DEWEY

working paper
department
of economics

A THEORY OF COLLECTIVE REPUTATIONS
with applications to the persistence of corruption and to firm quality

Jean Tirole

No. 93-13 June 1993

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

A THEORY OF COLLECTIVE REPUTATIONS
with applications to the persistence of corruption and to firm quality

Jean Tirole

No. 93-13 June 1993

8+0 7973

A Theory of Collective Reputations
with applications to the persistence of corruption and to firm quality

Jean Tirole*

June 7, 1993

*Institut d'Economie Industrielle, Toulouse, MIT and Ceras, Paris. This paper was prepared under a cooperative agreement between the Institute for Policy Reform (IPR) and Agency for International Development (USAID), Cooperative Agreement No.PDC#0095-A-00-1126-00. Views expressed in this paper are those of the author and not necessarily those of IPR or USAID. The author is grateful to Jacques Crémer, Drew Fudenberg, David Martimort, Patrick Rey, Paul Seabright, Lars Stole, and Barry Weingast for helpful comments.

Abstract

The topic of this paper, the phenomenon of group reputation, has been neglected in economic theory despite its importance for the social sciences. Because a group's reputation is only as good as that of its members, we focus on the interaction between individual incentives and collective reputation. Stereotypes are viewed as stemming from history dependence rather than from specific cultural or racial traits.

JEL n: 026,511

Keywords: Stereotypes, history dependence, corruption, firm quality

1 Collective reputations.

Collective reputations play an important role in economics and the social sciences. Countries, ethnic, racial or religious groups are known to be hard-working, honest, corrupt, hospitable or belligerent. Some firms enjoy substantial rents from their reputations for producing high-quality products. Some departments are reported to treat their faculty or students fairly. It seems futile to ascribe such stereotypes to intrinsic features of the populations. Rather, we view collective reputations as the outcome of group history. By definition, the collective reputation of a group reflects the average past behavior of its members. This implies that:

a) *A group's reputation is only as good as that of its members.* Each member is characterized by traits such as talent, diligence or honesty. Past individual behavior conveys information about these traits and generates individual reputations.

b) By contrast with group belonging, *individual past behavior is imperfectly observed.* If past individual behavior were fully unobserved, members of the group would have no incentive to sustain the reputation of the group. Conversely, the collective reputation would play no role if individual behaviors were perfectly observed. Imperfect observability of individual behavior thus underlies the phenomenon of collective reputation.

c) Hence, *the past behavior of the member's group is used to predict his individual behavior.* Each member's welfare and incentives are thus affected by the group's reputation.

d) And therefore, *the behavior of new members of a group depends on the past behavior of their elders.*

Despite their pervasiveness, collective reputations have not, to the best of my knowledge, been formally investigated. This paper is an attempt at filling this gap. Its key feature is the interplay between individual and collective reputations. We offer two variants of the same model. In the first, a member's individual reputation is imperfectly observed by his potential "trading partner" (who may or may not belong to the group). The incentive to sustain an individual reputation stems from the member's fear of *direct exclusion* by the trading partner. By this we mean that the individual reputation may induce the trading partner to behave in a way undesirable for the member (e.g. by not trading), while the belonging to the group is not affected. We apply the direct exclusion variant to the issue of corruption to explain why corruption is a societal phenomenon and why it tends to persist. (See section 2 for an overview of the application to corruption).

This direct exclusion variant does not seem appropriate when the trading partner has a low probability of knowing the member's past behavior. The buyer of a car does not even know the names of the worker, foreman and engineer who built the car. Yet brand image is an important factor in the car market. The reason why the car manufacturer's employee has an incentive to maintain quality is the fear of *delegated or internal exclusion*: It may be in the interest of the firm to fire employees who have demonstrated undesirable traits.

In this case, the worker is not concerned by the possibility that his supplying poor quality will have a significant impact on buyers' demand for his work, but rather by the possibility of being fired. So in the delegated exclusion variant the trading partner (the buyer) reacts to the collective reputation

and the group (the firm) excludes on the basis of individual reputation, while in the direct exclusion variant the trading partner reacts to both collective and individual reputations and the group does not necessarily control membership. Yet the two variants are formally very similar because imperfect observability of individual behavior plays the same central role[1].

Modeling in industrial organization has viewed the firm as a black box to study its quality choices and has ignored the question of why workers have individual incentives to defend the firm's collective reputation[2]. Our work offers one insight into this black box.

Before developing the formal analysis, it is worth noting that a *collective reputation is neither a convention nor a norm*[3]. A *convention* refers to the coordination on a particular Nash equilibrium in a situation of multiple Nash equilibria. Many models in economics have multiple equilibria, for example coordination games, repeated games, macroeconomic models with aggregate demand externalities or models of racial and sexual discrimination (which we will later discuss). The interpretation of a convention as the selection of a particular equilibrium is stressed for example in Cole et al. (1992), Kandori (1992), and Seabright (1992). Kreps (1990) compares corporate culture to

[1]If workers' individual behavior were not observed within the firm, there would be no incentive to sustain individual reputations, and firms could not build reputations for high quality. If workers' individual behavior were perfectly observed within the firm, workers would have no incentive to sabotage the firm's reputation, at least under the classic conditions under which individual reputation is sustainable.

[2]The works of Crémer (1986) and Kreps (1990) are exceptions to this rule. Among other things, our work departs from theirs in that the behavior of a firm's employees is truly history dependent. Crémer and Kreps develop repeated-game models of organizations with overlapping generations of workers in which future generations may punish current ones if these do not behave well. Kotlikoff et al (1988), in a similar spirit, show that in an overlapping generations framework, the young may refrain from taxing the olds' capital by fear that the next generation would tax their capital. In these models, the set of equilibria at each point of time is history-independent.

[3]We use the sociological definition of convention and norm (see, e.g., Elster (1989), Sugden (1989) and Ullman-Margalit (1977)). Economists often seem to call a norm what a sociologist would label a convention.

a convention, in that corporate culture in a firm is meant to communicate to its employees the (focal) behavior that they are expected to follow. The sociological notion of a *norm*, unlike that of a convention, seems to stress psychological factors at the expense of methodological individualism. That individuals are eager to be approved by others generates norms of etiquette or consumption norms[4]. Unlike convention behaviors, norm behavior need not be in one's self-interest (at least narrowly defined). Alternatively individuals want to be approved by themselves ("I will not litter in the park even if noone will see me"). Thus neither conventions nor norms refer to the interplay between individual and collective reputations described above.

