Preliminary draft. Nov 30, 2017

Fighting Corruption By Altering Equilibrium In An Assurance Game*

Avinash Dixit, Princeton University

INTRODUCTION

Game theory gives us two ways to think about corruption and how to fight it. One is to view corruption as the bad equilibrium of a Prisoners' Dilemma. When firms compete for a government contract or license, each can increase its chances by bribing a politician or bureaucrat with power over the award. But when they all give bribes, their chances may be no better than if they had all agreed to refrain, and in the aggregate they are giving up money. The business community can resolve this dilemma if it can establish a system of norms and sanctions on firms found to be corrupt. This is modeled theoretically by Kingston (2008) and Dixit (2015 a, b). However, successful resolution of the dilemma is difficult in practice, requiring a sufficiently long-run relationship in a stable community, and good information and communication about potential and actual cheating.

A more encouraging framework is the Assurance game. In this, if enough others refrain from corruption, it is better for any one individual not to be corrupt also, but if enough others are corrupt, it is in each individual's own interest to be likewise. Thus the game has two equilibria, one where everyone is corrupt and the other where everyone is clean. To achieve or sustain the good equilibrium, we have to create a convergence of expectations, or common knowledge, that others will be clean; then individual choices will make the expectation self-fulfilling. Such an equilibrium will also be robust to gradual changes in the composition of the business community; when the predominant expectation is one of good behavior, newcomers will find it in their own interest to conform, and a few departures will not alter the beliefs of those who remain.

Although it is not easy to create common knowledge to achieve convergence of expectations for attaining and sustaining the equilibrium without corruption, that is usually easier than establishing a system of norms of conduct and detection and punishment of misconduct that is required for resolving a prisoner's dilemma.[1] Therefore it is useful to know when the "corruption game" is likely to be one of assurance rather than one of prisoners' dilemma.

---

* I thank Sudipto Mundle, Saurabh Mukherjea, and participants in a World Bank conference on Corruption and Corporate Governance, Mumbai, March 30-31, 2016 for discussions that got us started on this idea, and especially Kaushik Basu for valuable discussions and Karla Hoff for useful comments.

[1] See Chwe (2001) for a study of how societies create common knowledge.

Such a game can arise for many reasons. Most simply, in some societies corruption is the way of life. If everyone is complicit in taking or giving bribes, or deriving other private benefit from one's official position, no one feels any guilt or shame about it. Even if laws exist on the books, no one will report corruption to a law enforcement agent; anyway these agents can themselves be bribed. In other societies corruption is widely regarded as evil. Corrupt acts are likely to be reported or detected quickly and punished. The corrupt also feel guilt and shame; these are often better deterrents than formal sanctions. Thus a whole culture can be either corrupt or clean.[2]

Another reason is that many workers and consumers have an ethical preference to work for, and buy from, firms with reputations for good governance and a clean public image. Therefore such firms can get their pick of more productive workers, and can sell their products or services at premium prices. In the U.S. Frank (2004, chapter 5) found that college graduates were willing to accept significant pay differentials in exchange for a more ethical work environment. Controlling for qualifications, average pay was 32 percent less in non-profit sectors than that in government employment, which in turn was 21 percent less than that in the for-profit sector. And law firms had to offer substantial opportunities for pro bono and public interest work to attract the best graduates of law schools. I have heard similar preferences expressed by bright young people in India. Turning to outcomes, in the U.S. Patagonia's environmental reputation, and Whole Foods' for selling organic produce and meat from more humanely treated animals, have been conducive to higher profitability. In India, research by Ambit Capital shows that firms with better quality of governance and accounting yield higher investment returns, and politically well-connected firms have not done well in recent years (Mukherjea 2016, see especially Exhibit 134 on p. 337 and Exhibit 136 on p.342).

All these forces are likely to be stronger when there is less prevailing corruption. If most firms are corrupt, the smart young workers and ethical customers have few opportunities to act on their preferences, and in such an environment any one firm does not benefit by being clean. But if most firms are clean, any corrupt firm is more likely to fare poorly in competition with the clean ones. Thus both the all-corrupt and none-corrupt configurations can be equilibria.

In this paper I construct a simple model to help pin down more rigorously the conditions that create the multiple equilibria, and to understand how a shift to the corruption-free equilibrium can be brought about.

