Political Corruption Traps*

MARKO KLAŠNJA, ANDREW T. LITTLE AND JOSHUA A. TUCKER

*A*cademics and policymakers recognize that there are serious costs associated with systemic corruption. Stubbornly, many countries or regions remain stuck in a high-corruption equilibrium—a "corruption trap." Most existing theories concentrate on mutually reinforcing expectations of corrupt behavior among a fixed set of bureaucrats or politicians, implying that changing such expectations can lead to lower corruption. We develop models that more fully characterize the political nature of corruption traps by also analyzing the behavior of voters and entrants to politics, as well as their interaction with incumbent politicians. We show that corruption traps can arise through strategic behavior of each set of actors, as well as through their interrelations. 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.

The study of corruption—its causes, effects, and persistence—is 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' domestic and foreign policies.

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 2007), as well as within countries; for example, there are large differences in corruption across American states (Alt and Lassen 2003; Glaeser and Saks 2006) and cities (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 very persistent: we rarely observe high-corruption countries or regions becoming low-corruption countries or regions, or vice versa (see Figure A1 in the Supplemental Appendix; see also Damania, Fredriksson and Mani 2004; Becker et al. 2016).[1] One plausible explanation for this persistence in corruption levels is that the dynamics underlying corruption allow for multiple equilibria.

[1] Two important exceptions are the small city-states of Singapore and Hong Kong (see Quah 1994). We should note that we are not taking a very long view—even countries like the United States had considerably more

The existence of multiple equilibria can lead to differences among similar countries and regions driven by whether they play the "low-corruption" or "high-corruption" equilibrium; if countries or regions consistently play the same equilibrium, these differences will persist over time.

This explanation has received significant theoretical attention. Most of the previous accounts focus on interactions between a fixed set of bureaucrats or politicians, proposing that incentives to engage in corrupt behavior may increase as government officials come to believe that other government officials are also corrupt. These mutually reinforcing beliefs create strategic complementarities which can then lead to multiple equilibria: a low-corruption equilibrium where bureaucrats or politicians 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. For example, in Cadot (1987), the high prevalence of corruption implies that higher-level bureaucrats are more tolerant of lower-level corruption. In Mishra (2005), the cost of compliance increases as others become increasingly non-compliant, lowering the benefits from being "clean" (see also Mauro 2004). 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 from engaging in corruption (see also Lui 1986).[2]

We build on this approach based on strategic complementarities, but 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, but also on voters and potential candidates for political office ("entrants"), as well as the interactions among these three sets of actors. In principle, bureaucrats are accountable to politicians, who in democracies are accountable to voters.[3] This electoral accountability should in principle make it easier to escape the high-corruption equilibrium.

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 voters react negatively to corruption regardless of the state of the economy, whereas in Moldova voters react negatively to corruption *only* when the state of the economy is also poor. Elsewhere, scholars have similarly found that in higher-corruption countries, voters are often insensitive to corruption (Banerjee and Pande 2009; Manzetti and Wilson 2007; Zechmeister and Zizumbo-Colunga 2013; Klašnja, Tucker and Deegan-Krause 2016), sometimes even when given precise information about it.[4] Voters in low-corruption countries tend to punish corruption more consistently (Welch and Hibbing 1997; Reed 1999; Hirano and Snyder 2012).[5] This suggests that voters' behavior may contribute to the multiplicity of equilibria in corruption prevalence.

(F'note continued)

corruption in the 19th century. However, very few countries have moved much on corruption relative to other countries in the past several decades.

[2] Relatedly, Corbacho et al. (forthcoming) show that citizens are more likely to engage in corruption when they are induced to perceive a higher overall level of corruption.

[3] We limit the applicability of our models to democracies with meaningfully competitive elections. For the study of anti-corruption efforts in autocracies, see, for example, Hollyer and Wantchekon (2015).

[4] For example, de Figueiredo, Hidalgo and Kasahara (2012) find no effect of an informational intervention about an incumbent's corruption on voters' electoral behavior in a mayoral race in Brazil. Some studies, however, do find more consistent evidence of voters punishing corruption in high-corruption countries (Ferraz and Finan 2008; Chong et al. 2015).

[5] However, even voters in low-corruption countries do not always sanction corruption (Klašnja 2015; Barberá, Fernández-Vázquez and Rivero 2016).

A second shortcoming of much of the existing literature is that high-corruption equilibria driven mainly by expectations of how corrupt others will be suggest a particular anti-corruption policy prescription: an intervention that rapidly changes the corruption expectations of a fixed set of actors. While such interventions are not costless, they seem feasible, and yet, they appear inconsistent with the fact that corruption persists across countries and regions. Instead, we argue that high-corruption equilibria may be hard to escape because they are driven by the prevalence of particular *types* of politicians, both those in office and those willing to enter politics, as well as the interaction among politicians and with voters. Our models thus exhibit more of the strategic complementarities that lead to multiple equilibria than past models suggest, highlighting why acting on bureaucrats' expectations alone may not be enough to reduce corrupt behavior.[6]

Our analysis builds on a simple model of corrupt behavior among politicians in office. Politicians vary based on their predisposition to engage in corrupt behavior (as well as ability), and choose a level of corrupt activity where the benefits from doing so are a function of the degree to which other politicians choose to engage in similarly corrupt behavior. This analysis has two important properties that drive our results about voter (and entrant) behavior. First, a given politician engages in more corrupt behavior when others in office are more predisposed to corruption. Second, the relative value of holding office for a politician is higher for those predisposed to corruption, and, in line with the above-mentioned work on multiple equilibria of corruption, the size of this "politician corruption differential" is increasing in the level of corruption. When combined with our assumptions about citizen preferences, the second point implies that as politics becomes more corrupt, citizens can become more tolerant of corrupt politicians, and thus less apt to vote them out of office.