2 An overview of the argument: The case of corruption.

It is commonplace to observe that corruption is a central issue faced by development policies. It affects all aspects of public life: enforcement of laws, collection of taxes and tariffs, management of public contracts, housing subsidies, police work, credit, building and business permits, and so forth. In many countries, corruption has become institutionalized. There are agreed scales of charges for public services, and markets for public offices are well developed (for instance, superintending engineers' posts on the coastal deltas in India cost up to 40 times the annual salary for that rank, for an expected duration on the job of two years [5]).

The large disparity in corruption patterns across countries and across epochs is puzzling. An African country (e.g., Zaire) will be completely corrupt while another (e.g., Kenya) will have kept a clean civil service. Most

[4]Such norms might also be rationalized by the economic theory of wasteful signaling, but the emphasis is rather on the eagerness to be approved.

[5]*The Economist*, May 4, 1991, India survey, pages 15-18.

LDCs are affected by the plague to a much larger extent than developed countries, the recent growth of corruption in France or the Japanese and Italian scandals notwithstanding. Corruption was pervasive in England and several other European countries two centuries ago and has much subsided since. While factors related to the social fabric such as a family-centered ethos or the existence of tightly-knit clans certainly play a role, it seems futile to ascribe corruption to particular cultures or racial groups. Rather, it is important to understand the historical reasons and institutional factors that make some societies more corrupt than others. Only then will we be able to have a good grip on how to tackle the issue[6].

We argue that *history matters*. In particular, a society in which corruption develops unfettered today is more likely to be corrupt tomorrow than an identical society that takes a better start. This conclusion will not come as a surprise to those who have observed the persistence of corrupt practices and witnessed the many unsuccessful attempts to eradicate them. It is nevertheless important to identify the causes of hysteresis.

Our purpose is not to investigate the many facets of corruption. Rather we want to build an abstract framework that can be used to study the dynamics of collective reputation in a wide range of circumstances. We first develop a stylized model in which economic activity requires trust between contracting parties that they will not engage in corrupt practices. The parties make inferences about the honesty of their potential trading partners on the basis of an imperfect observation of their track record, namely whether

[6]The topic of many articles and books [see, e.g., Gould (1980), Hager (1973), Klitgaard (1986, 1989, 1991), Myrdal (1970), Lui (1986), Noonan (1984), Rose-Ackerman (1978), and Theobald (1990)], corruption hasn't yet attracted much attention from economic theorists, and therefore its analysis lacks adequate foundations based on information economics and game theory (note that Robert Klitgaard's fascinating books on the topic constantly point at the relevance of information economics). A proper understanding of the phenomenon seems to require an examination of its microstructure.

they engaged in corrupt practices in the past. Because one's real track record is partially observed by potential trading partners, individuals may have an incentive to develop or maintain a reputation for honesty. On the other hand, because this track record is not perfectly observed, inferences are also based on the society's behavior as a whole. This combination of *individual and collective stigmas* is what in our model may create a scope for multiple equilibria. If society as a whole is honest, people are willing to trust individuals whom they have not heard to be corrupt. And because society will trust them in the future if they keep a clean record, individuals are willing to invest in a good reputation. In contrast, in a corrupt society, the general suspicion makes honesty a low-yield investment, and distrust is indeed justified (section 3).

We then study the issue of persistence of corruption by analyzing the sensitivity of equilibrium to initial conditions. In the benchmark, the economy is in a stationary equilibrium and has a low level of corruption. We then slightly perturb the economy by assuming that at the initial date (date 0), there is a one-shot increase in the gain to being corrupt (or a relaxation in the enforcement of anticorruption laws). The agents alive at date 0 engage in the corrupt activity at that date. The economy is otherwise unchanged at date $1, 2, \cdots$. We then ask whether the temporary increase in corruption necessarily has lasting effects, or whether the economy is able to go back to the low steady state level of corruption. Interestingly, we find that the economy must remain corrupt not only in the short run, but also in the long run. Our analysis unveils two effects: First, the agents who were alive at date 0 have smeared their reputation. In our model, they have more incentives to engage in corrupt activities than if they had always behaved honestly. They are thus *locked into corruption*. This idea explains the short-run persistence

of corruption. Shortly after date 0, there are lots of agents locked into corruption. This first effect however does not explain why the steady state is affected by this one-shot increase in corruption, since we assume that agents are progressively replaced by new ones (that is, our model is one of overlapping generations). Namely, why do the agents who arrive with an unsmeared (individual) reputation also necessarily engage in corrupt activities? Why do the young inherit the corrupt practices of their elders? The answer is that in the early periods after date 0, and because of imperfect observation of track records, the large number of agents who have been corrupt at date 0 and therefore remain corrupt raises a general suspicion. This suspicion affects new agents if their "age" (or more realistically, whether they had opportunities to get corrupt earlier) is not observed. Agents who arrive at date 1 are victims of this suspicion for at least a number T of periods and, if T is large enough (that is, if agents are not replaced very fast), have no incentives to remain honest. This implies that the number of agents with a smeared record does not decrease. In turn, agents who arrive at date 2 are victims of this suspicion for at least T periods, and decide to become corrupt. And so forth. We therefore obtain a *vicious circle of corruption*, where the new generations suffer from the original sin of their elders long after the latter are gone.

It should also be noted that in this model, *corruption ratchets up and not down*, in the sense that a one-shot reduction in corruption due, say, to tough enforcement of anticorruption laws has no lasting effect. It takes a minimum number of periods without corruption to upset the corrupt equilibrium. At this stage we have but an example, and no general result showing that the level of corruption in society increases faster than it decreases, but we find the example suggestive of why short run crackdowns on corruption often have

limited efficiency. We hope that further work will investigate the generality of this conclusion.

While sections 3 and 4 analyze the possible breakdown of desirable economic activity due to lack of trust and widespread corruption, section 5 uses similar modeling and ideas to study the development of other, *undesirable* activities. More precisely, we analyze the phenomenon of extortion. Suppose a foreign company wants to do business in a country and wonders whether it should bribe low-or high-level government employees to process goods through customs, issue work permits for company personnel or building permits for plants, grant a government contract or provide police protection. It has been well documented by Jacoby et al (1977) and many others that this is unfortunately one of the first questions business persons confront. Leaving aside any moral issue, we ask whether there can exist multiple equilibria with different levels of extortion. This is indeed the case. In a noncorrupt equilibrium, government officials do what they are meant to do even if they are offered no bribe, firms can get away by offering no bribe, and government officials have no incentive to give them trouble given that they will not be offered bribes in the future. In a corrupt equilibrium, firms attach a low probability of being able to conduct business without giving bribes, and they do offer bribes. Government officials are reluctant to do their job in the absence of a bribe because this might reveal their "softness" to future bribers. Again, the multiplicity of equilibria stems from the combination of individual and collective reputations (section 5).