---

[2] Such multiple equilibria of expectations and actions have been observed and studied in other social contexts, for example work vs. welfare; see Lindbeck, Nyberg and Weibull (1999).

THE MODEL

Some key ideas emerge from the above discussion to build into the model: (1) Firms can be, and can choose to be, either corrupt or clean. (2) Workers are differentiated by ability. (3) Workers prefer to work for clean firms. (3) Firms prefer more able workers. (4) Consumers prefer to buy from clean firms. (5) As usual, these preferences must be traded off against other considerations such as prices and wages. (6) Corrupt firms have an advantage in getting government business. (7) The relative strength of these advantages can depend on the proportions of clean and corrupt firms in the economy (which can itself be endogenous).

I construct a model that incorporates these ideas in a very simple way, namely by positing parameters to represent the strength of the various preferences and advantages. This allows me to exhibit possible outcomes easily, and highlights the rules of the game – information and order of moves in the interaction – that determine which outcome emerges as an equilibrium. That in turn suggests policies to design the game so as to achieve the desired corruption-free outcome.

The model has a continuum of firms over the interval [0,1]. All firms are ex ante identical, but each can choose to be clean or corrupt. I denote by x the fraction of corrupt firms, so $(1-x)$ is the fraction of clean firms. A corresponding continuum of workers extends over the interval [0,1]; worker θ has productivity θ. Each infinitesimal firm hires one infinitesimal worker.

I assume that the premium consumers are willing to pay for buying from a clean firm has functional form $(1+\pi x)$, where $\pi > 0$ is an exogenous parameter. The premium should be an increasing function of x, because when there are more corrupt firms, output from clean firms is scarcer and therefore more valued. The linear form is just for simplicity. However, a clean firm has to navigate its way in a corrupt environment. I assume that this takes up a fraction γx of its resources, leaving only $(1-\gamma x)$ for actual production. This is because when more firms are corrupt, a clean firm has to work harder to succeed against the prevailing culture as ll as against the many firms who have bought an inside track for government business. Again the linear form is only for simplicity, and $\gamma < 1$ is an exogenous parameter. The upper bound on possible values of γ ensures that even when x is close to 1 a clean firm can have positive output; otherwise we would have a tedious and unenlightening taxonomy of cases.

A corrupt firm gets a premium in its government contracts from cost padding or quality cutting. The premium will be lower when there are more corrupt firms because they are all chasing the same prize and everyone must offer a larger bribe or accept a smaller reward. Again assuming a simple linear form, write the premium as $[\lambda + \kappa(1-x)]$ where λ and κ are positive exogenous parameters. (λ need not be greater than 1; competition for rent in public contracts when x is close to 1 may be too destructive.) This premium accrues if the firm is not caught. Exposure and detection are less likely if many other firms are corrupt, so let the probability of detection be $\mu(1-x)$. The probability of getting away with corruption and receiving the premium is $[1-\mu(1-x)]$. If a corrupt firm is caught, it will lose all profit from the contract, and in addition pay a fine ϕ.

Two information conditions and orders of moves are conceivable, and the outcome is sensitive to that. In one, first firms choose whether to be clean or corrupt. This becomes public information through some process of independent and objective assessment; this could be analogous to the Michelin rating system for restaurants. Then the process of matching workers to firms takes place; thus workers in making their job acceptance decisions know the status of the firm making the offer. Finally consumers make their choices; they also know the type of every firm and that governs the price they are willing to pay. In the other game, the order is reversed. Firms hire workers; each firm gets to know the productivity of its worker; then it can decide whether to be clean or corrupt. When workers choose from job offers, they do not know the nature of the firm they will work for.[3]

Firms choose mode of conduct first

In this case, each firm calculates the consequences being corrupt and being clean, foreseeing the various possibilities and premia or penalties it may gather in the subsequent game, and chooses the mode of conduct with the higher expected payoff. If the firm chooses to be clean, it will have its pick among the best workers, but it does not know in advance the productivity of the actual worker it will get. If x firms are corrupt, a clean firm will therefore get a productivity draw in the interval $[x, 1]$, for an expected productivity of $(x+1)/2$. I assume, again for simplicity, that its payoff is proportional to the productivity of its worker; this is a reduced form of the subgame of production and wage payment not spelled out in the model. Recognizing the multiplicative factors for the premium the consumers are willing to pay and the fraction of resources the clean firm has to spend coping in a corrupt environment, the expected payoff is