We then proceed to show how these results affect the behavior of voters and potential candidates. In our main model, voters observe the ability and corruption predisposition of their current representative and decide whether to keep or replace her. If voters become more tolerant or even prefer corrupt representatives as others are more corrupt, there can be multiple equilibria: one equilibrium where voters are generally willing to retain corrupt politicians, and one where voters generally replace corrupt politicians. We call the former situation a "political corruption trap." Next, we draw on Caselli and Morelli (2004) in deriving a model that examines a similar complementarity between potential entrants, where those predisposed to corruption are more apt to run when they expect others to be corrupt as well. Importantly, we also show that when the dynamics from the voter and politician models are combined with the dynamics of the entry model, the potential for a political corruption trap is generally even greater.

Our analysis is related to but differs from several other important previous contributions that have also investigated the roles of political incentives and heterogeneity in politician types in persistence of corruption or quality of accountability more broadly. The closest model to our analysis of entry decisions is the model by Caselli and Morelli (2004), who argue that the current share of "bad" politicians will affect the relative returns to office-holding for other bad politicians. Our paper builds on this argument by primarily focusing on the strategic complementarities in voters' tolerance of corruption, while also exploring the interaction between voters' selection decisions and candidate entry dynamics.[7]

---

[6] Moreover, even changing the expectations of all three sets of actors (incumbents, entrants, and voters) simultaneously may be insufficient to reduce corruption unless it is possible to remove many corrupt politicians simultaneously. Thus, electoral institutions such as staggered elections may also play a role in sustaining political corruption traps.

[7] The details of our analysis also differ from Caselli and Morelli (2004). Most notably, the traps with bad politicians in Caselli and Morelli (2004) are hard to escape because politicians in office manipulate the value of

The possibility of multiple accountability equilibria arising from the interaction between voters and politicians has also been explored in recent related papers by Ashworth, Bueno de Mesquita and Friedenberg (forthcoming) and Svolik (2013). Both papers describe a trap due to voters' pessimistic expectations: a bad equilibrium where politicians exert little effort when little is expected of them. While we echo this result in our analysis as well, we focus more closely on how daunting this task of changing expectations may be if the goal is to bring about a reduction in corruption.[8]

This paper proceeds as follows. The second section lays out our main model of politician and voter behavior, which in turn can lead to a corruption trap. These corruption traps depend on a particular configuration of politician types; however, in the third section we describe traps arising from proportions of corrupt politicians that would be stable over time, independent of the particular distribution of politician types in any given election. The penultimate section briefly discusses our model of the behavior of entrants and corruption traps that arise therein, and more importantly, how corruption traps can arise through interactions between politicians, voters, and entrants. The last section summarizes our results and discusses their implications and directions for future research.

THE MODEL

Consider a country with N geographic units (e.g., cities, counties, or legislative districts). Each unit contains a voter (V), an incumbent politician (I), and a challenger (C). This environment could correspond to an electoral system where legislators are elected by plurality vote in single-member constituencies. In a proportional representation system (and elsewhere), the model could also describe the choice of directly elected local executives, such as governors or mayors.[9]

Politicians are characterized by a two-dimensional type which captures their general ability and predisposition to corruption. Let a_i^I and c_i^I be the ability and corruption predisposition for the incumbent politician in district i. The challenger's ability and predisposition are a_i^C and c_i^C. We assume that $a_i^J \geq \underline{a} > 0$ ($J \in \{I, C\}$), where $\underline{a}$ is the "minimum" ability level. The predisposition for corruption is binary, where $c_i^J = 0$ indicates no predisposition toward corruption ("clean"), while those with $c_i^J = 1$ have a predisposition toward corruption.

The voters decide which candidate to elect, and then the elected politicians decide how to allocate their time between corrupt and non-corrupt activities. The voter knows the type of the incumbent but not the challenger.[10] Let F be the cumulative distribution function for the challenger ability, and $q \in (0, 1)$ the probability that a challenger is corrupt. We assume that F is continuous and differentiable, and that a politician's predisposition to corruption and ability are independent. Once elected, the politician types become common knowledge. When referring to the elected politician, we drop the superscripts, so the type of the elected politician in district i is (a_i, c_i).

(F'note continued)

office-holding in the future. 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.

[8] Also, the details of our analysis differ from these models. Most notably, the multiplicity in both Ashworth, Bueno de Mesquita and Friedenberg (forthcoming) 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.

[9] In Section A2 of the Supplemental Appendix, we provide a number of real-world motivating cases.

[10] If voters also (partially) observe the type of challenger their re-election decision will depend on the type of challenger as well. However, citizens will still be more apt to keep incumbents who are more able and (sometimes) corrupt, which is the main force driving our results.

Politicians have an amount of effort to exert equal to their ability. Effort can be spent on corrupt activities $x_i \geq 0$ and other non-corrupt activities $y_i \geq 0$. Formally, politicians choose x_i and y_i subject to a "budget constraint" $x_i + y_i \leq a_i$. Let the utility for the elected politician (u^P) and voter (u^V) in district i be

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

$$u^V(x_i, y_i) = y_i + b g(x_i, \bar{x}).$$

(The non-elected politicians take no actions, and so their preferences do not affect the analysis.)

The first term of both utilities captures the returns to non-corrupt activity. For simplicity we set this equal to the effort allocated for both politician and voter. Importantly, we assume that the returns to non-corrupt activity do not have externalities for the politicians in other districts, as we will do for corrupt activities.[11]

Politicians with $c_i = 0$ get no benefit from corrupt behavior.[12] Those who are predisposed toward engaging in corrupt behavior get a partial payoff that is a function of their corruption choice (x_i) and the average level of corruption chosen by others ($\bar{x}$), captured by $g(x_i, \bar{x})$ term. We assume that g is continuous and twice differentiable, with $g(0, \bar{x}) = 0$, meaning that there is no return to corruption if the politician allocates no effort to this end.