Before turning to the analysis, I should point out that I do not subscribe to the view that corruption is a lesser harm. It has become fashionable in some academic circles in the last thirty years to argue that corruption is a market mechanism that frees the economy from the evils of excessive

bureaucracy. While this view has the merit of questioning the organization of bureaucracy, it ignores the substantial efficiency costs of corruption, not to mention moral and social effects and the implications for income distribution (including those due to the diversion of international aid). These efficiency costs include, among others, the selection of incompetent contractors and civil servants, the many barriers to entry into business, the shortage of tax and duty income, and the costs associated with tolerated pollution and job safety infringements. Accordingly, I will model corruption as a socially costly activity.

3 Individual and social stigmas: The case of trust.

This section develops a simple model in which the efficient organization of economic activity requires a minimum level of trust between contracting parties. More precisely, a principal (the buyer of a service) will contract with an agent (the supplier of the service) only if she is sufficiently confident that the agent will not engage in corrupt activities. The principal has some, albeit imperfect information about the agent's track record, namely about whether the agent has engaged in corrupt activities in the past.

Matching. We consider a stationary economy in which agents alive at date t remain in the economy up to (at least) date $t+1$ with probability $\lambda \in (0,1)$. With this "Poisson death process", we assume that each quit is offset by the arrival of a new agent, so that the population of agents is constant. The model is a matching model. At each date t, each (alive) agent is matched with a new principal[7]. The principal decides whether to offer task 1 or task 2 to the agent. Task 1 is the efficient task. Task 2 is a less efficient

[7]Principals can be either short lived or long lived.

task, but, for the principal, it is less sensitive to the agent's choosing to be corrupt. [In a slightly different version of the model, task 2 corresponds to the absence of a hire]. We will make an assumption guaranteeing that it is always optimal for the principal to at least offer task 2 to the agent rather than not hiring him. Once hired, the agent chooses whether to engage in the corrupt activity, that is whether "to cheat" (behave dishonestly). The principal's payoff from task 1 in the period is H if the agent behaves honestly and D if he cheats. Similarly her payoffs from task 2 are h and d. That task 1 is more sensitive to corruption than task 2 (given that the principal faces a nontrivial choice) means that

$$H > h \geq d > D.$$

We also assume that $d \geq 0$ so that it is optimal to hire the agent.

Agents' preferences. There are three types of agents: "honest", in proportion α, "dishonest", in proportion β, and "opportunistic", in proportion γ, where $\alpha + \beta + \gamma = 1$[8]. The proportions are the same for each cohort and therefore for the entire population. Honest agents have a strong distaste for and never engage in corrupt activities (alternatively, if corruption has a probability of being exposed and directly punished, "honest" agents might be ones for whom being punished is very costly). Dishonest agents always cheat, for instance because they derive a high benefit from it (alternatively, in a slightly different model, they might be transient agents who do not care about their reputation). Because honest and dishonest agents behave mechanistically (never and always cheat, respectively), the focus of our analysis is on opportunists. These have no aversion to being corrupt, but trade off the current benefit from corruption and the loss in reputation. Their benefits

[8]This formulation of preferences is standard in reputation models, see, e.g., Diamond (1991).

from being hired in tasks 1 and 2 and not cheating are B and b, respectively, where

$$B > b \geq 0.$$

They enjoy an additional short-run gain $G > 0$ from being corrupt in either task. That G is the same in both tasks simplifies the formal analysis. Note also that we do not model explicitly the role of anti-corruption campaigns. The simplest, albeit extreme interpretation of the model is that there is no hard evidence that could lead to the indictment of a corrupt agent. Alternatively, G could be an expected gain from being corrupt, which would allow a probability of confronting legal sanctions. Last the agents' discount factor is $\delta_0 \leq 1$. We will let $\delta \equiv \delta_0 \lambda$ denote the "relevant discount factor".

Information. Agents know their own preferences (that is, their types). Principals know the proportions α, β, γ and imperfectly observe the track record of the agent they are matched with. There are several ways of formalizing the imperfect observability of the track record. We choose a simple one in order to easily illustrate the main ideas. The principal has probability x_k of finding out that the agent has engaged in the past at least once in a corrupt activity when the agent has in fact cheated k times[9]. So the *observed* track record, that is the information of the principal the agent is matched with is binary. The principal knows that the agent has been corrupt at least once, or has no such knowledge. The assumption that the principal does not

[9]It would be interesting to extend the analysis to alternative information technologies. In particular it would seem reasonable to allow for forgetfulness (witnesses or evidence disappear over time). Our insights ought to carry over to such specifications, but new insights (such as the possibility of an individual's resuming an honest behavior after being corrupt) would arise.

We have performed a different check of robustness by assuming that once an individual is exposed a public file exposes him for the rest of his life. The expressions of Y and Z below are slightly altered, but the analysis goes through under the same assumptions 1 through 4. See also section 6 in which, once exposed, the agent is excluded for the rest of his life.

know the agent's age is important for the second effect unveiled in section 4 and giving rise to everlasting effects of a one-time shock in corruption. Of course this assumption should not be taken too literally. It is a metaphor for the idea that the principal may not be fully informed about the number of times the agent had an opportunity to be corrupt in the past.

Assumption 1: $x_0 = 0 < x_1 \leq x_2 \leq x_3 \leq \cdots < 1$

and

$$x_{k+1} - x_k < x_k - x_{k-1} \qquad \text{for all k.}$$

Assumption 1 says that the leakage of information about corruption becomes more likely when the agent has cheated more in the past; and that this increase occurs at a decreasing rate. This assumption simplifies the analysis by garanteeing that an individual is locked in corruption after having been corrupt a certain number of times.

