HIDDEN CORRUPTION

Marco Pani

No.506

WARWICK ECONOMIC RESEARCH PAPERS

THE UNIVERSITY OF WARWICK

DEPARTMENT OF ECONOMICS

HIDDEN CORRUPTION

Marco Pani
Department of Economics
University of Warwick
Coventry
ENGLAND

No.506

March 1998

Hidden corruption*

Marco Pani

University of Warwick

This version: March 1998

ABSTRACT

While most economic studies on corruption are mainly concerned with its actual occurrence, in this paper we claim that most economic effects rather derive from its mere possibility. When corruption is prevented through an expensive monitoring activity and by offering high incentives, its effects persist although corruption does not actually occur - a situation that may be described as "hidden corruption". In this paper, we formalise corruption as a particular principal-agent-client relation; we identify the essential economic characteristics of corruption, and we analyse different types of equilibria that arise under different circumstances. While corruption actually occurs in only one type of equilibria, its effects persist, although reduced in magnitude, in all the other types, where corruption is prevented but still potentially present; moreover, while equilibria of different types do not necessarily exhibit relevant differences in terms of expected utility, substantial differences appear instead when the possibility of corruption is removed. This paper also addresses other topics related to corruption, like the reasons that support the introduction of a legal norm prohibiting it, the effects of corruption on efficiency and social conflict, and the existence and implications of multiple equilibria.

KEY WORDS: Corruption; Principal-Agent Relation; Incentive Schemes.
JEL CLASSIFICATION: K1, K4, D8.

Correspondence should be addressed to: Marco Pani, Department of Economics, University of Warwick, GB - Coventry CV4 7AL E-mail: M.Pani@warwick.ac.uk

*I have benefited from comments and suggestions by Martin Cripps, Clive Fraser, Carlo Perroni and other staff of the Department of Economics of the University of Warwick, as well as participants in seminar presentations at the Midlands Economic Workshop. Research support from Warwick University is gratefully acknowledged.

1 Introduction

An aspect of corruption which does not appear to have received much attention in the literature is that most of its effects arise from the threat of its enactment, rather than from its actual occurrence. These effects derive from particular restrictions over the set of feasible contracts, which prevent or discourage the implementation of collusion-proof incentive schemes; when this occurs, corruption bears consequences on the expected utility of the parties involved even when it is successfully prevented and not actually observed - a situation that may be described as "hidden corruption". This paper focuses on these aspects, and shows how the effects of hidden corruption are generally comparable, and sometimes equivalent, to the effects of corruption actually occurring.

Following Rose-Ackerman (1978) and Klitgaard (1988), we analyse corruption within the framework of a principal-agent-client relation. In our model, corruption arises from the perturbation of a pre-existing principal-agent relation, where the principal is obliged to pay an exogenous basic salary to the agent in all situations, and can only apply non-negative variations to this reward, like awarding efficiency-related bonuses. While this is by no means the only type of restrictions that could generate corruption (other restrictions may derive from laws, regulations and other sources, or may be inherent to the particular relation that exists between the parties), this type of constraints may be frequently found whenever a minimum-salary legislation, professional rules, or collective bargaining agreements, impose a lower bound on the reward that is paid to civil service employees or other types of agents. In the presence of such constraints, collusion between the client and the agent cannot generally be prevented without reducing the utility of the principal, and the adoption of a legal norm prohibiting this collusion can be seen as a means used by the latter to reduce this loss in utility.

In her 1978 book on the subject, Rose-Ackerman described corruption as a particular agency relation where "*some third person, who can benefit from the agent's action, seeks to influence the agent's decision by offering him a monetary payment which is not passed to the principal*" (Rose-Ackerman, 1978, p. 6).[1] While basically accepting this definition, we agree with the prevalent opinion in considering corruption as essentially illegal, and we

[1] Along the same line, Barthan (1997) wrote that "*in a majority of cases [...] corruption ordinarily refers to the use of public office for private gains, where the official (the agent) entrusted with carrying out a task by the public (the principal) engages in some sort of malfeasance for private enrichment which is difficult to monitor by the principal*" (p. 1321). Similar definitions were provided by Friedrich (1989) and by Lowenstein (1989). Friedrich identified "*a corruption pattern [...] whenever [...] a responsible functionary or office holder, is by monetary or other rewards... induced to take actions which favour whoever provides the reward and thereby damage the group or organisation to which the functionary belongs, more specifically the Government*" (Friedrich, 1989, p. 15); for Lowenstein, corruption implies the involvement of a "public official", a "corrupt intent" by the parties involved, the transfer to the official of something of value, and the existence of a relationship between this thing of value and an official act; moreover, this relationship must include an intent to influence the act (Lowenstein, 1989, pp. 30-31).

therefore impose the additional requirement that the payment offered by the third party to the agent should contravene an existing norm.[2] Specifically, we characterise corruption as a particular type of principal-agent-client relation with the following characteristics: the principal delegates some discretionary power to the agent, requiring him to perform a particular task; the agent makes a contract with the client, whereby he agrees to deviate from the instructions of the principal to the benefit of the client, in exchange for some reward; but this contract is illegal, in the sense that it violates an existing norm. This norm may derive from a variety of sources, but is economically relevant insofar as its violation involves a positive risk of punishment for the parties who commit it; following Becker (1968), we assume that this punishment can be quantified in terms of an equivalent monetary fine.

Four economic relations thus characterise corruption and distinguish it from other types of action: (i) an agency relation between the principal and the agent; (ii) a contractual relation between the agent and the client; (iii) an obligation (resulting from some norm) on the part of the agent, of the client, or both, not to undertake such a contract; and (iv) a conflict of interests between the client and the principal. A direct contractual relation between the principal and the client is not an essential element of corruption and for this reason it will not be considered here, although this relation can be found in a variety of particular cases, like, for instance, when corruption is related to tax evasion or the control of crime.[3]

In comparisons with other types of action, the first of the above relations is absent when two partners collude to commit a theft, the second relation is absent in the case of mere abuse of power or nepotism, the third in the case of legal collusion (of the type discussed, for instance, by Kofman and Lawarrée, 1996); the fourth relation, finally, draws the distinction between corruption and extortion; remarkably, this distinction has not always been acknowledged by authors who analysed these topics, who often used a different terminology and adopted a different perspective (for instance, Beenstock, 1979; Cadot, 1987).[4] The

[2]Andvig and Moene (1990) claim that "*[a] member of a public organisation acts in a corrupt way if he [...] deals with a non-member and uses the organisation's resources [...] to acquire payment that is against the rules of the organisation or against the law*" (Andvig and Moene, 1990, p.65). In our opinion, however, corruption derives not so much from the improper use of some organisation's resources, as from the improper use of discretionary power; moreover, corruption is not confined to public organisations, and can be found also in the private sector (on this point see Rose-Ackerman, 1989).

[3]This relation has in fact been frequently formalised in works which analysed corruption in relation to tax evasion and crime control: see for instance Virmani (1983); Chander and Wilde (1992); Besley and McLaren (1993); Mookherjee and Png (1995).

[4]A clear illustration of the distinction between corruption and extortion can be found in Bardhan (1997): "*I am told that in Russia there is a clear terminological distinction between* mzdoimstvo, *taking a remuneration to do what you are supposed to do, and* likhoimstvo, *taking a remuneration for what you are not supposed to do*" (p.1323). Shleifer and Vishny (1993) distinguished between corruption "with theft", where

concept of corruption is however different from the concept of extortion; corruption stems from a bilateral agreement between the agent and the client, who illegally collude to obtain a benefit at the expense of the principal, while extortion derives from a unilateral abuse of discretion on the part of the agent, who threatens to deviate from the principal's instruction in order to obtain an undue benefit at the expense of the client. Even when it is the client who makes an offer to the agent, the situation is one of extortion whenever the offer is made in response to an implicit, but evident, threat. In a wider sense, the word "corruption" has been more or less explicitly used by some authors to include extortion as a particular case (Klitgaard, 1988; Hindricks *et al.*, 1996; Bliss and Di Tella, 1997); in this paper, however, we use it in a stricter sense and we limit our attention to our narrower definition of corruption.

We thus develop a model where all the above relations are present; we start with a basic principal-agent relation, and analyse how it is perturbed by the intervention of a client who offers a bribe to the agent; we analyse the utility implications of the new equilibria, and we examine how they are altered by the imposition of a norm banning corruption, supported by a non-negative probability of punishment in case of violation. We characterise the equilibria that arise in two different cases, when the implementation of the norm is monitored without costs at an exogenous rate, and when monitoring is costly to the principal who decides, in return, the probability of detection. This framework will also be used to analyse other features of corruption, like the attitudes of the parties towards the introduction of the above norm, and the implications of corruption in terms of efficiency and social conflict. Moreover, we use this model to show that multiple equilibria are possible even in a static context (their existence in a dynamic context was proved by Lui, 1986, Cadot, 1987, and Andvig and Moene, 1990); finally, we discuss some empirical and normative implications of our results.

The results that we obtain confirm our basic idea that most effects of corruption arise from its threat rather than from its actual occurrence. While corruption reduces aggregate utility and the expected utility of the principal, these effects generally persist - although reduced in magnitude - when corruption is prevented, but remains a real possibility. Indeed, while corruption actually takes place only in one type of equilibria, the expected utility of the parties in the other types of equilibria may not be too different, and, in particular circumstances, may even be the same. Relevant differences in utility appear instead between the equilibria that arise when corruption is possible and those that exist when corruption

an official who sells a good provided by the Government hides the sale and keeps the whole price of the good, and corruption "without theft", where the official illegally increases the price of the goods and keeps for himself the difference between the actual and the official price. Corruption with theft is similar to our narrower definition of corruption, since the official illegally provides the good to somebody who is not entitled to it, in exchange for an undue reward which he does not transfer to the principal; corruption without theft, instead, is similar to extortion, since the official provides the good to someone who is entitled to have it, and illegally demands a bribe on top of the price decided by the principal.

is not a feasible option (for instance, because the client does not exist at all).[5]

The introduction of a legal norm banning corruption allows in some cases to reduce the magnitude of these effects and leads in other cases to a different type of equilibrium; in general, however, it fails to restore the initial utility distribution. The adoption of this norm, moreover, is met with different attitudes in different circumstances; sometimes, it is unanimously supported, other times it is rejected without objections, while in other cases it becomes an issue of social conflict over the distribution of rent. The fact that certain types of behaviour (like Parliamentary lobbying) are legally allowed in some countries and constitute an offence of corruption in others may thus be interpreted as the result of a perfectly rational choice.

An *a priori* evaluation of the social effects of hidden corruption is also rather problematic; in some cases, the threat of corruption leads to a Pareto-inferior outcome and is thus a source of inefficiency, but in other cases it generates a redistribution of rent which benefits some parties at the expense of others, leading to a new equilibrium along the Pareto-efficient frontier.

On the empirical side, our results imply that a proper evaluation of the impact of corruption cannot merely rely on statistics of its observed occurrence; further information is required, in order to obtain a proper evaluation of the impact of the unobservable threat of corruption. Useful indicators in this sense might be provided by data concerning the reward structure of the agents, the probability of detection and the costs of monitoring, and some estimates of the amount that might be paid on bribes if corruption actually occurs. On the normative side, the design of effective anti-corruption strategies should extend beyond the adoption of merely repressive measures (like intensive monitoring and severe punishment), to include other types of intervention aimed at reducing the opportunities for corruption, including educational campaigns to increase the degree of honesty.

Our work builds on previous studies on corruption, particularly on the literature which analysed corruption in the framework of the theory of agency (Virmani, 1983; Chander and Wilde, 1992; Besley and McLaren, 1993; Mookherjee and Png, 1995; Hindricks *et al.*, 1996). Our model also bears some similarities with the work of Spiller (1990), who analysed a trilateral agency relation very similar to the principal-agent-client relation discussed here. The effects of the threat of corruption, however, do not appear to have received much attention in this literature, and the role played by restrictions on the set of feasible contracts has also been ignored. Our approach is somehow more general because we consider corruption in its essential structure rather than in relation with tax evasion or other specific situations; for this reason, we do not apply the principal-supervisor-agent model, that is generally used, preferring the more encompassing principal-agent-client formalisation, which does not require a direct contractual relation between the principal and the client.

[5]Throughout the paper, we shall ignore the expected utility of the client (a discussion of the reasons is provided in Section 3). Our results only apply to the effects of actual or hidden corruption on the particular subset of parties which does not include the client - namely, the principal and the agent.

