Supplementary Information for "Corruption as a Self-Fulfilling Prophecy: Evidence from a Survey Experiment in Costa Rica"

Ana Corbacho
International Monetary Fund
Daniel W. Gingerich
University of Virginia
Virginia Oliveros
Tulane University
Mauricio Ruiz-Vega
Macroeconomic Section-US Embassy in Jakarta, Indonesia

The Corrupting Influence of Expectations about Corruption: A Formal Model

Here we present a game-theoretic model that illustrates the essential role that the coordination of beliefs about corruption plays in determining the prevalence of corrupt behavior. Relative to other theoretical approaches to corruption, a distinguishing feature of our model rests with its emphasis on the importance of two-sided uncertainty among citizens and officials potentially inclined to engage in low-level, typically anonymous, corrupt exchanges. In particular, our framework for modeling corruption concentrates on the challenges inherent to initiating a corrupt exchange when both partners to a potential corrupt exchange have limited *a priori* information about their counterpart's inclination to engage in illicit behavior. In such circumstances, beliefs about aggregate levels of corruption may be crucial in defining the expected returns to initiating a corrupt exchange for the agents on both sides of the transaction, and *ipso facto*, crucial for determining how many potential corrupt transactions end up being instances of actual corruption.

In line with the traffic stops example in the text, we consider a game that takes place in a polity made up of two groups of individuals, drivers and police officers. In the game's only time period, all actors are organized into randomly matched pairs consisting of one driver and one police officer, with each actor assigned to only a single pair. Once the pairs have been established, each member of a driver-officer pair simultaneously decides whether or not to indicate a disposition to engage in a corrupt exchange.

Such an exchange takes place only if both members of the pair indicate a disposition towards corruption. If this happens, a generic driver i receives a return α_i, whereas a generic officer j receives return ϕ_j. Within each group, the returns to corruption are distributed uniformly, with returns to corruption having support $[\underline{\alpha}, \overline{\alpha}]$ among drivers and $[\underline{\phi}, \overline{\phi}]$ among police officers. The endpoints of these supports reflect exogenous characteristics of the polity, such as the level of government regulation of the economy or cultural aversion to corrupt practices, that systematically influence the individual returns to corrupt activity. We restrict our attention to settings in which corruption is at least potentially attractive to all actors, implying that $\underline{\alpha} > 0$ and $\underline{\phi} > 0$. Each agent's returns to corruption are private information; only the distributions of these quantities in each group is known publicly.

Indicating a disposition towards corruption when one's partner fails to do is assumed to be costly for all agents. The cost to a driver of indicating a willingness to engage in corruption when paired with an unwilling officer is $\sigma > 0$. The cost to an officer of indicating a willingness to engage in corruption when paired with an unwilling driver is $\tau > 0$. The values of these parameters reflect the quality of institutions as pertains to the monitoring and sanctioning of corruption in the polity. The higher the quality of said institutions, the greater the expected cost associated with an unreciprocated attempt to suborn (extort) one's partner. We permit these costs to vary across agent type, reflecting the fact that the legal sanctions assigned to actors in a corrupt exchange often differ according to whether they work in the public or private sector.[1] If each member of a driver-officer pair refrains from indicating a disposition towards corruption, then both agents receive a normalized return of zero. Any agent not indicating a disposition towards corruption when her partner does indicate such a disposition also receives a return of zero.

A driver's information about the officer she is matched with consists only of her knowledge of the distribution of returns to corruption among all officials. The equivalent is true about the information of an officer about the driver he is matched with. Thus, for a generic driver in our model, the expected returns to indicating a disposition towards a corrupt transaction are

$$p^e \alpha_i - (1 - p^e)\sigma, \tag{1}$$

where p^e represents the drivers' (collective) belief about the proportion of officers who are disposed towards corruption. For a generic officer, the expected returns to indicating a disposition towards a corrupt transaction are

$$q^e \phi_j - (1 - q^e)\tau, \tag{2}$$

where q^e represents the officers' (collective) belief about the proportion of drivers who are disposed towards corruption.

The above equations imply that any driver i for whom $\alpha_i \geq (1 - p^e)\sigma/p^e$ will indicate a disposition towards corruption, as will any officer j for whom $\phi_j \geq (1 - q^e)\tau/q^e$. Thus, the actual proportions of drivers and officers, respectively, who indicate a disposition towards corruption are as follows:

$$\begin{aligned} q &= \mathbb{P}\left(\alpha_i \geq \frac{(1-p^e)\sigma}{p^e}\right) \\ p &= \mathbb{P}\left(\phi_j \geq \frac{(1-q^e)\tau}{q^e}\right). \end{aligned} \tag{3}$$

The proportion of transactions between officers and drivers in the polity that result in corruption is simply the product of these two prevalence rates, $\Omega = pq$.

In equilibrium, the collective expectations of each type of actor about the other will be correct. This implies that $p^e = p = p^*$ and $q^e = q = q^*$, where (p^*, q^*) is an equilibrium pair of beliefs. Using the expression above and the fact that the distribution of returns to corruption within each group of agents is uniform, an equilibrium to the game is a pair (p^*, q^*) that satisfies the following two equations:

$$q^* = \begin{cases} 0 & \text{if} \quad p^* < \frac{\sigma}{\overline{\alpha}+\sigma} \\ \frac{\overline{\alpha}+(1-1/p^*)\sigma}{\overline{\alpha}-\underline{\alpha}} & \text{if} \quad p^* \in [\frac{\sigma}{\overline{\alpha}+\sigma}, \frac{\sigma}{\underline{\alpha}+\sigma}] \\ 1 & \text{if} \quad p^* > \frac{\sigma}{\underline{\alpha}+\sigma} \end{cases} \tag{4}$$

$$p^* = \begin{cases} 0 & \text{if} \quad q^* < \frac{\tau}{\overline{\phi}+\tau} \\ \frac{\overline{\phi}+(1-1/q^*)\tau}{\overline{\phi}-\underline{\phi}} & \text{if} \quad q^* \in [\frac{\tau}{\overline{\phi}+\tau}, \frac{\tau}{\underline{\phi}+\tau}] \\ 1 & \text{if} \quad q^* > \frac{\tau}{\underline{\phi}+\tau} \end{cases}$$

[1] In this respect, we follow the modeling strategy of Ryvkin and Serra (2012).