We now demonstrate the possibility of coexistence of two equilibria.

a) Low corruption equilibrium. Suppose that all opportunists always behave honestly. A principal offers task 2 to an agent who she knows has been corrupt in the past, since the agent is necessarily a dishonest agent and since $d > D$. In contrast, when the principal has no such information, the agent may be honest or opportunistic, or else be a dishonest agent with a deceivingly clean observed track record. The proportion of honest and opportunistic agents in the economy is $(\alpha + \gamma)$. The proportion of dishonest agents with a clean track record is βY where Y is the average probability that past corruption activities go unnoticed [10]:

$$Y = (1-\lambda)\left[1 + \lambda(1-x_1) + \lambda^2(1-x_2) + \cdots + \lambda^k(1-x_k) + \cdots\right].$$

The probability that the agent will not cheat given a clean observed record

[10]The proportion of "newborns"(who therefore have not yet cheated) is $(1-\lambda)$, the proportion of "one-period old"(who have cheated once) is $(1-\lambda)\lambda$, and so forth.

is $(\alpha+\gamma)/(\alpha+\gamma+\beta Y)$. The principal offers task 1 if and only if the following assumption holds:

Assumption 2:

$$\frac{\alpha+\gamma}{\alpha+\gamma+\beta Y}(H-h)+\frac{\beta Y}{\alpha+\gamma+\beta Y}(D-d)>0.$$

Do opportunists have an incentive not to become corrupt? By never being corrupt, they keep a clean (real and observed) record and are always offered task 1. Their payoff is therefore $B+\delta B+\delta^2 B+\cdots = B/(1-\delta)$. Suppose that they instead cheat today and keep cheating in the future. Their expected payoff is then

$$(B+G)+\delta(B+G)[1/(1-\delta)-Z]+\delta(b+G)Z,$$

where

$$Z = x_1+\delta x_2+\delta^2 x_3+\cdots$$

is the present discounted probability of being found out in the future given that one has cheated once and will continue cheating. So, a necessary condition for a low corruption equilibrium is:

Assumption 3: $G/(1-\delta)\leq\delta(B-b)Z$.

Appendix 1 shows that *the low corruption equilibrium indeed exists under assumptions 1 through 3.* The intuition is that from assumption 1, the agent has more incentive to be corrupt, the more he has been corrupt in the past. In this sense, *agents are locked into corruption once they start being corrupt.*

Note also that a low corruption equilibrium exists only if the principals are not poorly informed [11]. Agents must have enough incentives to maintain their reputation for honesty.

b) High corruption equilibrium. Suppose now that opportunists are always corrupt and principals always offer task 2. Because keeping a clean slate has

[11] If the x s are close to zero, Z is close to zero and assumption 3 is violated.

no value, it is indeed optimal for opportunists to be always corrupt. Is it optimal for a principal to offer task 2 to an agent with a clean slate? Such an agent is honest with probability $\alpha/[\alpha + (\beta + \gamma)Y]$ and either opportunistic or dishonest with probability $(\beta + \gamma)Y/[\alpha + (\beta + \gamma)Y]$. We thus make

Assumption 4:

$$\frac{\alpha}{\alpha + (\beta + \gamma)Y}(H - h) + \frac{(\beta + \gamma)Y}{\alpha + (\beta + \gamma)Y}(D - d) < 0.$$

The high corruption equilibrium exists if and only if assumption 4 holds. Note that assumption 4 holds when there are enough opportunistic and dishonest agents and when the principals' information is not very precise.

We conclude that the low and high corruption equilibria both exist when assumptions 1 through 4 hold [12] . The role of imperfect observability is highlighted by the facts that assumption 3 is violated if the principals' information is very bad and that assumption 4 is violated if the principals' information is very good.

Remark (comparison with the economic theory of discrimination): Our imperfect observability assumption is reminiscent of that made in Arrow's (1973) statistical theory of discrimination of minorities by employers [13] . Arrow looks a one-shot employment decision and assumes that workers first (secretly) invest in skills and then the employers run an imperfect test of the

[12]Sah (1991) has developed a theory of crime in which the multiplicity of equilibria has a different origin. In Sah's model, the probability of being caught and punished for a crime decreases with the number of other criminals, assuming that the budget for crime investigation is not very responsive to the level of crime. The individuals' choices of whether to commit a crime are therefore strategic complements: The more people commit crime, the more incentives the individual has to commit a crime. While the multiplicity of equilibria can be illustrated in a static framework, Sah's model is actually an intertemporal one in order to highlight the idea of osmosis; the focus is not on reputation as in the present paper, but on local learning about the probability of punishment. Individuals learn slowly about this probability by observing whether their neighbors get punished when they commit a crime.

[13]See also Akerlof (1976), Coate-Loury (1991), Kremer (1992), Lundberg-Startz (1983), Milgrom-Oster (1987), and Phelps (1972) for related ideas.

resulting ability. Because the test is imperfect, the employer uses the prior beliefs about whether the worker has invested in assessing the worker's true ability. If a higher prior belief that the worker has invested also makes it more profitable for the worker to invest, there is scope for multiple equilibria. The literature has interpreted the multiplicity of equilibria as the possibility of a differential treatment of workers based on their race, sex or other observable characteristics. There is an analogy between the theory of discrimination and the (more dynamic) theory of corruption developed here. In the corrupt equilibrium, agents face a general suspicion of corruption and do not gain from not becoming corrupt, in the same way that a discriminated against group has (under certain conditions) little incentive to invest in skills if the employer puts more weight on prior beliefs than on imperfectly measured ability.

There is however a sense in which the statistical discrimination theory is not about societal behavior; for, the multiplicity of equilibria in the discrimination model is independent of whether there are other employers or workers besides the employer and the worker in question[14]. Furthermore, the statistical discrimination theory, which is a static theory, is not about collective reputations, an intrinsically dynamic phenomenon, either.

4 Persistence of corruption.

We now investigate the effect of a one-time shock in corruption on the equilibrium. To keep the analysis simple, we specialize the model further by making

Assumption 5: $x_1 = x_2 = \cdots = x \quad \in (0,1)$.

[14]The low and high corruption equilibria in our model could similarly coexist with single long-lived principal and agent, but only under the implausible assumption that the principal does not observe her per- period payoff (otherwise the principal perfectly knows the agent's track record).

That is, the probability of exposure of corrupt activities is independent of the number of past corrupt acts. Assumption 5 implies in particular that an opportunist remains corrupt once he has started; it also implies that $Y = 1 - \lambda x$ and $Z = x/(1-\delta)$.

The low corruption equilibrium exists if and only if assumptions 2 and 3 hold, which we will assume. Suppose now that the economy faces a temporary shock at date 0. The gain from being corrupt at that date is very large, and so all agents alive at date 0 get corrupt. The parameters of the model (including the gain G from cheating) are unchanged at dates $1, 2, \cdots$. We show that under an additional assumption, the economy cannot go back to the low corruption equilibrium. Indeed, the unique equilibrium exhibits a high level of corruption forever.