To express the remaining assumptions on g, we use the common notation where g_1 is the partial derivative of g with respect to its first argument (in this case x_i), g_2 the partial derivative of g with respect to the second argument, g_{11} the second derivative with respect to the first argument, g_{12} the cross-partial derivative, etc. In particular, let the returns to corruption be increasing ($g_1 > 0$) and concave ($g_{11} < 0$) in the effort allocated; that is, there are diminishing marginal returns to corrupt behavior. In line with the literature on strategic complementarities in corrupt behavior we reviewed above, we assume that the return to corruption and the marginal return to more corrupt behavior are both increasing in how corrupt other politicians are ($g_2 > 0$ and $g_{12} \geq 0$).[13]

Now consider how corrupt behavior affects the voter (or, more precisely, the pivotal voter). For negative b, the citizen's payoff is decreasing in the corruption payoff earned by the politician. Voters may be directly harmed by the corrupt behavior of their representatives (in addition to the foregone effort the representatives could have made toward more productive ends) for a myriad of reasons, including negative effects of corruption on outcomes like investment (Mauro 1995) and human capital accumulation (Reinikka and Svensson 2004). However, even if corruption generally harms the pivotal voter, she may sometimes *benefit* when her representative is corrupt, implying $b > 0$. While $b \geq 1$ is unlikely (i.e., that the voter benefits above and beyond the politician's own corrupt payoff; we do not consider this case hereafter), $b \in (0, 1)$ implies that the voter's payoff is increasing in the corruption payoff earned by the politician, but scaled downward. There are numerous examples of re-election-seeking
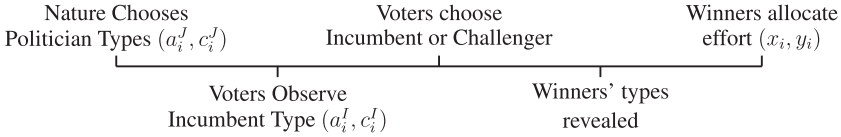
---

[11] The returns to effort in other political activities (constituency service, crafting policy, etc.) could in some contexts also exhibit strategic complementarities among elected officials. The important requirement for our model (and, more precisely, our interpretation with respect to corruption) is that whatever activities politicians do less of in order to engage in corruption generally exhibit weaker strategic complementarities.

[12] It may seem odd that we assume that clean politicians benefit even less from corruption than the voter. We employ this stark contrast to ensure politicians with $c_i = 0$ engage in no corruption, thus reducing the analysis to finding the corruption level of politicians with $c_i = 1$. The equilibrium analysis would be equivalent if politicians with $c_i = 0$ gain from engaging in corruption but also pay a sufficiently high cost (in addition to the opportunity cost of having less time for y_i) for engaging in corrupt behavior which makes them prefer to choose $x_i = 0$. This cost could be a psychological cost from engaging in illicit activities, or because politicians with $c_i = 0$ are not "sneaky" enough not to get caught.

[13] A class of functions meeting all of these requirements is $g(x, \bar{x}) = x^\alpha + (x\bar{x})^\beta$ for any $\alpha \in (0, 1)$ and $\beta \in (0, 1)$.

politicians funneling some of the corrupt rents to their electorate or political parties through vote-buying campaigns and patronage. For example, *coronéis*, local political bosses in Brazil, have long used corrupt revenues to broker votes (Mainwaring 1999; Gingerich 2014). We provide additional examples and further discussion of this assumption in Section A2 in the Supplemental Appendix.

The following timeline summarizes the sequence of moves:

Nature Chooses
Politician Types (a_i^J, c_i^J)

Voters choose
Incumbent or Challenger

Winners allocate
effort (x_i, y_i)

Voters Observe
Incumbent Type (a_i^I, c_i^I)

Winners' types
revealed

We first solve for the perfect Bayesian equilibria (PBE) of the model, and then search for a proportion of corrupt politicians that is stable in the long term.

Politician Stage

The politician strategies are a function mapping their corruption level and ability to an allocation of effort (x_i, y_i), where these strategies must be mutual best responses.

As politicians with $c_i = 0$ get no benefit from corruption, they allocate all of their effort to non-corrupt activities. Since the payoff for a corrupt politician is increasing in both x_i and y_i, they always use the entire effort budget: $x_i + y_i = a_i$. So, corrupt politicians choose an x_i that maximizes

$$u^P\left(x_i, \, a_i - x_i; c_i\right) = a_i - x_i + g(x_i, \, \overline{x}).$$

Let $\rho = \sum_i(c_i - 1)/(N - 1)$ be the proportion of the other politicians that are predisposed to corruption from the perspective of one who is corrupt. The equilibrium condition for a symmetric interior x^* is then given by

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

The marginal benefit to allocating an additional unit of effort to non-corrupt activities is always 1, 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^*. Under mild restrictions, there is a unique interior solution in the politician stage, where corrupt politicians allocate more of their time to illicit activities when many others are corrupt:

PROPOSITION 1 Given the assumptions on g stated above and (1) $g_{11}(x, \rho x) < -\rho g_{12}(x, \rho x)$, (2) $g_1(0, 0) > 1$, and (3) $g_1(\underline{a}, \rho \underline{a}) < 1$, the politician stage has a unique interior equilibrium level of corruption x^* chosen by politicians with $c_i = 1$ characterized by Equation 1. In equilibrium:

i. the level of corrupt activity for those predisposed to corruption is increasing in ρ, and
ii. the relative payoff of a corrupt politician ($u^P(c_i = 1; x^*) - u^P(c_i = 0; x^*)$) is positive and increasing in ρ.