An equilibrium level of corruption, in turn, is equal to $\Omega^* = p^* q^*$, where p^* and q^* belong to the same equilibrium pair.

Our model exhibits strategic complementarities between the actions of drivers and officers: the expected returns to corruption for a generic driver are increasing in the proportion of officers willing to be corrupt and vice-versa. These complementarities induce multiple equilibria. In particular, we prove below that there are three equilibria for this game: a high corruption equilibrium, (p^*_H, q^*_H), in which all drivers and officers indicate a willingness to engage in corruption, a low corruption equilibrium, (p^*_L, q^*_L), in which no drivers or officers indicate a willingness to engage in corruption, and an intermediate corruption equilibrium, (p^*_M, q^*_M), in which the proportions of drivers and officers who indicate a disposition to engage in corruption fall within the intervals $[\frac{\sigma}{\overline{\alpha}+\sigma}, \frac{\sigma}{\underline{\alpha}+\sigma}]$ and $[\frac{\tau}{\overline{\phi}+\tau}, \frac{\tau}{\underline{\phi}+\tau}]$, respectively.

Figure 1 in the main text presents the equilibria for this game. The proportion of drivers disposed towards corruption (a continuous function of the proportion of officers disposed towards corruption) is displayed in blue, whereas the proportion of officers disposed towards corruption (a continuous function of the proportion of drivers disposed towards corruption) is displayed in red. The three equilibria are the points of intersection between these two functions. They are the potential steady state patterns of corruption that could emerge given the premises of our game. By the definition of an equilibrium, once all actors coordinate their actions around any one of these points, no actor has an incentive to deviate from that point by changing her behavior.

Yet the equilibria of this game are not all equally plausible. In the language of dynamical systems, the high and low equilibrium points are so-called attracting fixed points, whereas the intermediate equilibrium point is a repelling fixed point. Conceptually, what this means is the following. If drivers and officers were to begin the game with beliefs about each other that were close to but slightly different from either of the extreme equilibria, i.e. located somewhere in a small neighborhood around p^*_H, q^*_H or p^*_L, q^*_L (but not on these specific points), actor beliefs and behavior would dynamically adjust until the high or low equilibrium, respectively, was eventually reached. However, if drivers and officers were to begin the game with beliefs about each other that were slightly different from the intermediate equilibrium, the aforementioned dynamical adjustment process would push them further away from this equilibrium. Indeed, only if drivers and officers began with the initial belief p^*_M, q^*_M would they wind up in this particular equilibrium. (See below for a formal proof). In this sense, we can think of the high and low equilibria as plausible equilibria for this game and the intermediate equilibrium as an implausible equilibrium, where a plausible equilibrium is one that can be reached through a reasonable dynamical process that begins off the equilibrium path.

Given that both the high and low corruption equilibrium are plausible representations of social behavior for our model, how likely is it that one or the other of the equilibria will actually obtain? An emerging theoretical literature on models with multiple equilibria suggests that the appropriate manner in which to address this question is to examine the relative size of the sets of initial beliefs that would generate each of the feasible equilibria (Medina 2007, 2013). Let us refer to any such set as the basin of attraction of an equilibrium. In Figure 1 in the main text, the basins of attraction for the two feasible equilibria are shaded in gray. More specifically, the basin of attraction for p^*_L, q^*_L is the set $B_L = [0, p^*_M) \times [0, q^*_M)$ whereas the the basin of attraction for p^*_H, q^*_H is the set $B_H = (p^*_M, 1] \times (q^*_M, 1]$. Any initial belief about drivers and officers located within the basin of attraction of an equilibrium will generate a dynamic adjustment process that eventually stabilizes at that equilibrium.

The sizes of the basins of attraction are determined by the quality of institutions and drivers' and officers' intrinsic tastes for corrupt activity. In particular, we prove below that the greater the severity of sanctions, the larger (smaller) the relative size of the basin for the low (high) corruption equilibrium. Similarly, we prove that the more intense tastes are for corruption, the smaller (larger) the relative size of the basin for the low (high) corruption equilibrium.

Let $\gamma \in \{\text{"low"},\text{"high"}\}$ represent one of the plausible corruption equilibria that could eventually emerge in the polity. Supposing that the probability of each equilibrium is directly proportional to the relative size of its basin of attraction, we can characterize the *expected* level of corruption, given our agents' tastes for corruption and the polity's technology for monitoring and punishing corruption, as follows:

$$\begin{aligned}
\mathbb{E}(\Omega) &= \mathbb{P}(\gamma = \text{"low"})(p_L^* q_L^*) + \mathbb{P}(\gamma = \text{"high"})(p_H^* q_H^*) \\
&= \mathbb{P}(\gamma = \text{"low"})(0) + \mathbb{P}(\gamma = \text{"high"})(1) \\
&= \frac{(1-p_M^*)(1-q_M^*)}{(1-p_M^*)(1-q_M^*) + p_M^* q_M^*}.
\end{aligned}$$

As is evident above, in the simple setting considered by our model the expected level of corruption and the probability of the high corruption equilibrium are one and the same. Thus, characterizing the expected scope of corruption in the polity boils down to characterizing the relative size of the basin of attraction of the high equilibrium relative to that for the low equilibrium. Since higher quality institutions and less permissive tastes for corruption reduce the relative size of the former vis-a-vis the latter, it follows immediately that both reduce the expected level of corruption in the polity.

