Corruption as a Self-Fulfilling Prophecy: Evidence from a Survey Experiment in Costa Rica

Ana Corbacho
Western Hemisphere Department, International Monetary Fund
Daniel W. Gingerich
Department of Politics, University of Virginia
Virginia Oliveros
Department of Political Science, Tulane University
Mauricio Ruiz-Vega
Western Hemisphere Department, International Monetary Fund

July 28, 2015

Abstract[1]

An influential literature argues that corruption behaves as a self-fulfilling prophecy. Its central claim is that the individual returns to corruption are a function of the perceived corruptibility of the other members of society. Empirically, this implies that if one were to exogenously increase beliefs about societal levels of corruption, willingness to engage in corruption should also increase. We evaluate this implication by utilizing an information experiment embedded in a large-scale household survey recently conducted in the Gran Área Metropolitana of Costa Rica. Changes in beliefs about corruption were induced via the random assignment of an informational display depicting the increasing percentage of Costa Ricans who have personally witnessed an act of corruption. Consistent with the self-fulfilling prophecy hypothesis, we find that internalizing the information from the display on average increased the probability that a respondent would be willing to bribe to a police officer by approximately 0.05 to 0.10.

JEL Classification: D73, D84, C83, C90
Keywords: corruption, spillover effects, survey experiments, sensitive question techniques

[1] Prepared for presentation at the Annual Meeting of the American Political Science Association, San Francisco, CA, September 3-6, 2015. Correspondence should be directed to Daniel W. Gingerich (dwg4c@virginia.edu) or Virginia Oliveros (volivero@tulane.edu). This paper was prepared when Ana Corbacho was Sector Economic Advisor at the Inter-American Development Bank.

1 Introduction

A question of enduring interest in the social sciences is under what conditions do individuals decide to engage in illicit and socially harmful forms of behavior. Responses to this question vary widely, but one can organize thinking on this issue into two main groups of arguments based on the role they assign to an individual's community environment. The role of community may be analytically relegated to the background, viewed as investing individuals with norms and a sanction schedule that 'prices' illicit behavior in particular ways, but otherwise not operating directly on individual choices. Alternatively, the role of community may be placed at the forefront, such that discussions about the individual returns to illicit behavior are deemed to be largely meaningless without reference to the choices of other community members about engaging in similar behavior. The difference here is between viewing illicit behavior as an individual phenomenon embedded within a particular societal context and viewing illicit behavior as an intrinsically social phenomenon.

Both views are encountered in the contemporary literature on corruption, which is the specific behavior upon which we focus here. A relatively new and growing micro-level empirical literature frames its analysis of corruption largely along the lines of the first view. An older but still vibrant game-theoretic literature explicitly adopts the second view, arguing that levels of corruption emerge endogenously from a society wide coordination game in which the individual returns to corrupt behavior are increasing in the inclination towards corruption of the other members of society. The two approaches have distinctive implications. The first implies that holding individual moral values and expectations about punishment constant, beliefs about the societal frequency of corruption should hold little sway over decisions about corrupt action. The second view holds the opposite: the higher individuals perceive the level of corruption in society to be, the more inclined they will be to engage in corrupt behavior themselves. In this latter scenario, corruption behaves as a self-fulfilling prophecy.

Does corruption corrupt? Does the level of perceived corruption in a society affect an individual's willingness to engage in corrupt behavior? We argue in this paper that the answer to this question is yes, that the decision to engage in corrupt behavior is crucially shaped by perceptions of what other actors in society are doing or are inclined to do. In this sense, we argue that individuals approach the choice to engage in corruption much as a game theorist would, taking strategic interdependence into account in assessing the costs and benefits of their actions. In particular, using data from an original survey we conducted in Costa Rica, we show that learning about increasing levels of corruption in society increases the likelihood that citizens would be willing to bribe a police officer in order to avoid paying a traffic ticket.

Our conclusions are based on a carefully crafted empirical research design. To deal with issues of social desirability bias, we employ a novel technique that combines the bias reducing advantages of sensitive survey techniques with direct questioning, which results in more precise

estimates. To deal with the problem of confounding, we induce exogenous variation in beliefs about corruption using a survey experiment that provides information about the rising levels of corruption in Costa Rica. In this way, our paper advances the literature on corruption in two ways. First, we provide the first experimental evidence about the effects of perceived corruption in society on an individual's willingness to engage in petty corruption. Although a number of formal papers have previously argued that corruption corrupts, our study is unique in providing convincing empirical evidence that this is actually the case. Second, we use a novel method to measure corruption that significantly reduces bias and increases precision—a technique that can easily be replicated in the study of other sensitive activities.

Costa Rica provides a propitious environment for studying corruption. Although corruption in Costa Rica is relatively low by regional standards, the country has experienced a substantial increase in perceived corruption in recent years. According to the 2013 Corruption Perception Index by Transparency International, Costa Rica is ranked 49/177, considerably above Honduras (140), Nicaragua (127), Guatemala (123), Mexico (106), and Panama (102). In fact, the only two Latin American countries that perform better than Costa Rica on this metric are Uruguay (19) and Chile (22). However, according to a nationally representative survey conducted by *Latinobarómetro*, the number of people who have witnessed an act of corruption has increased from 16% in 2006 to 24% in 2011.[2] At the elite level, the country has seen repeated political corruption scandals over the last decade, including the indictment of three different past presidents for bribery (and the conviction of two of them) as well as a number of forced resignations among cabinet officials. Corruption scandals of such magnitude have no precedent in earlier periods of the modern history of the country. Given these abrupt changes occurring in a country once characterized as the "Switzerland of Central America," beliefs about how deeply embedded corruption is in Costa Rica society are likely to be in flux for many individuals. For this reason, the country is a natural setting for exploring how information about the scope of corruption may drive corrupt behavior.

2 The Choice for Corruption: Two Views of Decisionmaking

Broadly speaking, there are two basic views of the decision process by which an individual, faced with the opportunity, chooses to engage in a corrupt act or refrains from doing so. The first view is a decision-theoretic one.[3] In this view, the choice of an actor to engage in corrupt behavior results from a fundamentally atomistic and societally non-contingent risk-return calculus. Presented with the opportunity to pursue an illicit action, an actor engages in an introspection exercise in which she considers the potential rewards and opportunity costs of corrupt action, her personal moral

[2] This perception of increasing corruption was corroborated by the focus groups we conducted in Costa Rica prior to the survey. See the on-line Appendix for more details.

[3] Here we use the phrase decision-theoretic (as opposed to game-theoretic) to denote the analysis of decisionmaking by individual agents whose decisions do not affect the returns to the decisions adopted by other agents.

views on the subject, the fixed, institutionally determined likelihood of detection, as well as the magnitude of the sanction such detection brings. A society wide level of corruption percolates up directly from the results of many such introspection exercises. Such a decision theoretic perspective is implied in a large body of empirical work that puts pride of place on the demographic characteristics and values of individuals in explanations of corrupt behavior. Recent studies of corruption emphasizing the explanatory role of gender (Dollar, Fisman, and Gatti 2001; Swamy et al. 2001; Sung 2003; Esarey and Chirillo 2013), age (Torgler and Valev 2006, 2010), attitudes inherited from one's nation of origin (Fisman and Miguel 2007; Cameron et al. 2009), and partisan preferences (Anderson and Tverdova 2003; Anduiza, Gallego, and Muñoz 2013) all fit within this rubric.

The second vision is an explicitly game-theoretic one. In this perspective, the choice of an actor to engage in corrupt behavior results from a fundamentally interdependent and societally contingent risk-return calculus. Presented with the same opportunity, an actor's introspection exercise incorporates all of the attributes and considerations described above, but now also hinges crucially on beliefs about what other actors, faced with similar decisions of their own, are likely to do. As before, a society wide level of corruption percolates up from these introspection exercises; however, here the resolution of each such exercise is contingent on the resolution of the others. This second view of decisionmaking typically holds that choice over corrupt acts is characterized by strategic complementarities. For our purposes, this means that the decisions actors make about engaging in corruption are complementary to one another in the sense that the expected return that any given actor associates with engaging in corruption is increasing in the expected number of other actors who do so. Consequently, all actors have a strong incentive to coordinate their behavior, be that partaking in corrupt action or abstaining from it.

The vast majority of theoretical work on corruption in the political economy tradition adopts a game-theoretic perspective with strategic complementarities.[4] Work in this vein has long emphasized that corruption is subject to coordination dilemmas and herd behavior, and, as a consequence, that high (or low) levels of corruption tend to feed upon themselves and persist over time. The specific mechanisms adduced to explain why this is so are many and varied.