The paper is organised as follows: Section 2 presents an outline of the model, where corruption is not legally prohibited; Section 3 analyses the causes and effects of the introduction of a legal prohibition, in the case when the probability of detecting its violations is exogenous (Subsection 3.1) and when it is endogenous and costly for the principal (Subsection 3.2). A discussion of the implications of the results is reported in Section 4, while Section 5 draws some final remarks and conclusions.

2 The model

A principal hires an agent to perform some task e, which can be carried out in two ways, e_H and e_L; e_H requires hard work and a high effort on the part of the agent, while e_L requires a lower effort and can be carried out working "lazily". Obviously, the agent prefers e_L, while the principal prefers e_H, because it implies, for instance, a more efficient organisation of some productive activity, bearing higher output at lower costs.[6] The principal cannot perfectly observe how the agent works, thus he cannot condition payment on the actual performance of the high effort. The principal only receives a random message m, whose probability distribution is correlated with the agent's work performance; when the agent works hard, the message is equal to 1 with probability $p > 1/2$ and to 0 otherwise; when the agent works lazily, the message is equal to 1 with probability $(1-p)$, otherwise it is equal to 0; in other words, the message is 'true' with probability p and 'false' with probability $(1-p)$. The principal thus conditions the amount of the payment on the realisation of this random binary message. If the message is 0, the principal pays the agent a minimum amount, which we assume to be exogenously given and normalised to 0; this amount may derive, for instance, from a legal provision binding all employers to pay in all situations a minimum basic salary to their employees, irrespective of their actual work performance. If the message is 1, the principal pays, in addition to the minimum salary, a performance-related bonus, which we shall simply call "the bonus" (or sometimes the "incentives"), whose amount t he declares in advance. A contract of this type is generally known as an *incentive scheme*, and under general conditions can induce the agent to work hard even if the principal cannot perfectly observe him.[7]

[6]This conflict of interest between the principal and the agent is not essential. Corruption may take place even when the agent is indifferent between alternative ways of performing this task, or when he strictly prefers to follow the instructions unless he receives an adequate payment in order to deviate. The case of a conflict of interest is however more interesting because it involves the strongest incentive to deviate and accept the bribe, and thus the highest opportunities for corruption. Our results extend however, with minor variations, to the other cases.

[7]The conditions required for an incentive scheme to be effective are the "participation constraint" and the "incentive compatibility" (or "individual rationality") constraint. The participation constraint ensures that the agent accepts to work for the principal at the agreed conditions and does not seek employment

Both principal and agent have quasi-linear utility functions, that depend on the work effort of the agent, e, and on the amount that is paid to him at the end. The utility of the principal is equal to $W(e,x) \equiv w(e) - x$, where x is the amount of the payment (equal to t when the bonus is paid and to 0 otherwise), while the utility of the agent is $U(e,x) \equiv u(e)+x$. The principal's preference for a high work effort is equal to $\Delta w \equiv w(e_H) - w(e_L)$ while the agent's preference for a low work effort is equal to $\Delta u \equiv u(e_L) - u(e_H)$.

Both parties are risk-neutral and take their decisions in order to maximise their expected utility.[8] The principal offers an incentive scheme that specifies the amount t of the bonus; the agent accepts the scheme and decides how to work. Once the work is done, a realisation of the random message m reaches the principal and is also observed by the agent; the bonus t is then paid, conditional on m being equal to 1.

The optimal choice for each party in equilibrium can be derived by backward induction. The agent responds to the offer of the principal, working hard if and only if the incentive scheme meets the *incentive compatibility constraint (ICC):*

$$t \geq \frac{\Delta u}{2p-1} \equiv t_a, \tag{1}$$

and working lazily otherwise.[9] The principal knows this constraint and the agent's response, and offers a bonus equal to t_a if and only if

$$\Delta w \geq p t_a; \tag{2}$$

otherwise, he offers no bonus at all, and the agent only receives his basic salary, deciding,

elsewhere; in this paper, we assume that the participation constraint is always satisfied, as a result of the legal requirement of a minimum salary. The incentive compatibility constraint is discussed in the text.

[8]Since the only random variable in the utility function is t, and since the utility function is linear in t, risk neutrality is here equivalent to a von Neumann-Morgenstern utility maximising behaviour.

[9]When the agent works hard, his expected utility is $U(H) \equiv u(e_H) + pt$, since he receives the bonus t whenever the message is 'true', which occurs with probability p (otherwise, he just receives his basic salary, equal to 0); when he works lazily, his expected utility is $U(L) \equiv u(e_L) + (1-p)t$, since this time he obtains the bonus only when the message is 'false', which occurs with probability $(1-p)$. The agent thus decides top work hard if and only if $U(H) \geq U(L)$, which implies (1).

in turn, to work lazily.[10] Substituting (1) into (2) yields

$$\Delta w \geq \frac{p}{2p-1}(\Delta u), \tag{3}$$

which is the condition under which the agent works hard in equilibrium. We assume hereafter that this condition is always satisfied.

At this point a third party intervenes, whom, in Klitgaard's terminology, we shall initially call "the client", and whom we shall later call "the corruptor" (the word "corruptor" is not appropriate at this stage, because no norm exists which forbids the client's intervention). Like the principal, the client is affected by the work effort of the agent, but, unlike him, he strictly prefers the low effort. Moreover, the client has perfect information about the way the agent works (unlike the principal),[11] and offers a payment to the agent conditional on his performance of the low effort. We shall refer to this payment as "the gift", and, at a later stage, as "the bribe".[12]

For instance, consider the case of an industry that wants to dispose illegally of its polluting waste. The industry is subject to controls by the Environmental Protection Agency, and these controls are carried out by inspectors whose actual effort the Agency cannot perfectly observe.[13] If the inspectors work hard, they pay frequent and accurate visits to the plants to check how the waste is being treated; if they work lazily, their visits are rare and inaccurate. The firm obviously prefers the inspectors to work lazily, and offers them a payment conditional on their visits being rare and superficial.[14] Alternatively, consider a

[10]The principal never sets $t > t_a$ since he can obtain the same behaviour from the agent at a lower cost by setting $t = t_a$; similarly, he never sets $t_a > t > 0$ since this would induce the agent to work lazily, in which case it would be more convenient to set $t = 0$. When the principal offers $t = t_a$, the agent works hard by (1) and the principal's expected utility is equal to $W(H) \equiv w(e_H) - pt_a$, since with probability p the message will be equal to 1 and the principal will have to pay the bonus t_a. When the principal sets $t = 0$, the agent works lazily and the principal only pays the basic salary irrespective of the realisation of the message, obtaining an expected utility equal to $W(L) \equiv w(e_L)$. The principal thus offers $t = t_a$ if and only if $W(H) \geq W(L)$, whence (2).

[11]The assumption that the client can perfectly observe the work effort of the agent is not essential; the client may for instance receive a second random message, s, correlated with the work effort of the agent, and may condition the payment of the gift on a particular realisation of this random message. In our model, however, this assumption allows to simplify the analysis, without loss of generality.

[12]Like the word "corruptor", also the word "bribe" is not yet appropriate, since no norm exists forbidding the gift.

[13]A similar case is described in Mookherjee and Png (1995).

[14]This payment needs not necessarily consist of a sum of money in a sealed envelope. The industry, for

taxpayer submitting false reports to the Tax Collecting Agency of the Exchequer.[15] The reports are examined by auditors on behalf of the Agency, but the Agency cannot control how the auditors carry out their work. The auditors may examine the reports carefully, cross-checking all declarations and matching them with additional information at their disposal, or they may merely perform a formal, quick check of the consistency of the declarations. The Agency obviously prefers the first option, but the taxpayer strongly prefers the second, and offers a payment to the auditors to convince them to adopt it.

The gift enters the agent's utility function in the same way as the payment given by the principal; the utility of the principal, on the contrary, is not directly affected by the gift. We assume that the value of the gift is decided by the client and is known by all parties; the client ignores how the agent is going to act if he does not offer the gift, and ignores both the agent's utility function and the incentives offered by the principal. As a result, the client offers a gift of fixed amount b, equal to his willingness to pay to have the agent work lazily rather than hard.[16] These assumptions greatly simplify the analysis, allowing to abstract from the client's maximisation problem and from a discussion of the bargaining process between the agent and the client, in order to focus on the effects of the client's intervention on the relation between the principal and the agent. While the preferences of the client may play a major role with respect to a number of aspects related to corruption, they are not essential to the purposes of this paper.[17] One implication of these assumptions is that the utility of the client is the same when the agent accepts the gift and works lazily as when he rejects it and works hard. A further assumption that we make is that the client has an established reputation for honouring his promises, which allows to abstract from any problems of enforcement that normally arise with illegal contracts.[18]

instance, may promise the inspectors well rewarded jobs once they leave the Agency. Moreover, we abstract from the fact that the firm could bribe the inspectors after their visit in order to obtain false and favourable reports; this possibility may in some cases be easily prevented, for instance, by including newly appointed junior members in the inspecting teams; while these members may be unable to decide the timing and number of the inspections, they may be able to inform higher authorities about the alteration of the reports, rendering ex-post corruption prohibitively risky.

[15] For a similar case see Chander and Wilde (1992).

[16] In other words, when the agent accepts the gift, he extracts all the client's rent from e_L.

[17] An exogenous value of the gift is also assumed by Bac (1996); Chander and Wilde (1992) assume instead that the gift is equal to a fixed, exogenous share of the client's rent, equal to a fraction γ of the tax evaded by the client. Other authors (for instance, Virmani, 1983; Basu *et al.*, 1992; Besley and McLaren, 1993; Mookherjee and Png, 1995) assume that the rent is divided according to a Nash bargaining solution.

[18] For a discussion see Paul and Wilhite (1994), and Kofman and Lawarrée's remarks about self-enforcing collusion agreements (Kofman and Lawarrée, 1996, p. 390). Bardhan (1997), in his survey of economic

The offer of the gift increases the agent's preference for working lazily; this in turn alters the incentives compatibility constraint, which becomes:[19]

$$t_b \equiv \frac{\Delta u + b}{2p-1} > t_a \tag{4}$$

The principal must now decide whether or not he should increase the bonus to meet the new constraint; if he does not, the agent will be lazy and the best choice will be to pay no bonus at all, irrespective of the message. In absence of restrictions on the set of contracts, the principal could offer an incentive scheme that satisfies the new incentive compatibility constraint yielding the same expected utility to the principal as he had before; for instance, he may raise the bonus to t_b and reduce the salary to $-p(t_b - t_a)$; in alternative, he could continue to offer the same salary as before, pay a bonus still equal to t_a when the message is 1, and convince the agent to work hard by paying a negative bonus equal to $-(t_b - t_a)$ whenever the message is 0; or, finally, he could resort to an appropriate mix of salary reductions and negative bonuses. All these arrangements are however prevented by the legal requirement of a minimum salary, which forbids both salary reductions and the imposition of a negative bonus; as a result, the principal can only choose between raising the bonus to t_b and paying no bonus at all, knowing that in this case the agent would work lazily. He chooses the first option if and only if[20]

$$\Delta w \geq pt_b > pt_a \tag{5}$$

Since (5) is more restrictive than (3), the intervention of the client may induce the principal to withdraw the bonus rather than increase it, leading from an equilibrium where the agent works hard to one where the agent works lazily. By (3), (4) and (5), this occurs whenever

$$\frac{p}{2p-1}(\Delta u + b) > \Delta w \geq \frac{p}{2p-1}(\Delta u). \tag{6}$$

When (6) holds instead, the expected utility of the principal is lower in the new equilibrium. In the following sections of this paper we shall assume that (6) always holds.

studies on corruption and development, writes that this problem has also been discussed in: Boycko, M.; Shleifer, A.; and Vishny, R.W. (1995), *Privatizing Russia*, Cambridge MA: MIT Press.

[19]The expected utility of the agent if he works hard is unaffected by the gift, which is not paid in this case; his expected utility if he works lazily accepting the gift becomes $U(L^b) \equiv u(e_L) + pt + b$; the agent thus works hard if and only if $U(H) \geq U(L^b)$, which implies $t \geq t_b$.

[20]If the principal offers a bonus equal to t_b, the agent works hard and the principal's expected utility is equal to $W(H') \equiv w(e_H) - pt_b$; if the principal offers no bonus at all, the agent works lazily and the expected utility of the principal is equal to $W(L) \equiv w(e_L)$. The principal thus sets $t = t_b$ if and only if $w(e_H) - pt_b \geq w(e_L)$, whence (5).