Let us perform the following thought experiment. Suppose that the opportunistic agents born at date 1 through t behave honestly before and at date t. This presumption gives the best chance to the existence of trust at date t. The probability of honest behavior at date t given an observed clean record and given that opportunists born at or before date 0 are locked into corruption is

$$\begin{aligned} p(t) &\equiv \frac{\alpha + \gamma(1-\lambda)\,(1+\lambda+\cdots+\lambda^{t-1})}{[\alpha+\gamma(1-\lambda)\,(1+\lambda+\cdots+\lambda^{t-1})] + [\beta Y + \gamma(1-x)(1-\lambda)\,(\lambda^t+\lambda^{t+1}+\cdots)]} \\ &= \frac{\alpha+\gamma(1-\lambda^t)}{[\alpha+\gamma(1-\lambda^t)] + [\beta Y + \gamma(1-x)\lambda^t]}. \end{aligned}$$

Suppose $p(1)(H-h) + (1-p(1))(D-d) < 0$. Recalling that $p(\infty)(H-h) + (1-p(\infty))(D-d) > 0$ (this is assumption 2) and noting that p is an increasing function, we let T denote the largest t such that

$$p(T)(H-h) + (1-p(T))(D-d) < 0.$$

That is, under the most optimistic assumption, principals still do not trust

agents with observed clean records at date T; thus $(T+1)$ is a minimum length for suspicion to phase out. Suppose now that

Assumption 6: $G\left(1+\delta+\cdots+\delta^{T-1}\right) \geq x\delta^{T}(B-b)/(1-\delta)$.

Assumption 6 states that it is a dominant strategy for an agent born at date 1 to cheat at date 1 (and therefore forever) given that the agent will not be trusted before (at best) date $(T+1)$. The left hand side of assumption 6 is the gain from cheating from date 1 through date T (discounted at date 1), and the right hand side is an upper bound on the cost of not being offered task 1 after date $(T+1)$. Note that assumption 6 requires T not to be too small, since with x_k constant for $k \geq 1$ assumption 3 is equivalent to $G \leq x\delta(B-b)$.

Consider now the generation born at date 2. All its elders have been corrupt in the past, and assumption 6 ensures similarly that cheating at date 2 and thereafter is a dominant strategy. By induction, the same is true for all generations. *Corruption has ratcheted up and does not subside even after the generation that has committed the original sin has by and large disappeared*[15].

[15]There exist other reasons than those unveiled here why corruption tends to persist once in place. Corruption may also persist because corrupt officials are likely to choose other corrupt officials to work with them and to succeed them. A benefit for a corrupt official from having a corrupt subordinate is that the official can extort the subordinate and obtain some of the bribes he collects. For example, the subordinate may be a tax collector who gives back a fraction of the bribes to his boss. Another benefit for a corrupt official from being surrounded by other corrupt officials is that these colleagues will be reluctant to denounce him by fear that they themselves might be exposed in a retaliation [see Andvig-Moene (1990) for a model in which a bureaucrat's cost of being corrupt decreases with the number of corrupt colleagues (such colleagues can be bribed not to report corrupt transactions).] Last, a corrupt official is likely to prefer having a corrupt successor, since a corrupt successor will not perform as well as an honest one and therefore will not disparage the departing official's performance. For these three reasons, corruption in hierarchies such as government, courts and political organizations is likely to have a life of its own. This explanation for persistence, if it is relevant, suggests that an anticorruption campaign is likely to be efficient if it fries big fish, since honest individuals cannot easily move up a hierarchy run by corrupt officials.

Another factor of persistence is the possibility of a low-budget trap (see, e.g., Klitgaard (1988)). A government official, like any economic agent, has an incentive to behave only if the cost of cheating (the probability of being punished times the extent of the punishment) exceeds the benefit of misbehaving (the bribe). The monetary punishment when caught

This simple model also illustrates the possible failure of a short-run anticorruption campaign. Suppose that at date 1 (or, equivalently at any later date) the government runs a tough anticorruption campaign that lasts one period and makes it unprofitable for opportunists to engage in corruption at that date. Suppose further that the following strengthening of assumption 6 holds:

$$G\left(1+\cdots+\delta^{T-2}\right) \geq x\delta^{T-1}(B-b)/(1-\delta).$$

Then it is a dominant strategy for generations born at dates 1 and 2 to cheat at date 2, and corruption prevails at all dates after date 1. *The anticorruption campaign only implies a decrease in corruption during the campaign and has no effect thereafter. Corruption does not ratchet down.*

5 Extortion.

We now apply similar ideas to study extortion [16]. Extortion occurs if the briber (the principal) is sufficiently convinced that the bribee (the agent) will

can be the loss of a well-paid job (plus, possibly, the confiscation of personnal assets). In particular, high wages for government officials may act as a potential deterrent to corruption. A country with a low level of tax collection or with high procurement expenditures pays low wages to its civil servants, who are then encouraged to become corrupt. Corruption in turn reduces tax collection and raises procurement expenditures, creating new budgetary problems. This yields a poverty cycle.

An objection to the previous reasoning is that the government could borrow internally or externally in order to give decent wages to the civil servants, get rid of corruption, escape the poverty trap and then reimburse its debt. Let us note however that it may not be easy to borrow internally substantial sums of private money in a poor country (in which, furthermore, the rich prefer to put their money abroad for safety and confidentiality reasons). Borrowing abroad is not easy either, if only because foreign creditors are worried by the possibility of repudiation of the debt. It is interesting to note in this respect that major international lenders often require tough budgetary discipline as a precondition for their loans. Future research ought to investigate the feasibility of an escape from the poverty trap in a situation of imperfect capital markets.

[16] See Strand (1990) for a different model of extortion. There, a bureaucrat asks for a bribe from a firm. The firm may accept the deal or report the attempt to extort to a government controller, who himself may or may not be corrupt. The firm is blacklisted if the controller is corrupt and receives a reward otherwise. A corrupt controller demands a bribe from the bureaucrat instead of punishing him.

not provide a service in the absence of a bribe. By analogy with the model of section 3 where the agent wanted to develop a reputation for trustworthiness, the agent here wants to look tough and convince principals that he will not provide services for them unless they give a bribe. The model shares a number of similarities with the previous one, and will purposedly share some of its notation.

