -\mu(2+\sigma) \frac{\partial \Pr(a=3)}{\partial \sigma} - \mu \Pr(a=3), \tag{18}$$

where

$$\frac{\partial \Pr(a=3)}{\partial \sigma} = \frac{\partial \Pr(a=3|\upsilon=1)}{\partial \sigma} q(\sigma) + \frac{\partial q(\sigma)}{\partial \sigma} \left[\Pr(a=3|\upsilon=1) - \Pr(a=3|\upsilon=0)\right]. \tag{19}$$

The above shows that change in utility $L$ gets from reporting crime due to a change in $\sigma$ is a decreasing function of her expectation about the likelihood of abuse at the hands of $P$. The change in the likelihood of abuse due to a change in $\sigma$, in turn, consists of the sum of incentive effect described in the main text (first expression in equation 19) and the selection effect (second

expression in equation 19). Both of these are positive, implying that the derivative of $\Pr(r = 0)$ with respect to $\sigma$ is also positive. (Note that $\Pr(a = 3|\upsilon = 0)$ is unrelated to $\sigma$ according to the specification of utilities in (2); consequently, the derivative of this quantity is zero and it drops out of the equation above).

Now consider the crime effect. This may be positive or negative, depending upon parameter values. To see this, decompose the derivative of $\Pr(s = 1)$ with respect to $\sigma$ as follows:

$$\frac{\partial \Pr(s=1)}{\partial \sigma} = \frac{\partial \Pr(s=1)}{\partial u_C(s=1)} \frac{\partial u_C(s=1)}{\partial \sigma}. \tag{20}$$

The first component of the above product is positive. The second component, the change in the indirect utility of $C$ from engaging in crime, is equal to

$$\begin{aligned} \frac{\partial u_C(s=1)}{\partial \sigma} = \frac{\partial \Pr(r=0)}{\partial \sigma} \left[\tau + \gamma + \beta\sigma(1 - \Pr(a=1))\right] \\ - (1 - \Pr(r=0))\beta \left[(1 - \Pr(a=1)) - \sigma \frac{\partial \Pr(a=1)}{\partial \sigma}\right]. \end{aligned} \tag{21}$$

The two terms in the sum above represent countervailing effects of $\sigma$ on $C$'s decision to engage in crime. The first expression in the sum is the *non-reporting effect*. This effect–which is positive–captures how the anticipated drop in the likelihood of crime reporting by $L$ due to an increase in $\sigma$ increases the utility that $C$ derives from engaging in crime. The second expression in the sum is the *sanction severity effect*. This effect–which is negative–captures how the disutility for the greater anticipated violence potentially meted out by $P$ due to an increase in $\sigma$ decreases the utility that $C$ derives from engaging in crime.

The crucial point to note about the sanction severity effect is that it is weighted by the probability of reporting, $1 - \Pr(r = 0)$. This captures the fact that $C$ will only face a sanction for his crime if $L$ decides to report the transgression in the first place. At low levels of $\sigma$, it is possible for the sanction severity effect to dominate, meaning that local increases in $\sigma$ reduce the probability of crime. This is because the probability of reporting will be relatively high, and

even though few officers will be violence prone, some non-violence prone office may nevertheless engage in violence (this latter scenario being more common the smaller is $\theta$). Also, with low $\sigma$ the potential costs to $L$ from police abusiveness are relatively low, so the reporting effect may be muted. However, as $\sigma$ increases, the disutility to $L$ from reporting increases rapidly, since high $\sigma$ implies both a high likelihood of interacting with a violence prone officer and a very bad outcome if abuse does take place. This has two effects: it weakens the sanction severity effect through the reduction in the probability of reporting and it increases the reporting effect, since higher $\sigma$ makes the stakes of abuse greater for $L$ and also increases the benefit (of foregone suffering from violence) $C$ receives when $L$ does not report. Eventually, the reporting effect comes to dominate the sanction severity effect and further increases in $\sigma$ increase the attractiveness of crime and increase the likelihood of crime without punishment.