When (5) holds, the principal raises the bonus to t_b, and the agent rejects the gift and continues to work hard. The intervention of the client does not affect the observable work performance of the agent, but it induces a reduction in the expected utility of the 'principal (who has to pay a higher bonus if the message is 1) which is exactly matched by an increase in the expected utility of the agent (who receives that bonus). Although no payment takes place between the client and the agent, the mere threat of it enables the agent to extract rent from the principal, inducing a redistribution in expected utility between those parties - a situation which we describe as "hidden corruption":

Proposition 1 (Hidden corruption) *When (5) holds, the agent rejects the client's offer but the mere possibility of accepting it enables him to extract rent from the principal. The intervention of the client thus induces a redistribution of expected utility between the principal and the agent, without affecting the work performance of the latter.*

PROOF: see Appendix.

When (5) fails, the principal offers no bonus, the agent works lazily and the gift is actually paid. The effects of the intervention of the client are observable and evident. Since $w(e_L) < w(e_H) - pt_a$, the expected utility of the principal is lower in the new equilibrium. Remarkably, the expected utility of the agent may also be lower; this occurs under the following condition.

Condition 1 *Inequality (6) holds; moreover,*

$$b < \frac{1-p}{2p-1}(\Delta u). \tag{7}$$

In this case, the benefit that the agent derives from working lazily accepting the gift does not offset the loss of the possibility of obtaining the bonus t_a. This leads to the following Proposition:

Proposition 2 *When Condition 1 holds, the intervention of the client reduces the expected utility of both principal and agent, leading to an equilibrium which is Pareto-inefficient from the point of view of those parties.*

PROOF: see Appendix.

The initial, Pareto-efficient equilibrium remains a feasible outcome, but it is no longer an equilibrium because it is not *subgame-perfect*: should the principal agree to offers the original bonus t_a in order to maintain the previous outcome, the agent would strictly prefer to work lazily and accept the gift, and the initial outcome would not be reached all the same. Any promise on the part of the agent to act otherwise, working hard and rejecting the gift, would not be credible unless it were supported by some reliable means of enforcement (which might include moral concerns or a renown reputation for honesty). A legal norm obliging the agent to reject the gift, involving some sort of punishment whenever he is caught accepting it, may thus be seen as a means to increase the credibility of this promise. Once this norm is introduced, accepting the gift constitutes a proper act of corruption, in line with the definition discussed in the previous section, and the gift itself can properly be called "a bribe".

As we shall see in the following section, the attitudes of the two parties towards the introduction of this norm vary under different circumstances. In general, we should expect the principal to support its introduction (since the intervention of the client always reduces his expected utility) while the agent would tend to support the introduction of the norm when Condition 1 holds, and to resist it in all other cases. In fact, the attitudes of both parties depend on how effective the norm would be to prevent corruption, and on the costs that are incurred to implement the norm and to detect its violations. What is remarkable is that even the introduction of this norm would generally fail to restore the initial equilibrium. More specifically, the introduction of the norm may successfully prevent the actual payment of the gift and may induce the agent to work hard, but the effects of hidden corruption on the expected utility of the parties would not generally be removed (although they may be reduced in magnitude).

3 Equilibria with corruption

In the previous section, the use of the words "corruptor" and "bribe" was not appropriate, because no legal norm existed forbidding the client to offer the gift or prohibiting the agent to accept it; the contract between these parties thus failed the essential requirement of illegality, that would qualify it as corruption.

In this section, we assume that this contract is forbidden, and we therefore use the words "corruptor" and "bribe". We discuss the reasons that justify the introduction of this provision, and we analyse the equilibria that are induced by its introduction and compare them with those that preceded both the intervention of the client and its prohibition. A situation where the client does not intervene will hereafter be described as "a world without corruption", while any situation where the client is present and active will be described as "a world with corruption", even when corruption remains hidden and does not actually occur. As is shown below, most effects on expected utility depend on whether or not we live in a world with corruption.

This section is divided in two subsections, that consider two different cases. First, we assume that the implementation of the norm is randomly checked at zero costs, with an exogenous probability of detecting a violation when it occurs; in the second part, we assume that the costs of monitoring are positive and increase with the probability of detection.

When the agent is caught accepting a bribe, he has to pay a fine of exogenous amount

F.[21] The fine enters the agent's utility function in the same way as all other payments, but with a negative sign. The fine, however, is not paid to the principal and does not figure as an argument of his utility function.[22] Assume, following Becker (1968), that the agent's decision on whether or not to violate the norm depends only on which alternative maximises his expected utility; in other words, the agent is not bound by moral or other non-utilitarian concerns.[23] Violations, however, are not detected with certainty; when they occur, they are detected with probability $\pi < 1$, while when they do not occur they are never detected (innocents are never found guilty). No fine is ever imposed on the corruptor.

Prosecution and punishment are public events, so the principal is perfectly informed about their occurrence; since the agent is never punished when innocent, and since the bribe is never paid when the agent works hard, the principal knows that, when the agent is fined, he must have been both guilty and lazy; as a result, the principal can improve the incentive scheme in the following way: if the agent is not fined, the principal proceeds as usual, paying the bonus if and only if the message is equal to 1; if the agent is fined, instead, the principal presumes that the agent has worked lazily and pays no bonus, irrespective of the message; by legal requirement, however, he always has to pay the minimum salary.[24]

The agent now can choose between three actions: reject the bribe and work hard (H); reject the bribe and work lazily (L); or accept the bribe and work lazily (C). He cannot accept the bribe and work hard, because in this case the corruptor would not pay the bribe. When the agent rejects the bribe, we shall say that he is "honest", while when he accept it we shall say that he is "corrupt".

The probability of the agent obtaining the bonus depends on which of these alternatives he chooses; when he is honest, the probability is equal to p if he works hard and to $1-p$ if he works lazily, but when the agent is corrupt the probability of obtaining the bonus is equal to $(1-p)(1-\pi)$, which is the joint probability of escaping detection and obtaining a

[21] If no fine is imposed, the norm has no economic significance and corruption is not distinguishable from legal collusion. A similar case is analysed in Marjit's model on corruptible hierarchies (Marjit, 1996), where no penalties are imposed on either of the parties involved in corruption.

[22] For instance, the fine may be paid to some independent enforcement authority like the judiciary; in alternative, the fine may actually be paid to the principal but it could be entirely spent to cover the fixed costs of prosecution.

[23] Some authors (for instance, Tirole, 1986) describe this type of behaviour as "opportunistic", as opposed to "honest" (when the agent always decides to abide by the norm) and "dishonest" (when he always violates it). In this paper, the agent is always opportunistic, and is said to be "honest" when he maximises his expected utility by turning down the offer of the bribe.

[24] Improvement in the quality of information occurs even when the agent could be found guilty when he is innocent, provided this occurs with lower probability than when he is indeed guilty. Without loss of generality, we rule out this possibility.

false favourable message (two events that are assumed to be independent). When the agent is corrupt, he obtains the bribe with certainty and he incurs the fine with probability π. The expected utility that he gains with each of these actions is thus equal to:

$$\begin{aligned} U(H) &\equiv u(e_H) + pt && \text{when he is honest and works hard;} \\ U(H) &\equiv u(e_H) + pt && \text{when he is honest and lazy, and} \\ U(C) &\equiv u(e_L) + (1-p)(1-\pi)t + b - \pi F && \text{when he is corrupt.} \end{aligned}$$

The agent chooses the action which maximises his expected utility conditional on the values of π and t; therefore, he works hard if and only if t satisfies the new incentive compatibility constraint:

$$t \geq t^*(\pi) \equiv \max(t_a, t_c(\pi)), \tag{8}$$

where[25]

$$t_c(\pi) \equiv \frac{\Delta u + b - \pi F}{2p - 1 + \pi(1-p)}. \tag{9}$$

Since the agent rejects the bribe when he works hard, $t^*(\pi)$ is a *corruption-proof* incentive scheme.[26] It can be shown (see Appendix) that $t_c(\pi)$ is a continuous, twice differentiable, strictly decreasing and convex function of π, and that $t^*(\pi) = t_c(\pi)$ if and only if π is not larger than some positive value π^*, which we assume to be lower than unity. Moreover, $t^*(\pi)$ is continuous, non-increasing and convex, and $t_c(0)$ is equal to t_b.

[25]The agent prefers (H) to (L) whenever $U(H) \geq U(L)$, which occurs if and only if $t \geq t_a$ (see the previous section); he prefers (H) to (C) whenever $U(H) \geq U(C)$, which yields:

$$u(e_H) + pt \geq u(e_L) + (1-p)(1-\pi)t + b - \pi F,$$

whence:

$$[p - (1-p)(1-\pi)]t \geq u(e_L) - u(e_H) + b - \pi F,$$

which is equivalent to:

$$t \geq t_c(\pi) \equiv \frac{\Delta u + b - \pi F}{2p - 1 + \pi(1-p)}.$$

As a result, the agent prefers (H) to any other alternative if and only if $t \geq \max(t_a, t_c(\pi))$. Q.E.D.

[26]Corruption-proof incentive schemes where the agent works lazily can be designed when $\pi > \pi^*$, but the principal has no reason to adopt them, since he is only concerned with the work effort of the agent, and not directly affected by the bribe.

Given the probability of detection, the agent works lazily whenever $t < t^*(\pi)$, and accepts the bribe only if $t < (b - \pi F)/[\pi(1-p)]$.[27] If $t \geq t^*(\pi)$, the agent works hard and rejects the bribe. Since $t^*(\pi)$ is decreasing when $\pi < \pi^*$ and constant thereafter, the incentives required to induce the agent to work hard diminish down to a minimum as the probability of detection increases; intuitively, the more likely the agent is to be caught when he accepts the bribe, the more inclined he will be to be honest and work hard. The minimum incentive-compatible bonus is reached when $\pi \geq \pi^*$ and is equal to t_a, which is the same bonus that is required to induce the agent to work hard in a world without corruption.

The optimal choice for the agent is illustrated in Figure 1. In Zone A, both π and t are too low, and the agent works lazily and takes the bribe (C); in Zone B, π is high enough to induce honesty but t is too low to induce hard work, and the agent chooses (L); in Zone C, finally, both π and t are sufficiently high to induce the agent to work hard and be honest (H).

3.1 Equilibria without monitoring costs

The optimal choice for the principal depends on which variables he controls and on which costs he has to pay. In this subsection, we assume that the principal can only decide the amount of the bonus, and has no control over monitoring. Monitoring is delegated to an independent authority like the judiciary, paid by the Government out of ordinary taxes. In the next subsection, we shall assume that the intensity of monitoring, and thus the probability of detection, are decided by the principal, who has to pay for their costs.

Since the utility of the principal does not directly depend on the bribe, the principal is only concerned with the quality of the agent's work, and with the amount of money he has to pay to achieve it; the principal offer the incentive-compatible bonus if and only if $pt^*(\pi) \leq \Delta w$, otherwise he offers no bonus and lets the agent work lazily and accept the bribe.[28]

When $\pi \geq \pi^*$ the incentive-compatible bonus $t^*(\pi)$ is equal to t_a, which is by assumption lower than $\Delta w/p$; therefore, the principal offers it; when $\pi = 0$, then $t^*(\pi) = t_b$, which is higher than Δw by assumption (6); therefore, the principal offers no bonus. Continuity and monotonicity of $t^*(\pi)$ in the interval $[0, \pi^*]$ thus ensure that there is a unique value $\pi' \in (0, \pi^*)$ such that $pt^*(\pi') = \Delta w$; the principal offers a positive bonus $t = t^*(\pi)$ if and

[27] The agent prefers (C) to (L) if and only if $U(C) < U(L)$, which yields:

$$u(e_L) + (1-p)t < u(e_L) + (1-p)(1-\pi)t + b - \pi F \tag{10}$$

whence: $\pi(1-p)t < b - \pi F$, which yields: $t < (b - \pi F)/[\pi(1-p)]$. Q.E.D.

[28] The expected utility of the principal if he offers $t = t^*$ is equal to $W^o(\pi) \equiv w(e_H) - pt^*(\pi)$, while his expected utility if he offers $t = 0$ is equal to $W(L) \equiv w(e_L)$. The principal thus offers $t = t^*$ if and only if $W^o(\pi) \geq W(L)$, which implies $pt^*(\pi) \leq \Delta w$.