As before, the model is one of matching. In each period, the agent (the government official, the bribee) is matched with a new principal (the firm, the briber). The timing within the period is as follows: First, the firm decides whether or not to offer a bribe to the official. For simplicity, we let B denote the size of the bribe. The firm gains $V > B$ if the agent provides the service. Second, the agent decides whether to provide the service. There are three types of agents: "honest", in proportion α, "corrupt", in proportion β, and "opportunist", in proportion γ, where $\alpha+\beta+\gamma = 1$. The proportions are the same for each cohort. Honest officials always provide the service. Corrupt officials never provide the service unless they receive bribe B. Opportunists, when they are offered no bribe, trade off a short-term cost $c > 0$ of not providing the service and the long-term loss of reputation for being tough. They provide the service if offered a bribe. One can think of c as coming either from scruples associated with not doing one's job or from a probability of being caught and punished. The probability of survival λ and the relevant discount factor δ are defined as before.

We again posit imperfect information about the agent. The principal has probability x_k of finding out that the agent has been weak at least once in

In Cadot (1987), a bureaucrat administers a test to grant a permit. There are two kinds of bureaucrats: "honest" (who grant the permit if and only if the candidate passes the test) and "corrupt" (who grant the permit if and only if they receive a bribe). The candidate, when asked for a bribe, can accept the deal or denounce the bureaucrat to a controller. Denunciation delays the permit (and, if the candidate does not know his ability, may not succeed).

the past, when the agent has in fact been weak k times, where "being weak" or "giving in" means that the agent provides the service to a principal who does not offer the bribe[17]. The x_k sequence satisfies assumption 1.

a) No extortion equilibrium. In a no extortion equilibrium, the firms never offer a bribe even when they don't know of any occurrence in which the government official gave in. In such an equilibrium, opportunists always give in, since they will never be offered a bribe in the future. Is this rational for a firm not to offer a bribe when it does not know its faces an honest or opportunistic agent? Let

$$Y \equiv (1-\lambda)\left[1+\lambda(1-x_1)+\lambda^2(1-x_2)+\cdots\right]$$

denote the average probability over the population of opportunists and honest agents that an opportunist or honest agent is not observed to have been weak in the past. The firm does not offer a bribe to an official whose type it does not know if and only if the following assumption holds:

Assumption 7:

$$B > \frac{\beta}{\beta+(\alpha+\gamma)Y}V.$$

Assumption 7 states that the size of the bribe exceeds the conditional probability that the official is corrupt times the value of the service to the firm. Note that the no extortion equilibrium exists only if the firm is not perfectly informed about the agent's track record (if λ and the xs are close to 1, Y is close to 0 and assumption 7 is violated).

*b) Extortion equilibrium.*Suppose now that the firms offer a bribe to those agents who are not known to have given in, and no bribe to those who are known to have given in; and that opportunists do not give in (unless they

[17]It would be worth investigating alternative assumptions on individual reputations. This restrictive, but simple assumption allows us to make direct use of the preceding analysis.

have already given in at least $k^* > 1$ times, in which case they give in) when offered no bribe.

If the firm knows that the official has given in at least once when offered no bribe, this official must be honest and therefore it is optimal for the firm not to offer a bribe. In contrast, if the firm does not know that the official has given in in the past, the firm optimally offers a bribe if the probability that the service will not be provided in the absence of a bribe times the value of the service exceeds the bribe:

Assumption 8:

$$B < \frac{\beta + \gamma}{\beta + \gamma + \alpha Y} V.$$

In an extortion equilibrium, it must also be the case that when offered no bribe an opportunist does not want to give in. Let z denote the present discounted expected number of bribes that the official receives by giving in and continuing to give in every time that he is not offered a bribe [18]. A necessary condition for the existence of the extortion equilibrium is that *Assumption 9*:

$$\delta\left(\frac{B}{1-\delta} - z\right) > c.$$

Conversely, the extortion equilibrium exists if assumptions 8 and 9 hold (the proof is almost identical to that in Appendix 1). Note that it can exist only if the principals' information is not too imprecise (if the xs are close to 0, z is close to $B/(1-\delta)$ and assumption 9 is violated).

We thus conclude that under assumptions 7, 8 and 9, the extortion and no extortion equilibria coexist. The formal analysis is almost identical to that of section 3. Yet the economics of trust (section 3) and extortion (this

[18] z is given by the following recursive equation: Let V_k denote the valuation of an opportunist who has given in k times in the past, and gives in whenever he has not been offered a bribe:

$$V_k = x_k \delta V_{k+1} + (1 - x_k)(B + \delta V_k).$$

Let $x = \lim_{k\to\infty} x_k$ and $V_\infty = (1-x)B/(1-\delta)$. Then $z \equiv V_1$.

section) differ in a few respects. In the extortion context, individuals want to build a reputation for the behavior that society tries to eradicate. In the trust context, they want to build a reputation for honesty. This distinction will have implications when adapting the design of anticorruption policies to the targeted form of corruption. A careful analysis of this conjecture falls outside the modest scope of this exploratory paper.

6 Exclusion from the group: The case of a firm's reputation for quality.

When the trading partner hardly observes the past individual behavior of the member, the latter's incentive to behave well can only come from the threat of retaliation by the group itself. We now assume that belonging to the group generates a rent but is no longer a fait acccompli. While we develop the analysis in the context of a firm's reputation for quality, *it applies* equally well *to any organization or group that coopts its members and can freely exclude them.*

We consider a stylized model of a firm as a workers' cooperative. Each period the workers share and consume the firm's profit. [We abstract from issues such as unequal treatment, hierarchies and delayed compensation in order to better focus on that of collective reputation. The assumptions driving the results are that incentive problems are not perfectly solved by alternative methods and that the workers enjoy a higher rent in a higher reputation firm.] Demand is inelastic and constant over time; the (large) number of workers is constant and, by normalization, equal to one worker per unit of good produced in a period. Workers who either quit (which occurs, as earlier, with Poisson probability $(1 - \lambda)$) or are fired are immediately replaced. No screening among workers at the hiring stage is feasible in our model. We also

assume for the moment that there is no cost for the firm of firing a worker and hiring a new one.

The consumers in each period observe the *firm*'s track record, namely the average quality of items produced in each past period. An item's quality is equal to H (high) or L (low), with $H > L > 0$. The consumers' (common) reservation price at date t is equal to the expected quality produced by the firm conditional on the firm's track record. Consumers do not observe the individual track record of the worker who has produced the particular item they buy. So, if ν_t is the consumers' posterior probability of buying a high quality item given the track record, the firm charges price $p_t = \nu_t H + (1-\nu_t)L$.

Producing a low-quality item costs nothing to the worker in charge of the item, while producing a high-quality unit involves a disutility of effort. There are three types of workers. Using the same notation and terminology as earlier, a worker is "honest" with probability α, meaning that this worker has no disutility of effort and always produces a high-quality item. With probability β, the worker is "dishonest"; he then has a very high disutility of effort for producing high-quality and always produces a low-quality item. Last, with probability γ, the worker has disutility of effort G of producing high-quality and behaves opportunistically.

