Political Corruption Traps

Marko Klašnja*
Andrew T. Little†
Joshua A. Tucker‡

November 12, 2014

Abstract

Academics and policy makers recognize that there are serious costs associated with systemic corruption. Stubbornly, countries or regions with high levels of corruption often seem resistant to attempts to mitigate corruption, remaining stuck in "corruption traps." Most existing theory concentrates on mutually reinforcing expectations of corrupt behavior among a fixed set of bureaucrats or politicians. We develop several models that more fully characterize the *political* nature of corruption traps by also analyzing the behavior of voters and entrants, as well their interaction with incumbent politicians. By linking politician, voter, and entrant behavior, we provide an explanation for why simply trying to change expectations among one set of actors is likely insufficient for eliminating corruption traps, and point instead to the importance of features such as electoral institutions.

Key Words: Corruption, Voting, Multiple Equilibria, Persistence, Corruption Traps

*Wilf Family Department of Politics, New York University mk3296 at nyu dot edu
†Department of Government, Cornell University andrew dot little at cornell dot edu
‡Wilf Family Department of Politics, New York University joshua dot tucker at nyu dot edu

1 Introduction

The study of corruption – its causes, effects, and persistence – has become a matter of great interest to both the scholarly and policy communities. Corruption is recognized as a threat to economic development, democratic consolidation, and human dignity. Consequently, the desire to reduce corruption has become an important goal of international financial institutions, civil society groups, and individual nations' foreign policies.[1]

Two stylized facts related to corruption have been influential in how academics, policymakers, and others think about whether and how it is possible to attain these goals. First, the prevalence of corruption varies significantly across different contexts. This variation can be seen cross-nationally (e.g. Treisman, 2000; Montinola and Jackman, 2002), but also *within* countries as well: for example, there are large differences in corruption across American states (Alt and Lassen, 2003; Glaeser and Saks, 2006) and cities (Banfield and Wilson, 1963; Menes, 2006); between Southern and Northern Italy (e.g. Banfield and Banfield, 1958; Golden and Picci, 2005); and across states in India (Transparency International India, 2005). Such variation gives reformers hope that corruption can indeed be reduced, that "high corruption" countries or regions can be turned into "low corruption" countries or regions.

However, it is also the case that high or low levels of corruption tend to be remarkably persistent: we rarely observe high corruption countries or regions becoming low corruption countries or regions, or visa-versa (Becker et al., 2011; Damania, Fredriksson and Mani, 2004; Fisman and Miguel, 2007).[2] One potential and plausible explanation for this persistent variation in corruption levels cross-nationally and within countries is that the dynamics underlying corruption allow for multiple equilibria: some countries or regions may consistently

[1] For example, the World Bank President Jim Yong Kim recently called corruption "Public Enemy Number One." `http://www.worldbank.org/en/news/press-release/2013/12/19/corruption-developing-countries-world-bank-group-president-kim`

[2] Two important exceptions are the small city-states of Singapore and Hong Kong; see Quah (1994).

play either the "low-corruption" or "high-corruption" equilibrium.

This explanation has received significant theoretical attention. Most of the previous accounts focus on the interactions among a fixed set of bureaucrats or politicians, and argue that incentives to engage in corrupt behavior may increase as government officials come to believe that other government officials are also corrupt (e.g. Andvig and Moene, 1990; Lui, 1986).[3] These mutually reinforcing beliefs can then lead to multiple equilibria: a low-corruption equilibrium where bureaucrats refrain from corruption when they believe that others will do so, and a high-corruption equilibrium where they engage in corruption in the belief that others will be corrupt as well.

We build on this type of argument and highlight two important shortcomings of the existing literature. First, we argue that to describe the politics of multiple equilibria of corruption more fully, it is important to focus not just on a fixed set of politicians or bureaucrats, as in most of the previous models, but also on voters and potential candidates for political office ("entrants"), as well as the interaction among these three sets of actors. In principle, corrupt politicians can be voted out of office, but the empirical work on voting behavior has shown that in some contexts voters are willing to condone or overlook corruption (Banerjee and Pande, 2007; Klašnja and Tucker, 2013) while in other contexts corrupt politicians are punished (Ferraz and Finan, 2008; Klašnja, Tucker and Deegan-Krause, 2014). This body of evidence suggests that voters themselves may be contributing to multiple equilibria of corruption. Moreover, multiplicity of equilibria may derive not just from the variation in incentives of incumbent politicians to be corrupt, as in most previous models, but in the variation of the type of candidates willing to run for office. When the majority of entrants are likely to be corrupt, voters may have little choice.

The second shortcoming of most of the existing literature is that high-corruption equilibria driven mainly by expectations of how corrupt others will be suggest a particular policy

[3] There are several notable exceptions, which we discuss in detail in Section 2 below.

prescription of how to root out corruption: an intervention that rapidly changes the corruption expectations of a fixed set of politicians – such as an increased probability of detection – should be effective, as the induced belief that others will behave less corruptly will be self-enforcing. While such interventions are not costless, they seem feasible.[4] And yet, this policy implication seems inconsistent with the fact that corruption is so persistent across countries and regions. By focusing on voters and entrants as well as politicians, and on the interaction among them, our models exhibit more of the strategic complementarities that lead to multiple equilibria than past models suggest, highlighting why acting on politicians' expectations alone may not be enough to reduce corrupt behavior.

We begin by presenting a simple model where politicians, who vary based on their predisposition to engage in corrupt behavior, choose a level of corrupt activity when the benefits from doing so are a function of the degree to which other politicians choose to engage in similar corrupt behavior (i.e. the "corruption level"). This model has three important properties that motivate our subsequent analysis. First, the relative value of holding office for a politician is higher for those predisposed to corruption, and the size of this "politician corruption differential" is increasing in the level of corruption. Second, it is possible for citizens to prefer having a corrupt representative if the aggregate corruption level is sufficiently high (and citizens enjoy some of the benefits from corruption). Third, as politics becomes more corrupt, citizens can become less apt to vote corrupt politicians out of office, and eventually may actually prefer a corrupt representative.

We then proceed to show how these results affect behavior by potential candidates for

[4] Abbink and Serra (2012, p.6) summarize this common policy prescription well: "As a result of the strategic complementarities existing in the costs associated with corruption, multiple corruption equilibria exist and a country could converge to the bad equilibrium, i.e., a "corruption trap" from which it is extremely difficult to escape. The main policy implication of these theories is that when corruption becomes systemic, i.e., when the bad equilibrium is reached, a "big push" is required to escape the trap, that is, a combination of policies acting on the actual probability of detection, the penalty associated with corruption, the wages of public officials, and individuals' general beliefs about the extent of corruption among both public officials and ordinary citizens."

office and voters.[5] Our model of potential entrants shows that those predisposed to be corrupt are more apt to run when they expect others to be corrupt as well. This can lead to multiple equilibria in the model, which we call a "political corruption trap", whereby either mostly corrupt candidates run for office with the expectation that politics will be corrupt, or mostly clean entrants run with the expectation that corruption will be low.

Finally, we present a model where voters observe the ability and corruption predisposition of their current representative and decide whether to keep or replace her. If the relative value of having a corrupt representative is increasing in how many others are corrupt, there can again be multiple equilibria and a political corruption trap: one equilibrium where voters are generally willing to retain corrupt politicians (as it is either not harmful or even beneficial to have a corrupt representative), and one where voters generally replace corrupt politicians (as it is harmful to have a corrupt representative if few others are corrupt). We further show that when the dynamics from the voter model are combined with the dynamics of the entry model, the potential for a political corruption trap is generally even greater.

These results highlight that political corruption traps arising in our models may be harder to escape than the high-corruption equilibria in most of the existing literature. When the corruption trap is driven by the political forces we model, policy interventions would need to act not just on politicians' expectations, as prescribed by most previous models, but also on the expectations of voters and potential entrants to the political arena. We believe that this is a considerably harder problem than coordinating expectations of bureaucrats or politicians alone. Moreover, our results also suggest that even changing expectations of all three sets of actors simultaneously may be insufficient to reduce corruption unless it is possible to remove many corrupt politicians simultaneously. This implies that political and electoral institutions that affect the way politicians are elected may play a part in sustaining political

[5] As will be evident in the following section, the model most naturally applies to countries with electoral systems with different representatives for different districts.

corruption traps. For example, staggered elections and multiple tiers of representation may make it harder to replace a corrupt political elite whole-sale, potentially perpetuating a high-corruption equilibrium through the dynamics we model; the model therefore also has important policy implications for anti-corruption efforts.

2 Literature

In most theoretical studies that have previously examined the variation in the level and persistence of corruption through the multiplicity of corruption equilibria, the multiplicity arises through strategic complementarities stemming from the interplay between heterogeneous incentives to engage in corruption, some form of externality from committing a corrupt act, and the beliefs about whether others will engage in corruption. In Andvig and Moene (1990), the high incidence of corruption lowers the search costs of the briber for a willing bribee, and lowers the probability of detection to the bureaucrats from engaging in corruption (see also Lui, 1986). In Mishra (2005), the cost from compliance increases as others become increasingly non-compliant, lowering the benefits from being "clean" (see also Mauro, 2004). In Cadot (1987), the high prevalence of corruption implies that higher-level bureaucrats are more tolerant of lower-level corruption; this in turn further encourages lower-level bureaucratic corruption since getting caught is less likely to result in being fired and the loss of future monetary and bribe income.[6]

We build on and depart from these models in two important ways. First, as mentioned above, these models do not fully describe the politics of multiple corruption equilibria. Multiple equilibria arising through the strategic complementarities among bureaucrats (Cadot,

[6]Others, such as Tirole (1996) and Hauk and Saez-Marti (2002) also focus on a limited set of actors (firms and buyers), but model the prevalence of corruption as path-dependent. They show that if the actions and reputations of individuals are not perfectly observable then past corruption can create a general sense of suspicion toward individuals and the groups to which they belong; this in turn decreases the incentives for new group members to remain honest or to transmit a norm of honesty to the new generations, thus "ratcheting up" corruption in the future.