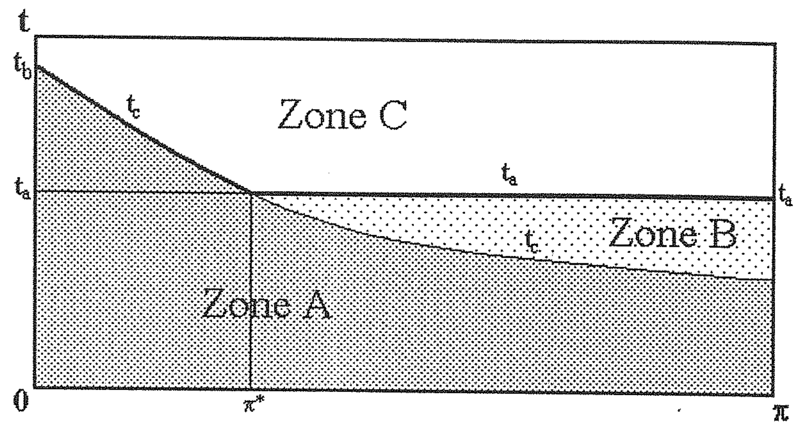
t
t_b
t_c
Zone C
t_a
t_a
t_a
Zone B
t_c
Zone A
0
π^*
π

Fig. 1

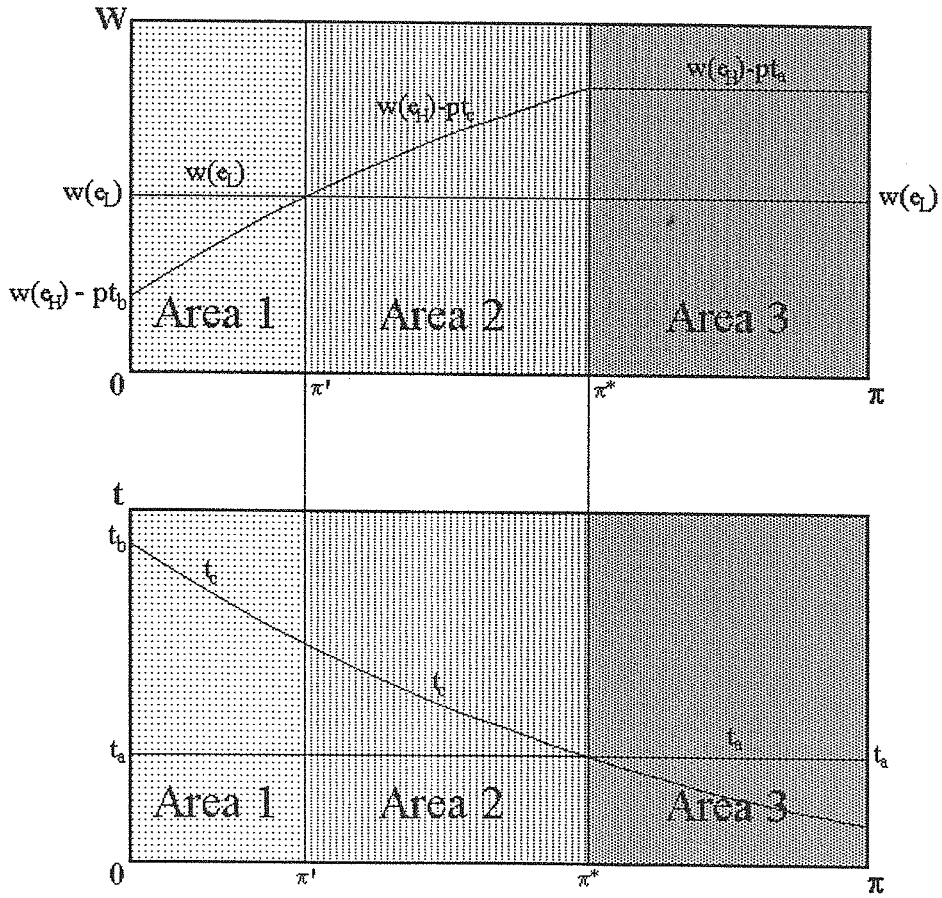
W
$w(e_H)-pt_a$
$w(e_H)-pt_c$
$w(e_L)$
$w(e_L)$
$w(e_L)$
$w(e_H) - pt_b$
Area 1
Area 2
Area 3
0
π'
π*
π
t
t_b
t_c
t_c
t_a
t_a
t_a
Area 1
Area 2
Area 3
0
π'
π*
π

Fig. 2

only if $\pi \geq \pi'$, and no bonus otherwise.

The interval $[0, 1]$ of feasible values of the probability of detection can thus be divided into three areas, corresponding to three different strategies and equilibria (see Figure 2). When the probability of detection is lower than π' (Area 1), the cost of inducing the agent to work hard is too high, and the principal prefers to offer no bonus and let him work lazily; following the suggestion of Besley and McLaren (1993), we may call this a "capitulation equilibrium", because the principal gives up all attempts to induce the agent to work hard; in this equilibrium corruption actually occurs. When the probability of detection lies between π' and π^* (Area 2), the cost of obtaining hard work from the agent is sufficiently low and the principal accepts to bear it; the bonus, however, is higher than t_a, which enables the agent to extract rent from the principal; we shall call this the "stick-and-carrot equilibrium", since the agent is induced to work hard by a mix of threat of punishment and increases in the bonus; corruption does not occur, but its effect is to enrich the agent at the expense of the principal; in this case, we say that corruption is "hidden". When π is not lower than π^*, finally, a bonus equal to t_a is sufficient to induce the agent to work hard; we call this the "deterrence equilibrium", because the threat of punishment is sufficiently high to convince the agent to reject the bribe without obtaining any rent. Corruption is "repressed", because its theoretical possibility does not enable the agent to obtain a higher utility.

In the discussion that follows, we qualify these equilibria in terms of the expected utility of the principal and the agent. We ignore the expected utility of the corruptor, for three different reasons. First of all, our purpose is to examine the effects of corruption (actual or hidden) on the initial relation between the principal and the agent; we are only marginally interested in the role played by the corruptor, and we are only concerned about whether or not he intervenes in the relation. Furthermore, in our model the bribe is equal to the corruptor's willingness to pay, which leaves the latter indifferent about whether or not the agent works lazily and accepts the bribe. The corruptor would enjoy a higher utility only if the agent worked lazily rejecting the bribe, but this outcome does not occur in equilibrium. In addition, the benefits that the corruptor receives as a result of the bribe may be considered unworthy of social protection; for instance, if corruption increases the opportunities for illegal pollution, the benefit of the company which pays the bribe and pollutes could hardly be considered as a positive component of social welfare.

Figure 3 shows the expected utility of the principal and the agent (hereafter "the parties") in the three areas of π. The Diagram (a) shows the expected utility of the principal, when he offers no bonus ($t = 0$: straight horizontal line) and when he offers a bonus equal to $t^*(\pi)$ (curve); as the diagram shows, the principal maximises his expected utility by setting $t = 0$ when $\pi < \pi'$ (capitulation equilibrium) and $t = t^*(\pi)$ otherwise. Diagrams (b.1), (b.2) and (b.3) show the expected utility of the agent, respectively, when Condition 1 holds, when Condition 2 holds and when both Conditions fail, Condition 2 being as follows:

Condition 2 *Inequality (6) holds; moreover, Condition 1 fails and*

$$b > (\pi' F) \frac{p}{(1-p)(1-\pi')} - \Delta u. \tag{11}$$

Diagram (a)

Expected utility of the principal

Diagram (b.1)

Expected utility of the agent when Condition 1 holds

Diagram (b.2)

Expected utility of the agent when Condition 2 holds

Diagram (b.3)

Expected utility of the agent when Conditions 1 and 2 fail

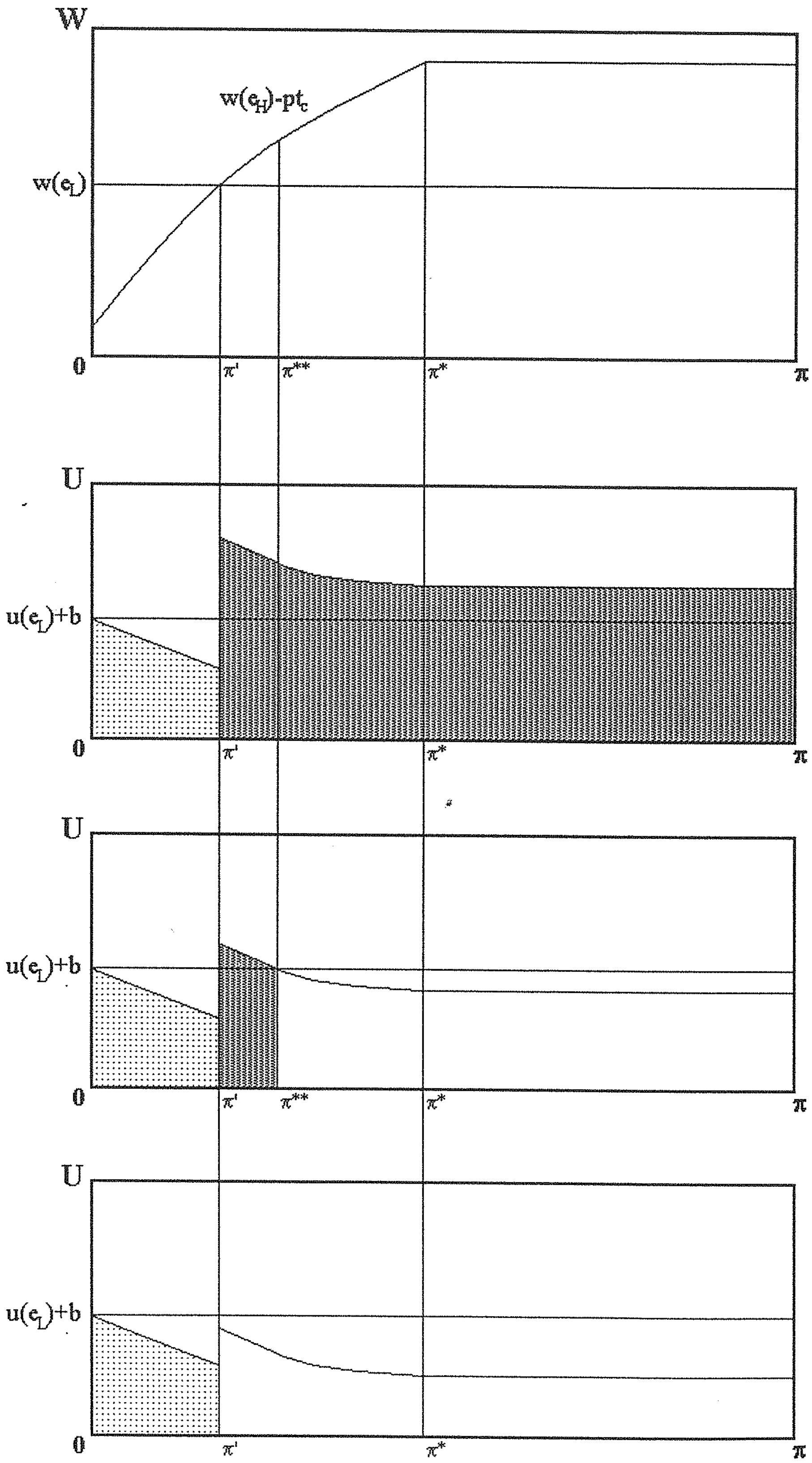

Fig. 3

The implications of this condition are discussed in the Appendix. The meaning of these Conditions is evident from the diagrams, where the straight, downward-sloping line shows the expected utility of the agent when $t = 0$, while the curve and horizontal line represent his expected utility when $t = t^*(\pi)$. When Condition 1 holds, the agent strictly prefers a situation with hidden or repressed corruption ($\pi > 0$ and $t = t^*(\pi)$) to a capitulation equilibrium where corruption occurs but is not monitored and no bonus is offered ($\pi = t = 0$); when Condition 2 holds, the agent prefers hidden or repressed corruption if and only if π is sufficiently low, otherwise he prefers the capitulation equilibrium; when both conditions fail, the agent strictly prefers the capitulation equilibrium.

Since the expected utility of the parties in a capitulation equilibrium is the same as when corruption is legally allowed (corruption is neither monitored nor punished, therefore the norm forbidding it is economically irrelevant), one result is immediately evident: while the principal weakly prefers to ban corruption, and strictly prefers to ban it whenever $\pi \geq \pi'$, the agent may agree or disagree with him depending on the two above conditions and on the exogenous value of the probability of detection; in particular, the agent will always support the ban when Condition 1 holds, and will continue to do so under Condition 2 provided that π is lower than some positive value $\pi^{**} > \pi'$, such that $u(e_L) + b = u(e_H) + pt_c(\pi^{**})$. In all other cases, the agent will oppose the ban, strictly preferring a capitulation equilibrium.

These results prove the following Proposition:

Proposition 3 *The introduction of a norm prohibiting corruption is opposed by the agent and is not supported by the principal when $0 < \pi < \pi'$, while it is unanimously supported under Condition 1 when $\pi \geq \pi'$, and under Condition 2 when $\pi' \leq \pi \leq \pi^{**}$. In all other cases, the principal supports the norm and the agent opposes it.*

PROOF: see Appendix.