Some accounts concentrate on how the existence of corruption undermines sanctioning mechanisms, thereby furthering the incidence of corruption in the first place (Lui 1986; Cadot 1987; Andvig and Moene 1990; Mishra 2006). Search costs have also been invoked to explain how corruption corrupts: the more certain are parties to a corrupt exchange about the corruptibility of their partners, the less costly it may be to find a counterpart willing to consummate a corrupt bargain (Ryvkin and Serra 2012; see also Andvig and Moene 1990). Recent work from a psychological perspective emphasizes the role of guilt aversion in generating corruption spillovers (Bal-

[4] See Aidt (2003) for a review of this literature.

afoutas 2011). Theoretical frameworks taking a long term view have emphasized mechanisms such as the allocation of talent between productive activities and rent-seeking (read: corruption) (Açemoglu 1995), reputational lock-in for collectivities (Tirole 1996), the intergenerational transmission of cultural values (Hauk and Saez-Marti 2002), and imitative processes of strategy selection (Accinelli and Sánchez Carrera 2012). Finally, several explicitly political accounts of corruption have argued that corruption spillovers emerge via the selection mechanism determining who holds public office (Caselli and Morelli 2004; Dal Bó, Dal Bó, and Di Tella 2006) or by affecting the bargaining power of political machines vis-a-vis rank-and-file members in the bureaucracy (Gingerich 2009). In all of this work, society wide corruption is envisioned as an inherently emergent phenomenon.

Both the decision-theoretic and game-theoretic approaches toward choice over corrupt acts are plausible on their face, as becomes clear when they are applied to police stops for traffic infractions, one of the most common arenas of petty corruption in the developing world and the focus of this paper. In such situations, the potential for a mutually beneficial transaction that defrauds the public fisc is clearly present. For example, the driver could pay the police officer a bribe in some amount lower than the official sanction for the infraction, and, in so doing, both agents would be better off than if they had followed the letter of the law. Alternatively, the police officer could insinuate his willingness to accept a bribe in lieu of applying the officially mandated sanction, again providing both actors with a financial benefit.

According to the decision-theoretic approach, a corrupt transaction would be consummated if both the driver and the police officer independently assessed that the financial returns to the exchange were sufficiently high, that their normative qualms–if any–were sufficiently minor, and that the risk of detection and its associated sanction were both sufficiently small. Seen in this light, the proportion of police stops in a polity that would result in corrupt transactions would be completely determined by the distribution of utilities generated by the pecuniary returns to the transaction, the distribution of moral tastes for or against corruption among drivers and police officers, and the quality of institutions insofar as concerns the monitoring and sanctioning of corruption.

A game-theoretic perspective would analyze the same situation differently. It might begin emphasizing the point that actually executing a corrupt transaction is not easy. For instance, if the driver offers a bribe to an unwilling police officer, she runs the risk of sanction for attempted bribery in addition to that for the initial infraction. Likewise, if the police officer begins the process of extorting a bribe from a driver strongly disposed against corruption, he runs the risk of being reported. Given the limited information that the driver and police officer can glean about each other during the traffic stop, each will have to make a decision to initiate or not initiate a corrupt transaction based largely upon their beliefs about what the typical driver and typical police officer is likely to do in such a scenario. As a consequence, expectations of social behavior are now

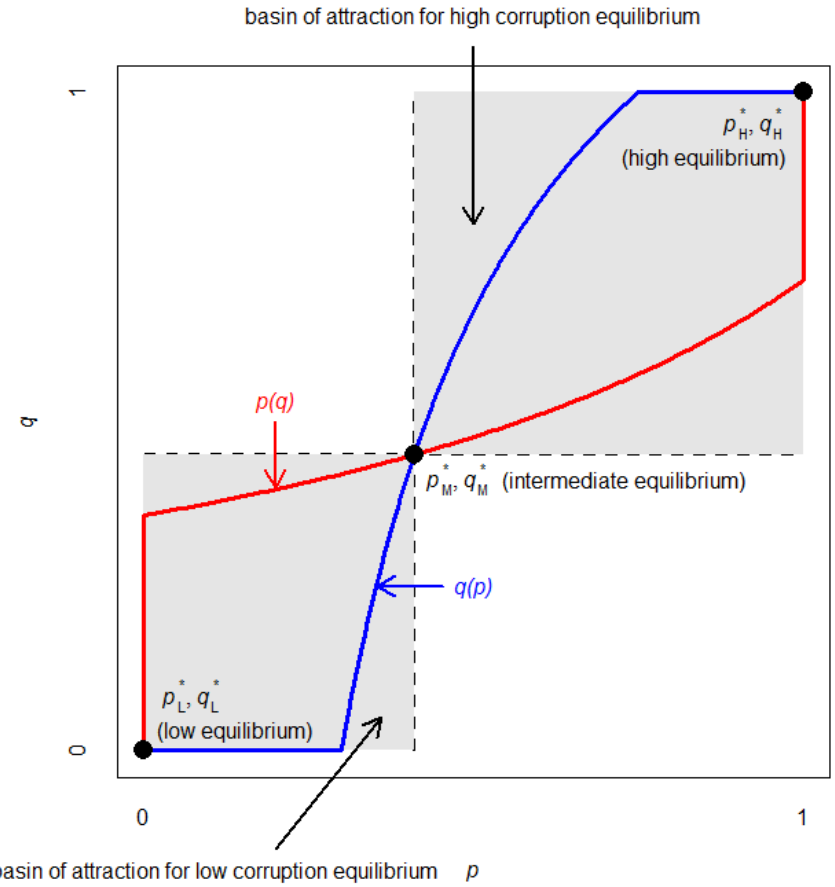
central to the choice problem of each agent. In such a setting, any given driver will be more inclined to initiate a corrupt transaction the greater the proportion of corruptly inclined police officers she believes there to be, since–in the absence of detailed information about the particular officer with whom she is dealing–she views said officer as a random draw from the population of officers. Similarly, and for the same reason, any given police officer would be more inclined to extort a bribe the greater the proportion of corruptly inclined drivers he believes there to be. The coordination of beliefs about the likely actions of others is central in this perspective, and, as such, it plays a crucial role in determining the proportion of police stops that ultimately result in corrupt transactions. Pecuniary returns, tastes for corruption, and institutions all remain relevant, but they alone are far from decisive in determining the prevalence of corrupt transactions.

The strategic complexity inherent to such exchanges is non-trivial. Thus, in order to properly consider the implications of a game-theoretic decision process marked by strategic complementarity for corruption during traffic stops, we developed a formal model specifically tailored to capture what we view as the central elements of such encounters: anonymity, two-sided uncertainty, and belief conditionality (for both drivers and police officers) of the returns to initiating a corrupt exchange. All details of the game–players, actions and timing, informational conditions, and formal proofs–are provided in the on-line appendix of this paper.

The central intuitions of our game can be gleaned from an examination of Figure 1. The proportion of drivers willing to denote a disposition towards corruption during a traffic stop, labeled q, is represented by the blue line displayed on the x-axis. This proportion is a function of the (expected) share of police officers disposed towards corruption. The proportion of police officers willing to denote a disposition towards corruption, labeled p, is represented by the red line displayed on the y-axis. This proportion is a function of the (expected) share of drivers disposed towards corruption. The equilibria for the corruption game are the points of intersection between the two lines (depicted as large black circles).

The first item to note is the fact that there are multiple equilibria. Indeed, there are three equilibria: a high corruption equilibrium (p^*_H, q^*_H) in which all drivers and police officers indicate a disposition towards corruption (all police stops result in a corrupt transaction), a low corruption equilibrium (p^*_L, q^*_L) in which no drivers and police officers indicate a disposition towards corruption (no police stops result in a corrupt transaction), and an intermediate equilibrium (p^*_M, q^*_M) in which the proportions of drivers and officials who indicate a disposition to engage in corruption fall within an interval between zero and one (some police stops result in a corrupt transaction). It is precisely the incentive for coordination that generates the multiplicity of equilibria.

Figure 1. Equilibria for the corruption game



Note: The proportion of citizens willing to denote a disposition towards corruption is represented by the blue line. The proportion of officials willing to denote a disposition towards corruption is represented by the red line. The equilibria for the corruption game are the points of intersection between the two lines (depicted as large black circles).

Yet not all the equilibria merit the same consideration. Of the three, only the low equilibrium and the high equilibrium are stable.[5] In particular, if drivers and police were to have 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. The same would not occur for the intermediate equilibrium. Thus, the two extreme equilibria are robust to small perturbations in beliefs, whereas the intermediate equilibrium is not.

In this framework, pecuniary returns, moral tastes for corruption, and institutions shape the likelihood that one or the other of the two stable corruption equilibria will occur. More specifically,

[5] To be specific, these equilibria satisify a criterion for plausibility called dynamical stability; the intermediate equilibrium does not satisfy this criterion. See the on-line appendix for details.

these items affect the relative size of the basins of attraction of the two equilibria (shaded in gray in Figure 1), defined as the set of initial beliefs which would ultimately lead a particular equilibrium to prevail. The severity of sanctions for corruption affects the relative size of the basins of the two plausible equilibria in a highly intuitive way: the greater the severity of sanctions, the larger (smaller) the relative size of the basin for the low (high) corruption equilibrium. Similarly intuitive is the influence of economic arrangements and cultural norms that determine tastes for corruption: the more intense said tastes, the smaller (larger) the relative size of the basin for the low (high) corruption equilibrium. Thus, the quality of institutions and moral tastes shape the scope of the gravitational pull of each plausible equilibrium, in so doing making driver-police coordination around one equilibrium point more or less likely than coordination around the other.[6]

However, the relevance of moral tastes and institutions notwithstanding, there is a clear sense in which the model implies that the influence of these factors is subsidiary to the role of expectations. Even in a polity whose institutions strongly sanction corruption and whose citizenry finds it distasteful, it will generally be the case that sufficiently cynical beliefs can lead drivers and police to coordinate around the high corruption outcome. By the same token, a polity whose institutions are highly permissive to corruption and whose citizens are generally tolerant of it may still wind up on the low corruption equilibrium should drivers' and police officers' beliefs about one another be sufficiently sanguine. Expectations about corruption act as self-fulfilling prophecies. As such, constitutive beliefs about the typical behavior of members of society play a central explanatory role in game-theoretic models of corruption with strategic complementarities; they play no such role in analogous decision-theoretic frameworks.