1987; Lui, 1986), between bureaucrats and citizens (Andvig and Moene, 1990), or among firms and buyers (Mishra 2005; see also Tirole 1996; Hauk and Saez-Marti 2002), miss the fact that the control of corruption is fundamentally a political process. Bureaucrats are ultimately accountable to politicians, who in democracies are accountable to voters.[7]

The fact that corrupt politicians can be voted out of office could in principle make it easier to eliminate the high corruption equilibria. However, empirical research on voters' reactions to corruption shows that not all electorates unconditionally punish corruption. Klašnja and Tucker (2013) conduct a survey experiment examining voters' reactions to economic performance and corruption in a low-corruption country (Sweden) and a high-corruption country (Moldova). They find that in Sweden (the low corruption country) voters react negatively to corruption regardless of the state of the economy, whereas in Moldova (the high corruption country) voters react negatively to corruption *only* when the state of the economy is also poor; when economic conditions are improving, corruption is considerably less important. Elsewhere, scholars have similarly found that in higher-corruption countries, voters are often insensitive to corruption (Banerjee and Pande, 2007), even if given precise information about it (Chong et al., 2011; de Figueiredo, Hidalgo and Kasahara, 2011; Humphreys and Weinstein, 2012).[8] Voters in low corruption countries tend to punish corruption more consistently (Hirano and Snyder, 2012; Peters and Welch, 1980; Reed, 1999; Welch and Hibbing, 1997).[9] These results suggest that for their part voters themselves may be contributing to the politically-driven multiple equilibria of corruption.

The common feature among the theoretical arguments cited above is that the multiple

[7]We limit the applicability of the model in this paper to functioning democracies with meaningful elections. Future extensions could involve semi-competitive elections as well. For the study of anti-corruption in autocracies, see for example Hollyer and Wantchekon (2015).

[8]However, other studies do find evidence of voters punishing corruption in high-corruption countries (Ferraz and Finan, 2008; Klašnja, 2014; Klašnja, Tucker and Deegan-Krause, 2014; Weitz-Shapiro and Winters, Forthcoming).

[9]However, even voters in low-corruption countries do not always sanction corruption (Barbera, Fernandez-Vazquez and Rivero, Forthcoming; Chang, Golden and Hill, 2010; Klašnja, 2013).

equilibria of corruption are driven by the heterogeneity of *incentives* to engage in corruption (e.g., the cost of finding a bribee or the probability of detection). The second important departure in the models we present below is that we focus on how the multiplicity of corruption equilibria is driven by the *types* of politicians – both those in office and those willing to enter politics. The behavior of voters and politicians in office affects the overall level of corruption, which in turn is likely to affect the kinds of candidates attracted to office. A low level of corruption implies that few incumbent politicians are corrupt and that voters are likely sensitive to corruption, making it less appealing for corrupt entrants to run for office. In turn, fewer corrupt entrants implies fewer corrupt candidates who win elections, fewer politicians who engage in corrupt behavior when in office and – in the case when corrupt politicians still manage to win – a higher likelihood of subsequently being voted out of office.[10] The opposite should be true when the level of corruption is high. In the models below, we seek to build on these intuitions and findings from empirical work.

Several important previous contributions have also investigated the importance of political incentives and heterogeneity in politician types for persistence of corruption or the quality of accountability more broadly. Most similarly to our argument about the type of politicians entering politics, Caselli and Morelli (2004) argue that the current share of "bad" politicians will affect the relative returns to office-holding for other bad politicians. As a result, there can be two equilibria: one where politics is dominated by good politicians, encouraging other good politicians to run, and one where bad politicians dominate politics. Our paper builds on this argument by: (1) also considering the effects of voter behavior, and (2) exploring the interaction between voters' selection decisions and entry dynamics.[11]

[10]Consider for example the "Toblerone Affair," during which Swedish MP Mona Sahlin was forced to withdraw her candidacy for leader of her party after a series of revelations that grew out of the fact that she had purchased a *candy bar* (we can only assume for personal consumption) with her official government credit card. (`http://blogs.reuters.com/global/2009/05/11/expenses-they-order-this-matter-differently-in-sweden/`). This may or may not have been the basis for the scandal in the opening episodes of the critically acclaimed Danish television show Borgen.

[11]Also, the details of our argument differ from Caselli and Morelli (2004) in various other ways. Most

The possibility of multiple accountability equilibria arising out of the interaction between voters and politicians has also been explored in recent related papers by Ashworth, Bueno de Mesquita and Friedenberg (2014) and Svolik (2013). Both papers show that multiple equilibria can arise due to a trap of bad or pessimistic expectations on the part of voters: a bad equilibrium where politicians exert little effort when little is expected of them, and a good equilibrium where politicians exert high effort when it is expected.[12] Svolik (2013) also considers the effect of voter and incumbent behavior on entry decisions, showing that in a good equilibrium, bad entrants are discouraged from running for office. A central conclusion of these models is that, like much of the literature on corruption, accountability could be improved by changing expectations. While we echo this result in our analysis as well, as mentioned above, we focus more closely on how daunting this task of changing expectations may be to bring about a reduction in corruption. By linking our study more directly with the literature on strategically-induced multiplicity of equilibria in models of corruption mentioned above, we generate predictions and policy implications not addressed by these studies.[13] We discuss further how our study differs from these other important contributions when we summarize our insights in Section 3.4 below.

significantly, the reason why traps with bad politicians are hard to escape in Caselli and Morelli (2004) is that politicians in office manipulate the value of office-holding in the future, and hence good politicians will choose a value that encourages more good politicians to run while bad politicians choose a value conducive to more bad politicians running. Our model does not rely on this type of argument; instead, entry decisions in our model depend primarily on the existing share of corrupt incumbent politicians.

[12] Meirowitz and Tucker (2013) make a related argument about protest against "bad" leaders: once citizens are sufficiently pessimistic about the pool of candidates they simply conclude that politicians are crooks and do not pay the cost to protest in an effort to replace the current incumbent with a different candidate.

[13] Also, the details of our analysis differ in other ways from these papers. Most notably, the multiplicity in both Ashworth, Bueno de Mesquita and Friedenberg (2014) and Svolik (2013) derives in large part from a voter's ability or willingness to use a politician's action to infer their hidden type. In our analysis, incumbent's type is observed by the voter, and so multiplicity arises from other sources. Also, the model in Ashworth, Bueno de Mesquita and Friedenberg (2014) considers voter-politician relations in isolation; our results discussed below show that an accountability trap in one constituency can change the type of politician running in other constituencies. Further, in Svolik (2013), an important feature that gives rise to multiplicity is the infinite horizon of the interaction between the actors, which is not the case in our analysis.

3 Formal Models

3.1 Politicians

We first present a model of the decision to engage in corrupt behavior among politicians. The model distills some of the ideas from the extant literature on corruption into a simple form, which will allow us to build more complex models that also account for candidate entry and voter behavior in subsequent sections.

Suppose there are $N > 1$ politicians indexed by i. Politicians are characterized by an ability $a_i > \underline{a} > 0$, where $\underline{a}$ is the "minimum" ability level, and a predisposition for corruption $c_i \in \{0, 1\}$. Politicians with $c_i = 0$ are not predisposed towards corruption (or "clean"), which we model by giving them no incentive to engage in corrupt behavior, while those with $c_i = 1$ will have incentives to be corrupt. Denote the proportion of $c_i = 1$ politicians with $\rho \in [0, 1]$. The distribution of ability does not affect the equilibrium analysis in this section.

Politicians have an amount of effort to exert equal to their ability, which they allocate between corrupt activities $x_i \geq 0$ and other, non-corrupt, activities $y_i \geq 0$. That is, they choose x_i and y_i subject to a "budget constraint" $x_i + y_i \leq a_i$. Let the utility for politician i be:

$$u^p(x_i, y_i; c_i) = y_i + c_i g(x_i, \overline{x})$$

The first term captures the benefits to non-corrupt activity, which for simplicity are equal to the effort spent. Politicians with $c_i = 0$ get no benefit from corrupt behavior, while those who are predisposed for corruption get a partial payoff that is a function of their corruption choice (x_i) and the average level of corruption chosen by others ($\overline{x}$), captured by the $g(x_i, \overline{x})$ term. Using notation where g_i is the derivative of g with respect to argument i and g_{ij} is the second derivative with respect to the i'th and j'th argument, we place the following restrictions on g:

Assumption 1. *g is continuous and twice-differentiable, with (i) $g(0, \overline{x}) = 0$, (ii) $g_1 > 0$, (iii) $g_{11} < 0$, (iv) $g_2 > 0$, and (v) $g_{12} > 0$*

Part (i) implies there is no return to corruption if the politician allocates no effort. Parts (ii)-(iii) imply the returns to corruption are increasing and concave in the effort allocated; that is, there are diminishing marginal returns to corrupt behavior. Following the literature on strategic complementarities in corrupt behavior cited in the previous section, parts (iv)-(v) imply that the return to corruption and the marginal return to more corrupt behavior are both increasing in how corrupt other politicians are.

We search for a symmetric Nash Equilibrium of this game, which is characterized by a function mapping the corruption level and ability of each politician to a level of corruption, where these strategies are mutual best responses.