All proofs can be found in the Supplemental Appendix.

Condition (1) implies that the diminishing returns effect of higher corruption levels ($g_{11}(x, \rho x)$) is stronger (i.e., more negative) than the strategic complementarity effect ($-\rho g_{12}(x, \rho x)$). If this does not hold, there may be multiple equilibria in the politician stage by the standard logic of corruption

traps in the existing literature presented above. To bring attention to a mechanism through which high- and low-corruption equilibria emerge *for other reasons*, we assume this does *not* occur.

Conditions (2)–(3) prevent a corner solution where some corrupt politicians either allocate all or none of their effort to corrupt behavior, which reduces the number of cases to consider. Therefore, conditions (1)–(3) imply a unique interior corruption level. Because of the strategic complementarities, adding more corrupt politicians makes those predisposed to corruption allocate even more effort to corruption. Further, since $g_2 > 0$ (and by the envelope theorem), office-holding becomes more valuable for corrupt politicians as ρ increases, a convenient fact when considering endogenous entry.

Voter Stage

Now consider voters' decisions given politician behavior. A voter with a non-corrupt politician who allocates all of her effort to other activities attains utility a_i. A voter with a corrupt politician who allocates x^* to corrupt activities and $a_i - x^*$ to other activities attains $a_i - x^* + bg(x^*, \rho x^*)$. Define the *voter corruption differential* as the utility difference between having a corrupt versus non-corrupt politician:

$$d^V(\rho, x^*) = a_i - x^* + bg(x^*, \rho x^*) - a_i = bg(x^*, \rho x^*) - x^*.$$

Conveniently, this is not a function of a_i: voters always prefer more able politicians, but the ability does not affect the relative value of having a corrupt or clean representative.

The expected payoff to choosing a challenger is then $E[a] + qd^V(\rho, x^*)$, and so for a fixed ρ the voter prefers a replacement to an incumbent politician with (a_i^I, c_i^I) if

$$a_i^I \le E[a] + (q - c_i^I)d^V(\rho, x^*).$$

The voter's decision is always monotone in the ability of their politician, but the threshold in ability will be different for a corrupt and a non-corrupt politician. If voters in other districts become more or less tolerant of corruption, this can change the calculus of a voter through changing expectations about ρ.

Most relevant for characterizing the equilibrium is determining the shape of the $d^V(\rho, x^*)$ function. It will be particularly important to know when the voter corruption differential can be increasing in ρ, which will broadly imply strategic complementarities among voters in the degree to which they tolerate corruption. More precisely, where $d^V(\rho, x^*)$ is positive (i.e., the voter's utility from a corrupt politician is higher than from a clean politician), $d^V(\rho, x^*)$ increasing in ρ implies that voters prefer a corrupt politician and that the magnitude of this preference is increasing in the general prevalence of corruption. It is also possible that $d^V(\rho, x^*)$ is *negative* but increasing in ρ, which means that voters prefer a clean politician but become relatively more tolerant of corruption as it becomes more prevalent. In either case, the total derivative of the voter corruption differential (accounting for the effect on x^*) is

$$\frac{d}{d\rho}\left[d^V(\rho, x^*)\right] = -\frac{\partial x^*}{\partial \rho} + b\frac{d}{d\rho}[g(x^*, \rho x^*)].$$

Since the emphasis in characterizing the voter behavior will be on how their decisions affect ρ, for the remainder of the paper we suppress the x^* argument in the voter corruption differential, writing it $d^V(\rho)$. Thus, when we say this function is increasing (decreasing), we mean that as more (fewer) politicians are corrupt, the relative value of having a corrupt representative increases (decreases), accounting for the change in x^*.

There are two competing effects on the voter corruption differential as the prevalence of corruption increases. First, corrupt representatives allocate less effort to non-corrupt activities

(the $-\frac{\partial x^*}{\partial \rho}$ term), reducing the relative value to the voter of having a corrupt politician. However, the returns to corruption increase, and if citizens enjoy some of the fruits of corruption, they can benefit from this as well. Whether more corrupt politics makes having a corrupt representative more or less valuable for the voter depends on which of these two effects dominates.

When $b \leq 0$, 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, the voter corruption differential is decreasing. On the other hand, if $b = 1$, the citizens payoff from corruption is exactly the same as their politician's, and part (ii) of Proposition 1 is equivalent to stating that $d^V(\rho)$ is increasing. In between these extremes, 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 2 If $b > \tilde{b}$ for some $\tilde{b} \in (0, 1)$, the voter corruption differential is increasing in ρ.

The intuition behind Proposition 2 is that even if citizens only get a small benefit from their representatives' corrupt behavior, they may prefer this fraction of the proceeds to having a completely clean politician in a highly corrupt environment.[14]

When can this complementarity lead to multiple equilibria? As long as at least one voter replaces their politician, all the voters become uncertain about the number of corrupt politicians who will end up in office. So, the details of which politicians are retained in equilibrium depend on the distribution of ability among the corrupt and non-corrupt politicians. Still, a standard existence argument ensures that there will be some (potentially mixed strategy) PBE to the model. Rather than present all of the cases here, we highlight two sufficient conditions under which there are multiple equilibria with different levels of tolerance of corruption:

PROPOSITION 3 The model has at least one PBE. Further:

i. if $d^V(\rho)$ is locally increasing between two feasible corruption levels (formally, if there exists an integer n' such that $d^V(n'/(N-1)) > d^V((n'-1)/(N-1))$), then for some distributions of incumbent types the model has multiple equilibria where a different number of corrupt politicians are replaced, and[15]

ii. if $d^V(\rho)$ is increasing for all ρ, then for any number of corrupt incumbents there exist distributions of incumbent types that result in multiple equilibria where either all of the corrupt politicians are kept or all are replaced.