It would be hard to overstate how different the policy implications are that emerge from a decision-theoretic versus game-theoretic view of decisionmaking over corruption. From a decision-theoretic perspective, polities that have high levels of corruption are the way they are because economic arrangements generate high pecuniary returns to corrupt transactions, citizens and officials have "bad" preferences, institutions that monitor and sanction corruption are weak and ineffective, or some combination of the above. Improving any one of these items will directly reduce the incidence of corruption in any given polity. From a game-theoretic perspective, polities suffering from high levels of corruption may very well have "good" preferences and institutions but are the way they are because citizens and officials have coordinated around a set of highly pessimistic beliefs about one another. Changing the underlying fundamentals without altering the coordination of beliefs may not solve the problem.

In the pages that follow, we provide an explicit empirical evaluation of the relevance of beliefs about the incidence of corruption in society for individual choices about engaging in corrupt behavior. Our analysis clearly demonstrates that the inclination to act corruptly is contingent

[6] See Medina (2007) for a general discussion of the probability of equilibria in coordination games.

on beliefs about the scope of corruption in society. For nearly thirty years, game-theoretic models of corruption with strategic complementarities have emphasized the importance of this belief contingency at a theoretical-level. Our paper is the first to empirically establish its relevance using experimental evidence.

3 Empirical Studies of Corruption Spillovers

The empirical literature examining the degree to which corrupt behavior exhibits strategic complementarities is still at a fairly early stage. Several papers have used aggregate data to show that corruption might be contagious (an empirical implication of complementarity). For example, Dong and Torgler (2012) use province-level data for China from 1998 to 2007 to show that social interaction has a significant positive effect on corruption. Using state-level US data from 1995 to 2004, Goel and Nelson (2007) also find evidence of neighboring corruption: a 10% increase in corruption in neighboring states increases corruption in a state by about 4 to 11%. Lopez-Valcarcel, Jiménez and Perdiguero (2014) similarly find evidence that corruption is contagious using a dataset of local Spanish municipalities from 2001-2010. Studies by Becker, Egger, and Seidel (2009) and Goel and Saunoris (2014) utilize cross-national data to estimate the degree to which corruption in one country affects its neighbors. In both cases, the authors find evidence of spillover effects. However, contrary to the findings of these studies, Marquez, Salinas-Jiménez, and Salinas-Jiménez (2011) find no evidence of corruption spillovers.

Articles that explore the contagiousness or self-fulfilling prophecy hypothesis using individual-level data as we do in this paper are few. Using data from the European Values Survey, Dong, Dulleck and Torgler (2012) show that the more respondents perceive others as being corrupt, the more tolerant they are towards someone accepting a bribe in the course of their duties. Similarly, a recent report from the Latin American Public Opinion Project (LAPOP), based on data from 24 Latin American countries, finds a positive correlation between beliefs that corruption is widespread among public officials and the likelihood of considering paying a bribe to be justified (Plata 2012). Recent work by Simpser (2013) provides individual-level evidence on one mechanism by which corruption can corrupt, namely, the intergenerational transmission of attitudes towards corruption. Using data from the US and European General Social Surveys, he finds that holding current institutional context constant, attitudes towards corruption in a respondent's ancestral country of origin strongly predict current attitudes towards the same. In laboratory experiments, Innes and Mitra (2013) find that information treatments indicating a peer propensity towards dishonesty in a game setting increased the likelihood of dishonest play, a finding that they interpret as evidence for the contagiousness of corruption hypothesis.

In a broader reading, our paper can be seen as a contribution to the literature on how social context affects individuals' willingness to engage in crime. Empirical evidence has shown that the

decision to commit a crime is affected by the behavior of others. For instance, using data from US cities and New York City neighborhoods, Glaeser, Sacerdote, and Scheinkman (1996) show that individuals are more likely to commit crimes when crime around them is widespread, especially less serious crimes such as larceny and auto theft. Using survey data from the US, Sheffrin and Triest (1992) find that perceiving other taxpayers as dishonest increases the likelihood that an individual will evade taxes. Other studies have shown that neighborhood and peer effects are similarly important in explaining related behaviors such as academic cheating (Carrell, Malstrom, and West 2008), fraud in emissions testing (Pierce and Snyder 2008), and shirking on the job (Ichino and Maggi 2000).

Finally, our paper is part of a growing literature that uses experimental and quasi-experimental methods in order to understand how information about corruption shapes the behavior of citizens. Most of this work has concentrated on how information about corruption affects vote choice and political participation. For example, taking advantage of a natural experiment generated in Brazil by the randomized federal auditing of local governments, Ferraz and Finan (2008) show that mayors revealed to be corrupt lose electoral support. Focusing on the case of Mexico, Chong et al. (2015) found that distributing information about a corrupt incumbent decreases incumbent support as well as turnout. Using an information experiment conducted in Brazil, Winters and Weitz-Shapiro (2013) present findings that indicate that information about corruption decreases support for an hypothetical corrupt politician, even when said politician performs well in office. The current study extends upon this work by explicitly considering the role that citizens–as opposed to politicians or other officeholders–may play in actively propagating corruption throughout their societies. Moreover, our paper is unique in the literature in that it provides micro-level experimental evidence on how "bad news" about corruption may lead citizens to perpetuate a vicious behavioral circle.

4 Measuring Citizens' Willingness to Engage in Corruption: The Joint Response Model

Accurately measuring whether a citizen would be willing to bribe (or has done so in the past) has long been recognized as one of the great challenges of empirical scholarship on corruption (e.g. Treisman 2007). Recognizing both the potential of social surveys to study corruption as well as the biases they invite when applied in standard form to sensitive issues, a number of scholars have begun to employ sensitive survey techniques (SSTs) in studies of this topic (Gingerich 2010, 2013; Malesky, Gueorguiev, and Jensen 2015). Following this lead, we utilize an SST-based approach in the current paper. However, we do so in a novel way, by utilizing individual responses about corruption based on *both* a specific SST as well as upon direct questioning. We refer to the statistical framework we utilize to analyze the joint protected and direct responses as the joint

response model. In a recent contribution, Gingerich, Oliveros, Corbacho, and Ruiz-Vega (2014) show that utilizing a joint response approach provides all of the bias reducing advantages of pure SST questioning while at the same time greatly enhancing the precision of parameter estimates.

4.1 Survey Questioning Format

The questioning strategy we utilized is easy to describe. First, survey respondents were presented with a question about willingness to bribe a police officer in order to avoid a traffic ticket using a particular SST format called the crosswise model (Tan, Tian, and Tang 2009). This technique provides anonymity to respondents via the commingling of responses about a sensitive issue (corruption) with responses about an innocuous question.[7] Next, at a later stage of the survey respondents were queried directly about willingness to bribe a police officer (in the exact same context), with the explicit option of "choose not to respond directly" provided to them in case they deemed a direct response to be uncomfortable. Observed responses about willingness to bribe were thus a discrete combination of responses under the protection afforded by the SST and the absence of protection under direct questioning.

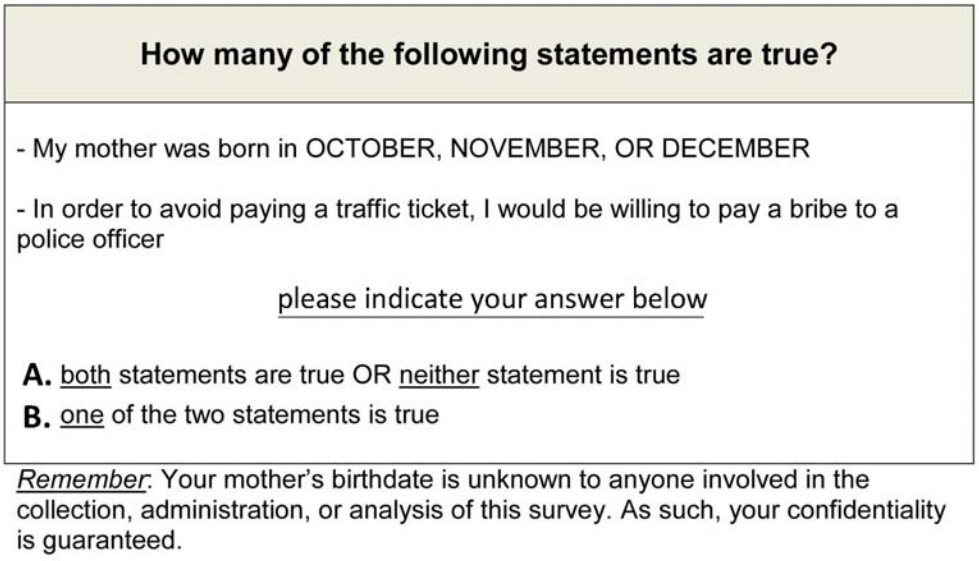
Figure 2 presents the question about willingness to bribe based on the crosswise model. The respondent was presented with two statements and asked how many were true. The first statement simply states that the respondent's mother was born in October, November, or December. One can conceptualize affirmative responses to this statement as indicating membership in a non-sensitive group. The second statement, the one of primary interest, denotes a willingness to pay a bribe. The privacy of the respondent was protected by constraining the manner in which she was allowed to respond. In particular, there are only two potential responses: one response (A) indicating that either both statements are true OR neither statement is true and another response (B) indicating that only one of the two statements is true (but not specifying which is true). Since neither of the two responses necessarily indicates willingness to bribe, the respondent's anonymity is guaranteed. For this reason, the respondent may be liberated from social desirability concerns that might otherwise prevent her from giving an honest answer about corruption.