Proofs for the Formal Model

Derivation of equilibria. It is straightforward to show that there are two corner solutions to this game, one where $p^* = q^* = 0$ (implying $\Omega^* = 0$) and another where $p^* = q^* = 1$ (implying $\Omega^* = 1$). In the first case, note that if the collective belief of drivers is that no officer is corrupt ($p^e = 0$), then the optimal response for each driver is to refrain from indicating a willingness to bribe ($q^*(p^e = 0) = 0$). If no driver is willing to indicate a disposition towards bribery, then the optimal response for each officer is also to refrain from indicating a disposition to engage in a corrupt exchange ($p^*(q^* = 0) = 0$). The logic for the second corner solution is directly analogous to the logic for the first, and it follows from fact that $\underline{\alpha} > 0$ and $\underline{\phi} > 0$.

To check for interior solutions, note that any such solution can be written

$$\begin{aligned}
q^* &= a - \frac{b}{p^*} \quad \text{for } p^* \in [\frac{\sigma}{\overline{\alpha}+\sigma}, \frac{\sigma}{\underline{\alpha}+\sigma}] \\
p^* &= c - \frac{d}{q^*} \quad \text{for } q^* \in [\frac{\tau}{\overline{\phi}+\tau}, \frac{\tau}{\underline{\phi}+\tau}],
\end{aligned} \tag{A1}$$

where $a = \frac{\overline{\alpha}+\sigma}{\overline{\alpha}-\underline{\alpha}}$, $b = \frac{\sigma}{\overline{\alpha}-\underline{\alpha}}$, $c = \frac{\overline{\phi}+\tau}{\overline{\phi}-\underline{\phi}}$, $d = \frac{\tau}{\overline{\phi}-\underline{\phi}}$.

Using the quadratic formula, the above system has two solutions:

$$\underline{\text{solution 1}} \quad : \quad \begin{cases} q^* = \frac{1}{2c}\left(-b+d+ac+\sqrt{(b-d-ac)^2-4adc}\right) \\ p^* = \frac{1}{2a}\left(b-d+ac+\sqrt{(b-d-ac)^2-4adc}\right) \end{cases}$$

$$\underline{\text{solution 2}} \quad : \quad \begin{cases} q^* = \frac{1}{2c}\left(-b+d+ac-\sqrt{(b-d-ac)^2-4adc}\right) \\ p^* = \frac{1}{2a}\left(b-d+ac-\sqrt{(b-d-ac)^2-4adc}\right) \end{cases}$$

Of the two solutions above, only solution 2 falls within the admissible range. This is the intermediate equilibrium. Thus, the three equilibria for the game are the two corner solutions and solution 2 as defined above.

Properties of the equilibria. *Claim 1: The equilibrium point p^*_H, q^*_H is an attracting fixed point.* Let t index an iteration of the dynamical adjustment process described by equation (4). Specifically, the endogenous variables on the RHS of (4) are indexed by t and those on the LHS are indexed by $t+1$. Set initial values $p(t) = 1-\epsilon$ and $q(t) = 1-\epsilon$, where ϵ is an arbitrarily small constant. For any ϵ sufficiently small such that $1-\epsilon > \frac{\sigma}{\underline{\alpha}+\sigma}$ and $1-\epsilon > \frac{\tau}{\underline{\phi}+\tau}$, the subsequent values of the endogenous quantities are $p(t+1) = 1$ and $q(t+1) = 1$. Thus, for off-the-path beliefs in this ϵ-neighborhood, the high equilibrium is reached in a single step. *Claim 2: The equilibrium point p^*_L, q^*_L is an attracting fixed point.* Using the same notation, set initial values $p(t) = \epsilon$ and $q(t) = \epsilon$. For any ϵ sufficiently small such that $\epsilon < \frac{\sigma}{\overline{\alpha}+\sigma}$ and $\epsilon < \frac{\tau}{\overline{\phi}+\tau}$, the subsequent value of the endogenous quantities are $p(t+1) = 0$ and $q(t+1) = 0$. Thus, for off-the-path beliefs in this ϵ-neighborhood, the low equilibrium is reached in a single step. *Claim 3: The equilibrium point p^*_M, q^*_M is a repelling fixed point.* According to the theory of dynamical systems, a necessary and sufficient condition for p^*_M, q^*_M to be a repelling fixed point is that each eigenvalue of the Jacobian of the system described in equation (4) must be greater than 1 in absolute value when the Jacobian is evaluated at p^*_M, q^*_M (cf. Alligood, Sauer, and Yorke 1996, p.70). The two eigenvalues of the Jacobian (at this point) are as follows:

$$\boldsymbol{\lambda} = \begin{pmatrix} \sqrt{\left(\frac{\tau}{\overline{\phi}-\underline{\phi}}/q^{*2}_M\right)\left(\frac{\sigma}{\overline{\alpha}-\underline{\alpha}}/p^{*2}_M\right)} \\ -\sqrt{\left(\frac{\tau}{\overline{\phi}-\underline{\phi}}/q^{*2}_M\right)\left(\frac{\sigma}{\overline{\alpha}-\underline{\alpha}}/p^{*2}_M\right)} \end{pmatrix}. \tag{A2}$$

Now note that at any interior solution, it must be the case that $\frac{\tau}{\overline{\phi}-\underline{\phi}} > q^*_M$ and $\frac{\sigma}{\overline{\alpha}-\underline{\alpha}} > p^*_M$, implying that both eigenvalues are greater than 1 in absolute value.