---

[14] Aravind Adiga, a Booker prize winning Indian writer, put this notion poignantly in the context of Indian politics:

> Even worse than the fake honest politician, is the genuinely honest one. In India, corruption is a form of patronage; a politician or bureaucrat who takes the bribe then has to let it cascade among a series of lesser bureaucrats and elected officials, who will make sure that the pet projects are completed. An honest politician has no goodies to toss around. This limits his effectiveness profoundly [...] Would you really want an honest man representing you in parliament? The neighbouring constituency may get that new cricket stadium you were supposed to ('Indians' Worst Fear: The Honest Politician', *The Guardian*, 30 July 2008, http://www.theguardian.com/commentisfree/2008/jul/30/india.scamsandfraud). In Section A2 of the Supplemental Appendix, we provide a number of additional examples consistent with this result.

[15] Recall N is the total number of districts (and therefore incumbents) so $N-1$ refers to the number of "other" districts (i.e., every district not including the voter's own district).

The intuition behind part (i) is that if $d^V(\rho)$ is increasing locally around a particular corruption level, then if one other constituency were to remove their corrupt politician, a voter who was previously slightly in favor of keeping their own corrupt incumbent would vote him out as well. So, if there are two politicians who are just above the threshold for retention, then there is one equilibrium where both are retained, and one where both are replaced. Part (ii) follows a similar logic: if voters are always more tolerant of corruption when others are corrupt, and all of the corrupt politicians are just above the ability threshold where the voter keeps them if every other voter with a corrupt politician does so, then there is one equilibrium where all of the corrupt politicians are retained and one where they are all removed from office.[16]

Even if an increasing voter corruption differential is plausible, the logic behind Proposition 3 requires a very particular configuration of politician types, which may not always be likely. To search for more robust results, we now consider the corruption traps arising from conditions in which the proportion of corrupt politicians are stable over time, independent of the particular distribution of ability at any given election.

LONG-RUN STABLE PREVALENCE OF CORRUPTION

We formalize the notion that the proportion of corrupt politicians is stable as follows:

DEFINITION   A long-run equilibrium to the voter model comprises a symmetric corruption choice x^* and ability thresholds $\hat{a}_0$ and $\hat{a}_1$ that jointly solve

$$x^* = \arg \max_x \, a_i - x + g(x, \rho \bar{x}), \tag{2}$$

$$\hat{a}_1 = E[a] - (1-q)d^V(\rho), \tag{3}$$

$$\hat{a}_0 = E[a] + q d^V(\rho), \tag{4}$$

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

Equations 2–4 are just the equilibrium conditions from the model in the previous section when there is a fixed and known proportion of corrupt politicians. Equation 5 implies that the proportion of corrupt politicians is equal to the conditional probability of being corrupt for politicians who are able enough to be retained (given their predisposition to corruption).

Finding an equilibrium can be reduced to finding a fixed point ρ^* which allows for a joint solution to 2–5. To find this fixed point, first define the probability of a randomly drawn politician having a high enough ability to be kept if clean (q_0^k) and corrupt (q_1^k) as

$$q_0^k(\rho) = (1-q)(1-F(E[a] + q d^V(\rho))),$$

$$q_1^k(\rho) = q(1-F(E[a] - (1-q)d^V(\rho))).$$

---

[16] As elaborated in the proof, this logic does not account for the fact that voters with *non-corrupt* politicians may also adjust their behavior when voters with corrupt politicians change their retention strategy, but this can be avoided by only looking at distributions where all of the clean politicians are of sufficiently high ability.

We can then collapse the equilibrium condition to finding a ρ which solves

$$\rho = \frac{q_1^k(\rho)}{q_0^k(\rho) + q_1^k(\rho)} \equiv \overline{q}(\rho). \qquad (6)$$

The $\overline{q}(\rho)$ function, which captures the relative likelihood of a politician being corrupt given she is good enough to be kept, is continuous and bounded by (0, 1), and so there must be at least one $\rho^* \in (0, 1)$ that satisfies Equation 6. To determine whether the intersection (and hence the stable level of corruption) is unique, differentiating the right-hand side with respect to ρ (and substituting the derivatives of the $q_i^k(\rho)$ terms) gives

$$\frac{\partial \overline{q}}{\partial \rho} = \frac{d}{d\rho}\left[d^V(\rho) \right] \frac{q(1-q)(q_0^k(\rho)f(E[a]-(1-q)d^V(\rho)) + q_1^k(\rho)f(E[a]+qd^V(\rho)))}{(q_0^k(\rho) + q_1^k(\rho))^2}. \qquad (7)$$

All terms other than the derivative of $d^V(\rho)$ are positive, so the sign of this expression is the same as the sign of $\frac{d}{d\rho}[d^V(\rho)]$. If $d^V(\rho)$ is decreasing, which as before implies that when more politicians are corrupt, the voter becomes worse-off with a corrupt representative, then $\frac{\partial \overline{q}}{\partial \rho}$ is negative, implying that Equation 6 must have a unique solution. However, if $d^V(\rho)$ is increasing and with a high enough slope, indicating that the relative benefit of having a corrupt politician increases quickly as the prevalence of corruption rises, then the slope of this equation can be greater than 1, raising 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:

PROPOSITION 4 There is at least one long-run stable equilibrium to the voter model, and if $\frac{d}{d\rho}[d^V(\rho)] \leq 0$, the equilibrium is unique. If $\frac{d}{d\rho}[d^V(\rho)] > 0$ there may be multiple equilibria to the voter model.