$$C(x) = (1+\pi x)(1-\gamma x)\frac{1+x}{2}.$$

A corrupt firm will have to pick a worker with productivity in the interval $[0, x]$ for an expected productivity $x/2$. Factoring in the premium in government contracts and the probability of being caught, the expected payoff is

$$D(x) = [\lambda + \kappa(1-x)][1-\mu(1-x)]\frac{x}{2} - \mu(1-x)\phi.$$

Now $C(0) = 0.5 > -\mu\phi = D(0)$. Therefore when $x = 0$ (no other firms are corrupt), it is better for any one firm to stay clean also; all-clean is always an equilibrium in this case of information and order of moves. All-corrupt can be an

[3] Think of the law graduate in Grisham (1991).

equilibrium if $C(1) < D(1)$, or $(1+\pi)(1-\gamma) < 0.5\,\lambda$; this depends on values of the parameters.

All equilibria can be found by graphing the functions $C(x)$ and $D(x)$ over $x \in [0,1]$. As well as the "pure" end-point equilibria defined by the inequalities above, there can be interior equilibria with a mix of corrupt and clean firms, defined by solutions to $C(x) = D(x)$. For general functional forms there can be multiple equilibria of this kind. For the linear forms above, there can be at most three. If we postulate a conventional adjustment process where x increases if $D(x) > C(x)$ and decreases if $C(x) > D(x)$, equilibria will be alternately stable and unstable, starting with the all-clean equilibrium which is stable.

Figure 1 shows a numerical example. The parameter values are chosen purely for purpose of illustrating the ideas:

Clean: $\pi = 0.1,\ \gamma = 0.7$
Corrupt: λ=1.2, κ=2, μ=0.2, ϕ=0.2

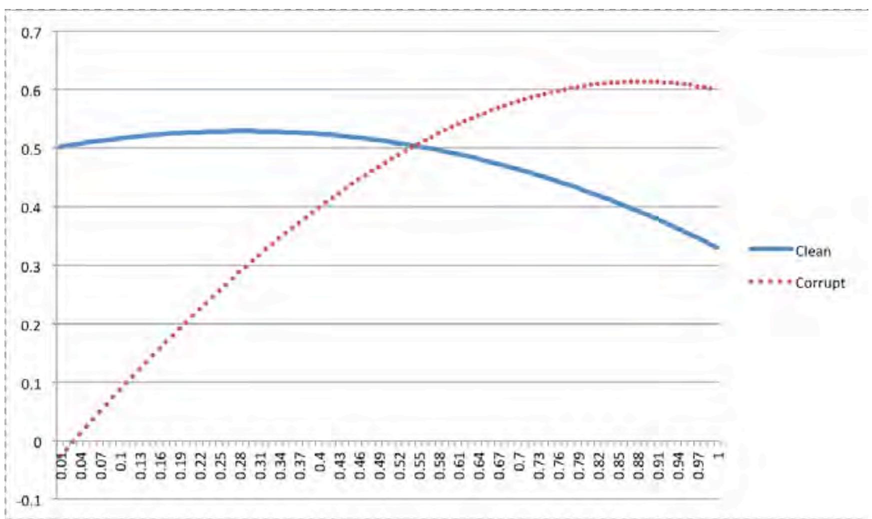

Fig. 1: Equilibria when firms choose mode of conduct first

There are three equilibria, the two end-point ones and an interior one where about 56% of the firms are corrupt. The all-clean and all-corrupt ones are stable and the interior one is unstable. If the economy starts near the all-corrupt point and we attempt reforms, initially we will encounter resistance as each firm's benefit from being corrupt is large. The policy measures will have to include better detection and punishment to overcome this temptation. But if we persist and succeed in reducing the proportion of corrupt firms below 55%, the balance of advantage will shift toward being clean, and a virtuous circle will begin, where fewer and fewer firms are corrupt and each one benefits more and more from becoming clean.

With more general functional forms, there may be multiple equilibria with some corruption, and policy will have to push the economy leftward past the last of them before the momentum leading to the all-clean equilibrium takes over.

Note that each of the payoff curves has an interior peak. A clean firm benefits from the existence of some corrupt firms, because it gets to pick workers from a higher productivity interval and enjoys a higher premium from consumers. A corrupt firm benefits from the existence of some clean firms, because it then has fewer corrupt competitors and therefore enjoys a higher premium in dealings with the government. However, these interior peaks cannot be sustained as equilibria.