In using the crosswise model, it is important to note that membership in the non-sensitive group (i.e. an unobserved, affirmative response to the first statement) is special in that it: 1) must be known to each respondent but unknown to survey administrators (and known by each respondent to be unknown to administrators); 2) must be statistically independent of the sensitive trait of interest (willingness to bribe); 3) must have a proportion in the population of interest that is known in advance; and 4) must have proportion which is different from 1/2 (otherwise the crosswise responses would provide no information). Using the birth month of one's mother, as we did for this question, helps ensure that nearly all respondents would know their own group

[7] The crosswise model is mathematically identical to the Warner version of the well-known randomized response technique, but it is administered without the use of a randomizing device.

assignment and that they would also be aware that the survey enumerator did not know their group assignment. Moreover, there is no plausible mechanism by the birth month of one's mother should be tied to willingness to bribe, so the group indicator and the sensitive item are surely independent of one another. Finally, the population proportion of individuals belonging to the non-sensitive group is verifiable based on census records, meaning that such a group can be easily chosen such that the probability of membership differs arbitrarily from $1/2$.[8]

Figure 2. The crosswise survey item on willingness to bribe



The direct question about willingness to bribe, presented to respondents at the very end of the survey, asked them to respond only to the second statement presented in Figure 1. In this case, response options were "True", "False", and "I prefer not to respond." Detailed information about the enumerator scripts used for questioning about willingness to bribe is provided in the on-line appendix.

In order to calculate the probability of having one's mother born in the indicated interval of months, we conducted a nationally representative telephone survey of 1,200 Costa Ricans during July 15-20, 2013. The survey queried respondents directly about the birthday of their mother and father. To assess the veracity of the survey reports, these responses were checked against data from Costa Rica's National Institute for Statistics and Censuses (INEC) on month of birth for newborns for the 2000-2011 period.[9] The figures from the self-reports and census data were essentially

[8] Recent studies have reported very good performance with the crosswise model in substantive applications ranging from cheating by undergraduates to tax evasion (Jann, Jerke, and Krumpal 2012; Körndorfer, Krumpal, and Schmukle 2014; Kundt 2014).

[9] Appendix Table 4 in the on-line appendix compares the proportion of birthdays falling into the indicated months from the survey self-reports to the actual proportions for newborn births produced by INEC.

identical. This implies that it is highly unlikely that recall bias or similar problems were present in such a way as to pose a threat to the use of birthdays as employed in this paper.

4.2 Statistical Framework

Following the discussion above, we consider a setting in which each respondent i in a randomly selected sample of size n is first queried about her (unobservable) willingness to bribe $\theta_i \in \{0$ ("unwilling"), 1 ("willing")$\}$ using the crosswise method then later asked if she would be willing to respond directly to a question about the same topic. If the respondent responds affirmatively to the latter question, she is then prompted to respond directly to the statement about bribery. The (observable) combined response of respondent i to the two questions is denoted by the vector $Y_i = (y_i^D, y_i^A)$, where $y_i^D = \{0$ ("False"), 1 ("True"), $\varnothing$ ("unwilling to respond directly")$\}$ is the observed response when i is asked to respond directly and $y_i^A \in \{0$ ("B"), 1 ("A")$\}$ is the observed ("anonymous") response when i is queried about bribery using the crosswise model. The observed response set is thus an array with six distinct elements, $\mathcal{Y} = \{(0,0),(0,1),(1,0),(1,1),(\varnothing,0),(\varnothing,1)\}$, with $k \in \mathcal{Y}$ representing an arbitrary element in this set. Without loss of generality, we can relabel responses as $Y_i \in \mathcal{Y} = \{1,2,...5,6\}$, where each natural number $1,..,6$ represents one of the six distinct response combinations. For the responses using the crosswise technique, $z \neq 1/2$ will denote the probability that the first statement is true (e.g. the probability that the respondent's mother was born in the indicated interval of months). This quantity is known prior to collecting the data. (For the question displayed in Figure 2, $z = 0.264$.)

Our primary interest in this paper resides in estimating the parameters of a model of the conditional probability of being willing to bribe given a respondent's experiences and observed characteristics. Let $\pi_i \equiv \mathbb{P}(\theta_i = 1|\mathbf{X}_i) = (1+\exp(-\mathbf{X}_i^\top \boldsymbol{\beta}))^{-1}$ where $\mathbf{X}_i$ is a vector of background characteristics and/or a treatment assignments recorded in the social survey along with a constant and $\boldsymbol{\beta}$ is the parameter vector. Since $\boldsymbol{\beta}$ reflects the influence of the experiences or characteristics of a respondent on her willingness to bribe, we refer to the elements of this vector as the explanatory parameters of our statistical model.

Our statistical framework rests on two key assumptions. The first is called *honesty given protection*: given the protection afforded by the crosswise model, all respondents are assumed to respond honestly and as prompted by the technique (cf. Gingerich 2010; Blair and Imai 2012). In other words, lying is assumed to occur *only* when respondents are prompted to respond directly about their willingness to bribe. The second assumption is called *one-sided lying*. It holds that individuals who do not bear the sensitive trait never falsely claim that they do. Rather, the set of potential liars is limited to those respondents who do bear the sensitive trait.

The statistical model parameterizes patterns of evasiveness under direct questioning. In particular, let λ_θ^T, λ_θ^L, and $1 - \lambda_\theta^T - \lambda_\theta^L$ denote the probability that, when queried directly, a respondent whose status is θ tells the truth about her willingness to bribe, lies about her willingness,

Table 1. Probability table for observed data under assumption of honesty given protection and one-sided lying

Y_i	outcome	probability
1	$(y_i^D = 0, y_i^A = 0)$	$z\lambda_0^T(1-\pi_i) + (1-z)\lambda_1^L\pi_i$
2	$(y_i^D = 0, y_i^A = 1)$	$(1-z)\lambda_0^T(1-\pi_i) + z\lambda_1^L\pi_i$
3	$(y_i^D = 1, y_i^A = 0)$	$(1-z)\lambda_1^T\pi_i$
4	$(y_i^D = 1, y_i^A = 1)$	$z\lambda_1^T\pi_i$
5	$(y_i^D = \varnothing, y_i^A = 0)$	$z(1-\lambda_0^T)(1-\pi_i) + (1-z)(1-\lambda_1^T-\lambda_1^L)\pi_i$
6	$(y_i^D = \varnothing, y_i^A = 1)$	$(1-z)(1-\lambda_0^T)(1-\pi_i) + z(1-\lambda_1^T-\lambda_1^L)\pi_i$

or refuses to answer the question, respectively. Formally, one-sided lying implies that $\lambda_0^L = 0$. The assumption reflects the presumed direction of social desirability bias in sensitive surveys. If concerns about social desirability make it difficult for respondents with a sensitive trait to publicly divulge their status, those same concerns should ensure that respondents without the sensitive trait would have no incentive to falsely state that they bear the trait. Since this second set of parameters captures potential biases in responses generated by direct questioning, we refer to these as the diagnostic parameters of our statistical model.

The probability that a given respondent exhibits each combination of responses in the observed response set is presented in Table 1. Each cell of the table expresses the probability of observing the particular response combination represented by that cell.

Let $I(.)$ be an indicator function equal to 1 if its argument is true, 0 otherwise, $\mathbb{P}_Y(k|\mathbf{X}_i)$ be the probability that respondent i's observed joint response is in category k given her background characteristics, the model for observed responses (e.g. the probabilities presented in Table 1), and the model for the conditional probability of being willing to bribe, and let $\boldsymbol{\xi} = (\lambda_1^T, \lambda_1^L, \lambda_0^T, \boldsymbol{\beta})^\top$ be the vector of parameters to be estimated. The log-likelihood function for the parameters given the observed data is written:

$$\ln L(\boldsymbol{\xi}|Y, \mathbf{X}) = \textstyle\sum_{i=1}^{n}\sum_{k=1}^{6} I(Y_i = k)\ln \mathbb{P}_Y(k|\mathbf{X}_i). \tag{1}$$

Note that if one simply wishes to calculate the (unconditional) proportion of individuals willing to bribe, one can write $\pi_i = \pi = \mathbb{P}(\theta_i = 1)$. In this case, $\boldsymbol{\xi} = (\pi, \lambda_1^T, \lambda_1^L, \lambda_0^T)^\top$ and the log-likelihood function simplifies to

$$\ln L(\boldsymbol{\xi}|Y) = \textstyle\sum_{k=1}^{6} n_k \ln \mathbb{P}_Y(k), \tag{2}$$

where $n_k = \sum_{i=1}^{n} I(Y_i = k)$ is the number of respondents exhibiting response category k.