Figure 1 illustrates this result. In each panel, the solid curve is the right-hand side of Equation 6, and the dashed line is the left-hand side (i.e., the 45° line). Therefore, an intersection corresponds to a stable level of corruption. In the left panel, $d^V(\rho)$ is decreasing (i.e., having a corrupt politician makes the voter worse-off as corruption becomes more prevalent), and as a result there is a unique stable proportion of corrupt politicians. In the middle panel $d^V(\rho)$ is increasing, but at a slow rate, and so there is still a unique intersection and hence only one stable corruption level. In the right panel, $d^V(\rho)$ is increasing more sharply, and as a result there are two stable corruption levels, one where nearly all politicians are clean, and one where nearly all politicians are corrupt.

In sum, the possibility of multiple equilibria in the voter model provides an 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 tolerant 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 out of office, voters in district j may not be willing to do so either, if in a generally corrupt system the corrupt politicians are regarded as more effective. We discuss several examples in Section A2 of the Supplemental Appendix.
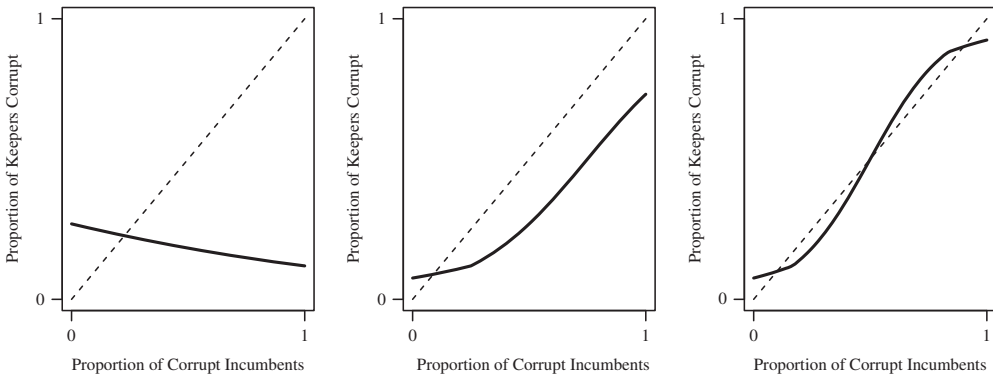
Fig. 1. *Examples with one (left and middle panels) and multiple (right panel) long-run stable corruption levels*
Note: In all three panels, $q = 1/2$ and ability follows an exponential distribution with rate parameter $\lambda = 1$. The panels vary in $d^V(\rho)$, which is decreasing in the right panel ($d^V(\rho) = -1 - \rho$), increasing but at a relatively slow rate in the middle panel ($d^V(\rho) = -3 + 4\rho$), and increasing at a faster rate in the right panel ($d^V(\rho) = -3 + 6\rho$).

VOTER BEHAVIOR WITH ENDOGENOUS ENTRY

Similar dynamics can lead to multiplicity of equilibria due to strategic complementarities among potential entrants to the political arena. Section A4 of the Supplemental Appendix contains a model where in each district a number of potential entrants decide whether or not to run for office, with the expectation that the winners will play the politician stage described above (i.e., there are no voters). Since the general logic of this model is close to our main model (and the results are of a similar flavor as those in Caselli and Morelli 2004), here we only briefly describe the results from this model and their implications for our main model.

The two key results from endogenizing entry decisions are as follows: (1) the proportion of entrants who are predisposed to corruption is increasing in the expectation of how many others will be corrupt, and (2) this complementarity can lead to a multiplicity of equilibria for similar reasons as in our main model. Therefore, corruption traps can arise individually at the level of incumbents, voters, and entrants.

However, we want to stress here how even if the complementarities in the politician, voter, and entry stages do not separately lead to a political corruption trap, the *interaction between the three* can lead to multiple equilibria and hence a corruption trap.

To show this, we now let the proportion of entrants that are corrupt be a function of ρ, that is, $q(\rho)$. As is predicted in Proposition A1 in the Supplemental Appendix, we assume that q is increasing, implying that more potential entrants are corrupt when more incumbents are corrupt.

The equilibrium condition is the same as before, with $q(\rho)$ replacing the constant ρ. Figure 2 shows how the examples from Figure 1 change when there is endogenous entry. As in Figure 1, in the left panel $d^V(\rho)$ is negative, meaning that voters are more apt to punish corruption when it is more prevalent. However, unlike in Figure 1, now the proportion of retained corrupt politicians is increasing in ρ because of the increase in corrupt candidates in the pool of entrants. Nonetheless, these complementarities between politicians and entrants are not strong enough to induce multiplicity, and so there is still only one stable—and relatively low—corruption level. In the middle panel (which previously had one stable corruption level despite strategic complementarities among both politicians and voters), the combination of strategic
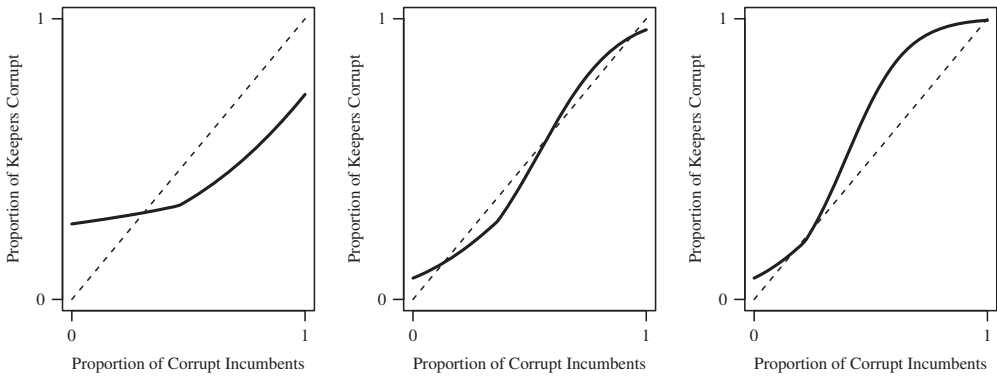
Fig. 2. The equilibrium condition for the examples in Figure 1 with endogenous entry
Note: In all three panels, $q(\rho) = 1/2 + 2\rho/5$, and the ability and $d^V(\rho)$ functions are the same as the corresponding panels in Figure 1.

complementarities of all three actors is sufficient to generate a corruption trap. In the right panel, as before, there are multiple stable corruption levels, but the proportion of corrupt incumbents who are retained in the high-corruption equilibrium is even higher than in Figure 1.