Size of basins of attraction. Here we establish that the size of the basins of attraction of the two plasuible equilibria are a function of the severity of sanctions for corruption and the intensity of drivers' and officers' intrinsic tastes for corruption. In particular, we prove that: 1) the size of the basin of attraction of the high corruption equilibrium decreases with increases in the severity of sanctions whereas the size of the basin of attraction for the low corruption equilibrium increases in the severity of sanctions; 2) the size of the basin of attraction of the high corruption equilibrium increases with increases in the taste of drivers and officers for corruption whereas the size of the basin of attraction for the low corruption equilibrium decreases in the intensity of tastes for corruption.

To structure the proof, note that the size of each basin of attraction is uniquely defined by the intermediate equilibrium p^*_M, q^*_M. Any change in a parameter that shifts both p^*_M and q^*_M upwards will reduce the size of the basin of attraction for the high corruption equilibrium and increase the size of the basin of attraction for the low corruption equilibrium. Similarly, any change in a parameter that shifts both p^*_M and q^*_M downwards will increase the size of the basin of attraction for the high corruption equilibrium and decrease the size of the basin of attraction for the low corruption equilibrium. Thus, to prove that the size of the basins change with sanctions and intrinsic tastes in the manner suggested it suffices to show that the derivatives of p^*_M and q^*_M with respect to σ and τ are all positive and that the derivatives of p^*_M and q^*_M with respect to $\overline{\alpha}$ and $\overline{\phi}$ are all negative.

We begin by noting that, according to (A1), one has:

$$\begin{aligned} \frac{\partial p^*_M}{\partial \sigma} &= \frac{d}{q^{*2}_M}\frac{\partial q^*_M}{\partial \sigma} \\ \frac{\partial p^*_M}{\partial \overline{\alpha}} &= \frac{d}{q^{*2}_M}\frac{\partial q^*_M}{\partial \overline{\alpha}} \\ \frac{\partial q^*_M}{\partial \tau} &= \frac{b}{p^{*2}_M}\frac{\partial p^*_M}{\partial \tau} \\ \frac{\partial q^*_M}{\partial \overline{\phi}} &= \frac{b}{p^{*2}_M}\frac{\partial p^*_M}{\partial \overline{\phi}}, \end{aligned} \tag{A3}$$

which implies that the sign of $\frac{\partial p^*_M}{\partial \sigma}$ is the same as $\frac{\partial q^*_M}{\partial \sigma}$, the sign of $\frac{\partial p^*_M}{\partial \overline{\alpha}}$ is the same as $\frac{\partial q^*_M}{\partial \overline{\alpha}}$, the sign of $\frac{\partial q^*_M}{\partial \tau}$ is the same as $\frac{\partial p^*_M}{\partial \tau}$, and that the sign of $\frac{\partial q^*_M}{\partial \overline{\phi}}$ is the same as $\frac{\partial p^*_M}{\partial \overline{\phi}}$.

Differentiating, we get:

$$\begin{aligned} \frac{\partial q^*_M}{\partial \sigma} &= \frac{1}{2c(\overline{\alpha}-\underline{\alpha})}\left[(c-1)+\frac{(-b+d+ac)(c-1)+2dc}{\sqrt{(b-d-ac)^2-4adc}}\right] \geq 0 \\ \frac{\partial p^*_M}{\partial \tau} &= \frac{1}{2a(\overline{\phi}-\underline{\phi})}\left[(a-1)+\frac{(-b+d+ac)\,(a-1)+2a(c+d)}{\sqrt{(b-d-ac)^2-4adc}}\right] \geq 0 \\ \frac{\partial q^*_M}{\partial \overline{\alpha}} &= \frac{1}{2c}\left[-(\frac{\partial b}{\partial \overline{\alpha}}-c\frac{\partial a}{\partial \overline{\alpha}})+\frac{(-b+d+ac)(\frac{\partial b}{\partial \overline{\alpha}}-c\frac{\partial a}{\partial \overline{\alpha}})+2dc\frac{\partial a}{\partial \overline{\alpha}}}{\sqrt{(b-d-ac)^2-4adc}}\right] \leq 0 \\ \frac{\partial p^*_M}{\partial \overline{\phi}} &= \frac{1}{2a}\left[-\frac{\partial d}{\partial \overline{\phi}}+a\frac{\partial c}{\partial \overline{\phi}}+\frac{(-b+d+ac)\left(-\frac{\partial d}{\partial \overline{\phi}}+a\frac{\partial c}{\partial \overline{\phi}}\right)+2a\left(c\frac{\partial d}{\partial \overline{\phi}}+d\frac{\partial c}{\partial \overline{\phi}}\right)}{\sqrt{(b-d-ac)^2-4adc}}\right] \leq 0, \end{aligned} \tag{A4}$$

where the signs of the derivatives follow from the facts that $c > 1$, $a > 1$, $b < ac$, $\frac{\partial a}{\partial \overline{\alpha}} < \frac{\partial b}{\partial \overline{\alpha}}$, and $\frac{\partial c}{\partial \overline{\phi}} < \frac{\partial d}{\partial \overline{\phi}}$.

Survey Methodology

The household survey consisted of face-to-face interviews of 4200 residents (18 year old and older) of the Gran Area Metropolitana (GAM), which includes 30 cantons in the provinces of Alajuela, Cartago, Heredia, and San José. The GAM is the principal urban center in Costa Rica. It contains approximately 2.6 million 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, between October 2013 and April 2014. On average, interviews lasted 25 minutes.

A two-stage clustered random sample based on the 2000 national census was generated (with fixed proportions for age and gender). 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, interviewers were selected based on quotas by gender and age, 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 pilot consisting of 48 cases, administered in October 15 and 16. 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, especially the different experimental treatments and the crosswise questions.