The case when principal and agent agree to reject a norm against corruption is shown by the dotted areas in Figure 3; the case when they agree to introduce the norm is shown by the shaded areas. In both cases, the adoption of a ban can be decided on efficiency considerations; the parties agree to introduce or reject it according to which alternative yields a Pareto-efficient outcome. In the remaining areas, an agreement cannot be reached on mere efficiency grounds, since the introduction of the norm induces a movement along the Pareto-efficient frontier, benefitting one party at the expense of the other; the norm banning corruption thus becomes a political issue, involving a conflict over the distribution of rent between different members of society; its resolution depends on the relative bargaining power of the two parties, or on their comparative influence over the decision-making authority.

Figure 3 also allows a comparison between the equilibria that arise when corruption is possible (and detected with positive probability), and the equilibrium that arises in a world without corruption, which was discussed in Section 3 and is equivalent to the deterrence equilibrium in a world with corruption.

As Figure 3 shows, when $\pi < \pi^*$ the principal is always worse off in a world with corruption. If he offers a bonus equal to $t^*(\pi)$, he obtains the same work effort at a higher expected cost; if he offers no bonus at all, he obtains a lower work effort; in both cases, his expected utility is lower than in a world without corruption. The agent, on his part, is always better off in a situation of hidden corruption ($\pi' \leq \pi < \pi^*$), where he performs the same work effort and obtains a higher bonus than in a world without corruption. In this case the agent gains what the principal loses, and aggregate utility is the same as in a world without corruption; hidden corruption induces a redistribution of expected utility between the two parties, which is decreasing with the probability of detection:

Proposition 4 *Compared to the equilibrium in a world without corruption, hidden corruption with costless monitoring enables the agent to extract rent form the principal, without affecting the agent's work effort or the aggregate utility of the two parties; moreover, the extent of the redistribution in expected utility is decreasing in the probability of detection.*

PROOF:

Hidden corruption occurs if and only if $\pi' \leq \pi < \pi^*$, in which case the principal offers a bonus equal to $t_c(\pi)$ and the agent works hard and obtains a rent equal to $p(t_c - t_a) > 0$; the agent's expected utility is equal to $u(e_H) + pt_c$, which is higher than his expected utility in a world without corruption; the principal's expected utility, in turn, is equal to $w(e_H) - pt_c$, which is lower that his expected utility in a world without corruption. The agent thus increases his expected utility at the expense of the principal, while the sum of their expected utility remains equal to $w(e_H) + u(e_H)$, the same as in a world without corruption. The extent of the redistribution of expected utility from the principal to the agent is equal to $p[t_c(\pi) - t_a]$, which is a decreasing function of π.

Q.E.D.

When $\pi < \pi'$, corruption actually occurs, but the agent and the principal would be both weakly better off if the norm banning corruption were abolished. In this situation, the agent might in some cases prefer to be living in a world without corruption; in other cases, however, he prefers the existing corrupt world (the alternative depending on Conditions 1 and 2 as well as on the actual value of π, as shown in Figure 3). Aggregate utility, however, is lower when corruption actually occurs, being equal to $w(e_L) + u(e_L)$, which is by assumption lower than $w(e_H) + u(e_H)$.[29]

3.2 Equilibria with positive monitoring costs

Assume, now, that monitoring corruption is costly, and that its costs are related to the probability of detection by a positive, continuous, twice differentiable function $c(\pi)$, with the following properties: $c(0) = 0$ (no fixed costs); $c'(\pi) > 0$ (positive marginal costs);

[29] Assumption (3), given that $p/(2p-1) < 1$, implies that $\Delta W \geq \Delta U$, which in turn implies $w(e_H) + u(e_H) > w(e_L) + u(e_L)$.

$c''(\pi) > 0$ (diminishing returns). These costs are borne by the principal, and enter his utility function in the same way as the payment he makes to the agent; on the other hand, the principal can now choose the intensity of monitoring, and thus the probability of detection.

The principal is now confronted with two decisions, concerning the optimal probability of detection and the optimal amount of the bonus. While a higher probability of detection involves higher monitoring costs, it also reduces the incentive-compatible bonus that induces the agent to work hard.

Let $W^*(\pi) \equiv w(e_H) - pt^*(\pi) - c(\pi)$ be the expected utility of the principal when he agrees to pay this bonus and let $W(\pi) \equiv w(e_L) - c(\pi)$ be his expected utility when he pays no bonus at all. It can be shown that $W^*(\pi)$ is continuous and concave, and that $W^*(\pi) > W(\pi)$ if and only if $\pi > \pi'$ (see Appendix).

Concavity of $W^*(\pi)$ in the interval $[\pi', 1]$ ensures that it is maximised by a unique value $\pi^o \in [\pi', 1]$; moreover, π^o cannot be larger than π^*, since $W^*(\pi)$ is strictly decreasing in the interval $[\pi^*, 1]$ (a higher probability of detection in this interval only involves higher monitoring costs without allowing any reductions in the incentive-compatible bonus).[30] π^o therefore lies within the interval $[\pi', \pi^*]$; if it is an internal point, it is implicitly identified by the FOC:

$$c'(\pi) = -p\frac{dt_c}{d\pi} \equiv \frac{p}{2p - 1 + \pi(1-p)}[F + (1-p)t_c(\pi)] \tag{12}$$

(see Appendix).

$W(\pi)$ is instead strictly decreasing in π, because when the agent works lazily an increase in the probability of detection involves higher monitoring costs without any benefits for the principal; the maximum of $W(\pi)$ thus lies at $\pi = 0$. The expected utility of the principal is therefore maximised either at $\pi = \pi^o$, or at $\pi = 0$, or at both values. These three possibilities are shown in Figure 4. The top diagram shows the first, when the principal sets π equal to π^o and offers a bonus equal to $t_c(\pi^o)$, while the agent works hard and corruption remains hidden; the middle diagram shows the second possibility, when the principal sets $\pi = 0$ and offers no bonus, while the agent works lazily and corruption occurs; the bottom diagram, finally, shows the third possibility, when $W(0) = W^*(\pi^o)$ and the principal is indifferent between the two alternatives, choosing either with equal probability although they involve a completely different work effort and corruption pattern; in this case, the system exhibits multiple equilibria (namely, two), as is described by the following Proposition:

Proposition 5 (Multiple equilibria) *When*

$$\Delta w = (\Delta u + b - \pi^o F)\frac{p}{2p - 1 + \pi^o(1-p)} + c(\pi^o)$$

[30] In this interval, $t^*(\pi)$ is constant and equal to t_a, while $c(\pi)$ is increasing in π.

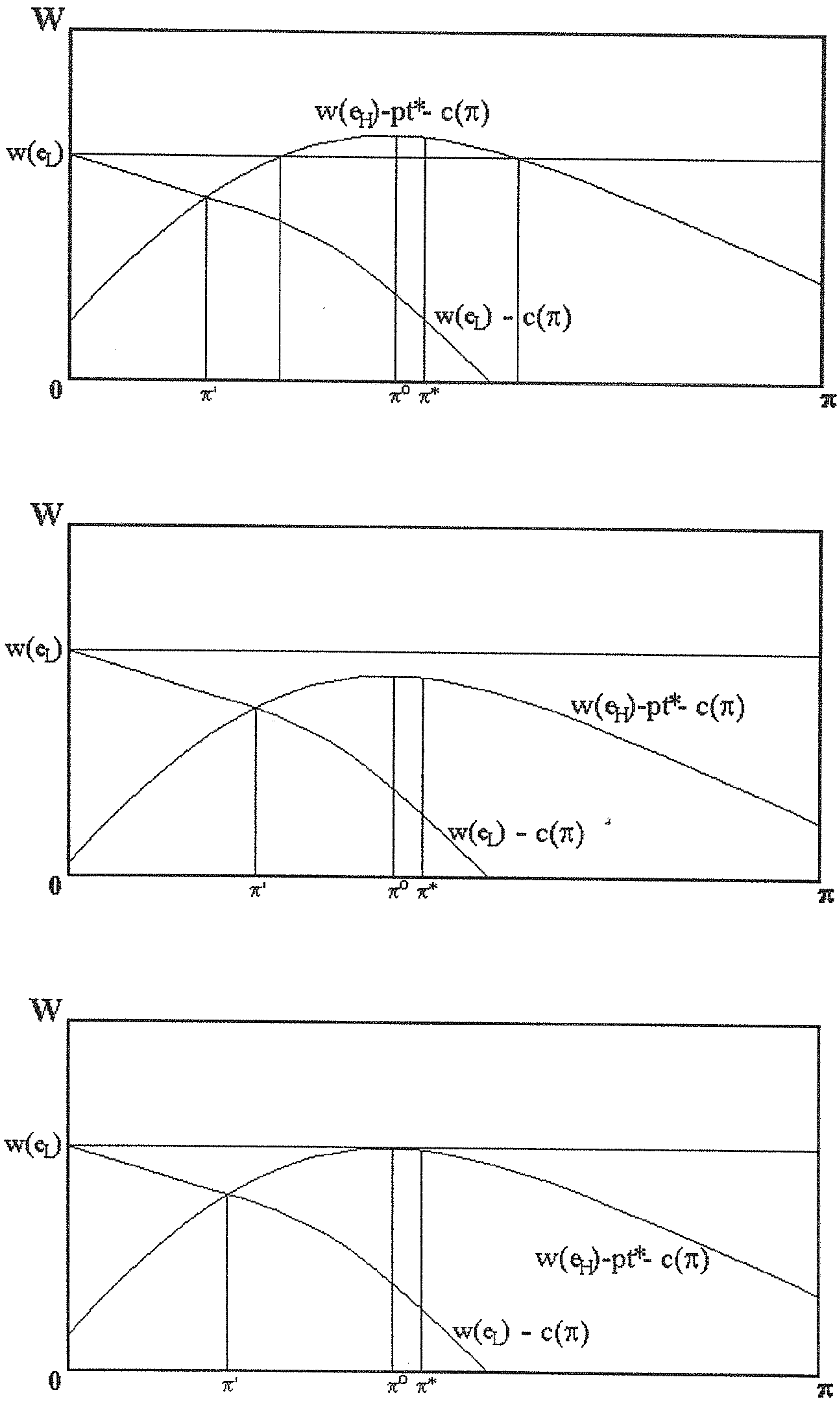
W
w(e_L)
w(e_H)-pt*- c(π)
w(e_L) - c(π)
0
π'
π^0 π*
π
W
w(e_L)
w(e_H)-pt*- c(π)
w(e_L) - c(π)
0
π'
π^0 π*
π
W
w(e_L)
w(e_H)-pt*- c(π)
w(e_L) - c(π)
0
π'
π^0 π*
π

Fig. 4

the system exhibits two equilibria, one of which involves hard work and hidden or repressed corruption while the other involves a lazy work effort while corruption actually occurs.

PROOF:

By equation (9), the equality in Proposition (5) is equivalent to: $\Delta w = pt_c(\pi^o) + c(\pi^o)$, which implies: $W^*(\pi^o) \equiv w(e_H) - pt_c(\pi^o) - c(\pi^o) = w(e_L) \equiv W(0)$. In this case, therefore, the expected utility of the principal is the same if he sets $\pi = \pi^o$ as if he sets $\pi = t = 0$; the principal may choose either of these alternatives, and two equilibria thus exist. In one of these equilibria, the principal pays the incentive-compatible bonus and the agent works hard and rejects the bribe, while in the other the principal pays no bonus and the agent works lazily and accepts the bribe.

Q.E.D.

Using the terminology adopted in the previous subsection, these equilibria can be distinguished into three types: a deterrence equilibrium (when $\pi = \pi^o = \pi^*$), stick-and-carrot equilibria (when $\pi' \leq \pi = \pi^o < \pi^*$) and a capitulation equilibrium (when $\pi = 0$). The deterrence equilibrium occurs when $\pi^o = \pi^*$ and $W^*(\pi^*) \geq w(e_L)$, since the principal sets π equal to π^* and t equal to t_a. In this equilibrium, corruption does not actually occur (since the threat of punishment is too high), and is repressed rather than hidden because its threat does not allow the agent to extract rent from the principal; the principal, however, still suffers from the mere possibility of corruption, because this possibility induces him to pay for a high level of monitoring. The stick-and-carrot equilibria occur when $\pi^o < \pi^*$ and $W^*(\pi^*) \geq w(e_L)$, since the principal sets π below π^* and t equal to t_c, greater than t_a; in these equilibria, corruption is hidden and the agent is able to extract rent. The capitulation equilibrium finally occurs $W^*(\pi^*) \leq w(e_L)$; in this case, the principal sets $\pi = t = 0$ whatever the value of π^o, the agent works lazily and corruption occurs. Notice that the expected utility of the principal is always maximised in equilibrium, since it is the principal who chooses the type of equilibrium by deciding the level of monitoring.