Keeping with the notation of the paper, let x_k denote the probability that the firm, i.e., the workers, find out that a worker has produced at least one low-quality item in the past when the agent has in fact produced k low-quality items in the past. The firm may find out either directly through word of mouth or observation of past decisions of the worker, or indirectly through consumers' complaints about the durability of the product. The sequence $\{x_k\}$ satisfies assumption 1. For consistency with the previous section, we assume that the firm does not know the "age" of its workers (or,

more realistically, the number of times they had a good opportunity to shirk); however the analysis would not be affected if the firm knew the workers' age, because, in the derivations below, the firm fires a worker anytime it has evidence of a low-quality production. As before, we also assume that, in each period, conditional on their not being fired, workers stay in the firm with probability λ (the survival rate), and we let δ denote the relevant discount factor, namely λ times the workers' discount factor.

We look at steady states and define a high- (low-) reputation firm as a firm in which opportunists always produce high- (low-) quality items. We follow section 3 in proving the possibility of existence of a good and a bad equilibria. In both cases, the firm keeps workers for whom it has no evidence of wrongdoing and fires the others.

a) *High-reputation firm.* The expected present discounted tenure in the firm of a worker who produces low-quality in each period is:

$$\tilde{Y} = (1-\lambda)\sum_{t=0}^{\infty}\lambda^t(1-x_0)(1-x_1)\cdots(1-x_t),$$

(where $x_0 = 0$): λ^t is the probability of not quitting the firm before age t, and $(1-x_0)\cdots(1-x_t)$ is the probability of not being caught. Note that $\tilde{Y}$ differs from $Y = (1-\lambda)\sum_{t=0}^{\infty}\lambda^t(1-x_t)$ only to the extent that exclusion when it happens is permanent and not temporary. Assuming that new workers are drawn in a pool with proportions (α, β, γ) of honest, dishonest and opportunistic workers[19] , the steady-state proportions in a high-reputation firm are $(\alpha, \beta\tilde{Y}, \gamma)$. Consumers therefore

[19]In a general equilibrium context, this assumption means that, once fired by a firm, workers are not rehired by another firm; otherwise dishonest and (depending on the equilibrium) opportunistic workers would be overrepresented in the labor market compared to the prior distribution. We could alternatively have conducted the analysis under the assumption that firms cannot learn (or infer the) previous behavior of workers in other firms.

pay per unit:

$$p_H = \frac{\alpha+\gamma}{\alpha+\gamma+\beta\tilde{Y}} H + \frac{\beta\tilde{Y}}{\alpha+\gamma+\beta\tilde{Y}} L.$$

A necessary condition for an opportunist to produce high quality is that he prefers to always produce high quality rather than always producing low quality:

$$\frac{p_H - G}{1-\delta} \geq p_H \sum_{t=0}^{\infty} \delta^t (1-x_0)\cdots(1-x_t) \equiv p_H \left[\frac{1}{1-\delta} - \delta\tilde{Z}\right],$$

where $\delta\tilde{Z}$ is the expected present discounted reduction in tenure due to producing low quality. Again, $\tilde{Z}$ differs from Z only to the extent that exclusion is permanent. We thus make:

Assumption 3': $G/(1-\delta) \leq \delta p_H \tilde{Z}$.

Following the reasoning in Appendix 1 shows that assumption 3' (together with assumption 1) is also sufficient for the existence of a high-reputation equilibrium.

b) *Low-reputation firm.* In a low-reputation firm only the honest workers produce high quality. In particular, because opportunists don't produce high quality, there is more firing and therefore ***more turnover in a low-reputation firm than in a high-reputation firm.*** Consumers pay per unit:

$$p_L = \frac{\alpha}{\alpha+(\beta+\gamma)\tilde{Y}} H + \frac{(\beta+\gamma)\tilde{Y}}{\alpha+(\beta+\gamma)\tilde{Y}} L < p_H.$$

A necessary and sufficient condition for the existence of a low-reputation-firm equilibrium is:

Assumption 6': $G/(1-\delta) \geq \delta p_L \tilde{Z}$.

Assumption 6' states that the rent attached to working in a low-reputation firm is too small to dissuade the worker from shirking. Because assump-

tions 3' and 6' are not mutually inconsistent, there may exist multiple equilibria.

Remark 1 (hysteresis): The analysis of section 4 suggests that reputation is a very valuable asset in the sense that it may be impossible for the firm to rebuild its reputation after having lost it. Things however depend on the existence of costs of firing the whole labor force. Suppose that a firm goes through a period of lax management in which the opportunists produce low quality and that, as in section 4, these are locked into producing low quality in the future. If the cost of mass firing is large, so that the firm relies on quits and firings based on evidence to renew its labor force, we know from section 4 that the firm will never be able to (re)build a reputation for high quality even long after the period of negligent management. Mass firing (implying firing without evidence and therefore firing even the honest workers) is the firm's only chance to recover, if it is doable at a reasonable cost.

Remark 2 (Labor market externalities): As we already noted, we were able to take the proportions (α, β, γ) in the population of unemployed workers as given for our partial equilibrium analysis. It would be interesting to extend the analysis to study the labor market equilibrium. Indeed the proportions in the pool of unemployed workers depend on how many firms have a high reputation, and therefore fire only dishonest workers.

7 Conclusion.

We all belong to several organizations, cultures, and racial groups. Our welfare and our incentives depend not only on our own reputation but also on that of the groups we are associated with. This paper has shown that individual reputations are determined by collective reputations, and vice versa. Technically, collective reputations seem to often give rise to "strategic com-

plementarities": A member's incentive to maintain an individual reputation is stronger, the better the group's reputation. When discipline is sustained by the threat of exclusion from the group, low rents attached to being in a low-reputation group create low individual incentives to remain in that group and therefore perpetuate the group's bad reputation. When group belonging is an unalterable trait, distrust in the group may make good behavior a low-yield investment and thus distrust may be a self-fulfilling prophecy. Unsurprisingly, there may (although there need not) exist multiple, Pareto-ranked equilibrium collective reputations[20].

Perhaps even more fascinating is the history dependence of collective reputations. In our view, stereotypes are long lasting because new members of a group at least partially inherit the collective reputation of their elders. We have seen that a one-time, non recurrent shock on the behavior of a population can prevent the population from ever returning to a satisfactory state even long after the members affected by the original shock are gone. A more general study of history dependence involves non-steady-state statistical inference techniques, but should be high on our research agenda.