As non-corrupt politicians ($c_i = 0$) get no benefit from corruption, it is immediate that they allocate all of their effort to non-corrupt activities. The payoff for a corrupt politician is increasing in both x_i and y_i, so they always use all of their effort budget: $x_i + y_i = a_i$. Hence we can write the optimization problem for a corrupt politician as selecting x_i to maximize:

$$u^p(x_i; c_i, a_i) = a_i - x_i + g(x_i, \overline{x})$$

Treating the corruption level of others as fixed, this gives a first order condition:

$$g_1(x_i, \overline{x}) = 1$$

Note that this is independent of a_i, so as long as the corrupt politicians always choose an interior $x_i \in (0, a_i)$ (which is ensured by the assumptions below), we can characterize the equilibrium strategy solely by the corruption level chosen by those with $c_i = 1$, and call this

x^*. Since proportion ρ of politicians are corrupt, the equilibrium condition for an interior x^* is such that it solves:

$$g_1(x^*, \rho x^*) = 1 \tag{1}$$

The marginal benefit to allocating an additional unit of effort to non-corrupt activities is always one, so this condition states that corrupt politicians get the same marginal benefit from allocating more effort to corruption, given the fact that other corrupt politicians also choose x^*. The derivative of $g_1(x, \rho x)$ with respect to x is:

$$\frac{\partial g_1(x, \rho x)}{\partial x} = g_{11}(x, \rho x) + \rho g_{12}(x, \rho x)$$

which may be positive or negative since $g_{11} < 0$ but $g_{12} > 0$. In words, this is because there are diminishing returns to corrupt behavior when keeping the behavior of others fixed, but in a symmetric equilibrium increasing the corruption level of others also increases the returns to corruption. If the second effect dominates for some part of the parameter space, it is possible in some cases for this expression to be increasing. In the extreme, if this expression is increasing everywhere, there will never be an interior corruption level: corrupt politicians will either allocate none or all of their time to corrupt activities. In fact, both of these allocations may be an equilibrium choice due to these strategic complementarities, which is analogous to how other models result in a "corruption trap" with multiple equilibria.

To bring attention to a mechanism through which high and low corruption equilibria emerge for other reasons, we assume that $g_{11}(x, \rho x) < -\rho g_{12}(x, \rho x)$, so that $\frac{\partial g_1(x,\rho x)}{\partial x} < 0$. To reduce the cases to consider, we also assume $g_1(0, 0) > 1$ (which ensures $c_i = 1$ types will engage in some corrupt behavior even if no other politician does) and $g_1(\underline{a}, \rho \underline{a}) < 1$ (which ensures $c_i = 1$ types do not allocate all of their time to corruption) for all ρ, where $\underline{a}$ is the minimum ability level. This gives the following:

Proposition 1. *The politician model has a unique equilibrium level of corruption x^* chosen by the $c_1 = 1$ types characterized by equation 1. In equilibrium:*

(i) corrupt politicians get a higher payoff than non-corrupt politicians,

(ii) the optimal level of corrupt activity for those predisposed to corruption is increasing in ρ, and

(iii) the relative payoff of a corrupt politician is increasing in ρ

Proof Part (i) follows from the fact that corrupt politicians could always choose $x_i = 0$ and get the same payoff as a non-corrupt politician, but choose a higher level and hence must get a higher payoff. For part (ii), implicitly differentiating the equilibrium condition gives:[14]

$$\frac{\partial x^*}{\partial \rho} = -\frac{x^* g_{12}(x^*, \rho x^*)}{\frac{\partial g_1(x^*, \rho x^*)}{\partial x^*}}$$

The numerator is positive and the denominator is negative, so the expression is positive. For part (iii), the payoff for a non-corrupt politician is constant in ρ, and the envelope theorem and the fact that $g_2 > 0$ implies the equilibrium payoff for a corrupt politician is increasing in ρ. ∎

For later derivations, it will be useful to express the difference between the equilibrium payoff to a corrupt politician and non-corrupt politician – or the *politician corruption differential* – as:

$$d^p(\rho) = [a_i - x^* + g(x^*, \rho x^*)] - a_i = g(x^*, \rho x^*) - x^*$$

Proposition 1 implies that $d^p(\rho)$ is positive and increasing in ρ.

[14] Technically ρ can only take on values $0, 1/N, 2/N, ...1$, but the second argument of g can be any real number, so this derivative is well-defined.

As we will focus on voter behavior in the next section, now consider the payoff to a citizen represented by politician i. Suppose the payoff to being represented by a politician making effort allocation x_i, y_i is:

$$u^v(x_i, y_i) = y_i + bg(x_i, \overline{x})$$

for some $b \in (0, 1)$. That is, the citizen's (or, mnemonically, "(v)oter's") payoff is also linearly increasing in the non-corrupt activities as well as increasing in the corruption payoff earned by the politician, but scaled downward. One way to interpret this is that for every $b + 1$ dollars earned by the politician from engaging in corruption, they keep 1 to themselves and distribute $b < 1$ to the citizens. More precisely, we assume the *relative* returns to corrupt activity are lower for citizens, but that they still may profit from corruption. This assumption warrants some elaboration. Corruption is typically defined as a misuse of public office for personal gain. We believe that assuming that politicians take a majority of the returns for themselves is consistent with this definition of corruption. At the same time, the portion of the returns b distributed to citizens can be interpreted in several ways, such as clientelism (Vicente and Wantchekon, 2009; Vicente, 2013), vote buying (Brusco, Nazareno and Stokes, 2004; Nichter, 2008) or spoils from a corrupt but able politician who knows how to bend the rules to "get things done" (Munoz, Anduiza and Gallego, 2013; Rundquist, Strom and Peters, 1977), especially when politics is generally corrupt (Barbera, Fernandez-Vazquez and Rivero, Forthcoming).

In equilibrium, the payoffs from having a non-corrupt and corrupt representative are then:

$$\begin{aligned} u^v(c_i = 0; x^*) &= a_i \\ u^v(c_i = 1; x^*) &= a_i - x^* + bg(x^*, \rho x^*) \end{aligned}$$

Analogous to the politician corruption differential, define the *voter corruption differential* as:

$$d^v(\rho) = u_i^v(c_i = 1) - u_i^c(c_i = 0) = bg(x^*, \rho x^*) - x^*$$

As $b \to 0$, the citizen gets a higher payoff from having a non-corrupt politician (i.e., $d^v < 0$), as she only benefits from non-corrupt activities and corrupt politicians allocate less time to these. On the other hand, as $b \to 1$, the citizen payoff is the same as the politician payoff, and since corrupt politicians get a higher payoff, this means the citizen prefers a corrupt representative. By the continuity of the payoff function:

Proposition 2. *There exists a* $\hat{b} \in (0, 1)$ *such that the payoff to having a corrupt politician is higher than having a non-corrupt politician if and only if* $b > \hat{b}$.

Proof Set $\hat{b} = \frac{x^*}{g(x^*, \rho x^*)}$, which is positive, and, since $d^p(\rho) > 0$, so $x^* < g(x^*, \rho x^*)$, and hence $\hat{b} < 1$. ∎

Next we consider how the relative value of having a corrupt representative changes as ρ increases. Consider the following equality:

$$\frac{\partial d^v(\rho)}{\partial \rho} = -\frac{\partial x^*}{\partial \rho} + b\frac{\partial g(x^*, \rho x^*)}{\partial \rho}$$

There are two competing effects here on voter welfare. First, as the level of corruption increases, $c_i = 1$ representatives allocate less effort to non-corrupt activities (the $-\frac{\partial x^*}{\partial \rho}$ term), reducing the relative value of having a corrupt politician. However, the returns to corruption increase, and, since citizens enjoy some of the fruits of corruption, they benefit from this as well. Whether making politics more corrupt makes having a corrupt representative better or worse depends on which of these effects dominates.

When $b = 0$, citizens get no benefit from corruption, and so the only effect of increasing ρ is that corrupt politicians allocate less effort to non-corrupt activities, which hurts their constituents. As a result, $d^v(\rho)$ is decreasing. On the other hand, if $b = 1$, the citizens payoff is exactly the same as the politician, and part iii of proposition 1 is equivalent to stating that $d^v(\rho)$ is increasing. In particular, the relative value of having a corrupt representative is increasing in ρ if and only if citizens get a high enough benefit from corrupt activity:

Proposition 3. *If $b > \tilde{b}$ for some $\tilde{b} < 1$, the citizen payoff to having a corrupt politician is increasing in ρ*

Proof See the Supplemental Appendix.

It is also plausible that when corruption is rare citizens benefit little from corruption (i.e., the b term is lower), but when corruption is more pervasive and open citizens benefit more from it. Such an environment would only reinforce that having a corrupt representative is bad when few others are corrupt, but result in the relative value of having a corrupt representative increase in ρ. This will be an important case in the selection model of voter behavior we present in Section 3.3 below.[15]

3.2 Entry

Next, we present a model where the composition of the pool of those running for office changes as a function of how corrupt they expect politics to be.

To build on the politician model presented in the previous section, suppose that for each of the N districts there are $M > 1$ candidates. As in Section 3.1, each candidate has an ability a_i, and is either predisposed to corruption, denoted with $c_i = 1$, or averse to corruption, denoted $c_i = 0$. Assume the prior probability of being corrupt is given by q, and

[15] Nonetheless, we show in the Supplemental Appendix that it is not necessary that voters prefer corruption politicians – only to tolerate them – for multiple equilibria and hence political corruption traps to emerge.

that ability is independent of corruption and drawn from a cumulative distribution function F. Potential candidates do not know the type of other potential candidates.

Candidates pay a cost k to run for office. Those who do not enter office get a payoff normalized to zero, and those who win office observe the type of other winners and choose a corruption level x_i with the same utility function as in the politician model solved in Section 3.1. Let $\overline{a}$ be the highest possible ability level, and assume $k < \overline{a}$, which will imply that a non-corrupt potential candidate with the highest possible ability would run if guaranteed victory.

