# Lemons by Design: Sowing secrets to curb corruption[*]

[Latest version **here**]

Andrew Clausen[†], Christopher Stapenhurst[‡]

February 26, 2022

## Abstract

We study a problem in which a firm can bribe an inspector to conceal evidence of illegal pollution. We find that the cheapest way to deter bribes is (i) to secretly select either the firm or the inspector to 'win' a reward whenever evidence is reported; and (ii) to give both the firm and the inspector a secret clue about who won will. If the the inspector conceals the evidence, then the winner forgoes their reward — i.e. they 'get a lemon'. The distribution of clues is carefully constructed to engineer the worst possible lemons problem in the market for concealment: player $i$ only enters the market if her clue is strong enough to make her believe that player $j$ is the winner, despite knowing that player $j$ only enters if his clue indicates that player $i$ is the winner. But then higher order reasoning leads neither players to enter the market, no matter what clue they receive. Hence, bribery never takes place in equilibrium. As well as deterring bribes cheaply and robustly, this result demonstrates the full extent of contagious adverse selection in bilateral trades.

**Keywords:** corruption, mechanism design, information design, adverse selection, contagion.

**JEL codes:** D82, D86, D73.

[*]Acknowledgments: We would like to thank Alex Dickson, James Dunham, Marina Halac, Ed Hopkins, Tatiana Kornienko, Raghav Malhotra, John Moore, Alfonso Montes, Mariann Ollar, Christian Lippitsch, Rachel Scarfe, Ina Taneva, Rafael Veiel, Rakesh Vohra, Yaoyao Xu, and Gabriel Ziegler for helpful comments and suggestions.

[†]The University of Edinburgh. andrew.clausen@ed.ac.uk

[‡]The University of Edinburgh. c.stapenhurst@ed.ac.uk

# 1 Introduction

The threat of corruption is a serious concern in many incentive design problems. For example firms bribe politicians to grant them contracts without competitive tender; nations bribe international inspectors to ignore or destroy evidence of greenhouse gas emissions; and criminals bribe judges to grant light sentences. Corruption not only reallocates welfare unfairly, but also reduces aggregate welfare by creating perverse incentives. The UN Security Council estimates that corruption in the form of bribes, money laundering and tax evasion, directly reduces world GDP by 5% annually (United Nations, 2018). However this figure only measures realised corruption, it does not measure the cost of resorting to 'second best' contracts which anticipate the possibility of corruption. In the U.S., approximately 10% of wages are paid worker's whose primary responsibility is monitoring, directly or indirectly, the actions of someone else.[1]

In this paper we focus on corruption relating to the suppression of evidence by colluding parties. For example Duflo et al. (2013) find evidence that factory owners in India pay pollution inspectors to report that their factories are compliant with regulations, when in fact they are not. Corruption occurs because the factory and the inspector can generate joint surplus by reporting compliance rather then non-compliance. If corrupt behaviour is detectable and punishable with high enough fines, then the threat of large punishment can deter corruption for free (e.g. Becker (1968); von Negenborn and Pollrich (2020)). But in cases where large punishments are not available, deterring corruption is costly because it requires the regulator to pay rewards for non-corrupt behaviour. For example, Duflo et al. (2013) incentivise factory inspectors to report truthfully by paying them a reward (efficiency wage) large enough to subsume the joint surplus generated by colluding with the factories. They estimate that it would cost \$1300 per year to enforce compliance for a single small to medium scale factory with high pollution potential.[2] Reducing the cost of these rewards is important both on the intensive margin, because existing regulations that are already being enforced can be enforced at lower cost; and on the extensive margin, because regulations that are currently to expensive to be enforced, can be made affordable.

Our contribution is to show how information design (together with mechanism or 'transfer' design) can be used to reduce the cost of paying rewards, whilst still implementing compliance. We *engineer a market failure* in the market for bribes by giving the players private information about their transfer so as to create *the worst possible lemons problem*. In this mechanism, each player receives a random number between 0 (lemon) and 1 (peach). The player with the higher number receives a reward (or amnesty, in the case of the firm) for reporting incriminating evidence, so they do not want to conceal evidence. Hence, neither player wants to conceal evidence if they believe that the other player has the lower number (i.e. the lemon). Although it is common knowledge that

[1] Using data from the Occupational Employment and Wage Statistics (OEWS) programme and ONET occupation descriptions.

[2] Reported capital investment less than US \$ 2 million.

there is always surplus from corrupt deals, each player is too paranoid to agree to any corrupt deal because higher order reasoning leads them to believe that the other player would only accept a given deal if they received a 0. Hence no concealment takes place.