We utilize the Expectation-Maximization (EM) algorithm to obtain the maximum likelihood estimates (MLEs) of the parameters our statistical model. The EM algorithm, typically applied in incomplete-data settings, is particular apposite for the setting studied in this paper due to the (partial) unobservability of our outcome of interest.[10]

5 The Information Experiment

To study the effects of perceived corruption in society on an individual's willingness to engage on corrupt behavior, we combined an information experiment embedded in a household survey with the modeling framework developed above. The survey consisted of face-to-face interviews of 4200 residents (18 year old and older) of the Gran Área 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 the firm *Borge y Asociados* between October 2013 and April 2014.[11]

Rather than basing our analysis on the observed correlation between perceptions of corruption and willingness to engage in corrupt behavior, a strategy likely to suffer from potentially severe problems of confounding, we induce exogenous variation in beliefs about corruption via the random assignment of respondents to distinct informational treatments. Three informational treatments were employed in our experiment: a corruption treatment, an inefficiency treatment, and a control condition. In the corruption treatment, respondents were presented with a flyer depicting the increasing percentage of Costa Ricans who have directly observed an act of corruption (from 2006 to 2011). A second treatment, the inefficiency treatment, was introduced as a "placebo" in order to test whether respondents were affected by the information included in the corruption treatment or just the fact that they were given a flyer with negative information about the capacity of the Costa Rican state to deal with illicit behavior. In the inefficiency treatment, respondents were presented with a flyer presenting the (lack of) productivity of the legal system in dealing with a particular crime: assault with a deadly weapon. In the control condition, respondents were not presented with any flyer. Randomization of treatment assignment was programmed directly into the portable digital assistants (PDAs) the enumerators used to conduct the survey. Random assignment to different types of information ensured that, on average, groups of respondents were indistinguishable on both observable and unobservable characteristics. Appendix Table 5 in the on-line appendix provides evidence on balance in observable respondent characteristics across treatment groups. Excluding the treatment assignments, respondents were asked questions from otherwise identical questionnaires.

[10] A detailed exposition on the use of the EM-algorithm for models of this kind can be found in Gingerich, Oliveros, Corbacho, and Ruiz-Vega (2014).

[11] See the on-line appendix for more information on the survey methodology and execution.

Figure 3. The two treatments, corruption and judicial inefficiency

The two flyers are presented in Figure 3. The flyer on the left hand side is the corruption treatment. It states, "Did you know that corruption in Costa Rica has increased?" Below the statement appears a bar graph showing the increase in the percentage of Costa Ricans who had personally witnessed an act of corruption from 2006 (16%) to 2011 (24%). At the bottom right hand side of the flyer, the source of the information, a nationally representative survey conducted by *Latinobarómetro*, is displayed. The flyer on the right hand side is the inefficiency treatment. It states, "In 2011, 6,812 cases of assault with a deadly weapon entered into the judicial system. However, only 47 individuals were sent to jail for this crime." Below the statement appears a bar graph showing the relative magnitudes of the number of cases filed for this crime (6812), the number of judicial decisions made on cases of the crime (333), and the number of individuals actually sent to jail (47). The source of the information, National Judicial Statistics, is displayed on the bottom right corner.

In the latter third of the survey (well after exposure to one of the three experimental conditions), respondents assigned to the two informational treatments were presented with a verification question that asked them to describe what the informational graphic they received was about. Subsequent to this, all respondents were prompted to respond to the aforementioned question about

willingness to bribe, first in crosswise format and then, at very the end of the survey, in direct questioning format.

The purpose of the verification question was to identify individuals who were assigned to the information treatments but who failed to fully internalize the information they were given. We classify an individual as a full recipient of an information treatment if she was assigned to said treatment and could recall the basic content of the treatment according to her response on the verification question. Respondents assigned to the corruption treatment were categorized as full recipients if they stated on the verification question that the informational graphic they received was about how "corruption has increased in recent years" or that it dealt with "something about corruption." According to this standard, seventy six percent of respondents assigned to the corruption treatment were full recipients (1065 out of 1393). Respondents assigned to the judicial inefficiency treatment were categorized as full recipients if they stated on the verification question that the informational graphic they received was about how "there are many reports of crime but few people go to jail" or that it dealt with "something about how bad/inefficient the judicial system is." Sixty five percent of respondents assigned to the judicial inefficiency treatment were full recipients according to this standard (904 out of 1385).

6 Corruption Does Corrupt

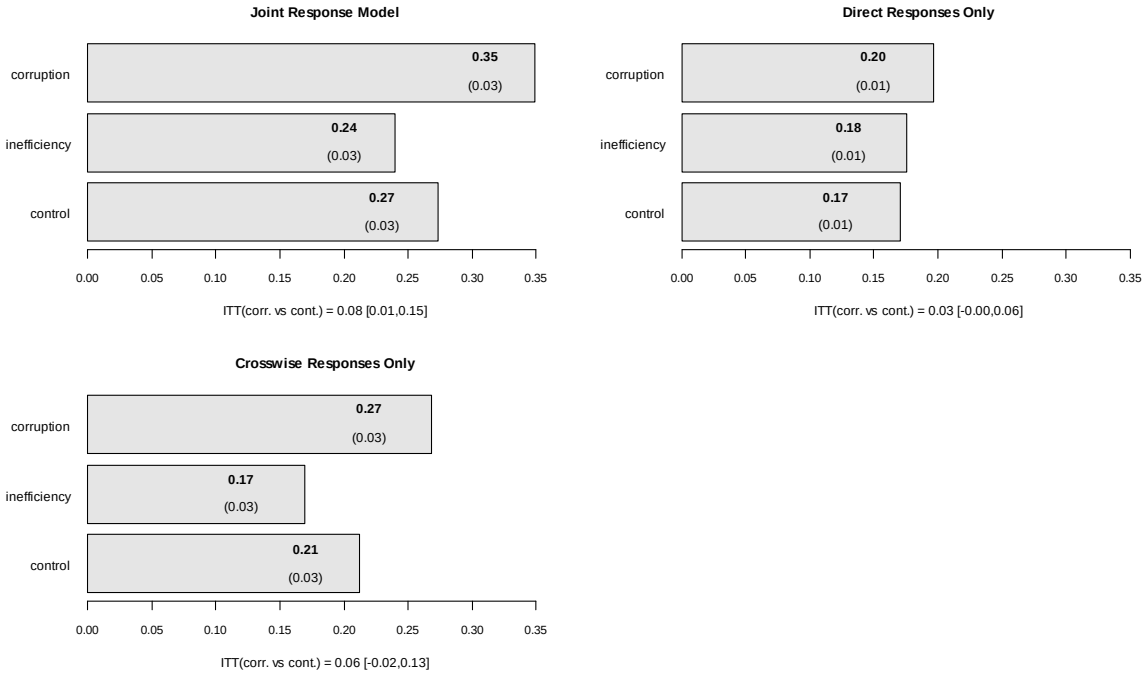
6.1 Intent-to-Treat Estimates

We begin our analysis by examining differences in the estimated proportion of respondents willing to bribe according to treatment assignment. These differences are the basis of our estimates of the intent-to-treat (ITT) effect: the average impact of random assignment to one of the three experimental conditions (but not necessarily the impact of internalizing the information contained in the graphical displays.) Results are presented utilizing our joint response approach, the direct survey responses by themselves, as well as the crosswise responses by themselves. The findings are presented in a series of barplots displayed in Figure 4.

All told, the results provide support for the self-fulfilling hypothesis that an individual's willingness to engage in corrupt behavior is affected by her exposure to information about the level of corruption in society. According to the estimates based on the joint response model, the proportion of respondents assigned to the corruption information treatment who would be willing to bribe a police officer was 0.35, whereas the proportion of respondents assigned to the control condition willing to do the same was only 0.27. Thus, the average effect of exposure to information about the increasing scope of corruption was equal to 0.08. This is a substantively large effect: exposure to the corruption treatment was estimated to increase the proportion of respondents willing to bribe by 28%. In addition to being large in magnitude, the effect was statistically significant

by conventional standards. Our placebo treatment–information about judicial inefficiency–did not have a statistically significant impact on willingness to bribe.

Figure 4. Estimated proportion of respondents willing to bribe by treatment condition and questioning method



Note: Standard errors in parentheses. Ninety-five percent confidence intervals for the average treatment effect in square brackets.

In addition providing evidence that corruption corrupts, the figure also presents differences in response patterns across questioning techniques. In every treatment condition, estimates of willingness to bribe based on direct responses were below those based on the joint response model and crosswise responses only. Moreover, it would appear that desirability bias substantially attenuated downward the estimate of the effect of exposure to information about corruption: the impact estimate based only on direct responses was about half that based only on the crosswise responses and just about one third as large as that based on the joint responses. The use of only the crosswise responses resulted in an estimated effect equal to 0.06, below but roughly similar to that based on the joint responses. However, this estimate was not statistically significant at conventional standards, due in part to the fact that it is based on an inefficient statistical approach that failed to incorporate the potentially useful information available from the pattern of direct responses.

The next step of our analysis consisted of the use of the joint response-modified logistic regression framework described earlier.[12] Table 2 depicts coefficient estimates showing the impact

[12] Unlike the joint response estimates presented in Figure 3, this approach pools the estimation of the diagnostic parameters across all three treatment conditions.