All survey enumerators utilized PDAs (personal digital assistants) to conduct the survey. An initial set of questions in the survey asked respondents about their personal experiences with crime and corruption, and perceptions of efficiency and corruption within the police and judicial authorities. After these questions, the experimental component of the survey began. The randomization of treatment conditions was programmed directly into the PDA. The PDA indicated to the enumerator which informational flyer, if any, should be given to the respondent. If one of the two informational flyers was selected, the enumerator gave a laminated sheet containing the display to the respondent and asked that she read the information contained therein. After the respondent had read the sheet, she returned it to the enumerator. Subsequent questions contained in the survey queried respondents about the perceived credibility of the informational display (if one was assigned), beliefs about the police and judicial authorities, overall sense of security, willingness to report crime and collaborate with the criminal justice system, tolerance for police violence, tolerance of corruption, beliefs about the scope of corruption, and the socio-demographic characteristics of the respondent herself.

Survey enumerators were recruited by Borge y Asociados and were mostly experienced with the administration of surveys. They went through extensive training on the details and administration of the survey instrument, especially on the execution of the crosswise questions and the administration of the different treatments. The training for the crosswise component of the survey consisted of a thorough explanation of the logic and functioning of the technique, as well as live practice sessions in which each enumerator practiced her delivery of this section of the survey both in front of members of the research team and administrators from Borge y Asociados. By contract, only enumerators that had gone through these training sessions participated in the administration of the survey. Any potential enumerator demonstrating insufficient mastery in the delivery of this component of the survey—the most challenging— in the training sessions was removed from the team of enumerators. An important feature of the delivery of this component of the survey consisted of a script describing to respondents how a hypothetical individual with a particular value on a sensitive item and a mother born in a particular month would respond to a given crosswise item. This script was given to all respondents prior to the commencement of the sensitive questions of interest.

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 subgroup of the sample (30% percent of the total) by phone, after which this information was destroyed. Team leaders also conducted verification in the field by randomly selecting households for verification the same day that the interview was conducted. If mistakes were found using 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.[2]

Focus Groups

Focus groups were conducted in San José with residents of varied backgrounds on August 6, 7, and 8, 2013, prior to fielding the household survey. The goal of these focus groups was threefold. First, the purpose was to get a general sense of individuals' perceptions of the main topics covered in the survey: corruption and inefficiency in different areas of the government, crime and issues of citizen security, and reporting of crime. Second, we tested different versions of the treatments to be used on the household survey. Finally, we evaluated each group's understanding of the logic of the crosswise questions.

Phone Survey

Prior to conducting 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 was twofold. First, we used the survey to evaluate our questions, the questions' wording, and the order of questions for the household survey. Second, we wanted to collect information about respondents' recollections of their parents' birthdays in order to be able to use that information for the crosswise questions in the household survey. We did so by asking respondents directly about the day of birth of their mother and father in the telephone survey. To check the veracity of these self-reports, these were checked against statistical tables produced by Costa Rica's National Institute for Statistics and Censuses (INEC) on month of birth for newborns for the 2000-2011 period (the period for which the data was available). Since there should be no systematic differences in month and day of birth across sex of child, responses for mothers and fathers were pooled together. The comparison in Appendix Table 4 shows that self-reported parent's birthdays were almost identical to the actual information obtained from INEC.

[APPENDIX TABLE 4 ABOUT HERE]

Enumerator Scripts

Main outcome variable - Crosswise Format

Interviewers explained first the logic of the crosswise format with the following script:

"Now I am going to ask you a series of questions with a special format. These are questions especially design to protect the privacy of your answers. To be able to answer them you will have to remember (and do not tell me) the birthday of some of your relatives. The technique is based precisely on the fact that neither me, nor any of the people involved on the survey know the birthday of your relatives. This is what guarantees that we cannot not know exactly what your answer was. "Let me show you an example… " (Ahora le voy a hacer una serie de preguntas que tienen un formato especial. Son preguntas diseñadas especialmente para proteger la privacidad de

[2]Rates calculated according to the American Association of Public Opinion Research.

sus respuestas. Para poder contestarlas usted va a tener que recordar (y no decirme) el día del cumpleaños de algunos de sus parientes. La técnica se basa justamente en que ni yo, ni ninguna de las personas involucradas en la encuesta, conocen el cumpleaños de sus parientes. Esto es lo garantiza que no podamos saber cual fue exactamente su respuesta. Le muestro con un ejemplo…). Interviewers then showed respondents the example card presented in Appendix Figure 5.

[APPENDIX FIGURE 5 ABOUT HERE]

Continuation of the script:

"In my case, my mother's birthday is in the month of December and I WOULD be willing to tell a lie to avoid a family conflict. So, my answer to the question: 'How many of the following statements are true?' is "A" ("Both or neither of the statements are true"). Now let's suppose that my mother's birthday was in January and I've already told you that I would be willing to tell a lie, so my answer in this case would be "B" ("Only one of the statements is true"). Finally, if my mother's birthday was in January and I would NOT be willing to tell a lie, then my answer would be "A" because neither of the statements would be true. Since nobody knows the date of my mother's birthday, it is not possible to identify my answer to the specific statement about lying. Did I explain myself clearly? Would you like me to repeat the example? (En mi caso, mi madre cumple años en el mes de diciembre y yo SI estaría dispuesto a decir una mentira para evitar un conflicto familiar. Por lo tanto, mi respuesta a la pregunta "¿Cuántas de las siguientes afirmaciones son ciertas?" es la "A" ("Las dos o ninguna de las dos afirmaciones son ciertas"). Ahora supongamos que mi madre cumpliese años en enero, y ya le dije que yo estaría dispuesto a decir una mentira, entonces mi respuesta sería la "B" ("Una sola de las afirmaciones es cierta"). Por último, si mi madre cumpliese años en enero y yo NO estuviese dispuesto a decir una mentira, mi respuesta sería la "A" porque ninguna de las afirmaciones es cierta. Como nadie sabe cuándo es el cumpleaños de mi madre, no es posible saber realmente cual es mi respuesta a la pregunta sobre mentiras. ¿Me explico? ¿Le gustaría que le repita el ejemplo?)