This approach builds on earlier work by Ortner and Chassang (2018) and von Negenborn and Pollrich (2020). They develop endogenous one-sided informational frictions as a means of deterring collusion, but our informational friction is two-sided, and hence more severe. Moreover, our solution accommodates imperfect monitoring (unlike Ortner and Chassang (2018)) and limited liability (unlike von Negenborn and Pollrich (2020)). Our result also provides answers to some open problems identified by Carroll (2016): we find that contagion in higher order beliefs *does* limit the amount of surplus that a pair of agents can obtain when they have transferable utility (in contrast to the case without transferable utility), and we provide an example of a 'worst case' information structure for the transferable utility case. We discuss the relevance of our results to existing literature in more detail in section 6.

The outline of the paper is as follows. Section 2 describes the role of the firm and the inspector without intervention from a regulator. Section 3 shows how a regular can use transfers to incentivise the firm to comply with the law, first when bribery is impossible, and then in the case where the firm can bribe the inspector. We introduce the information design in section 4 by means of a simple example. Section 5.1 formally describes the regulator's incentive design problem when both transfer design and information design are available as tools. Section 5.2 presents our main result: a characterisation of the cheapest scheme that provides adequate incentives for the firm to comply when bribery is possible. The proof of our main results is given in section 5.3. Finally, Section 6 compares our results with the related literature and section 7 concludes with some extensions for future research.

# 2 Environment

Finn the risk neutral firm chooses whether to pollute, $p$, or comply $c$. Other things equal, polluting yields Finn a payoff of 1, whereas complying yields a payoff of 0. Isla the risk neutral inspector inspects Finn's firm. If Finn chooses to pollute then she obtains evidence of pollution with probability $\pi_p$; otherwise she obtains no evidence. If Finn chooses to comply then Isla obtains evidence of pollution with probability $\pi_c < \pi_p$. The fact that $\pi_p$ can be strictly less than 1 implies that her monitoring technology is imperfect — she may fail to obtain evidence of pollution even though Finn has been polluting. Similarly, if $\pi_c$ is strictly greater than 0 then she may find evidence that Finn has been polluting, even though he has been compliant. The requirement that evidence is more likely to arise when Finn does pollute than when he complies ensures that evidence is indicative of pollution. If Isla obtains evidence of pollution then she can either report it to Rose the regulator, or she can keep silent. We assume that evidence cannot be fabricated, so if Isla does not obtain evidence then she has no choice but to keep silent.

Rose the regulator wants to incentivise Finn to comply, and she does so by fining

Finn an amount $f_e > 0$ whenever Isla reports evidence of pollution. If Isla doesn't report evidence then Rose fines Finn some amount $f_0$. We refer to the pair $(f_e, f_0)$ as a *scheme*. The fine $f_0$ is necessary because Rose doesn't want Finn to shut down completely, so she needs to compensate him for the risk of being fined by mistake when he chooses to comply. More precisely, she must satisfy his voluntary participation constraint,

$$-\pi_c f_e - (1-\pi_c) f_0 \geq 0,$$

which says that his expected payoff must be weakly positive if he chooses to comply. Since $f_e > 0$, this constraint can only be satisfied by fining Finn a negative amount $f_0 < 0$ whenever Isla reports nothing.

Rose must also ensure that Finn prefers to comply than to pollute. This is embodied in Finn's incentive compatibility constraint,

$$-\pi_c f_e - (1-\pi_c) f_0 \geq 1 - \pi_p f_e - (1-\pi_p) f_0,$$

which says that his expected payoff must he higher if he chooses to comply than if he chooses to pollute. It is more insightful to rewrite this constraint as

$$f_e - f_0 \geq \Pi.$$

where $\Pi := \frac{1}{\pi_p - \pi_c}$. We refer to the difference $f_e - f_0$ as *Finn's incentive to comply* because this is by how much his expected fine falls when he complies. The quantity $\Pi$ is Finn's benefit from polluting divided by the marginal risk of being caught, which we refer to as his risk-adjusted benefit of polluting. Thus Finn's incentive compatibility constraint says that the size of his incentive to comply must be exceed his *risk-adjusted* benefit of polluting.

Finally, since Finn's incentive compatibility constraint will ensure he complies in equilibrium, Rose's expected cost of a scheme $(f_e, f_0)$ is $-\pi_c f_e - (1-\pi_c) f_0$. Her objective is to minimise her cost subject to the voluntary participation and incentive compatibility constraints.

## 3 Benchmark: transfer design (complete information)

Here we show that incentivising compliance is free when bribes are not possible, but costly when bribes are possible. Thus, the need to deter bribes is the only inefficiency in our model.