The genesis of collective reputations is a complex phenomenon. The modest object of this paper has been to shed light on some of its facets. We hope that the topic will soon receive from economic theorists the attention it deserves.

[20]See, e.g., Milgrom-Roberts (1990) and Vives (1990) for general results for games with strategic complementarities.

References

Akerlof, G. (1976) "The Economics of Caste and of the Rat Race and other Woeful Tales", *Quarterly Journal of Economics*, 91: 599-617.

Andvig, J.C. and K.O. Moene (1990) "How Corruption may Corrupt", *Journal of Economic Behavior and Organization*, 13:63-76.

Arrow, K. (1973) "The Theory of Discrimination", in O. Ashenfelter and A. Rees (eds.) *Discrimination in Labor Markets*, Princeton University Press.

Becker, G. (1968) "Crime and Punishment: An Economic Approach", *Journal of Political Economy*, 76:169-217.

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

Coate, S. and G. Loury (1991) "Will Affirmative Action Policies Eliminate Negative Stereotypes?", mimeo, University of Pennsylvania and Harvard University.

Cole, H., Mailath, G. and A. Postlewaite (1992), "Social Norms, Savings Behavior and Growth", *Journal of Political Economy*, 100: 1092-1125.

Crémer, J. (1986), "Cooperation in Ongoing Organizations", *Quarterly Journal of Economics*, 101:33-49.

Diamond, D. (1991) "Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt", *Journal of Political Economy*, 99: 689-721.

Elster, J. (1989) "Social Norms and Economic Theory", *Journal of Economic Perspectives*, 3:99-118.

Gould, D. (1980) *Bureaucratic Corruption and Underdevelopment in the Third World: The Case of Zaire*, New York: Pergamon Press.

Hager, M. (1973) "Bureaucratic Corruption in India: Legal Control of Maladministration" *Comparative Political Studies*, 6:197-219.

Jacoby, N., Nehemkis, P. and R. Eells (1977) *Bribery and Extortion in World Business: A Study of Corporate Political Payments Abroad*, New York: Mac Millan Publishing Co.

Kandori, M. (1992) "Social Norms and Community Enforcement", *Review of Economic Studies*, 59: 61-80.

Klitgaard, R. (1986) *Elitism and Meritocracy in Developing Countries: Selection Policies for Higher Education*, Baltimore: John Hopkins University Press.

Klitgaard, R. (1988) *Controlling Corruption*, Berkeley: University of California Press.

Klitgaard, R. (1991) *Adjusting to Reality*, San Francisco: ICS Press.

Kotlikoff, L., Persson, T. and L. Svensson (1988) "Social Contracts as Assets: A Possible Solution to the Time-Consistency Problem ", *American Economic Review*, 78:662-677.

Kremer, M. (1992) "The O-Ring Theory of Economic Development", Mimeo,

Kreps, D. (1990) "Corporate Culture and Economic Theory", in J. Alt and K. Shepsle ed. *Perspectives on Positive Political Economy*, Cambridge University Press, 90-143.

Krueger, A. (1974) "The Political Economy of the Rent-Seeking Society", *American Economic Review*, 64:291-303.

Lui, F. (1986) "A Dynamic Model of Corruption Deterrence", *Journal of Public Economics*, 31:215-236.

Lundberg, S. and R. Startz (1983) "Private Discrimination and Social Intervention in Competitive Labor Markets", *American Economic Review*, 73:340-347.

Milgrom, P. and S. Oster (1987) "Job Discrimination, Market Forces and the Invisibility Hypothesis", *Quarterly Journal of Economics*, 102:453-476.

Myrdal, G. (1970) "Corruption as a Hindrance to Modernization in South Asia" in *Political Corruption: Readings in Comparative Analysis*, ed. A. Heidenheimer, New York: Holt, Rinehart and Winston.

Noonan, J. (1984) *Bribes*, New York: Mac Millan.

Phelps, E. (1972) "The Statistical Theory of Racism and Sexism", *American Economic Review*, 62: 659-661.

Rose-Akerman, S. (1978) *Corruption: A Study in Political Economy*, New York: Academic Press.

Sah, R. (1991) "Social Osmosis and Patterns of Crime", *Journal of Political Economy*, 99:1272-1295.

Strand, J. (1990) "Bureaucratic Corruption in Government Contract Procurement: A Theoretical Model", mimeo, University of Oslo.

Seabright, P. (1992) "Is Cooperation Habit-Forming?", mimeo, Churchill College, Cambridge.

Sugden, R. (1989) "Spontaneous Order", *Journal of Economic Perspectives*, 3:85-98.

Theobald, R. (1990) *Corruption, Development and Underdevelopment*, Durham: Duke University Press.

Ullman-Margalit, E. (1977) *The Emergence of Norms*, Oxford: Oxford University Press.

Appendix 1
(Incentives to cheat in a low corruption equilibrium)

Let V_k denote an agent's expected present discounted value of present and future payoffs when the agent has cheated k times in the past. These are "continuation valuations". An agent who has cheated k times in the past will cheat again only if

$$G + \delta V_{k+1} \geq \delta V_k. \tag{A.1}$$

Suppose that the agent finds it optimal to cheat when he has cheated k times, and not to cheat when he has cheated $(k+1)$ times. Then

$$V_k = (1 - x_k) B + x_k b + G + \delta V_{k+1} \geq (1 - x_k) B + x_k b + \delta V_k, \tag{A.2}$$

and

$$V_{k+1} = (1 - x_{k+1}) B + x_{k+1} b + \delta V_{k+1}. \tag{A.3}$$

$(A2)$ and $(A3)$ yield

$$G \geq \delta (x_{k+1} - x_k)(B - b)/(1 - \delta). \tag{A.4}$$

On the other hand, the agent prefers stopping to cheat with record $(k+1)$ to cheating once more and then stopping. So

$$G + \delta \tilde{V}_{k+2} \leq \delta V_{k+1}, \tag{A.5}$$

where

$$\tilde{V}_{k+2} = (1 - x_{k+2}) B + x_{k+2} b + \delta \tilde{V}_{k+2} \tag{A.6}$$

$$\left(\tilde{V}_{k+2} \leq V_{k+2}\right). \tag{A.7}$$

Equations $A5$ and $A6$ yield

$$G \leq \delta (x_{k+2} - x_{k+1})(B - b)/(1 - \delta). \tag{A.8}$$

Inequalities $A4$ and $A7$ are inconsistent with assumption 1. So if it is optimal to cheat with record k, it is also optimal to cheat with any record $k' > k$.

2267 048