Enumerators were instructed to explain the technique and repeat the example as many times as was necessary for the respondents to understand the technique. Once this was achieved, enumerators handed out the card shown in Appendix Figure 6.

[APPENDIX FIGURE 6 ABOUT HERE]

Main outcome variable - Direct Questioning Format

Interviewers explained first why we were asking the same question twice with the following script:

"I've just asked you a series of questions about topics that were a little sensitive by using a technique that protects the privacy of the responses. Thanks to that technique, as I was explaining before, there is no way for us to identify your precise answer to those questions. However, we know that not everyone thinks that these topics are especially sensitive. Thus, in finishing with the survey we would like to ask you directly about these same topics. Of course, if you prefer not to answer any of these questions, please just let me know. For each of these questions, please tell me if the statement is true, false, or if you would rather not answer." (Hace un rato le hice una serie de preguntas sobre temas un poco sensibles utilizando una técnica que protege la privacidad de las respuestas. Gracias a esa técnica, como le explicaba antes, no tenemos forma de saber exactamente qué es lo que Ud. nos contestó. Sin embargo, sabemos que no todo el mundo considera esos temas tan sensibles así que para finalizar la encuesta nos gustaría preguntarle nuevamente en forma directa sobre esos mismos temas. Por supuesto, si Ud. prefiere no contestar a alguna de estas preguntas, simplemente me dice. En cada caso, dígame por favor si la afirmación es verdadera, falsa o prefiere no responder.)

After the enumerators provided this explanation they asked respondents: “In order to avoid paying a traffic ticket, I would be willing to pay a bribe to a police officer.” (Para evitar pagar una multa de tránsito, estaría dispuesto/a a pagar un soborno a un policía.) Response options were: "True", "False", and "I prefer not to respond" (Prefiero no contestar).

References

[1] Alligood, KT, TD Sauer, and JA Yorke. 1996. *Chaos: An Introduction to Dynamical Systems*. New York, NY: Springer-Verlag.

[2] Medina, LF. 2007. *A Unified Theory of Collective Action and Social Change*. Ann Arbor: University of Michigan Press.

[3] ______. 2013. The Analytical Foundations of Collective Action Theory: A Survey of Some Recent Developments. *Annual Review of Political Science* 16: 259-283.

[4] Ryvkin, D., and D Serra. 2012. "How Corruptible Are You? Bribery under Uncertainty." *Journal of Economic Behavior & Organization* 81(2), 466-477.

Appendix Table 4: Proportion of births falling into indicated months, telephone self-reports vs. census data

Births occurring in October, November, or December	
proportion of mother's and father's birthdays occurring in indicated months according to telephone survey self-reports: **0.264**	
actual proportion of newborn births occurring in indicated months (INEC)	
2000	0.268
2001	0.262
2002	0.268
2003	0.260
2004	0.264
2005	0.265
2006	0.268
2007	0.270
2008	0.266
2009	0.261
2010	0.262
2011	0.265
avg.	**0.265**

Appendix Table 5: Balance in Respondent Characteristics across Treatment Assignments

covariates	treatment groups						
	corruption		inefficiency		control		
	mean	s.e.	mean	s.e.	mean	s.e.	p.value
Demographics							
male	0.52	0.01	0.48	0.01	0.50	0.01	0.19
age	38.1	0.42	38.6	0.42	38.3	0.42	0.72
education							0.23
without study	0.01	0.00	0.01	0.00	0.01	0.00	
primary incomplete	0.07	0.01	0.06	0.01	0.07	0.01	
primary complete	0.19	0.01	0.20	0.01	0.22	0.01	
secondary incomplete	0.27	0.01	0.28	0.01	0.28	0.01	
secondary complete	0.24	0.01	0.24	0.01	0.23	0.01	
technical studies incomplete	0.02	0.00	0.01	0.00	0.01	0.00	
technical studies complete	0.03	0.00	0.02	0.00	0.03	0.00	
university incomplete	0.09	0.01	0.09	0.01	0.07	0.01	
university complete	0.08	0.01	0.09	0.01	0.07	0.01	
post-graduate	0.01	0.00	0.01	0.00	0.01	0.00	
head of household	0.49	0.01	0.46	0.01	0.48	0.01	0.32
cellphone	0.94	0.01	0.93	0.01	0.94	0.01	0.38
laptop	0.43	0.01	0.41	0.01	0.42	0.01	0.60
tablet	0.26	0.01	0.26	0.01	0.24	0.01	0.31
car	0.39	0.01	0.37	0.01	0.39	0.01	0.34
motorcycle	0.14	0.01	0.14	0.01	0.13	0.01	0.41
plasma, LCD, or LED TV	0.56	0.01	0.53	0.01	0.55	0.01	0.26
Cable or Satellite TV	0.70	0.01	0.71	0.01	0.72	0.01	0.73
internet	0.56	0.01	0.54	0.01	0.54	0.01	0.59
Costa Rican national	0.91	0.01	0.90	0.01	0.90	0.01	0.68
Prior beliefs							
In recent years, insecurity in the GAM has:							0.96
increased	0.72	0.01	0.71	0.01	0.73	0.01	
decreased	0.05	0.01	0.05	0.01	0.05	0.01	
stayed the same	0.23	0.01	0.24	0.01	0.22	0.01	
In recent years, corruption in Costa Rica has:							0.45
increased	0.83	0.01	0.84	0.01	0.83	0.01	
decreased	0.01	0.00	0.02	0.00	0.02	0.00	
stayed the same	0.16	0.01	0.14	0.01	0.16	0.01	
Of all the cases that enter the legal system, how many do you think are resolved?							0.38
the majority	0.08	0.01	0.08	0.01	0.07	0.01	
many	0.09	0.01	0.08	0.01	0.08	0.01	
few	0.37	0.01	0.41	0.01	0.38	0.01	
very few	0.41	0.01	0.38	0.01	0.41	0.01	
none	0.05	0.01	0.05	0.01	0.05	0.01	
Prior experiences							
direct contact with police or transit officer in previous year	0.25	0.01	0.26	0.01	0.26	0.01	0.54
knows personally a police officer	0.42	0.01	0.41	0.01	0.41	0.01	0.87
bribe solicited by police officer in previous year	0.03	0.00	0.03	0.00	0.03	0.00	0.81
knows personally someone accused, prosecuted, or sentenced by the criminal justice system	0.42	0.01	0.40	0.01	0.41	0.01	0.74