The expected utility of the agent, instead, is always higher in a stick-and-carrot equilibrium than in a deterrence equilibrium, while a comparison with the capitulation equilibrium depends on Conditions 1 and 2. When Condition 1 holds, the capitulation equilibrium is the worse for the agent; when Conditions 1 and 2 both fail, the capitulation equilibrium is the best; while when Condition 2 holds the capitulation equilibrium is preferred to the deterrence equilibrium, and it is also preferred to the stick-and-carrot equilibrium if and only if $\pi^o \geq \pi^{**}$ (see Figure 5).

Under particular conditions, the agent may be indifferent between a stick-and-carrot and a capitulation equilibrium; if the principal is also indifferent between them, two equilibria exist which yield the same expected utility to both parties but involve a completely different behaviour on the part of the agent:

Proposition 6 *When $W^*(\pi^o) = w(e_L)$ and $\Delta u + b = pt_c(\pi^o)$, two equilibria exist which are equivalent for both the principal and the agent; one of these equilibria involves a lazy work*

Diagram (a)

Expected utility of the principal

Diagram (b.1)

Expected utility of the agent when Condition 1 holds

Diagram (b.2)

Expected utility of the agent when Condition 2 holds

Diagram (b.3)

Expected utility of the agent when Conditions 1 and 2 fail

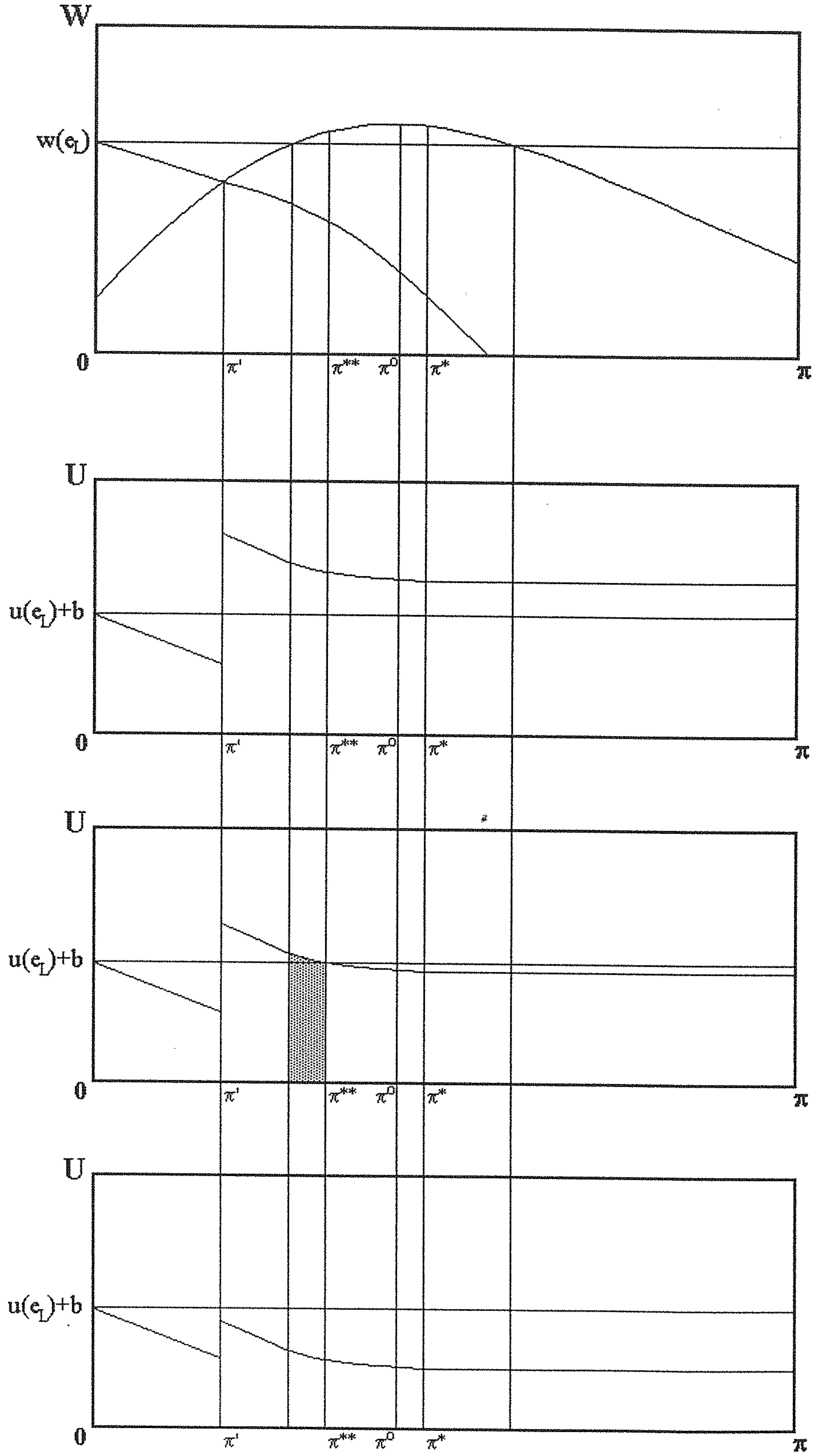

Fig. 5

effort and the actual occurrence of corruption, while in the other equilibrium corruption is hidden or repressed and the agent works hard.

PROOF: see Appendix.

Some of the implications of this proposition will be discussed in the following section. Consider now the attitudes of the two parties towards the introduction of the legal norm prohibiting corruption. Like in the previous subsection, we shall ignore the expected utility of the corruptor. Once corruption is legally prohibited, the level of monitoring is unilaterally decided by the principal, but at an earlier stage the agent may have some power in influencing the decision on whether or not corruption should be legally banned. As in the case where monitoring is costless, the equilibrium that is reached without a norm prohibiting corruption is equivalent to the capitulation equilibrium.

Four different cases are thus possible:

(i) when $w(e_L) > W^*(\pi^o)$, the introduction of the norm leads to a capitulation equilibrium and therefore produces no effects (since its violations are not monitored); both parties are thus indifferent towards it;

(ii) when $w(e_L) < W^*(\pi^o)$ and Conditions 1 and 2 fail, the introduction of the norm induces a stick-and-carrot or a deterrence equilibrium but the agent prefers a capitulation equilibrium; the principal is in favour of the norm and the agent is against it;

(iii) when $w(e_L) < W^*(\pi^o)$ and Condition 1 holds (or Condition 2 holds and $\pi^o \leq \pi^{**}$), the norm leads to a stick-and-carrot or a deterrence equilibrium which both parties prefer to a capitulation equilibrium; both parties are thus in favour of the norm;

(iv) when $w(e_L) < W^*(\pi^o)$, Condition 2 holds and $\pi^o > \pi^{**}$, the norm leads to a stick-and-carrot or a deterrence equilibrium which, from the point of view of the agent, is worse than a capitulation equilibrium; the agent opposes the norm and the principal supports it. In this case, however, if π were to be set at or below π^{**}, the agent would be better off with the norm than without it (in other words, stick-and-carrot equilibria exist which are preferred by the agent to a capitulation equilibrium); as a result, the two parties may agree to introduce the norm, provided that the principal commits to set $\pi \leq \pi^{**}$.

A legal norm prohibiting corruption would be ineffective in case (i) while in case (iii) it induces a Pareto-improvement; in cases (ii) and (iv), the introduction of the norm benefits the principal and reduces the expected utility of the agent, becoming an issue of social conflict. Case (iv) however is rather peculiar, because it involves the possibility of Pareto-improvements by compromise. The parties may thus engage in bargaining over the appropriate value of π in order to reach an agreement that would improve the expected utility of both, where π would be set equal to or below π^{**}, in the shaded area in Diagram (b.2) of Figure 5.

Compare, now, the three types of equilibria that arise in a world with corruption and costly monitoring, with the equilibrium that is reached in a world without corruption. Unlike when monitoring is costless, this equilibrium is no longer equivalent to a deterrence equilibrium, because the latter involves positive monitoring costs which reduce the expected utility of the principal and is, therefore, weakly Pareto-inferior. In a stick-and-carrot equilibrium, the agent perceives a higher expected utility than in a world without corruption while the expected utility of the principal is lower, as a result both of positive monitoring costs and of a higher level of the bonus; aggregate utility is also lower, due to the presence of positive monitoring costs. In the capitulation equilibrium monitoring costs are nil but the expected utility of the principal is lower than in a world without corruption; the agent's utility may be higher or lower depending on Condition 1, and aggregate utility is lower, as shown in the previous subsection.

This yields the following Proposition:

Proposition 7 *When monitoring is costly, the possibility of corruption always reduces aggregate utility even when corruption is hidden or repressed. When corruption is hidden, the agent is able to extract rent from the principal and his expected utility is higher than in a world without corruption, but his gain is lower than the loss incurred by the principal.*

PROOF:

In a world without corruption aggregate utility is equal to $w(e_H) + u(e_H)$. In a capitulation equilibrium, aggregate utility is equal to $w(e_L) + u(e_L)$, which is lower than $w(e_H) + u(e_H)$ by assumption (see (3)). In all other equilibria, aggregate utility is equal to $w(e_H) + u(e_H) - c(\pi)$, which is also lower than $w(e_H) + u(e_H)$. Hidden corruption occurs in a stick-and-carrot equilibrium, where the agent works hard and rejects the bribe but receives a higher bonus than in a world without corruption, perceiving a rent equal to $pt_c(\pi^o) - pt_a$ and thus enjoying a higher expected utility than in a world without corruption; the principal's loss however is larger than the agent's gain, because, on top of paying the bonus, the principal has to pay for the cost of monitoring.

Q.E.D.

4 Discussion

The results of the previous two sections support our initial statement, that most economic effects of corruption derive from its potential performance, rather than from its actual occurrence. Of the three types of equilibria discussed above, only the capitulation equilibrium involves the actual performance of corruption; the other equilibria, however, involve substantial economic effects, with the only exception of the deterrence equilibrium with costless monitoring, which guarantees an effective immunisation of the principal-agent relation from the effects of the corruptor's intervention.

This fact bears important empirical and normative implications. From an empirical point of view, particular care should be used in the interpretation of the statistics about corruption. Since the economic impact of corruption depends on its potential rather than on its actual occurrence, these statistics (or estimates) should be supplemented by additional information concerning the incentives facing the agent, the cost of monitoring and the probability of detection, or, in alternative, by some measure of the existing opportunities for corruption. Ignoring these aspects would generally lead to an underestimation of the true economic impact of corruption.

From a normative perspective, our results suggest that the design of effective anti-corruption strategies cannot only be centered on repressive measures, and should include substantial preventive interventions. Repressive actions, like the detection and prosecution of violations, can be quite effective at reducing the actual occurrence of corruption, but any long-term reduction in the overall social costs of corruption can only be achieved by reducing its opportunities, and this is more likely to be achieved by a preventive approach. Effective strategies in this sense may include organisational reforms and a redefinition of roles in order to reduce the areas of discretionary power, improvements in the design of regulations that would render their implementation less arbitrary, procedures and rules that would increase the accountability of the officials and bureaucrats, as well as educational initiatives aimed at strengthening the values of honesty and loyalty among the citizens.

An illustration of these points may be provided by the experience of the Independent Commission Against Corruption (ICAC) in Hong Kong, an independent monitoring authority that was set up in 1973 to combat the widespread corruption in the police force in the colony. The ICAC was articulated in three Departments, each responsible for a different type of interventions; the Operations Department (OD) carried out repressive actions, like investigations, arrests and prosecutions; the Corruption Prevention Department (CPD) operated preventive interventions like the evaluation of the organisational structure of the police and the elaboration of suggestions for organisational improvements and the design of new rules and procedures; the Community Relations Department (CRD), finally, pursued a strategy of sensibilisation and involvement of the citizens, including initiatives to inform the public and lectures in the schools. In the end, the preventive actions of the CPD and CRD played an essential role in reducing corruption to substantially lower levels (see Klitgaard, 1988).