Table 2. Relationship between information treatments and bribery of police (joint response model)

	model 1			model 2		
Parameters	estimate	s.e.	95% int.	estimate	s.e.	95% int.
diagnostic parameters						
$\widehat{\lambda}_1^T$	0.61	0.03	[0.55,0.67]	0.63	0.03	[0.58,0.68]
$\widehat{\lambda}_1^L$	0.36	0.03	[0.30,0.41]	0.34	0.03	[0.28,0.39]
$\widehat{\lambda}_0^T$	0.97	0.00	[0.96,0.98]	0.97	0.00	[0.96,0.98]
explanatory parameters						
constant	-0.98	0.09	[-1.17,-0.80]	2.81	0.46	[1.97,3.77]
corruption treatment	**0.24**	**0.10**	**[0.04,0.45]**	**0.23**	**0.11**	**[0.01,0.43]**
inefficiency treatment	**0.01**	**0.11**	**[-0.18,0.22]**	**0.03**	**0.12**	**[-0.19,0.24]**
male	_	_	_	0.89	0.10	[0.70,1.10]
log(age)	_	_	_	-1.29	0.12	[-1.54,-1.06]
education (base=some college)	_	_	_			
primary or less	_	_	_	0.12	0.16	[-0.17,0.40]
secondary incomplete	_	_	_	0.40	0.15	[0.13,0.68]
secondary complete	_	_	_	0.29	0.15	[-0.01,0.59]
some technical	_	_	_	0.05	0.28	[-0.51,0.62]
ITT (corruption vs. control)	**0.05**	**0.02**	**[0.01,0.09]**	**0.04**	**0.02**	**[0.00,0.08]**
	$n = 4193$			$n = 4192$		

of the two information treatments on a respondent's willingness to offer a bribe to a police officer to avoid paying a traffic ticket. Two regression models were estimated: one in which the informational treatments entered as the sole explanatory variables and one in which we included the age, gender, and education of the respondent. Previous findings in the literature suggest that women and older individuals might be less involved in corruption and/or less likely to condone corruption than men and younger individuals (Swamy et al. 2003; Torgler and Valev 2006, 2010) and that more educated individuals (or richer, the two variables often used as proxies) are more tolerant of corruption (Winters and Weitz-Shapiro 2013).

Both estimated models told a very similar story: exposing respondents to information about the growing scope of corruption in society made them significantly more likely to indicate a willingness to bribe a police officer to avoid paying a traffic ticket. Again, it thus appears that, as predicted by theory, corruption really does corrupt. As shown in the table, the average effect of exposure to information about the growing scope of corruption was estimated to be 0.05 in the baseline model and 0.04 in the model with additional covariates. In both cases, the ITT estimates were statistically significant by conventional standards. As above, in neither of the estimations did the judicial inefficiency treatment have any discernible impact on willingness to bribe, suggesting

that the effect of the corruption treatment was caused by the specific information contained in it and not by just the fact that respondents were exposed to some negative information about the capacity of the Costa Rican state to deal with illicit activity. In terms of background characteristics, men appeared to be substantially more inclined to bribe than women, younger respondents more inclined to bribe that older respondents, and individuals with incomplete secondary school educations more inclined to bribe than individuals with some exposure to college (the baseline education category).[13]

6.2 Local Average Treatment Effects

The estimates of the impact of information about corruption provided above understate the impact of actually internalizing information about the scope of corruption. This is so because some of the respondents assigned to the two information treatments, perhaps due to a momentary distraction at the time of the experiment, fatigue or ill health on the day of the survey, or some other factor, failed to fully consume the information to such an extent that they could recall it accurately later. Since these individuals cannot be said to have received the treatment in a meaningful way, their presence in a given treatment group deflates the impact estimate associated with that treatment.

In order to estimate the causal impact of actually internalizing the information embedded in the information treatments, we employ an instrumental variables approach. In our conceptualization, a respondent only receives an information treatment if she is assigned to it *and* has internalized the information in the treatment according to the criteria described earlier. Treatment assignment is thus an instrumental variable. Compliance with the instrument is perfect for respondents assigned to the control group (these individuals cannot internalize the information from a given information treatment because they have not received it) but assumed imperfect for respondents assigned to the two graphical displays. Thus, in the setting considered here the instrumental variables estimator for a given information treatment is equal to the ITT divided by the compliance rate for that treatment (e.g., the rate of internalization). As is well known, this estimator estimates the local average treatment effect (Angrist, Imbens, and Rubin 1996): in our case, this is the average effect of receiving an information treatment for those individuals who could be encouraged to internalize the information in the treatment based on their assignment.

Table 3 presents our estimates of the local average treatment effect associated with internalizing the information contained in the two information treatments. These estimates are based upon employing the joint response model in order to estimate respondents' willingness to bribe. We

[13] As with any other information experiment, there are good reasons to believe that different types of citizens might react differently to the information provided. Individual characteristics such as gender, age, education, weath, and prior beliefs about the level of corruption might interact with the information about rising levels of corruption. To study whether these characteristics have a conditioning effect on our experimental results, we estimated conditional intent-to-treat effects (CITTs) for various subgroups of the population based on these characteristics. None of the differences in the intent-to-treat estimates achieved statistical significance. See Tables 6-11 in the on-line Appendix.

Table 3. Local average treatment effects of internalizing information treatments (outcome measured using joint response model)

	LATE estimates		
information treatment	diagnostic parameters (unpooled vs. pooled)		
	unpooled	pooled - model 1	pooled - model 2
corruption	0.10	0.07	0.05
	(0.05)	(0.03)	(0.03)
	[0.00,0.20]	[0.00,0.13]	[0.00,0.11]
inefficiency	-0.05	0.00	0.01
	(0.06)	(0.03)	(0.03)
	[-0.16,0.05]	[-0.06,0.07]	[-0.06,0.07]

Note: Standard errors in parentheses and ninety-five percent confidence intervals in brackets. Standard errors and confidence intervals were calculated via the nonparametric bootstrap.

present results based on estimating prevalence rates and diagnostic parameters separately for each treatment group as well as results based on our modified-logistic regression framework (which pools the diagnostic parameters across treatment groups). In the case of the corruption treatment, the estimated local average treatment effects are substantively large and all statistically significant by conventional standards, supporting the notion that expectations about a widening societal scope of corruption drive willingness to bribe. The estimate of the local average treatment effect produced by the joint response approach without pooling diagnostic parameters was equal to 0.10; the modified-logistic regression estimates based on pooling were equal to 0.07 and 0.05 (the former based on employing an explanatory model that did not condition on covariates and the latter based on employing an explanatory model that did condition on covariates). Since the vast majority of respondents assigned to the corruption treatment internalized the information therein, the local average treatment effect estimates are only modestly larger in magnitude than the ITT estimates presented in the previous section. In terms of the judicial inefficiency treatment, we find that internalizing the information therein had no statistically significant impact on willingness to bribe. Again, our conclusion is that citizens are responding specifically to the information about the scope of corruption embedded in the corruption informational display and not to generically negative information about the capacity of the Costa Rican state.

7 Conclusion

This paper provides the first set of experimental findings in favor of the proposition that corruption corrupts. Utilizing data from a large-scale social survey conducted in Costa Rica, one that combined both a survey experiment and a novel format for asking sensitive questions (the joint re-

sponse model), we find that exposing citizens to information about the growing scope of corruption in their society made them individually more disposed to engage in corruption. In this way, our paper offers empirical support to a large theoretical literature on corruption that has long claimed that the phenomenon is characterized by strategic complementarities between economic agents. At a methodological-level, our paper provides an illustration of how survey responses generated by sensitive survey techniques and direct questioning can be combined in order to better understand the determinants of sensitive forms of political behavior.

One may wonder to what degree the results we obtained with subjects in Costa Rica reflect the dynamics of corruption in other countries in the region and beyond. As discussed in the introduction, in the last decade Costa Rica has witnessed a number of public scandals unusual for its political history, as well as rising levels of corruption and crime. All these factors might have contributed to making our respondents particularly susceptible to the information provided in our experiment. Could we expect similar results in other countries? As with any other information experiment, the information provided can only have an effect if the respondents find it credible. For that reason, we would not a priori expect similar results to emerge were our experiment conducted in countries with long lived and unchallenged reputations for cleanliness in government. Nor would we expect a strong effect of information in the opposite direction (decreasing scope of corruption) in countries with equally long lived and unchallenged reputations for widespread corruption. In both instances, citizens' priors about the scope of corruption would likely be quite difficult to dislodge. Countries where reputations for corruption appear to be in flux would seem to be the most likely candidates to exhibit informational effects of the kind reported in these pages.

In this respect, it is worth noting that Costa Rica's recent experience is not particularly unusual. According to the Latinobarómetro— the same source we used in our experiment— two other countries (out of the 18 in the sample) have experienced recent increases in corruption. While the percentage of citizens observing an act of corruption in Costa Rica went from 16 to 24 between 2006 and 2011, this percentage increased from 12 to 14 in Colombia, and from 17 to 21 in the Dominican Republic during the same period. Also in a similar period, the number of people who think that all or most of government officials are involved in corruption has risen in 13 out of the 20 African countries with comparable data (Afrobarometer, 2008/9 vs. 2011/3).