Appendix Table 6: Subgroup intent-to-treat estimates by gender}

	males			females		
Parameters	estimate	s.e.	95% int.	estimate	s.e.	95% int.
diagnostic parameters						
$\widehat{\lambda}_1^T$	0.62	0.03	[0.56,0.69]	0.58	0.06	[0.49,0.72]
$\widehat{\lambda}_1^L$	0.35	0.03	[0.28,0.42]	0.36	0.06	[0.21,0.46]
$\widehat{\lambda}_0^T$	0.96	0.01	[0.94,0.97]	0.98	0.01	[0.96,0.99]
explanatory parameters						
constant	-0.58	0.12	[-0.79,-0.34]	-1.46	0.17	[-1.84,-1.14]
corruption treatment	**0.20**	**0.14**	**[-0.08,0.48]**	**0.25**	**0.17**	**[-0.04,0.57]**
inefficiency treatment	**-0.04**	**0.16**	**[-0.36,0.25]**	**0.08**	**0.17**	**[-0.24,0.43]**
ITT (corruption vs. control)	**0.05**	**0.03**	**[-0.02,0.12]**	**0.04**	**0.03**	**[-0.01,0.09]**
		$n = 2096$			$n = 2097$	

Appendix Table 7: Subgroup intent-to-treat estimates by terciles of wealth index

	bottom tercile			middle tercile			top tercile		
Parameters	est.	s.e.	95% int.	est.	s.e.	95% int.	est.	s.e.	95% int.
diagnostic parameters									
$\widehat{\lambda}_1^T$	0.52	0.05	[0.42,0.65]	0.54	0.04	[0.46,0.64]	0.78	0.07	[0.67,0.93]
$\widehat{\lambda}_1^L$	0.44	0.06	[0.32,0.55]	0.43	0.04	[0.35,0.52]	0.18	0.07	[0.01,0.30]
$\widehat{\lambda}_0^T$	0.96	0.01	[0.94,0.98]	0.96	0.01	[0.95,0.98]	0.98	0.01	[0.97,1.00]
explanatory parameters									
constant	-1.11	0.20	[-1.52,-0.74]	-0.70	0.16	[-1.48,-0.84]	-1.10	0.17	[-1.46,-0.81]
corruption treatment	0.17	0.24	[-0.28,0.71]	0.26	0.18	[-0.10,0.60]	0.19	0.18	[-0.15,0.50]
inefficiency treatment	-0.07	0.23	[-0.51,0.38]	-0.19	0.20	[-0.61,0.14]	0.16	0.17	[-0.13,0.51]
ITT (corruption vs. control)	0.04	0.05	[-0.05,0.14]	0.06	0.04	[-0.02,0.14]	0.04	0.03	[-0.03,0.10]
		$n = 1272$			$n = 1400$			$n = 1391$	

Appendix Table 8: Subgroup intent-to-treat estimates by prior beliefs about changes in scope of corruption in Costa Rica in recent years

	increased			decreased or unchanged		
Parameters	estimate	s.e.	95% int.	estimate	s.e.	95% int.
diagnostic parameters						
$\widehat{\lambda}_1^T$	0.59	0.03	[0.53,0.66]	0.70	0.10	[0.54,0.93]
$\widehat{\lambda}_1^L$	0.37	0.04	[0.30,0.43]	0.29	0.09	[0.06,0.44]
$\widehat{\lambda}_0^T$	0.97	0.00	[0.96,0.98]	0.94	0.01	[0.93,0.97]
explanatory parameters						
constant	-0.94	0.10	[-1.14,-0.74]	-1.16	0.24	[-1.62,-0.68]
corruption treatment	**0.22**	**0.12**	**[-0.03,0.43]**	**0.30**	**0.27**	**[-0.23,0.82]**
inefficiency treatment	**-0.01**	**0.11**	**[-0.24,0.21]**	**-0.01**	**0.26**	**[-0.52,0.51]**
ITT (corruption vs. control)	**0.04**	**0.02**	**[-0.01,0.09]**	**0.06**	**0.05**	**[-0.05,0.16]**
		$n = 3472$			$n = 696$	