A strategy in the entrant model consists of a mapping from the potential candidate type (i.e., the pair (a_i, c_i)) to a decision to run, and a mapping from the type to a corruption level if elected. Our equilibrium definition (formalized below) will require that the strategies chosen are mutual best responses.

Let $Pr(\text{win}; \cdot)$ be the probability of winning, conditioned on the strategies used by others. For simplicity we assume that there is a lottery that determines which candidate wins, so the probability of winning if m_{-i} other candidates run in that district is $1/(1 + m_{-i})$. In equilibrium each candidate will be uncertain about how many others actually run; see the Supplemental Appendix for a derivation of the expected probability of winning.

For a fixed proportion of corrupt (winning) politicians, a candidate prefers to enter if:

$$Pr(\text{win}; \cdot)(a_i + d^p(\rho)c_i) \geq k,$$

where $d^p(\rho)$ is the difference between the equilibrium payoff for a corrupt versus clean politician as defined above. Since this payoff is increasing in a_i, any optimal entrance strategy can be characterized by a threshold for non-corrupt candidates $\hat{a}_0^p$ and corrupt candidates $\hat{a}_1^p$ such that potential candidates run if and only if $a_i \geq \hat{a}_{c_i}^p$.

The winners then choose corruption levels as in the politician model. As demonstrated in the previous section, politicians with $c_i = 0$ always select $x_i = 0$ and we restrict attention to equilibria where corrupt politicians choose a symmetric corruption level x^*. Formally, our equilibrium definition is as follows:

Definition An equilibrium to the entry model comprises a corruption level x^* and entry thresholds $\hat{a}_0^p$ and $\hat{a}_1^p$ that jointly solve:

$$x^* = \arg\max_{x_i} a_i - x + g(x_i, \overline{x}) \tag{2}$$

$$\hat{a}_0^p = k/Pr(\text{win}; \cdot) \tag{3}$$

$$\hat{a}_1^p + d^p(\rho) = k/Pr(\text{win}; \cdot) \tag{4}$$

where

$$\rho = \frac{q(1 - F(\hat{a}_1^p))}{q(1 - F(\hat{a}_1^p)) + (1 - q)(1 - F(\hat{a}_0^p))} \tag{5}$$

And $Pr(\text{win}; \cdot)$ is the expected probability of winning given the entry thresholds, derived in the Supplemental Appendix.

Condition 2 states that politicians choose a corruption level that is a mutual best response as derived in the previous section; as above this can be characterized by the level chosen by all corrupt politicians. Conditions 3-4 imply that given this behavior on the politician level, potential entrants with exactly the threshold ability level $\hat{a}_{c_i}^e$ (given their corruption disposition c_i) are indifferent between running and not. That is, those with ability level above this threshold prefer to run and those below the threshold prefer to stay out of politics. Condition 5 states that the proportion of corrupt politicians is equal to the probability of being corrupt given having a high enough ability to run.[16]

[16]That is, potential entrants behave as if the proportion of corrupt politicians will be exactly equal to the probability of being corrupt if running, while in fact when the number of districts is finite the number

While this is a non-linear system with multiple endogenous variables that interact in complex ways, we can solve the for the equilibria of the model with a relatively simple algorithm that focuses on the proportion of politicians with a predisposition for corruption:

1. Propose an equilibrium proportion of corrupt politicians $\rho^* \in [0, 1]$.

2. Derive the optimal corruption level x^* and the resulting politician corruption differential $d^p(\rho^*)$.

3. Given $d^p(\rho^*)$, compute the entry rules $\hat{a}_0^p$ and $\hat{a}_1^p$ that generate a proportion of corrupt entrants by equation 5.

4. If this proportion is equal to ρ^*, then ρ^* (and the corresponding strategies derived in steps 2-3) comprise an equilibrium; if not there is no equilibrium with proportion of corrupt politicians ρ^*.

The analysis in the previous section accomplishes step 2, so what remains is steps 3-4.

First, we show how the entry behavior changes when treating the proportion of corrupt politicians (and resulting politician behavior) as exogenous:

Proposition 4. *For all ρ:*

(i) There is a unique set of thresholds meeting equations 3-4, and

(ii) $\hat{a}_0^p$ is weakly increasing in ρ and $\hat{a}_1^p$ is weakly decreasing in ρ

The proof for this and all remaining propositions can be found in the Supplemental Appendix. Proposition 4 states that for a proposed level of corrupt incumbents, we can

of corrupt politicians is uncertain (following a binomial distribution). So, if the value of holding office is nonlinear in ρ and the number of districts is finite, this solution concept is not exactly the same as Perfect Bayesian Equilibrium (PBE), which would require the politicians to explicitly sum over the potential realized values of ρ, rendering the solution far more analytically complex. However, when N, is large, the calculation implicitly made by potential candidates approximates the PBE. A simpler justification for this assumption is a behavioral one: citizens simply make an approximation of how corrupt politics will be by equation 5.

determine a unique pair of entry decision rules and hence the (average) proportion of corrupt entrants. Further, as there are more corrupt incumbents, fewer non-corrupt candidates run and more corrupt candidates run. These entry rules imply a proportion of corrupt entrants, and one way to describe the equilibrium condition is that the proportion of corrupt entrants is equal to the proposed equilibrium level. Formally, a ρ^* meeting

$$\rho^* = \frac{q(1 - F(\hat{a}_1^p(\rho^*)))}{q(1 - F(\hat{a}_1^p(\rho^*))) + (1-q)(1 - F(\hat{a}_0^p(\rho^*)))} \tag{6}$$

and the corresponding x^*, $\hat{a}_0^p$, and $\hat{a}_1^p$ computed by equations 2-4 comprise an equilibrium. Our main technical result in this section is as follows:

Proposition 5. *There is at least one – and potentially more than one – ρ^* meeting equation 6 and hence at least one equilibrium and potentially multiple equilibria to the entry model.*

The potential for multiple equilibria means it is possible that there is a stable high level of corruption – where those predisposed to be corrupt are more apt to run for office expecting others to be corrupt as well – and a stable lower level of corruption – where corrupt politicians are less apt to run under the expectation that there will be fewer gains from corrupt behavior. This is loosely analogous to the central idea in Caselli and Morelli (2004) where there can be multiple equilibria, one where most politicians are "good types" and one where most are "bad types."

It is difficult to derive clear analytic results of when such a political corruption trap arises from the entry dynamics; in the Supplemental Appendix, we provide several illustrations. Loosely speaking, there can be multiple equilibria when there value of corrupt behavior is highly contingent on the number of others who are corrupt. That is, it must be the case that returns to corruption are substantially higher when there is a large pool of others predisposed to corruption. We examine when the forces we study lead to multiple equilibria and hence a political corruption trap in the final model presented next.

3.3 Voters

Our final model builds on the previous two but with a focus on voter behavior. As in Sections 3.1 and 3.2, there are N districts. Each district has an incumbent politician, a challenger, and a representative voter. In the first stage of the model, voters simultaneously decides whether to keep their incumbent or replace her with a challenger. Those who are elected then play the politician-level game described in Section 3.1.

Again, we assume politicians are characterized by an ability $a_i \geq \underline{a}$ and corruption level $c_i \in \{0, 1\}$. Assume politician abilities are drawn from a distribution F that admits a density f, such that $f(a) > 0$ for all $a > \underline{a}$. We assume that voters know the type of their incumbent representative but not the potential replacement (or "challenger"). Let the probability of a challenger being corrupt be given by a constant q.

The strategy for a politician is a mapping from their type to a corruption level x_i, and a strategy for a voter is a mapping from the type of their incumbent to a decision to keep the incumbent or vote them out in favor of the challenger.

Given the equilibrium politician behavior described above, we can write the citizen payoff as:

$$u^v(a_i, c_i; \rho) = a_i + d^v(\rho)c_i.$$

The first term implies that the voter's payoff is increasing in the ability of the politician they select. Recall that when d^v is negative, voters prefer a non-corrupt politician, and when d^v is positive they prefer a corrupt politician. Further, recall that d^v can be either positive or negative, and either increasing or decreasing in ρ.

The expected payoff to choosing a challenger is then $\mathbb{E}[a] + qd^v$, and hence the voter prefers a replacement to an incumbent politician with (a_i, c_i) if:

$$a_i + d^v(\rho)c_i \leq \mathbb{E}[a] + qd^v(\rho)$$

In particular, this implies that the voter prefers a corrupt incumbent to the challenger if $a_i \geq \mathbb{E}[a] - (1-q)d^v(\rho)$ and a clean incumbent to the challenger if $a_i \geq \mathbb{E}[a] + qd^v$.

Our equilibrium definition in this section is:

Definition An equilibrium to the voter model comprises a symmetric corruption choice x^* and voter thresholds $\hat{a}_0^v$ and $\hat{a}_1^v$ that jointly solve:

$$x^* = \arg\max_x a_i - x + g(x, \rho\overline{x}) \tag{7}$$

$$\hat{a}_1^v = \mathbb{E}[a] + (1-q)d^v(\rho) \tag{8}$$

$$\hat{a}_0^v = \mathbb{E}[a] + qd^v(\rho) \tag{9}$$

where

$$\rho = \frac{q(1 - F(\hat{a}_1^c))}{q(1 - F(\hat{a}_1^e)) + (1-q)(1 - F(\hat{a}_0^e))} \tag{10}$$

Analogous to the entry model, equation 7 implies that once in office politicians behave as derived in Section 3.1 (which can be summarized with a corruption choice x^* by all $c_i = 1$ types). Equations 8-9 imply that voters behave optimally given the strategies of politicians and other voters. Importantly, equation 10 implies that the proportion of corrupt politicians is equal to the proportion of retained politicians that are corrupt. One way to frame this condition is that in the long-run all the politicians in office are good enough to be retained. For example, if all districts start with a random politician and then continually replace their incumbent based on the decision rules elaborated above (and incumbents never leave for exogenous reasons), the long-run proportion of corrupt politicians will be given by equation 10.