Our results also provide some indications about the attitudes that the various parties may take towards the introduction of these measures. As Proposition 2 shows, there are circumstances in which both the principal and the agent are damaged by the intervention of the client, in which case, under the conditions outlined in Proposition 3, both parties would agree on the adoption of a legal norm prohibiting corruption. In other situation, however, corruption is an issue of social conflict over the distribution of rent, and any measures aimed at reducing it can be considered as an attempt by one of the parties to resolve the conflict in his favour. These considerations suggest an economic interpretation of the legal norm prohibiting corruption, as an instrument, used by either or both parties, to restore as much as possible the situation that preceded the intervention of the client. Remarkably, the need for such norm stems from a lack of credibility of any promise by the agent to reject the offer of the client, in the presence of restrictions over the set of feasible incentive schemes.

The possibility of multiple equilibria, shown by Proposition 5, proposes additional questions. While these equilibria involve different patterns of corruption and work effort, by definition they yield the same expected utility to the principal, and by Proposition 6 they may yield the same expected utility also to the agent. One implication is that the principal, and possibly the agent too, may perceive a similar level of expected utility in situations which are rather different from an observational point of view. Moreover, minor variations in some parameters inducing a discrete jump between these equilibria may have only a marginal effect on the expected utility of the parties.

Applied to the historical record of some countries, these results provide somewhat peculiar suggestions. Public institutions and agencies in various countries (the police force in Hong Kong; the Internal Revenue Service in the Philippines; the Customs and Excise Department in Singapore: see Klitgaard, 1988) have experienced a marked reduction in their observed corruption levels, following long periods during which corruption had been widespread and persistent. According to the above considerations, these changes may have derived not so much from a final success in correcting previous distortions and irrationalities, but from a perfectly rational decision to reduce the level of actual corruption (and replace it with hidden or repressed corruption) at a particular moment in time. In fact, a continuous, and unobservable, incremental evolution of some parameters may have triggered a jump from a capitulation equilibrium to a stick-and-carrot (or deterrence) equilibrium, which, by Proposition 6, may have had a much less impressive impact on individual utility or social welfare than may appear at first sight.

The most remarkable example is maybe that of Hong Kong. While until the early 1970s the Hong Kong police force was renown for its high degree of corruption, this picture changed completely after the establishment in 1973 of the above mentioned ICAC, an external monitoring authority endowed with special powers and with a generous budget. Why was the ICAC established so late, and why were previous attempts to reduce corruption not nearly as successful? While learning effects and technological improvements may have played their part in the final success, the results of this paper suggest a more interesting interpretation: as a result of a continuous, incremental evolution of the economic and social structure of the colony, in the 1970s Hong Kong may have switched from a capitulation equilibrium to a deterrence equilibrium. The Hong Kong Government, playing the role of the principal, had become increasingly concerned that widespread corruption in the police could reduce the support for the Government among the population and hamper the ability of the police to guarantee security in the colony, with potential destabilising effects in the long term. These concerns grew over time as these effects accumulated, until the authorities decided that time had come for an expensive, but effective, anti-corruption drive. Our work suggests however that this rapid reduction in corruption may have had, at least in the short term, a lower impact on the utility of the Hong Kong Government than one may be led to expect.[31]

The results discussed above also suggest a new interpretation of the observed differences in the levels of corruption in otherwise similar countries (see Cadot, 1987, p. 223; Andvig and Moene, 1990, p. 63; Bardhan, 1997, p.1330). According to our results, these differences may be explained in terms of the selection of different equilibria in situations where multiple equilibria are present, or, more generally, in terms of the adoption of different equilibria induced by small, unobservable differences in the values of some parameters. Once again, such differences need not imply substantial differences in the levels of welfare, which may turn out to be similar in countries with deeply different levels of corruption. On the contrary, countries which exhibit similar, low levels of corruption may enjoy different levels of welfare, if the opportunities for corruption are low in one country and high in another. In fact, it is the magnitude of the opportunities for corruption, rather than its actual occurrence, that determines the overall impact of corruption on welfare.

Some care should be taken, however, in interpreting these results, in relations with various limits of our model. First of all, our results apply only to a three-stage non-repeated game with three players, and cannot immediately be extended to situations involving a large number of parties and several repetitions of the game (as is the case with countries). Moreover, several assumptions have been made which may reduce the significance of our results. The role of the corruptor here is reduced to offering a payment to the agent conditional on a particular work effort; although this description is sufficient to configure corruption, more robust results would require a deeper analysis of this role, including an examination of the preferences of the client and of the quality of information that is available to him. Moreover, in our model both principal and agent have perfect information about the structure and parameters of the model, including the existence or non-existence of the client, the amount of the bribe, and the probability with which the actual payment of the bribe can be detected. In real life, information about these elements may be imperfect and asymmetric, and the formalisation of these aspects may enable to reach more significant conclusions.

Finally, a more complete analysis of the structure of corruption may be obtained by

[31] A discussion of the increasing concerns with corruption by the Hong Kong authorities can be found in Klitgaard (1988). Other examples discussed by Klitgaard include the successful effort by Justice Plana to reduce corruption in the Internal Revenue Service of the Philippines, and the reduction of corruption in the Customs and Excise Department in Singapore, brought about by the establishment of the Corrupt Practices Investigation Bureau (CPIB) which later inspired the ICAC in Hong Kong. More problematic would be an evaluation of the recent *mani pulite* scandal that took place in Italy in 1992, when a small group of public prosecutors brought to court eminent politicians, officials and businessmen on charges of corruption. While the lasting effects of this scandal are yet to appear, this case can be interpreted as a decision by the principal (consisting of several institutional authorities who had previously refrained from similar investigations, and of the population at large, who strongly supported the effort of the prosecutors), to switch from a capitulation equilibrium to a deterrence equilibrium; as in Hong Kong, this decision was probably triggered by a growing concern about several negative effects of corruption.

considering additional features, like the issue of the division of surplus between the agent and the corruptor, and the problems of enforcement of their illegal contract. In this paper, we assumed that the agent extracts all the surplus, and any problems of enforcement were assumed away by presuming that the corruptor had a credible reputation for honouring its promises; in other works, the division of surplus has been formalised in more detail, usually in terms of a Nash bargaining solution; further analysis may show that this issue could play a major role in determining the equilibrium outcome. Similar considerations may apply to the issue of enforcement, which is not much discussed in the literature; the adoptions of different methods to solve this problem may induce a qualitatively different behaviour in all the parties and lead to different outcomes. Further research along these lines may provide a deeper insight into corruption.

5 Conclusions

In this paper, we modelled corruption as a particular type of principal-agent-client relation. We showed that the potential for corruption arises when the client alters a pre-existing relation between the principal and the agent, leading to a Pareto-inefficient outcome or to a redistribution of rent in favour of the agent. Most economic effects arise from this perturbation, and persist even when the parties respond by banning the contract between the agent and the client. Although different types of equilibria arise in this case and only one of these types involves the actual performance of corruption, all these equilibria generally involve a different level of expected utility for each party with respect to the situation that preceded the intervention of the client.

In particular, although with a sufficiently high probability of detection corruption can be successfully prevented by a mix of incentives and punishment, the possibility of corruption enables the agent to extract rent from the principal, unless the risk of detection is particularly high. Even in this case, while the agent fails to extract rent, the principal incurs a loss if he has to pay for the cost of monitoring.

This bears relevant implications both on the normative and on the empirical side. On the normative side, repressive interventions aimed at discouraging corruption by increasing the threat of punishment appear to be less effective in reducing its economic impact, than preventive interventions aimed at reducing the opportunities for corruption, including educational programmes aimed at increasing loyalty and honesty. On the empirical side, cross-section or cross-country comparisons based on statistical data about the actual occurrence of corruption may yield inaccurate results, unless these data are properly interpreted and integrated by additional information.

This paper also shows how the introduction of a legal norm prohibiting corruption is sometimes supported by both principal and agent, while in other cases it involves a social conflict on the distribution of rent and a confrontation of bargaining power and political influence; in terms of efficiency and social conflict, an evaluation of corruption cannot be decided *a priori*, since corruption leads, in alternative situations, to Pareto-dominated outcomes or to a redistribution of rent between different parties.

The model shows that multiple equilibria may arise even in a static context, if monitoring corruption is costly for the principal. These equilibria always yield the same expected utility to the principal, and under certain conditions they yield the same expected utility also to the agent. These results provide interesting suggestions in the interpretations of the historical record of some countries. On one side, countries which have experienced remarkable reductions in corruption may have rationally decided to switch from an equilibrium with actual corruption to one where corruption is hidden or repressed, enjoying only minor increases in utility and social welfare; on the other side, countries with remarkable observable differences in the levels of corruption may in fact be enjoying similar levels of welfare, while countries with similar low levels of corruption may be affected by substantial differences in welfare, if corruption is completely absent in one country and costly prevented (hidden or repressed) in another.

References

Andvig, Jens Chr.; and Moene, Karl Ove (1990) "How Corruption May Corrupt," *Journal of Economic Behavior and Organization*; 13(1), 63-76.

Bac, Mehmet (1996) "Corruption and Supervision Costs in Hierarchies," *Journal of Comparative Economics*; 22(2), 99-118.

Bardhan, Pranab (1997) "Corruption and Development," *Journal of Economic Literature*; 35(3), 1320-1346.

Basu, Kaushik; Bhattacharya, Sudipto; and Mishra, Ajit (1992) "Notes on Bribery and the Control of Corruption," *Journal of Public Economics*; 48(3), 349-59.

Becker, Gray S. (1968) "Crime and punishment: An economic approach," *Journal of Political Economy*; 76, 169-217.

Becker, Gray S., and Stigler, G. J. (1974) "Law enforcement, malfeasance, and compensation of enforcers," *Journal of Legal Studies*, 3(1), 1-18.

Beenstock, Michael (1979) "Corruption and Development," *World Development*, 7(1), 15-24

Bernheim, B. Douglas; and Whinston, Michael D. (1986) "Common Agency," *Econometrica*; 54(4), 923-42.

Besley, Timothy; and McLaren, John (1993) "Taxes and Bribery: The Role of Wage Incentives," *Economic Journal*; 103(416), 119-41.

Bliss, Christopher, and Di Tella, Rafael (1997) "Does Competition Kill Corruption?" *Journal of Political Economy* 105(5), 1001-1023.

Cadot, Olivier (1987) "Corruption as a Gamble," *Journal of Public Economics*; 33(2), 223-44.

Chander, Parkash, and Wilde, Louis (1992) "Corruption in Tax Administration," *Journal of Public Economics*; 49(3), 333-49.

Friedrich, Carl J. (1989) "Corruption Concepts in Historical Perspective," in: Heidenheimer, Arnold J.; Johnston, Michael; and LeVine, Victor T. (eds.), *Political corruption: A handbook*, New Brunswick, N.J. and Oxford: Transaction Books, 1989, pages 15-24.

Hindricks, Jean; Keen, Michael; and Muthoo, Abhinay (1996) "Corruption, Extortion and Evasion," *mimeo.*

Klitgaard, Robert (1988) *Controlling corruption*, Berkeley and London: University of California Press, 1988, pages xiii, 220.

Kofman, Fred, and Lawarrée, Jacques (1996) "On the Optimality of Allowing Collusion," *Journal of Public Economics*; 61(3), 383-407.

Lowenstein, Daniel H. (1989) "Legal Efforts to Define Political Bribery," in: Heidenheimer, Arnold J.; Johnston, Michael; and LeVine, Victor T. (eds.), *Political corruption: A handbook*, New Brunswick, N.J. and Oxford: Transaction Books, 1989, pages 29-38.

Lui, Francis T. (1986) "A Dynamic Model of Corruption Deterrence," *Journal of Public Economics*; 31(2), 215-36.

Marjit, Sugata (1996) "A Note on Corruptible Hierarchy - Where do we Need the Honest Ones?," *Indian Statistical Institute Discussion Paper No. A-538*, July 1996, pages 12.

Mookherjee, Dilip, and Png, I. P. L. (1995) "Corruptible Law Enforcers: How Should They Be Compensated?," *Economic Journal*; 105(428), 145-59.

Paul, Chris, and Wilhite, Allen W. (1994) "Illegal Markets and the Social Costs of Rent-Seeking" *Public Choice*; 79(1-2), 105-15.

Rose-Ackerman, Susan (1989) "Corruption and the Private Sector," in: Heidenheimer, Arnold J.; Johnston, Michael; and LeVine, Victor T. (eds.), *Political corruption: A handbook*, New Brunswick, N.J. and Oxford: Transaction Books, 1989, pages 661-684.

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

Shleifer, Andrei; and Vishny, Robert W. (1993) "Corruption," *Quarterly Journal of Economics;* 108(3), 599-617.

Spiller, Pablo T. (1990) "Politicians, Interest Groups, and Regulators: A Multiple-Principals Agency Theory of Regulation, or 'Let Them Be Bribed,' " *Journal of Law and Economics;* 33(1), 65-101.