The numbers are deliberately chosen to illustrate one other possibility. Each firm's payoff in the all-corrupt equilibrium (0.6) is higher than that in the all-clean equilibrium (0.5). Thus reform is not in the interests of the business community; the premium from being corrupt in government contracting is too high. If society wants a clean equilibrium for other reasons (such as consumers' and taxpayers' interests, or harmful effects of corruption on future investment and growth, or a general benefit from having ethical behavior in society) policy will have to proceed against opposition from business.

This model is robust to many variations of the assumptions so long as the order of moves is maintained. Here are two examples.

First, consumer preference for buying from non-corrupt firms may be weaker if the whole culture is more corrupt. To capture this, change the π in the premium specification to $\pi(1-x)$, so the preference gets smaller as the fraction of corrupt firms increases. This turns out to make very little difference to Figure 1, except that the interior equilibrium shifts slightly to the left, from 55% to 53% of firms.

Second, suppose the preference to work for non-corrupt firms also vanishes in a very corrupt economy. If this preference did not exist, each type of firm would get a random draw of workers with expected productivity equal to ½. So let us suppose that the actual is a weighted average between this and the form stipulated above, with weights x and $(1-x)$ respectively. Then the expected productivity becomes

$$\text{Clean: } x\,\frac{1}{2} + (1-x)\,\frac{1+x}{2} = \frac{1+x-x^2}{2}$$

$$\text{Corrupt: } x\,\frac{1}{2} + (1-x)\,\frac{x}{2} = \frac{x(2-x)}{2}$$

Once again the general shape of the curves is the same as in Figure 1, but now we see a much bigger shift of the interior equilibrium, namely from 55% to 26%. So active policy must do more work before the self-sustaining process of convergence to a clean equilibrium kicks in.

Firms choose mode of conduct second

In this case, workers are matched with firms first. All firms are ex ante identical, so we can assume a random matching without loss of generality. Once a firm hires a worker, it gets to know the worker's productivity. Then it chooses whether to be clean or corrupt. To find the equilibria of this second-stage game, consider the best response of the firm with worker of productivity θ when x other firms are corrupt.

Using the expressions for the premia and penalties specified for the first case, the expression for this firm's expected payoff if it chooses the clean mode is

$$C(x,\theta) = (1+\pi x)(1-\gamma x)\theta,$$

and if it chooses the corrupt mode, the expected payoff is

$$D(x,\theta) = [\lambda + \kappa(1-x)][1-\mu(1-x)]\theta - \mu(1-x)\phi.$$

The firm will choose to be corrupt if $D(x,\theta) > C(x,\theta)$, or

$$\{\, [\lambda + \kappa(1-x)]\,[1-\mu(1-x)] - (1+\pi x)(1-\gamma x) \,\}\,\theta > \mu\phi(1-x).$$

Abbreviate the expression in the brackets on the left hand side as $F(x)$. Then $F(x)\theta > \mu\phi(1-x)$ defines a region in (x,θ) space; of course only the part in the unit square $0 \leq x \leq 1$, $0 \leq \theta \leq 1$ is meaningful.

If $F(x) < 0$, then the inequality cannot be satisfied for any $\theta \in [0,1]$, so all firms will choose to be clean. If $F(x) > 0$, define $\Theta(x) = \mu\phi(1-x)/F(x)$. If this is in (0,1), then the firms whose workers have productivities in the range $[\Theta(x), 1]$ will choose to be corrupt. If $x < 1 - \Theta(x)$, that is, $x + \Theta(x) < 1$, it means fewer firms are currently corrupt than want to be corrupt, so x will increase. If $x + \Theta(x) > 1$, then x will decrease. This process can lead to one of the extreme equilibria with all firms clean ($x = 0$) or all firms corrupt ($x = 1$), or an interior equilibrium defined by $x + \Theta(x) = 1$.

It is interesting to note that in a mixed equilibrium the most productive range of firms will be the ones who choose to be corrupt. The probability of detection μ and the fine ϕ are the same for all firms, and act as stronger deterrents to firms in the lower range of productivity. If policy-makers think this undesirable, they will have to enact and enforce laws so that detection risks or fines increase more than proportionately with productivity!