Equally relevant, Costa Rica is among various countries in Latin America and the Caribbean that have recently experienced large scale corruption scandals that have shaken the political establishment and have the potential to reshape citizens' views toward their government . For instance, over the course of the past decade, citizens in Brazil have been subjected to a seemingly endless litany of political corruption scandals reaching into the uppermost ranks of government. These scandals, from the *Mensalão* (Big Monthly) congressional vote buying scheme under the Luiz Inácio Lula da Silva administration to the *Lava Jato* (Car Wash) kickback scheme currently threaten-

ing to derail the administration of Dilma Rousseff, have imposed enormous losses to the Brazilian state and have implicated top party leaders, high ranking members of Congress, the heads of some the country's largest construction firms, and numerous members of the federal public bureaucracy. While revelations of political corruption are not uncommon in Brazil, the media attention allocated to corruption allegations in recent years has been continuous and overwhelming. Another country potentially at an inflection point is Chile. Even more so than Costa Rica, Chile has long had a reputation for probity in government and institutional strength. However, the last several years have produced repeated scandals involving the illicit financing of political campaigns by politicians across the ideological spectrum as well as serious instances of tax fraud, money laundering, and influence peddling by a small, politically connected elite. Without convincing countervailing governmental action, scandals such as those afflicting Brazil and Chile could in theory perturb beliefs in a direction conducive to the spread of corruption along the lines of what we found with our subjects in Costa Rica.

In closing, we would like to underline some important practical lessons to be taken from this paper. First and foremost, we believe our findings encourage the adoption of additional nuance in policy-oriented discussions about the role of transparency in reducing corruption. Greater transparency is often heralded by those in the international development community as one of the most important antidotes for reducing corruption. There is a solid academic basis for this. A number of influential papers, cited earlier, suggest that providing voters with information on corruption can promote a virtuous response in which citizens attempt to vote corrupt politicians out of office. Informed by these findings, we do not doubt that transparency has an important role to play in the fight against corruption, especially at election time. However, the results we encountered in this study lead us to believe that transparency about corruption might be more of a double-edged sword than many have previously thought.

In particular, our findings lead us to be concerned about the potentially damaging effects of shaming campaigns that attempt to galvanize public opinion against corruption by widely disseminating the message that large swaths of public officials are on the take or that the accumulated losses of corruption are enormous. Examples of such campaigns range from the "I paid a bribe" websites in India, Kenya, and Pakistan (which provide real time information on anonymous, self-reported bribery payments across a country), the installation of the so-called *abusometro*, an electronic billboard located in a busy intersection of Mexico City that gives citizens a running tally of estimated public education funds lost to corruption, and the various corruption bus tours that expose citizens to the sites of corrupt exchange and the fruits of corruption in cities such as Monterrey, Mexico, and Prague, Czech Republic. Although these efforts are well intentioned, our findings suggest that constantly nailing inconvenient facts into citizens' heads about the scope of corrup-

tion can shape expectations about the behavior of public officials in such a way as to perpetuate the very problem such campaigns are designed to solve.

At the end of the day, many forms of corruption–especially the most virulent forms of petty corruption that erode the living standards of the poorest segments of the population in developing countries–are not simply a matter of what officials decide to do to citizens, but rather a matter of what certain officials and certain citizens conspire to do *together* against the interests of the public at large. As the saying goes, it takes two to tango, and the dance of corruption requires at least two willing partners inclined to risk their reputation, patrimony, and possibly their liberty in order to cement a corrupt exchange with an individual about whom they typically know fairly little. In such contexts, societal expectations about what typical citizens and officials are inclined to do will be central in delimiting the risk-reward calculus of both parties to the exchange. Transparency of a kind that conveys the message that there is little to be risked but much to be gained from pursuing such an exchange is a type of transparency that most polities are best left without.

References

Accinelli, E., and Carrera, E. J. S. 2012. "Corruption Driven by Imitative Behavior." *Economics Letters* 117(1), 84-87.

Açemoglu, D. 1995. "Reward Structures and the Allocation of Talent." *European Economic Review* 39 (1): 17-33.

Aidt, T. S. 2003. "Economic Analysis of Corruption: A Survey." *The Economic Journal* 113 (491): F632-F652.

Anderson, CJ and YV Tverdova. 2003. "Corruption, Political Allegiances and Attitudes Towards Government in Contemporary Democracies." *American Journal of Political Science* 47 (1): 91-109.

Anduiza, E., Gallego, A., & Muñoz, J. (2013). "Turning a Blind Eye: Experimental Evidence of Partisan Bias in Attitudes Toward Corruption." *Comparative Political Studies* 46 (12): 1664-1692.

Andvig, JC and K. O. Moene. 1990. “How Corruption May Corrupt.” *Journal of Economic Behavior & Organization* 13(1): 63-76.

Angrist, J. D., G.W. Imbens, and D.B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American statistical Association* 91(434): 444-455.

Balafoutas, L. 2011. "Public Beliefs and Corruption in a Repeated Psychological Game." *Journal of Economic Behavior & Organization* 78(1): 51-59.

Becker, S, PH Egger and T. Seidel. 2009. "Common Political Culture: Evidence on Regional Corruption Contagion." *European Journal of Political Economy* 25: 300-310.

Blair, G. and K. Imai. 2012. “Statistical Analysis of List Experiments.” *Political Analysis* 20(1): 47-77.

Cadot, O.. 1987. “Corruption as a Gamble.” *Journal of Public Economics* 33(2): 223-244.

Cameron, L., Chaudhuri, A., Erkal, N., and L Gangadharan. 2009. "Propensities to Engage In and Punish Corrupt Behavior: Experimental Evidence from Australia, India, Indonesia and Singapore." *Journal of Public Economics* 93 (7): 843-851.

Carrell, S. E., FV Malmstrom, and JE West. 2008. “Peer Effects in Academic Cheating.” *Journal of Human Resources* 43(1): 173-207.

Caselli, F., and M Morelli. 2004. "Bad Politicians." *Journal of Public Economics* 88(3): 759-782.

Chong, A., A De La O., L., D Karlan, and L Wantchekon. 2015. "Does Corruption Information Inspire the Fight or Quash the Hope? A Field Experiment in Mexico on Voter Turnout, Choice, and Party Identification." *Journal of Politics* 77 (1): 55-71.

Dal Bó, E., P Dal Bó, and R Di Tella. 2006. "'Plata o Plomo?': Bribe and Punishment in a Theory of Political Influence." *American Political Science Review* 100(1): 41-53.

Dollar, D., Fisman, R., & Gatti, R. (2001). “Are women really the “fairer” sex? Corruption and women in government.” *Journal of Economic Behavior & Organization* 46(4), 423-429

Dong, B., and B Torgler. 2012. “Corruption and social interaction: Evidence from China.” *Journal of Policy Modeling* 34(6): 932-947.

Dong, B., Dulleck, U., & Torgler, B.. 2012. “Conditional corruption.” *Journal of Economic Psychology* 33(3): 609-627.

Esarey, Justin, and Gina Chirillo. 2013. “‘Fairer Sex’ or Purity Myth? Corruption, Gender, and Institutional Context.” *Politics and Gender* 9(4): 390–413.

Ferraz, C, and F Finan. 2008. “Exposing Corrupt Politicians: The Effects of Brazil’s Publicly Released Audits on Electoral Outcomes.” *The Quarterly Journal of Economics* 123(2): 703-745.

Fisman, R and E Miguel. 2007. "Corruption, Norms, and Legal Enforcement: Evidence from Diplomatic Parking Tickets." *Journal of Political Economy* 115 (6): 1020–1048.

Gingerich, D. W. 2013. *Political Institutions and Party-Directed Corruption in South America: Stealing for the Team*. Cambridge: Cambridge University Press.

_____. 2010. “Understanding Off-the-Books Politics: Conducting Inference on the Determinants of Sensitive Behavior with Randomized Response Surveys.” *Political Analysis* 18: 349-380.

_____. 2009. "Ballot Structure, Political Corruption, and the Performance of Proportional Representation." *Journal of Theoretical Politics* 21(4): 509-541.

_____, V. Oliveros, A Corbacho, and M Ruiz-Vega. 2014. "When to Protect? Using the Crosswise Model to Integrate Protected and Direct

Responses in Surveys of Sensitive Behavior." Unpublished manuscript. (http://danielgingerich.weebly.com/uploads/3/1/4/6/31460749/when_to_protect_oct_23_2014.pdf)

Glaeser, E., B Sacerdote, and J. Scheinkman. 1996. “Crime and Social Interactions.” *Quarterly Journal of Economics* 111 (2): 507-548.

Goel, R. K., and MA Nelson. 2007. "Are Corrupt Acts Contagious?: Evidence from the United States." *Journal of Policy Modeling* 29(6): 839-850.

Goel, R. K., and JW Saunoris. 2014. "Global Corruption and the Shadow Economy: Spatial Aspects." *Public Choice* 161: 119-139.

Hauk, E., and M Saez-Marti. 2002. "On the Cultural Transmission of Corruption." *Journal of Economic Theory*, 107(2), 311-335.

Ichino, A. and G. Maggi. 2000. “Work Environment and Individual Background: Explaining Regional Shirking Differentials in a Large Italian Firm.” *Quarterly Journal of Economics* 115: 1057-1090.