SUMMARY AND DISCUSSION

The overarching conclusion of the formal models we have 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, that is, the presence of a high-corruption as well as a low-corruption equilibrium.

Figure 3 summarizes these interactions. For each of the three levels, we can think of the actors in question interdependently behaving in a way that may lead to more corruption: (1) politicians choosing to engage in more corrupt behavior, (2) voters being more tolerant of or even approving of corrupt politicians, and (3) those predisposed to corruption being more apt to enter politics. Solid arrows indicate relationships which we have formalized in the previous sections or the Supplemental Appendix. The dashed arrows represent the relationships we informally discuss below.

In our models, these three levels primarily interact through the proportion of politicians predisposed toward corruption (ρ). As ρ increases, politicians choose a higher level of corruption (Proposition 1; solid arrow from ρ to politician), voters can be more tolerant of corruption (Equation 7; solid arrow from ρ to voter), and corrupt candidates are relatively more likely to enter than clean entrants (Proposition A1 in the Supplemental Appendix; 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 greater voter tolerance of 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^*) encourages corrupt entrants and induces greater voter lenience toward corruption. Finally, the insights from the Voter Behavior with Endogenous Entry section show
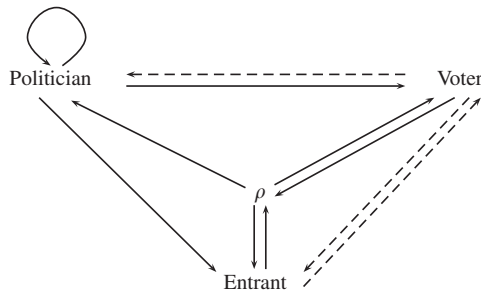
Fig. 3. Relationships between the three levels of analysis
Note: ρ is the proportion of politicians predisposed toward corruption. Solid arrows indicate relationships which we have formalized in the previous sections or the Supplemental Appendix. The dashed arrows represent the relationships we informally discuss in the text.

the combined effect of ρ on the behavior of entrants and voters (combining solid arrows from ρ to voter and entrant).[17]

There are other potential sources of positive feedback among these levels of analysis not modeled here that could make corruption traps even harder to escape. 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). Candidates 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, they will have no incentive to vote based on corruption (dashed line from entrant to voter).[18]

The most common implication of the existing accounts of strategically induced corruption traps is that a way to escape to a lower corruption equilibrium is to change the expectations among a fixed group of politicians or bureaucrats. In our model, if all politicians were up for re-election at once, the corruption trap may also be solved by coordinating expectations as in the previous literature. However, an important point we make is that this coordination requires that all (pivotal) 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. Moreover, if endogenous entry is also considered, then potential entrants to politics would need to simultaneously change their expectations about what other kinds of people will run for office as well.

We think this extent of coordination is unrealistic for several reasons. First, 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 voters to abandon the monitoring of politicians, discourage them from protesting against the regime, and even question the merits of democratic rule.

Second, given the importance of the prevalence of corruption among incumbents for the possibility of a corruption trap in our models, coordinating the expectations of all the actors (and voter patience) may not be enough, unless the entire political elite in power is replaced. While

---

[17] Most of the above-mentioned work on multiple corruption equilibria looks at the circular arrow from politician to politician, which we directly assume with $g_{12} \geq 0$.

[18] However, see Klašnja (2016) where voters may prefer challengers *even if* they are also corrupt, when incumbents' gains from corruption increase over time spent in office.

our equilibrium definition—as with any standard solution concept—assumes that such coordination of expectations occurs, inducing rapid shifts in expectations across a wide variety of actors is likely to be very difficult. As Golden 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 (2010, 80–1).

This suggests that the persistence of corruption could in part be a consequence of political and electoral institutions that influence how many politicians are up for re-election at any point in time. While rigorously formalizing this in our models is difficult, staggered election timing, multiple tiers of representation, and the many incumbency advantages may make it even more difficult to replace politicians *en masse*.[19] 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., Pellegrini and Gerlagh 2008). This is potentially a valuable area for future research.

REFERENCES

Alt, James E., and David Dreyer Lassen. 2003. 'The Political Economy of Institutions and Corruption in American States'. *Journal of Theoretical Politics* 15(3):341–65.

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. forthcoming. 'Accountability and Information in Elections'. *American Economic Journal: Microeconomics*. Available at https://www.aeaweb.org/articles?id=10.1257/mic.20150349&from=f

Banerjee, Abhijit, and Rohini Pande. 2009. 'Parochial Politics: Ethnic Preferences and Politician Corruption'. Unpublished manuscript, Massachusetts Institute of Technology.

Banfield, Edward C., and Laura Fasano Banfield. 1958. *The Moral Basis of a Backward Society*. New York: Free Press.

Barberá, Pablo, Pablo Fernández-Vázquez, and Gonzalo Rivero. 2016. 'Rooting Out Corruption or Rooting for Corruption? The Electoral Consequences of Corruption Scandals in Spain, 2003-2011'. *Political Science Research and Methods* 4(2):379–97.

Becker, Sascha, Katrin Boeckh, Christa Hainz, and Ludger Woessmann. 2016. 'The Empire is Dead, Long Live the Empire! Long-Run Persistence of Trust and Corruption in the Bureaucracy'. *The Economic Journal* 126(590):40–74.