As with the entry model, we search for equilibria by proposing an equilibrium proportion of corrupt politicians ρ^*, use this to compute x^*, $d^v(\rho^*)$, $\hat{a}_0^v$ and $\hat{a}_1^v$, and then check if equation 10 is met. To express this concisely, first define the probability of a randomly drawn politician being clean and of high enough ability to be "kept" as:

$$q_0^k(\rho) = (1-q)(1 - F(\mathbb{E}[a] + qd^v(\rho)))$$

and the analogous probability for a corrupt politician:

$$q_1^k(\rho) = q(1 - F(\mathbb{E}[a] - (1-q)d^v(\rho)))$$

We can then write the equilibrium condition as:

$$\rho^* = \frac{q_1^k(\rho^*)}{q_0^k(\rho^*) + q_1^k(\rho^*)} \equiv \overline{q}(\rho^*). \tag{11}$$

The right-hand side of equation 11 is continuous and bounded by $(0,1)$, and so there must be at least one $\rho^* \in (0,1)$ that meets equation 11. To determine whether intersection (and hence equilibrium) is unique, differentiate the right-hand side with respect to ρ, giving:

$$\frac{\partial \overline{q}}{\partial \rho} = \frac{(q_0^k(\rho) + q_1^k(\rho))\frac{\partial q_1^k}{\partial \rho} - q_1^k(\rho)\left(\frac{\partial q_1^k}{\partial \rho} + \frac{\partial q_0^k}{\partial \rho}\right)}{(q_0^k(\rho) + q_1^k(\rho))^2} = \frac{q_0^k(\rho)\frac{\partial q_1^k}{\partial \rho} - q_1^k(\rho)\frac{\partial q_0^k}{\partial \rho}}{(q_0^k(\rho) + q_1^k(\rho))^2} \tag{12}$$

Referring back to the definitions of $q_i^k(\rho)$, this can be written as:

$$\frac{\partial \overline{q}}{\partial \rho} = \frac{\partial d^v(\rho)}{\partial \rho} \frac{q(1-q)(q_0^k(\rho) f(\mathbb{E}[a] - (1-q)d^v(\rho)) + q_1^k(\rho) f(\mathbb{E}[a] + qd^v(\rho)))}{(q_0^k(\rho) + q_1^k(\rho))^2}$$

All terms other than $\frac{\partial d^v(\rho)}{\partial \rho}$ are always positive, so the sign of this expression is the same as the sign of $\frac{\partial d^v(\rho)}{\partial \rho}$. If $\frac{\partial d^v(\rho)}{\partial \rho}$ is negative – which again implies that when more politicians are corrupt, it is worse to have a corrupt representative, – then $\frac{\partial \overline{q}}{\partial \rho}$ is negative, which implies equation 11 must have a unique solution. However, if $\frac{\partial d^v(\rho)}{\partial \rho}$ is large enough – which implies that the relative benefit of having a corrupt politician increases quickly as there are more additional corrupt politicians – then the slope of this equation can be greater than one, which raises the possibility of multiple intersections and hence multiple equilibria, one with a relatively low share of corrupt politicians and one with a relatively high share of corrupt politicians. Summarizing:

Proposition 6. *There is at least one equilibrium to the voter model, and if* $\frac{\partial d^v(\rho)}{\partial \rho} \leq 0$ *the equilibrium is unique. If* $\frac{\partial d^v(\rho)}{\partial \rho} > 0$ *there may be multiple equilibria to the voter model.*

The possibility of multiple equilibria in the voter model provides another explanation of corruption traps that is inherently political. When there are equilibria with both a low and high level of corruption and the higher corruption equilibrium is played, the accountability mechanism does not necessarily solve problems of corruption and can perpetuate political corruption traps: voters are either accepting of corrupt politicians or avoid clean politicians under the mutual expectation that voters in other districts will do the same. That is, the accountability mechanism requires that citizens are willing to take costly actions to remove "bad" politicians from office. In the case of corruption, if voters in district i are unwilling to vote corrupt politicians from office, voters in district j may not be willing to do so either. In the case where it is better to have a corrupt representative when others are corrupt as well (i.e., d^v is increasing in ρ), voters may even "hold accountable" *clean* politicians who avoid corruption.

3.3.1 Illustrations

It is difficult to provide additional analytical results about when such multiplicity of equilibria is possible. Therefore, we conclude this section with illustrations that highlight conditions under which the multiplicity occurs. In all of the illustrations that follow, ability is drawn from an exponential distribution with scale parameter 1.[17] In the initial illustrations, the proportion of corrupt challengers is $q = .5$; later we allow it to be a function of ρ. What varies in the figures is the voter corruption differential d^v, which we take as exogenous.[18]

In the main text we consider the case where d^v is linear, and hence can be written $d^v(\rho) = b_0 + b_1\rho$. In the Supplemental Appendix, we show illustrations with a non-linear form of d^v.[19] In this formulation, b_0 can be interpreted as the voter corruption differential when no others are corrupt, and b_1 is how the relative value of having a corrupt politician changes as ρ increases. That is, if $b_1 < 0$, the relative value of having a corrupt politician becomes worse as there are more corrupt politicians, while if $b_1 > 0$ the relative value of having a corrupt representative increases as there is more corruption. As proposition 6 states, there will always be a unique equilibrium if $b_1 \leq 0$, and hence we only include illustrations where $b_1 > 0$.

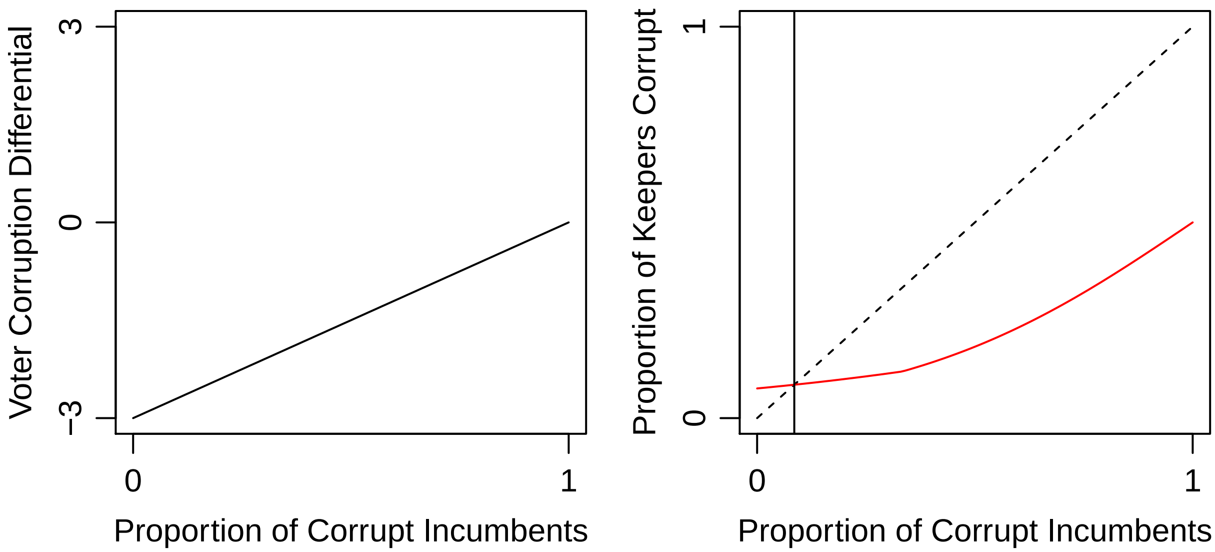
In the first illustration (Figure 1), $b_0 = -3$ and $b_1 = 3$, which implies that it is always bad to have a corrupt politician, but the degree to which it is bad diminishes as there are additional corrupt politicians. In this case, the proportion of kept politicians that are corrupt is increasing in the proportion of others that are corrupt, but the slope of this curve is never steep enough for there to be multiple intersections. So, there is a unique equilibrium with a low level of corruption.

[17] Some of the assumptions assume there is an upper bound on the ability level, which is not the case with the exponential distribution, but all of the results hold when this upper bound is infinite.

[18] That is, we do not specify a primitive g function and solve the politician model, as it is the resulting d^v that matters for voter behavior.

[19] The linear form does not have a natural micro-foundation from the previous section, but allows us to explore the equilibrium condition for a range of functions characterized by only two parameters.