Tirole, Jean (1986) "Hierarchies and Bureaucracies: On the Role of Collusion in Organizations," *Journal of Law, Economics and Organization*; 2(2), 181-214.

Virmani, A. (1983) *The Microeconomics of a Corrupt Tax Bureaucracy*, World Bank Discussion Paper n. 54, pages 37.

A Appendix

A.1 Proof of Proposition 1

Equation (5) implies that the principal prefers to increase the bonus to t_b and thus induce the agent to work hard, than to offer no bonus at all and let the agent work lazily, accepting the client's gift. The expected utility of the agent thus increases to $U(H') \equiv u(e_H) + pt_b > u(e_H) + pt_a \equiv U(H)$, while the expected utility of the principal is reduced to $W(H') \equiv w(e_H) + pt_b < w(e_H) + pt_a \equiv W(H)$; the rent extracted by the agent is equal to:

$$pt_b - pt_a = \left(\frac{p}{2p-1}\right) b > 0.$$

The agent's work effort is not affected by the intervention of the client, since the agent continues to work hard; aggregate utility also remains unchanged at $w(e_H) + u(e_H)$. The intervention of the client thus causes a redistribution of expected utility between the principal and the agent. Q.E.D.

A.2 Proof of Proposition 2

If the principal responds to the possibility of corruption by raising t from t_a to t_b, corruption obviously reduces his expected utility, since the expected value of his payment to the agent increases, while the agent continues to work hard. If he responds setting $t = 0$, the agent works lazily and the expected utility of the principal becomes equal to $w(e_L)$, which is by assumption lower than $w(e_H) - pt_a$ (this is implied by inequality (2)). As a result, the intervention of the client always has a negative effect on the expected utility of the principal. When Condition 1 holds, the principal sets $t = 0$, since (6) implies that $w(e_H) - pt_b < w(e_L)$. The agent thus works lazily and takes the bribe, and his expected utility is equal to $u(e_L)+b$, which is lower than $u(e_H)+pt_a$ because Condition 1 requires that $b < (1-p)[\Delta u/(2p-1)] \equiv (1-p)t_a$, which implies: $u(e_L)+b < u(e_L)+(1-p)t_a = u(e_H)+pt_a$. Both the principal and the agent thus suffer a reduction in expected utility as a result of the intervention of the client; the new equilibrium is thus Pareto-inferior. Since the old equilibrium outcome is still feasible, the new equilibrium is inefficient. Q.E.D.

A.3 Properties of $t_c(\pi)$ in the interval $[0,1]$

By definition, $t_c(\pi)$ is equal to:

$$\frac{\Delta u + b - \pi F}{2p - 1 + \pi(1-p)}. \tag{13}$$

t_c is continuous in π in the interval $[0, 1]$, since both numerator and denominator in (13) are continuous in π, and the denominator is strictly positive in this interval. t_c is twice differentiable with respect to π; its first derivative is equal to:

$$t_c'(\pi) \equiv \frac{-F[2p-1+\pi(1-p)] - [\Delta u + b - \pi F](1-p)}{[2p-1+\pi(1-p)]^2} = \frac{-F - t_c(\pi)(1-p)}{2p-1+\pi(1-p)}, \qquad (14)$$

which is continuous in $[0,1]$; the second derivative is equal to:

$$t_c''(\pi) \equiv \frac{-(1-p)t_c'(\pi)[2p-1+\pi(1-p)] - [-F - t_c(\pi)(1-p)](1-p)}{[2p-1+\pi(1-p)]^2} =$$

$$= \frac{2(1-p)[F + t_c(\pi)(1-p)]}{[2p-1+\pi(1-p)]^2}, \qquad (15)$$

which is also continuous in $[0,1]$. Since t_c' is strictly negative in the interval $[0,1]$ (the denominator in (14) is negative and the numerator is positive in that interval), t_c is strictly decreasing in π. Moreover, since t_c'' is strictly positive, t_c is convex. Finally, $t_c(0)$ is equal to $(\Delta u + b)/(2p-1) \equiv t_b$. Q.E.D.

A.4 Properties of $t^*(\pi)$ in the interval $[0,1]$

By definition, $t^*(\pi)$ is equal to $\max(t_c(\pi), t_a)$. Since t_c is strictly continuous and monotonic and since $t_c(0) = t_b > t_a$, while $\lim_{\pi \to \infty} t_c(\pi) = -F/(1-p) < 0 < t_a$, some value π^* exists such that $t_c(\pi^*) = t_a$ and $t_c(\pi) > t_a$ if and only if $\pi < \pi^*$. The function $t^*(\pi)$ is thus equal to $t_c(\pi)$ in the interval $[0, \pi^*]$ and to t_a in the interval $[\pi^*, 1]$. $t^*(\pi)$ is continuous in the interval $[0,1]$, since it is continuous in the interval $[0,\pi^*]$ (where it is equal to $t_c(\pi)$), in the interval $[\pi^*, 1]$ (where it is equal to t_a), and at $\pi = \pi^*$ (where $t_c(\pi) = t_a$). $t^*(\pi)$ is non-increasing in π, since it is strictly decreasing in the interval $[0,\pi^*]$ and constant in the interval $[\pi^*, 1]$. $t^*(\pi)$ is twice differentiable in the interval $[0,1]$, with the exception of the point $\pi = \pi^*$, since both $t_c(\pi)$ and t_a are twice differentiable but their derivatives at $\pi = \pi^*$ are different. $t^*(\pi)$ is weakly convex since $t_c(\pi)$ is strictly convex and t_a is constant. Q.E.D.

A.5 Implications of Condition 2

Inequality (6) implies that the principal does not want to set t equal to $t^*(0) = t_b$ when $\pi = 0$, and would rather prefer a capitulation equilibrium; as a result, π' is strictly positive. Since Condition 1 fails, the agent strictly prefers a capitulation equilibrium where he works lazily and takes the bribe, getting no bonus and never being punished, to a deterrence equilibrium, where he works hard and rejects he bribe, getting a bonus equal to t_a. The inequality in (11) implies that, for sufficiently low values of π, the agent prefers a stick-and-carrot equilibrium to a capitulation equilibrium. The inequality in (11) can be rewritten:

$$\left(\frac{(1-p)(1-\pi')}{p-(1-p)(1-\pi')}\right) b > \left(1 - \frac{(1-p)(1-\pi')}{p-(1-p)(1-\pi')}\right)(\pi' F) - \left(\frac{(1-p)(1-\pi')}{p-(1-p)(1-\pi')}\right)(\Delta u), (16)$$

which implies:

$$\left(\frac{(1-p)(1-\pi')}{p-(1-p)(1-\pi')}\right)(\Delta u + b - \pi' F) > \pi' F.$$

Recalling that $t_c(\pi')$ is equal to $(\Delta u + b - \pi' F)/[p - (1-p)(1-\pi']$, this last inequality implies:

$$(1-p)(1-\pi')t_c(\pi') - \pi' F > 0,$$

which in turn implies:

$$u(e_L) + (1-p)(1-\pi')t_c(\pi') + b - \pi' F > u(e_L) + b. \tag{17}$$

The LHS in (17) is the expected utility that the agent obtains when he works lazily and accepts the bribe in a stick-and-carrot equilibrium where $\pi = \pi'$ and $t = t_c(\pi')$; by definition of t_c, this utility is the same as the agent would obtain if he worked hard rejecting the bribe under the same conditions. The RHS in (17) is the agent's expected utility when he works lazily and accepts the bribe in a capitulation equilibrium. In the end, (11) implies that, although the agent prefers the capitulation equilibrium to the deterrence equilibrium, at least one stick-and-carrot equilibrium exists that is preferred by the agent to any other type of equilibria. Since the expected utility of the agent is continuous in π, there is some value $\pi^{**} \in [\pi', \pi^*]$ such that the agent prefers a stick-and carrot equilibrium to a capitulation equilibrium if and only if $\pi \leq \pi^{**}$. Notice that Condition 1 implies (11), since it requires that $u(e_L) + b < u(e_H) + pt_a$, which in turn implies $u(e_L) + b < u(e_H) + pt_c$. Q.E.D.

A.6 Proof of Proposition 3

When $0 < \pi < \pi'$ the principal sets $t = 0$ and the agent works lazily. The introduction of a norm banning corruption thus reduces the expected utility of the agent by πF (which is the expected value of punishment), without increasing the expected utility of the principal. The agent thus opposes the norm, while the principal does not support it. When $\pi \geq \pi'$, the principal sets $t = t^*(\pi)$ and the agent works hard. The expected utility of the principal is equal to $w(e_H) - pt^*(\pi)$, which is equal to $w(e_L)$ when $\pi = \pi'$ (by definition) and is larger than $w(e_L)$ when $\pi \geq \pi'$ (since $t^*(\pi)$ is non-increasing in π and strictly decreasing when $\pi < \pi^*$). The principal thus weakly supports the norm, and strictly supports it if π is strictly larger than π'. The agent's expected utility is equal to $u(e_H) + pt^*(\pi)$. When Condition 1 holds, this value is larger than $u(e_L) + b$ (since $t^*(\pi) \geq t_a$ and Condition 1 implies that $u(e_L) + b < u(e_L) + (1-p)t_a \equiv u(e_H) + pt_a \leq u(e_H) + pt^*(\pi)$); as a result, also the agent supports the norm. When Condition 2 holds, the agent's expected utility is higher than $u(e_L) + b$ if π is larger than some value $\pi^{**} \in [\pi', \pi^*]$, and lower otherwise (see above); the agent thus supports the norm if and only if $\pi \leq \pi^{**}$. Q.E.D.

A.7 Properties of $W(\pi)$ and $W^*(\pi)$

(i) $W(\pi)$ is continuous, strictly decreasing and concave in π, since $c(\pi)$ is by assumption continuous, strictly increasing and convex.

(ii) $W^*(\pi)$ is continuous and concave in π, since $t^*(\pi)$ and $c(\pi)$ are both continuous and convex. Moreover, $W^*(\pi)$ is strictly decreasing in the interval $[\pi^*, 1]$ since in this interval $t^*(\pi)$ is equal to t_a, which is constant, while $c(\pi)$ is strictly increasing by assumption. $W^*(\pi)$ is two times continuously differentiable in the interval $[0, \pi^*)$ and in the interval $(\pi^*, 1]$; its first derivative in the interval $[0, \pi^*)$ is equal to:

$$-pt'_c(\pi) - c'(\pi), \tag{18}$$

which is continuous since $t'_c(\pi)$ and $c'(\pi)$ are both continuous, while its second derivative is equal to:

$$-pt''_c(\pi) - c''(\pi), \tag{19}$$

which is also continuous since $t''_c(\pi)$ and $c''(\pi)$ are also continuous. In the interval $(\pi^*, 1]$, the first derivative is equal to $-c'(\pi)$, which is continuous and negative, while the second derivative to $-c''(\pi)$, which is also continuous and negative.

(iii) $W^*(\pi) > W(\pi)$ if and only if $\pi > \pi'$, since $W^*(\pi) > W(\pi)$ implies $w(e_H) - pt^*(\pi) > w(e_L)$, which by definition occurs if and only if $\pi > \pi'$.

A.8 Solutions for π^o

Since $W^*(\pi)$ is continuous, concave and twice differentiable in the interval $[0, \pi^*)$, an internal maximum in that interval must satisfy the FOC:

$$-pt'_c(\pi^o) - c'(\pi^o) = 0, \tag{20}$$

which implies $c'(\pi) = -pt'_c(\pi)$. By the global concavity of $W^*(\pi)$, the SOC is always satisfied and the value of π^o is unique. If the FOC is not satisfied by any internal point, $W^*(\pi)$ is maximised by the corner solution $\pi^o = \pi^*$. The value of π^o cannot be larger than π^* because $W^*(\pi)$ is strictly decreasing in the interval $[\pi^*, 1]$.

A.9 Proof of Proposition 6

When $W^*(\pi^o) = w(e_L)$, the principal is indifferent between an equilibrium where $\pi = \pi^o$ and $t = t^*(\pi^o) = t_c(\pi^o)$ (since $t^*(\pi^o) = t_c(\pi^o)$ given that $\pi^o \leq \pi^*$), that induced the agent to work hard, and a capitulation equilibrium where $\pi = t = 0$, where the agent works lazily. Two equilibria thus exist, which yield the same expected utility to the principal. The expected utility of the agent in the first equilibrium is equal to $u(e_H) + pt_c(\pi^o)$, while in the second equilibrium it is equal to $u(e_L) + b$. If $\Delta u + b = pt_c(\pi^o)$, the two expected utilities are the same, and the agent is also indifferent between the two equilibria. Q.E.D.