Appendix Table 9: Subgroup intent-to-treat estimates by age

	18-30 yrs			31-50 yrs			>50 yrs		
Parameters	est.	s.e.	95% int.	est.	s.e.	95% int.	est.	s.e.	95% int.
diagnostic parameters									
$\widehat{\lambda}_1^T$	0.65	0.04	[0.58,0.75]	0.62	0.05	[0.53,0.75]	0.42	0.10	[0.30,0.66]
$\widehat{\lambda}_1^L$	0.31	0.04	[0.21,0.39]	0.35	0.06	[0.22,0.44]	0.54	0.10	[0.27,0.69]
$\widehat{\lambda}_0^T$	0.97	0.01	[0.96,0.99]	0.96	0.01	[0.95,0.98]	0.97	0.01	[0.96,0.99]
explanatory parameters									
constant	-0.61	0.14	[-0.88,-0.35]	-1.14	0.16	[-1.48,-0.84]	-1.53	0.32	[-2.19,-0.95]
corruption treatment	0.23	0.16	[-0.07,0.56]	0.31	0.17	[-0.03,0.63]	0.05	0.33	[-0.66,0.66]
inefficiency treatment	-0.01	0.18	[-0.41,0.32]	0.11	0.18	[-0.25,0.46]	-0.07	0.34	[-0.71,0.66]
ITT (corruption vs. control)	0.06	0.04	[-0.02,0.13]	0.06	0.03	[-0.01,0.13]	0.00	0.05	[-0.09,0.09]
		$n = 1654$			$n = 1633$			$n = 905$	

Appendix Table 10: Subgroup intent-to-treat estimates by education (part 1)

	primary school or less			secondary school incomp.		
Parameters	estimate	s.e.	95% int.	estimate	s.e.	95% int.
diagnostic parameters						
$\widehat{\lambda}_1^T$	0.59	0.08	[0.48,0.76]	0.55	0.04	[0.47,0.65]
$\widehat{\lambda}_1^L$	0.34	0.08	[0.16,0.47]	0.44	0.04	[0.35,0.52]
$\widehat{\lambda}_0^T$	0.98	0.01	[0.96,0.99]	0.94	0.01	[0.92,0.97]
explanatory parameters						
constant	-1.27	0.20	[-1.68,-0.91]	-0.55	0.17	[-0.88,-0.20]
corruption treatment	**0.36**	**0.22**	**[-0.06,0.79]**	**0.11**	**0.21**	**[-0.29,0.52]**
inefficiency treatment	**-0.21**	**0.23**	**[-0.69,0.22]**	**-0.03**	**0.20**	**[-0.44,0.34]**
ITT (corruption vs. control)	**0.07**	**0.04**	**[-0.01,0.15]**	**0.02**	**0.05**	**[-0.07,0.12]**
		$n = 1160$			$n = 1162$	

Appendix Table 11: Subgroup intent-to-treat estimates by education (part 2)

	secondary school compl.			some technical or university		
Parameters	estimate	s.e.	95% int.	estimate	s.e.	95% int.
diagnostic parameters						
$\widehat{\lambda}_1^T$	0.66	0.06	[0.55,0.79]	0.66	0.09	[0.52,0.89]
$\widehat{\lambda}_1^L$	0.30	0.06	[0.16,0.41]	0.29	0.09	[0.05,0.44]
$\widehat{\lambda}_0^T$	0.98	0.01	[0.96,1.00]	0.97	0.01	[0.95,0.99]
explanatory parameters						
constant	-0.90	0.19	[-1.30,-0.56]	-1.32	0.25	[-1.87,-0.81]
corruption treatment	**0.17**	**0.22**	**[-0.21,0.59]**	**0.31**	**0.26**	**[-0.19,0.83]**
inefficiency treatment	**0.14**	**0.21**	**[-0.25,0.52]**	**0.04**	**0.27**	**[-0.49,0.55]**
ITT (corruption vs. control)	**0.04**	**0.05**	**[-0.04,0.13]**	**0.06**	**0.05**	**[-0.03,0.16]**
		$n = 986$			$n = 885$	

How many of the following statements are true?
(¿Cuántas de las siguientes afirmaciones son ciertas?)

- My mother was born in OCTOBER, NOVEMBER, or DECEMBER. (*El cumpleaños de mi **madre** es en octubre, noviembre o diciembre*)

- In order to avoid a family conflict, I would be willing to tell a lie. (*Para evitar un conflicto familiar, estaría dispuesto/a a decir una mentira*)

Mark the appropriate answer
(Indique la respuesta correcta):

A. **BOTH OR NEITHER** of the statements are true. (***LAS DOS O NINGUNA*** *de las afirmaciones son ciertas*)
B. **ONLY ONE** of the statements is true. (***UNA SOLA*** *de las afirmaciones es cierta*)

Remember: No one from the research team knows your mother's birthday. This guaranties the anonymity of your response. *(Recuerde: Nadie en el grupo de investigación sabe cuándo es el cumpleaños de su pariente. Esto garantiza la privacidad de su respuesta)*

Appendix Figure 5: The practice crosswise survey item

How many of the following statements are true?
***(¿Cuántas de las siguientes afirmaciones son ciertas?*)**

- My mother was born in OCTOBER, NOVEMBER, or DECEMBER. (*El cumpleaños de mi **madre** es en octubre, noviembre o diciembre*)

- In order to avoid paying a traffic ticket, I would be willing to pay a bribe to a police officer. (*Para evitar pagar una multa de tránsito, estaría dispuesto/a a pagar un soborno a un policía*)

Mark the appropriate answer
(Indique la respuesta correcta):

A. BOTH OR NEITHER of the statements are true. (***LAS DOS O NINGUNA*** *de las afirmaciones son ciertas*)
B. ONLY ONE of the statements is true. (***UNA SOLA*** *de las afirmaciones es cierta*)

Remember: No one from the research team knows your mother's birthday. This guaranties the anonymity of your response. *(Recuerde: Nadie en el grupo de investigación sabe cuándo es el cumpleaños de su pariente. Esto garantiza la privacidad de su respuesta)*

Appendix Figure 6:The actual crosswise survey item