Figure 1: Illustration of voter model with one stable proportion of corrupt incumbents

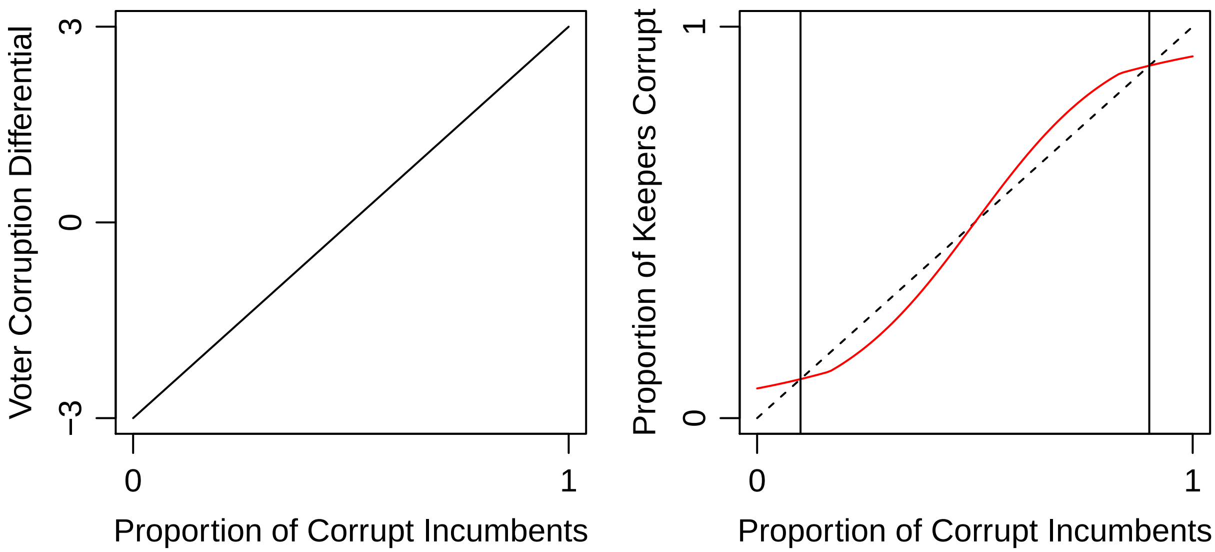


Next, consider a case where $b_0 = -3$ and $b_1 = 6$ (Figure 2). This implies that citizens prefer a non-corrupt representative when less than half of the other politicians are corrupt, but prefer a corrupt representative when more than half of others are corrupt. Now the slope of the right-hand side of the equilibrium condition is steep enough that there is an equilibrium with a very low proportion of corrupt politicians (and corrupt politicians are generally voted out of office) and one with a very high proportion of corrupt politicians (and *non-corrupt* politicians are generally voted out of office).[20]

To generalize beyond these examples, Figure 3 illustrates when there are multiple equilibria as a function of b_0 and b_1 for linear corruption effects. The parameters with multiple equilibria are represented by the shaded area. For this set of parameters, there are multiple equilibria when (1) the slope b_1 is highly positive, and (2) the intercept is negative and

[20]The "middle" equilibrium is unstable, in the sense that a slight increase in the proportion of corrupt politicians would lead there to be even more corruption. On the other hand, in the stable equilibrium, an increase in the proportion of corrupt politicians to $\rho^* + \epsilon$ would lead to a proportion of re-elected corrupt politicians that is less than $\rho^* + \epsilon$.

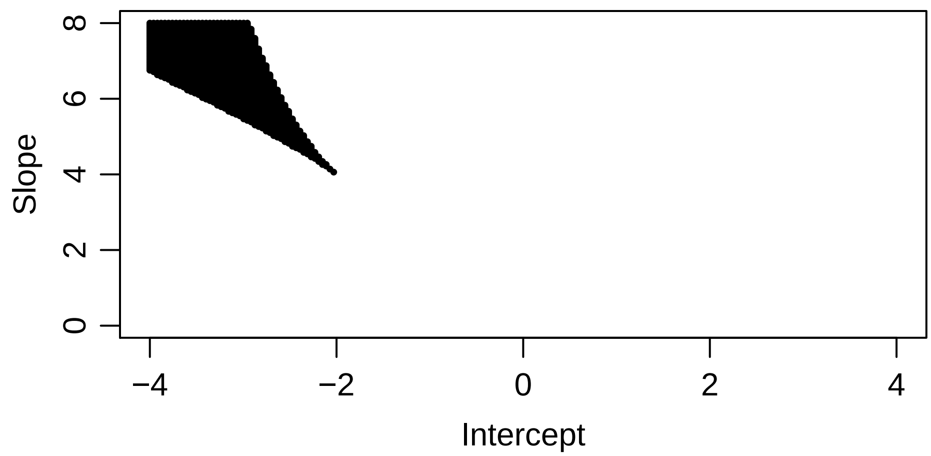
Figure 2: Illustration of voter model with multiple equilibria



around half the slope. This implies that the relative value of having a corrupt representative must be quickly increasing in how many others are corrupt, and the effect of having a corrupt politician must be close to zero when half of politicians are corrupt. Where this intercept lies depends on the prior probability of corruption q, but the general fact that the value of having a corrupt politician must sometimes be positive and sometimes negative always holds in the linear case.

The Supplemental Appendix contains additional illustrations where the voter corruption differential is non-linear, which reinforces the finding that $d^v(\rho)$ must be steep at some point for there to be a political corruption trap. It is important to note that a non-linear d^v raises a possibility that there can be multiple equilibria even if voters always prefer a clean politician; see the Supplemental Appendix for more details.

Figure 3: Range of linear voter corruption differentials that result in multiple equilibria



3.3.2 Non-Constant Entry

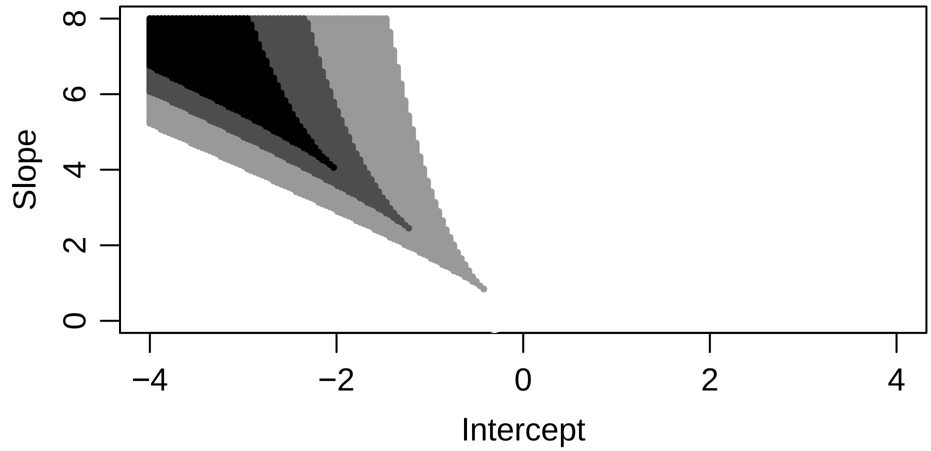
Up to now we have treated the entry and voter dynamics as sharing a common but *separate* underpinning in politician behavior. Our final set of illustrations shows that taking insights from the entry model and incorporating them into the voter model can be another force for multiple equilibria, and hence a political corruption trap.

To do so, we now let the proportion of entrants that are corrupt be a function of ρ (as is predicted in proposition 4). That is, we now write $q(\rho)$, and assume q is increasing. The Supplemental Appendix shows how this changes the analytic solution to the model; here we skip to an illustration of how this makes political corruption traps more likely.

In particular, Figure 4 looks at when there are multiple intersections with a linear q for three cases: first, for $q = .5$ (as before), then for $q = .3 + .4\rho$, and finally for $q = .1 + .8\rho$. Note that for all three of these entry functions, $q(.5) = .5$, however for the latter two the proportion of corrupt incumbents has an increasingly large effect on the proportion of corrupt entrants.

As the proportion of corrupt entrants becomes more heavily dependent on the number

Figure 4: Range of linear corruption effects with multiple equilibria when entry is a function of ρ



The black area corresponds to a case where p_c is constant in ρ, and the dark and light grey areas correspond to the cases when p_c is slightly increasing in ρ and sharply increasing in ρ, respectively.

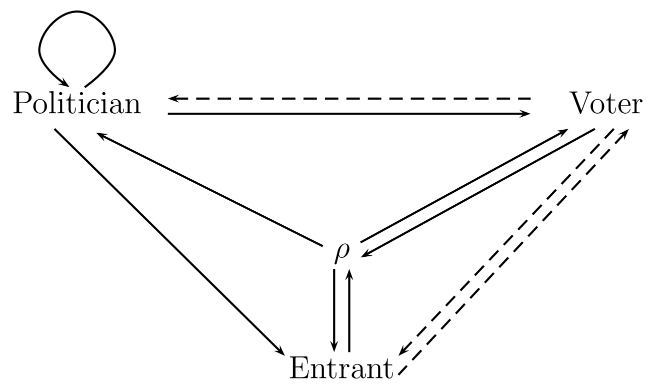
of corrupt incumbents (lighter shaded areas in Figure 4), the range of voter corruption differential functions with multiple equilibria increase. This captures the idea that when both voters and potential entrants condition their behavior on expectations of how many other politicians are corrupt, it is easier to construct cases where there are equilibria with both very few corrupt politicians and nearly pervasive corruption.

3.4 Summary

The overarching conclusion of the formal models we presented is that decisions made by incumbent politicians, potential entrants to politics, and voters all interact in a way that can lead to political corruption traps, i.e., the presence of a high-corruption as well as the low-corruption equilibrium.

Figure 5 summarizes these interactions. For each of the three levels, we can think of the actors in question inter-dependently behaving in a way that may lead to more corruption:

Figure 5: Relationships between the three levels of analysis



ρ is the proportion of politicians predisposed towards corruption.

(1) politicians explicitly choosing to engage in more corrupt behavior, (2) those predisposed to corruption being more apt to enter politics, and (3) voters being more tolerant of or even approving of corrupt politicians. Solid arrows indicate relationships which we have formalized in the previous sections. The dashed arrows represent the relationships we informally discuss below.

In our models, these three levels primarily interact through the channel of what proportion of politicians are predisposed towards corruption (ρ). As ρ increases, politicians choose a higher level of corruption (proposition 1; solid arrow from ρ to politician), voters can be more accepting of corruption (Equation 12; solid arrow from ρ to voter), and more corrupt potential entrants are relatively more likely to enter than clean potential entrants (Proposition 4; solid arrow from ρ to entrant). The solid arrows from entrant/voter to ρ follow from the definition of ρ in these models: more corrupt entrants and more voter tolerance toward corruption directly lead to more corrupt politicians.