Innes, R and A Mitra. 2013. "Is Dishonesty Contagious?" *Economic Inquiry* 51 (1): 722-734.

Jann, B., J. Jerke, and I Krumpal. 2012. “Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism.” *Public Opinion Quarterly* 76(1): 32-49.

Korndörfer, M., I Krumpal, and SC Schmukle. 2014. "Measuring and Explaining Tax Evasion: Improving Self-reports Using the Crosswise Model." J*ournal of Economic Psychology* 45: 18-32.

Kundt, T. C. 2014. "Applying ‘Benford’s Law’ to the Crosswise Model: Findings from an Online Survey on Tax Evasion." Unpublished manuscript. (http://ssrn.com/abstract=2487069)

Lichbach, MI and AS Zuckerman, eds. 1997. *Comparative Politics: Rationality, Culture, and Structure*. Cambridge: Cambridge University Press.

Lopez-Valcarcel, B. G., Jiménez, J. L., & Perdiguero, J. 2014. “DANGER: LOCAL CORRUPTION IS CONTAGIOUS!” Unpublished manuscript. (http://www.alde.es/encuentros/anteriores/xviieea/trabajos/g/pdf/179.pdf)

Lui, F.T., 1986. “A Dynamic Model of Corruption Deterrence.” *Journal of Public Economics* 31: 215-236.

Malesky, E. J., Gueorguiev, D. D., and NM Jensen. 2015. "Monopoly Money: Foreign Investment and Bribery in Vietnam: a Survey Experiment." *American Journal of Political Science* 59 (2): 419-439.

Márquez, M. A., Salinas-Jiménez, J., & Salinas-Jiménez, M. D. M. 2011. "Exploring Differences in Corruption: The Role of Neighboring Countries." *Journal of Economic Policy Reform*, 14(1): 11-19.

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

Mishra, A. (2006). "Persistence of Corruption: Some Theoretical Perspectives." *World Development*, 34(2), 349-358.

Pierce, L., and J Snyder. 2008. “Ethical Spillovers in Firms: Evidence from Vehicle Emissions Testing.” *Management Science* 54(11): 1891-1903.

Plata, JC. 2012. “To Bribe or Not to Bribe.” *AmericasBarometer Insights* 72: 1-8.

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

Sheffrin, S. M., and R. K. Triest. 1992. “Can Brute Force Deterrence Backfire? Perceptions and Attitudes in Taxpayer Compliance,” In *Why People Pay Taxes: Tax Compliance and Enforcement*, edited by J. Slemrod. Ann Arbor: University of Michigan Press: 193–218.

Simpser, A. 2013. “The Intergenerational Persistence of Attitudes Toward Corruption.” Unpublished manuscript. (http://ssrn.com/abstract=2241295).

Sung, H.-E. 2003. “Fairer sex or fairer system? Gender and corruption revisited.” *Social Forces* 82(2), 703–23.

Swamy, A., S Knack, Y Lee, and O Azfar. 2003. "Gender and Corruption." In *Democracy, Governance and Growth*, edited by Stephen Knack: pp. 191-224.

Tan, M. T., G. L. Tian, and M. L. Tang. 2009. “Sample Surveys with Sensitive Questions: A Nonrandomized Response Approach.” *The American Statistician* 63(1): 9-16.

Tirole, J. 1996. “A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality).” *Review of Economic Studies* 63 (1): 1-22.

Torgler, B, and NT Valev. 2010. “Gender and Public Attitudes Toward Corruption and Tax Evasion.” *Contemporary Economic Policy* 28(4): 554-568.

_____. 2006. “Corruption and Age.” *Journal of Bioeconomics* 8(2): 133-145.

Treisman, D. 2007. "What Have We Learned about the Causes of Corruption from Ten Years of Cross-National Empirical Research?" *Annual Review of Political Science* 10: 211-244.

Winters, MS, and R Weitz-Shapiro. 2013. “Lacking Information or Condoning Corruption: When Do Voters Support Corrupt Politicians?” *Comparative Politics* 45(4): 418-436.

On-line Appendix for "Corruption as a Self-Fulfilling Prophecy"

1 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.[14] 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

$$p^e \alpha_i - (1 - p^e)\sigma, \tag{3}$$

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

$$q^e \phi_j - (1 - q^e)\tau, \tag{4}$$

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:

$$\begin{aligned} q &= \mathbb{P}\left(\alpha_i \geq \frac{(1 - p^e)\sigma}{p^e}\right) \\ p &= \mathbb{P}\left(\phi_j \geq \frac{(1 - q^e)\tau}{q^e}\right). \end{aligned} \tag{5}$$

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

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

the following two equations:

$$q^* = \begin{cases} 0 & \text{if} \quad p^* < \frac{\sigma}{\overline{\alpha}+\sigma} \\ \frac{\overline{\alpha}+(1-1/p^*)\sigma}{\overline{\alpha}-\underline{\alpha}} & \text{if} \quad p^* \in [\frac{\sigma}{\overline{\alpha}+\sigma}, \frac{\sigma}{\underline{\alpha}+\sigma}] \\ 1 & \text{if} \quad p^* > \frac{\sigma}{\underline{\alpha}+\sigma} \end{cases} \tag{6}$$

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

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 $\gamma \in \{\text{"low","high"}\}$ 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, char-

acterizing 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.

2 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:

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

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:

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

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 $\frac{\partial c}{\partial \overline{\phi}} < \frac{\partial d}{\partial \overline{\phi}}$.

3 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.[15]

4 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.

5 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,

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

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**

responses for mothers and fathers were pooled together. The comparison in Appendix Table 3 shows that self-reported parent's birthdays were almost identical to the actual information obtained from INEC.

6 Enumerator Scripts

6.1 Main outcome variable - Crosswise Format

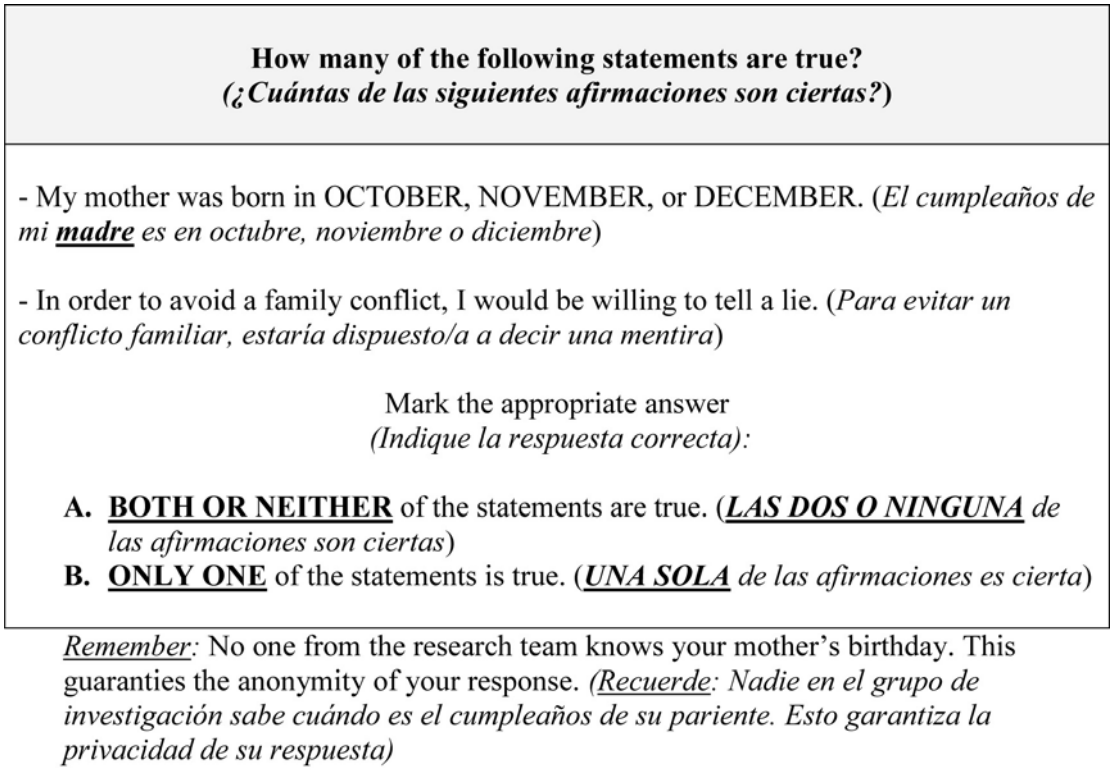
Interviewers explained first the logic of the crosswise format with the following script:

"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 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 below:

Continuation of the script:

Appendix Figure 5. The practice crosswise survey item



"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 5.

Appendix Figure 6. The actual 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)*

6.2 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

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

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

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

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

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

Note: Note: P-values based on chi-square test of homogeneity for categorical covariates and an ANOVA test for equality of group means for continuous covariates (e.g., age).

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]
	$n = 1272$			$n = 1400$			$n = 1391$		

Note: The wealth index consists of the factor scores generated by a two parameter item response theory model in which the observable inputs indicated ownership of material possesions including a cellphone, laptop, tablet, car, motorcycle, flat screen television, cable or satellite TV hookup, and an internet connection.

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]**
	$n = 986$			$n = 885$		