This case leads to many possible outcomes, and I illustrate just two. For the parameter numbers that were used for Figure 1, we find $F(x) > 0$ and $x + \Theta(x) < 1$

for all $x \in [0,1)$, so the all-corrupt equilibrium is the only possibility. For a different set of parameter values

$$\text{Clean: } \pi = 0.1,\ \gamma = 0.7$$
$$\text{Corrupt: } \lambda=1.2,\ \kappa=0.3,\ \mu=0.5,\ \phi=0.5$$

we have the following ranges:

$0 \leq x \leq 0.21$	$F(x) < 0$
$0.22 \leq x \leq 0.42$	$F(x) > 0$ and $x + \Theta(x) > 1$
$0.43 \leq x < 1$	$F(x) > 0$ and $x + \Theta(x) < 1$

If initially $x \geq 0.43$, then x will continue to increase, converging to the all-corrupt outcome. If initially $x \leq 0.42$, then x will continue to decrease, and once it crosses 0.22 all firms will prefer to be clean, leading to the all-clean outcome.

COMPARISONS AND POLICY IMPLICATIONS

The theory and numerical work suggest that an ethical preference to work for non-corrupt firms is key for the existence of a non-corrupt equilibrium and the ease of attaining it. For this to be possible, workers need to know the type of firm when they consider offers of employment. That in turn requires a game structure where firms choose their mode of conduct first, and this choice needs to be publicly known. With these conditions, all-clean is always an equilibrium. If the economy starts near an all-corrupt status-quo, reforms will have to push against resistance for a while, but eventually a self-sustaining process of move toward the all-clean situation will take over and gather momentum. With the opposite order of moves, all-clean may not be an equilibrium; in fact all-corrupt may be the only equilibrium.

How can policy or deliberate action select the first of these two orders of moves? By monitoring and publicizing the conduct of firms, helping the clean ones acquire and publicize a good reputation with the public, and naming and shaming the corrupt ones, if not explicitly then by their omission from the list of the good. Think of this as a rating system, like the credit score system for individuals in the U.S. and other countries, or the Michelin star system for restaurants.

Once such a system is in place and acquires its own reputation, firms will find that being cleaner enables them to attract and retain more productive workers, and improve their attraction to customers. Firms will then strive to be rated highly, just as restaurants work hard to get and maintain Michelin stars, and consumers struggle to improve their credit scores at Equifax and other agencies.

Of course the proviso that the rating system needs its own reputation is important. It should be based on publicly stated and objective criteria, have its own adequate and competent staff to research the companies, and an oversight board of respected citizens such as retired high-level civil servants and judges, respected senior people from the business community and academia. The rating system should also have some protection from unfounded attempts to malign its reputation that corrupt firms will surely make. So long as it can show that the stated criteria were followed, courts should quickly dismiss with prejudice any frivolous or nuisance lawsuits from poorly rated firms.

Finally, note that if a corrupt outcome is the equilibrium of a prisoners' dilemma game as in Kingston (2008) and Dixit (2015 a, b), shifting to a good outcome using a community-based system of norms and sanctions requires acquisition and public dissemination of accurate information about any cheating. These requirements are very similar to those in the assurance game above. Thus the importance of creating a rating institution like the one outlined above is not specific to any one type of the strategic game of corruption among business firms.

REFERENCES

Chwe, Michael. 2001. *Rational Ritual: Culture, Coordination and Common Knowledge.* Princeton, NJ: Princeton University Press.

Dixit, Avinash. 2015 a. "Corruption: Supply-side and demand-side solutions." In *Development in India: Micro and Macro Perspectives,* eds. S. Mahendra Dev and P.G. Babu, Springer, pp. 57-68.

Dixit, Avinash. 2015b. "How Business Community Institutions Can Help Fight Corruption." *World Bank Economic Review,* 29 (suppl 1): S25-S47.

Frank, Robert H. 2004. *What Price the Moral High Ground: Ethical Dilemmas in Competitive Environments.* Princeton, NJ: Princeton University Press.

Grisham, John. 1991. *The Firm*. New York: Random House.

Kingston, Christopher. 2008. "Social Structure and Cultures of Corruption." *Journal of Economic Behavior and Organization* 67(1): 90-102.

Lindbeck, Assar, Sten Nynerg and Jörgen W. Weibull. 1999. Social norms and economic incentives in the welfare state. *Quarterly Journal of Economics* 114(1): 1-35.

Mukherjea, Saurabh. 2016. *Unusual Billionaires.* New Delhi: Penguin Random House India.