The solid arrows from politician to entrant and politician to voter indicate that even for a fixed proportion of incumbents predisposed to corruption, more corrupt behavior (i.e.

greater x^*) has analogous effects on voters and entrants. That is, while the voter and entrant models primarily focused on the proportion of corrupt incumbents (ρ), what really matters is the aggregate level of corruption (ρx^*). So, we could equivalently write the voter and politician corruption differentials as $d^p(\rho x^*)$, and the same results that formalize the ρ to entrant and ρ to voter lines imply the lines from politician to entrant and politician to voter. Finally, the insights from Section 3.3.2 show the combined effect of ρ on the behavior of entrants (solid arrow from ρ and entrant) and voters (solid arrow from ρ to voter).

There are other potential sources of positive feedback among these levels of analysis not modeled here. For example, if voters are less apt to punish corrupt behavior, then politicians will be more free to engage in corruption (dashed line from voter to politician). Further, those predisposed to corruption are more apt to run if behaving in a corrupt manner is less likely to lead to voter punishment (dashed line from voter to entrant). Finally, if voters think all entrants will be corrupt, there is no incentive to vote out corrupt incumbents for this reason (dashed line from entrant to voter).[21]

We can also relate our argument to the existing literature on multiple corruption and accountability equilibria in the context of Figure 5. Most of the existing formal work on multiple corruption equilibria looks at the circular arrow from politician to politician (also formalized here in Proposition 1; Andvig and Moene 1990; Cadot 1987; Lui 1986; Mauro 2004).[22] The main model in Caselli and Morelli (2004) presents a somewhat different argument for the arrows between entrant and ρ, where candidates that are higher quality in general are more likely to run for office when other high quality candidates run; they also argue that politicians choose policy decisions that make candidates more like them (good or bad) run in the future, capturing an alternative channel from politician to entrant. Ash-

[21]However, see Klašnja (Forthcoming) where voters may prefer challengers *even if* they are also corrupt, when incumbents' gains from corruption increase over time spent in office, possibly due to learning or the time it takes to develop rent-extraction networks.

[22]Formally, even for a fixed ρ the incentives to commit fraud are increasing in how much fraud others commit, i.e., x^*.

worth, Bueno de Mesquita and Friedenberg (2014) and Svolik (2013) characterize differently the arrows between politician and voter, arguing that voters' pessimistic expectations can fail to incentivize politicians who could otherwise be induced to behave. Svolik (2013) also models the arrows between voter and entrant by showing that when bad entrants are discouraged from running when voters pay the cost of monitoring politicians and potential entrants' outside options are good.

In sum, the findings from our models, summarized in Figure 5, suggest that there are strong strategic complementarities in many of the decision among voters, entrants and incumbents that determine the level of corruption in a polity. This raises the problematic possibility that interactions between these three sets of actors create even more opportunities for a political corruption trap. We conclude by informally discussing why we believe such political corruption traps may be particularly hard to escape.

4 Discussion

The most common implication of the existing account of strategically-induced corruption traps is that a way to escape to a lower corruption equilibrium is to change the expectation among the fixed group of politicians (or bureaucrats). That is, believing that all others will switch to a lower corruption equilibrium can be a self-fulfilling prophecy. In our models, if all politicians are up for re-election at once, the type of corruption trap that exists when there are multiple equilibria can also be interpreted as solvable by coordinating expectations as in the canonical literature. That is, if all the voters anticipate that all other voters – including those in other constituencies – will switch to using the selection rules associated with the low corruption equilibrium rule and are patient enough to wait until the corruption level reaches its lower steady state, it will be optimal to do so.[23] If endogenous entry is also

[23]Recall we assume challengers are drawn at random and may not be worth keeping by any decision rule, so it may take multiple replacements to obtain a sufficiently able representative.

considered, then potential entrants to politics would need to coordinate their expectations as well.

We think this is unrealistic for three reasons. First, it requires coordinating expectations of many actors that have no direct contact with each other, rather than a fixed and relatively small group of incumbent politicians or bureaucrats. Coordinating expectations of voters alone can be long and unpredictable; for example, voters in Italy were routinely electing corrupt politicians for over forty years before the breakdown of the complex system of kick-backs and electoral corruption (Chang, Golden and Hill, 2010).

Second, it may require voters to be very patient, as it may take multiple replacements to find a representative that is both competent and non-corrupt. The results of Svolik (2013) and Meirowitz and Tucker (2013) suggest that such patience may be unrealistic, since repeated poor performance by incumbents may lead the voters to abandon the monitoring of politicians, discourage them from protesting against the regime, and even question the merits of democratic rule.

Third, given the importance of the prevalence of corruption among incumbents (ρ) for the possibility of a corruption trap in our models, coordinating the expectations of all actors and voter patience may not be enough, unless the entire political elite in power is replaced. This is likely to be very difficult. As (Golden, 2006) notes:

> the implosion of the Italian party system on the heels of the Clean Hands investigations [in the early 1990s] is the only known historical instance in any democratic nation of the electoral repudiation by voters of an entire corrupt national political elite 2006, pp.80-81.

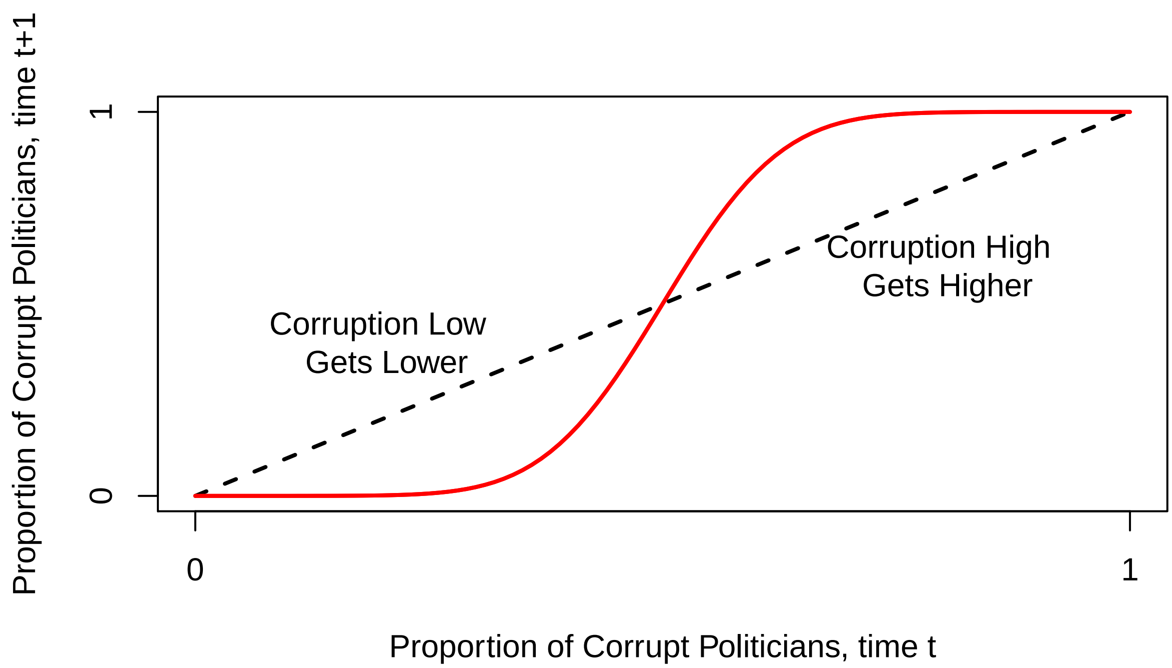
This suggests that the persistence of corruption may in part be a consequence of political and electoral institutions that influence how many politicians are up for reelection at any point in time. For example, staggered election timing and the many incumbency advantages may make it even more difficult to replace politicians *en masse*. Therefore, our results suggest that exploring the effects of electoral and political institutions on the *persistence* of corruption may be a valuable way to learn more about how countries can escape these traps. Existing studies of the institutional antecedents of corruption primarily focus on explaining the effects on the *level* of corruption, rather than its persistence over time (e.g. Lambsdorff, 2006; Pellegrini and Gerlagh, 2008).

Finally, political corruption traps can be even stronger when these dynamics interact over time. A geometric intuition for when a political corruption trap can be strong is given by Figure 6. This figure shows a more general version of what we formalized above, drawing the proportion of corrupt politicians in the next "period" as a function of the proportion of corrupt politicians today. In this hypothetical curve, when the preponderance of corrupt politicians is low (specifically, below a half), the curve lies below the 45 degree line, indicating that even fewer corrupt politicians will be around in the next period. When the preponderance of corruption is high, the opposite is true: politics gets even more corrupt. As a result, there are three intersections corresponding to two stable equilibria: one where nearly all politicians are corrupt and one where nearly all politicians are clean.

This can help explain why at any given moment in time, we observe what looks like two very stable equilibria – low corruption and high corruption societies – despite the fact that in the past there may have been moderate levels of corruption in what are now the low corruption countries/regions. What is particularly interesting about this intuition is that the explanation for why moderately corrupt countries become less so has nothing to do with a changing norm regarding corruption (or "culture of corruption"), but rather with a dynamic that emerges over time from the behavior of both politicians and voters.[24] Similarly, the logic can explain why countries that are fairly corrupt may become more so over time, eventually arriving at a "high corruption" stable equilibrium. Svolik (2013) formalizes an even more

[24]For the cultural transmission of attitudes towards corruption, see Simpser (2013).

Figure 6: A Strong Political Corruption Trap



pessimistic version of this argument, where a sufficiently large proportion of "bad" politicians leads citizens to abandon monitoring and holding their politicians accountable entirely. If these forces mean more that a sufficiently high proportion of politicians will be corrupt in the future – as is the case in Figure 6 – a political corruption trap will ensue.