Cadot, Olivier. 1987. 'Corruption as a Gamble'. *Journal of Public Economics* 33(2):223–44.

Caselli, Francesco, and Massimo Morelli. 2004. 'Bad Politicians'. *Journal of Public Economics* 88(3–4): 759–82.

Chong, Alberto, Ana L. de la O, Dean Karlan, and Leonard 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.

---

[19] The main challenge in incorporating these features in our models is that our solution concepts assume a long-run distribution where all politicians are up for election at the same time. Having staggered elections, for example, requires voters (and politicians and entrants) to make more complex forward-looking choices, which we leave to future research.

Corbacho, Ana, Daniel W. Gingerich, Virginia Oliveros, and Mauricio Ruiz-Vega. forthcoming. 'Corruption as a Self-Fulfilling Prophecy: Evidence from a Survey Experiment in Costa Rica'. *American Journal of Political Science*.

Damania, Richard, Per G. Fredriksson, and Muthukumara Mani. 2004. 'The Persistence of Corruption and Regulatory Compliance Failures: Theory and Evidence'. *Public Choice* 121(3–4):363–90.

de Figueiredo, Miguel F. P., F. Daniel Hidalgo, and Yuri Kasahara. 2012. 'When Do Voters Punish Corrupt Politicians? Experimental Evidence from Brazil'. University of California, Berkeley.

Ferraz, Claudio, and Frederico S. Finan. 2008. 'Exposing Corrupt Politicians: The Effects of Brazil's Publicly Released Audits on Electoral Outcomes'. *Quarterly Journal of Economics* 123(2): 703–45.

Gingerich, Daniel W. 2014. 'Brokered Politics in Brazil: An Empirical Analysis'. *Quarterly Journal of Political Science* 9(3):269–300.

Glaeser, Edward L., and Raven E. Saks. 2006. 'Corruption in America'. *Journal of Public Economics* 90(6):1053–072.

Golden, Miriam A. 2010. 'Puzzles of Political Corruption in Modern Advanced Democracies'. In Hideko Magara (ed.), *Democracy and Accountability: Globalized Political Responsibility*, 184–99. Tokyo, Japan: Fukosha.

Golden, Miriam A., and Lucio Picci. 2005. 'Proposal for a New Measure of Corruption, Illustrated with Italian Data'. *Economics & Politics* 17(1):37–75.

Hirano, Shigeo, and James M. Snyder. 2012. 'What Happens to Incumbents in Scandals?'. *Quarterly Journal of Political Science* 7:447–56.

Hollyer, James R., and Leonard Wantchekon. 2015. 'Corruption and Ideology in Autocracies'. *Journal of Law, Economics and Organization* 31(3):499–533.

Klašnja, Marko. Forthcoming. 'Uninformed Voters and Corrupt Incumbents'. *American Politics Research*.

Klašnja, Marko. 2016. 'Increasing Rents and Incumbency Disadvantage'. *Journal of Theoretical Politics* 28(2):225–65.

Klašnja, Marko, and Joshua A. Tucker. 2013. 'The Economy, Corruption, and the Vote: Evidence from Experiments in Sweden and Moldova'. *Electoral Studies* 32(3):536–43.

Klašnja, Marko, Joshua A. Tucker, and Kevin Deegan-Krause. 2016. 'Pocketbook vs. Sociotropic Corruption Voting'. *British Journal of Political Science* 46(1):67–94.

Lui, Francis T. 1986. 'A Dynamic Model of Corruption Deterrence'. *Journal of Public Economics* 31(2):215–36.

Mainwaring, Scott P. 1999. *Rethinking Party Systems in the Third Wave of Democratization: The Case of Brazil*. Stanford: Stanford University Press.

Manzetti, Luigi, and Carole J. Wilson. 2007. 'Why Do Corrupt Governments Maintain Public Support?'. *Comparative Political Studies* 40(8):949–70.

Mauro, Paolo. 1995. 'Corruption and Growth'. *The Quarterly Journal of Economics* 110(3): 681–712.

Mauro, Paolo. 2004. 'The Persistence of Corruption and Slow Economic Growth'. *IMF Staff Papers* 51(1):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):478–90.

Menes, Rebecca. 2006. 'Limiting the Reach of the Grabbing Hand: Graft and Growth in American Cities, 1880 to 1930'. In Edward L. Glaeser and Claudia Goldin (ed.), *Corruption and Reform: Lessons from America's Economic History*, 63–93. Chicago, IL: University of Chicago Press.

Mishra, Ajit. 2005. 'Persistence of Corruption: Some Theoretical Perspectives'. *World Development* 34(2):349–58.

Pellegrini, Lorenzo, and Reyer Gerlagh. 2008. 'Causes of Corruption: A Survey of Cross-National Analyses and Extended Results'. *Economics of Governance* 9(3):245–63.

Quah, Jon S. T. 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 Ofer Feldman (ed.), *Political Psychology in Japan: Behind the Nails Which Sometimes Stick Out (and Get Hammered Down)*, 131–48. Commack, NY: Nova Science.

Reinikka, Ritva, and Jakob Svensson. 2004. 'Local Capture: Evidence from a Central Government Transfer Program in Uganda'. *Quarterly Journal of Economics* 119(2):679–705.

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.

Transparency International India. 2005. 'India Corruption Study'. Transparency International India and Center for Media Studies, New Delhi, India.

Treisman, Daniel. 2007. 'What Have We Learned About the Causes of Corruption from Ten Years of Cross-National Empirical Research?'. *Annual Review of Political Science* 10:211–44.

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):226–39.

Zechmeister, Elizabeth J., and Daniel Zizumbo-Colunga. 2013. 'The Varying Political Toll of Concerns About Corruption in Good Versus Bad Economic Times'. *Comparative Political Studies* 46(10):1190–218.