It is worth noting that this logic highlights the importance of regime transitions as a key moment in determining a country's future trajectory in terms of quality of governance. Usually, transitions are seen as important because they are an opportunity to impose a new set of norms on state actors, but they can be undermined by an underlying "culture of corruption" among bureaucrats that is hard to change, especially if incentives still remain for strategic interactions among bureaucrats that lead to corrupt behavior. If our explanation is correct, then the key is to reduce the number of corrupt politicians, and to do so quickly and thoroughly enough so that countries get on the dynamic path to a "low corruption equilibrium".

References

Abbink, Klaus and Danila Serra. 2012. Anti-Corruption Policies: Lessons from the Lab. In *New Advances in Experimental Research on Corruption*, ed. Danila Serra and Leonard Wantchekon. Vol. Research In Experimental Economics Volume 15 Emerald Group Publishing.

Alt, James E and David Dreyer Lassen. 2003. "The political economy of institutions and corruption in American states." *Journal of Theoretical Politics* 15(3):341–365.

Andvig, Jens Chr and Karl Ove Moene. 1990. "How corruption may corrupt." *Journal of Economic Behavior & Organization* 13(1):63–76.

Ashworth, Scott, Ethan Bueno de Mesquita and Amanda Friedenberg. 2014. "Accountability Traps." Manuscript.

Banerjee, Abhijit and Rohini Pande. 2007. "Parochial Politics: Ethnic Preferences and Politician Corruption.".

Banfield, Edward C and James Q Wilson. 1963. *City politics*. Vol. 335 Harvard University Press Cambridge, MA.

Banfield, Edward C and Laura Fasano Banfield. 1958. *The moral basis of a backward society*. Free Press New York.

Barbera, Pablo, Pablo Fernandez-Vazquez and Gonzalo Rivero. Forthcoming. "Rooting Out Corruption or Rooting For Corruption? The Electoral Consequences of Corruption Scandals in Spain, 2003-2011." *Political Science Research and Methods* . Manuscript.

Becker, Sascha, Katrin Boeckh, Christa Hainz and Ludger Woessmann. 2011. "The Empire is Dead, Long Live the Empire! Long-run Persistence of Trust and Corruption in the Bureaucracy.".

Brusco, Valeria, Marcelo Nazareno and Susan C. Stokes. 2004. "Vote Buying in Argentina." *Latin American Research Review* 39(2):66–88.

Cadot, Olivier. 1987. "Corruption as a Gamble." *Journal of Public Economics* 33(2):223–244.

Caselli, Francesco and Massimo Morelli. 2004. "Bad politicians." *Journal of Public Economics* 88(3–4):759 – 782.

Chang, Eric C.C., Miriam A. Golden and Seth J. Hill. 2010. "Legislative Malfeasance and Political Accountability." *World Politics* 62(2):177–220.

Chong, Alberto, Ana L. De La O, Dean Karlan and Leonard Wantchekon. 2011. "Looking Beyond the Incumbent: The Effects of Exposing Corruption on Electoral Outcomes." NBER Working Paper No. 17679.

Damania, Richard, PerG. Fredriksson and Muthukumara Mani. 2004. "The Persistence of Corruption and Regulatory Compliance Failures: Theory and Evidence." *Public Choice* 121(3-4):363–390.

de Figueiredo, Miguel FP, F Daniel Hidalgo and Yuri Kasahara. 2011. "When do voters punish corrupt politicians? Experimental evidence from Brazil." *Unpublished Manuscript, University of California Berkeley* .

Ferraz, Claudio and Fred S. Finan. 2008. "Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes." *Quarterly Journal of Economics* 123(2):703–745.

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

Glaeser, Edward L and Raven E Saks. 2006. "Corruption in America." *Journal of Public Economics* 90(6):1053–1072.

Golden, Miriam A. 2006. "Some Puzzles of Political Corruption in Modern Advanced Democracies." Paper presented at the 2005 Annual Meetings of the Japan Political Science Association, Tokyo, and at a meeting on political accountability, Waseda University.

Golden, Miriam A and Lucio Picci. 2005. "Proposal for a new measure of corruption, illustrated with Italian data." *Economics & Politics* 17(1):37–75.

Hauk, Esther and Maria Saez-Marti. 2002. "On the cultural transmission of corruption." *Journal of Economic Theory* 107(2):311–335.

Hirano, Shigeo and James M Snyder. 2012. "What Happens to Incumbents in Scandals?" *Quarterly Journal of Political Science* 7:447–456.

Hollyer, James R. and Leonard Wantchekon. 2015. "Corruption and Ideology in Autocracies." *Journal of Law, Economics and Organization* Forthcoming. Manuscript.

Humphreys, Macartan and Jeremy M. Weinstein. 2012. "Policing Politicians: Citizen Empowerement and Political Accountability in Uganda." Paper presented at the Annual Bank Conference on Development Economics at the World Bank.

Klašnja, Marko. 2013. "Uninformed Voters and Corrupt Incumbents." Manuscript.

Klašnja, Marko. 2014. "Corruption and the Incumbency Disadvantage: Theory and Evidence." Manuscript.

Klašnja, Marko. Forthcoming. "Increasing Rents and Incumbency Disadvantage." *Journal of Theoretical Politics* .

Klašnja, Marko and Joshua A. Tucker. 2013. "The Economy, Corruption, and the Vote: Evidence from Experiments in Sweden and Moldova." *Electoral Studies* 32:536–543.

Klašnja, Marko, Joshua A. Tucker and Kevin Deegan-Krause. 2014. "Pocketbook vs. Sociotropic Corruption Voting." *British Journal of Political Science* Forthcoming.

Lambsdorff, Johann Graf. 2006. Consequences and Causes of Corruption: What Do We Know from a Cross-Section of Countries? In *International Handbook on the Economics of Corruption*, ed. Susan Rose-Ackerman. Northampton, M.A.: Edward Elgar.

Lui, Francis T. 1986. "A Dynamic Model of Corruption Deterrence." *Journal of Public Economics* 31:1–22.

Mauro, Paolo. 2004. "The persistence of corruption and slow economic growth." *IMF staff papers* pp. 1–18.

Meirowitz, Adam and Joshua A. Tucker. 2013. "People Power or a One-Shot Deal? A Dynamic Model of Protest." *American Journal of Political Science* 57(2):478–490.

Menes, Rebecca. 2006. Limiting the Reach of the Grabbing Hand: Graft and Growth in American Cities, 1880 to 1930. In *Corruption and Reform: Lessons from America's Economic History*, ed. Edward L. Glaeser and Claudia Goldin. Chicago, IL: University of Chicago Press.

Mishra, Ajit. 2005. "Persistence of Corruption: Some Theoretical Perspectives." *World Development* 34:349–358.

Montinola, Gabriella A. and Robert W. Jackman. 2002. "Sources of Corruption: A Cross-Country Study." *British Journal of Political Science* 32:147–170.

Munoz, Jordi, Eva Anduiza and Aina Gallego. 2013. "Why do Voters Forgive Corrupt Politicians? Implicit Exchange, Noise and Cynicism." Manuscript.

Nichter, Simeon. 2008. "Vote Buying or Turnout Buying? Machine Politics and the Secret Ballot." *American Political Science Review* 102(1):19–31.

Pellegrini, Lorenzo and Reyer Gerlagh. 2008. "Causes of Corruption: a Survey of Cross-National Analyses and Extended Results." *Economics of Governance* 9:245–263.

Peters, John G. and Susan Welch. 1980. "The Effects of Charges of Corruption on Voting Behavior in Congressional Elections." *The American Political Science Review* 74(3):697–708.

Quah, Jon ST. 1994. "Controlling Corruption in City-States: A Comparative Study of Hong Kong and Singapore." *Crime, Law and Social Change* 22(4):391–414.

Reed, Steven A. 1999. Punishing Corruption: The Response of the Japanese Electorate to Scandals. In *Political Psychology in Japan: Behind the Nails which Sometimes Stick Out (and Get Hammered Down)*, ed. Ofer Feldman. Commack, NY: Nova Science pp. 131–148.

Rundquist, Barry S., Gerald S. Strom and John G. Peters. 1977. "Corrupt Politicians and Their Electoral Support: Some Experimental Observations." *The American Political Science Review* 71(3):954–963.

Simpser, Alberto. 2013. "The Inter-generational Persistence of Attitudes toward Corruption." Manuscript.

Svolik, Milan. 2013. "Learning to Love Democracy: Electoral Accountability, Government Performance, and the Consolidation of Democracy." *American Journal of Political Science* 57(3):685–702.

Tirole, Jean. 1996. "A Theory of Collective Reputations (with Applications to the Persistence of Corruption and to Firm Quality)." *The Review of Economic Studies* 63(1):1–22.

Transparency International India. 2005. "India Corruption Study." Transparency International India and Center for Media Studies.

Treisman, Daniel. 2000. "The Causes of Corruption: a Cross-National Study." *Journal of Public Economics* 76(3):399–457.

Vicente, Pedro C. 2013. "Is Vote-Buying Effective? Evidence from a Field Experiment in West Africa." Manuscript.

Vicente, Pedro C. and Leonard Wantchekon. 2009. "Clientelism and Vote Buying: Lessons from Field Experiments in African Elections." *Oxford Review of Economic Policy* 25(2):292–305.

Weitz-Shapiro, Rebecca and Matthew S. Winters. Forthcoming. "Lacking Information or Condoning Corruption? Voter Attitudes Toward Corruption in Brazil." Comparative Politics.

Welch, Susan and John R. Hibbing. 1997. "The Effects of Charges of Corruption on Voting Behavior in Congressional Elections, 1982-1990." *The Journal of Politics* 59(1):226–239.