The case where Rose is unconstrained by the possibility of bribery is the 'first best' case:

**Proposition 1** (First best)**.** *Suppose bribery is not possible. For any constant $k \geq \Pi$, the scheme $(f_e^{FB}, f_0^{FB})$ defined by*

$$\begin{aligned} f_e^{FB} &= (1-\pi_c)k \\ f_0^{FB} &= -\pi_c k, \end{aligned}$$

1. *satisfies Finn's voluntary participation and incentive compatibility constraints;*
2. *costs* $c^{FB} := 0$*;*
3. *costs (weakly) less than any other scheme that satisfies Finn's voluntary participation and incentive compatibility constraints.*

*Proof.*

1. Finn's expected payoff is $-\pi_c(1-\pi_c)k - (1-\pi_c)(-\pi_c k) = 0$ so his voluntary participation constraint is satisfied. His incentive is equal to $(1-\pi_c)k - (-\pi_c k) = k \geq \Pi$ so his incentive compatibility constraint is also satisfied.
2. The cost of the scheme is $c^{\text{FB}} := -\pi_c(1-\pi_c)k - (1-\pi_c)(-\pi_c k) = 0$.
3. Rose's expected cost is exactly equal to Finn's expected payoff and voluntary participation requires Finn's expected payoff to be greater than 0. Therefore no scheme that satisfies voluntary participation can cost less.

□

The schemes described in proposition 1 gives a lower bound on the cost of schemes that incentivise compliance. However, they are not robust to bribes. If incriminating evidence is realised, then bribing Isla to stay silent decreases Finn's fine from $f_e$ to $f_0$. Therefore, Finn will be willing to pay Isla any bribe $b \leq f_e - f_0 = k$. Isla is indifferent between reporting evidence and staying silent, so she will be willing to accept any bribe $b \geq 0$. Thus Isla and Finn can agree to any bribe $0 \leq b \leq k$. If Finn anticipates that he will be able to bribe Isla some amount $b < \Pi$ to stay silent, then polluting and bribing Isla will give him a higher expected payoff than complying, so he will choose to pollute. Consequently, Rose needs to deter bribes if she wants to incentivise compliance.

We postpone detailed description of what it means to be bribery-proof until section 5. For the moment, it suffices to say that Rose cannot deter bribes by either reducing the size of Finn's incentive (without violating his incentive constraint), or by paying Isla a reward $r \geq 0$ for reporting evidence. We assume that Isla cannot fabricate evidence, and that she cannot be punished for failing to report evidence (for instance she may be an employee with limited liability, or whistle blower acting of her own volition). Clearly, Rose will never want to reward Isla for staying silent, so it is without loss to assume that she pays Isla a reward equal to 0 when she stays silent. Thus a scheme is now defined by a triplet $(f_e, f_0, r)$ and has expected cost $\pi_c(r - f_e) + (1-\pi_c)(-f_0)$.

Suppose Rose continues to impose the first best fines $f_e = (1-\pi_c)k$ and $f_0 = -\pi_c k$ for some $k \geq \Pi$ and additionally offers Isla a reward $r = k + \epsilon$, for some $\epsilon > 0$. We saw in proposition 1 that these fines satisfy Finn's voluntary participation and incentive compatibility constraints. This scheme also deters bribery because Isla's opportunity cost of staying silent is equal to her reward, so she demands a bribe of at least $k + \epsilon$. But Finn is willing to pay a bribe of at most $k$, so there are no bribes that are mutually agreeable to both Isla and Finn. The cost of this scheme is equal to Isla's expected reward, which is $\pi_c(k + \epsilon)$. This cost is minimised by choosing $k = \Pi$ and $\epsilon$ to be as small as possible. There is no 'smallest' $\epsilon > 0$ so the optimal scheme does not exist. To avoid this technical difficulty, we will be content to say that a scheme deters bribes if it can be made to deter bribes by adding any $\epsilon > 0$ to Isla's reward — we address this point in more detail in section 5.1. We refer to the resulting scheme as 'third best' because it is the cheapest scheme among all transfer-only schemes that satisfy Finn's voluntary participation and incentive compatibility constraints; but it is not as cheap as the schemes that utilise information design in sections 4 and 5.

**Proposition 2** (Third best). *The scheme* $(f_e^{TB}, f_0^{TB}, r^{TB})$ *defined by*

$$f_e^{TB} = (1 - \pi_c)\Pi$$
$$f_0^{TB} = -\pi_c \Pi$$
$$r^{TB} = \Pi,$$

1. *deters bribes and satisfies Finn's voluntary participation and incentive compatibility constraints;*
2. *costs* $c^{TB} := \pi_c \Pi$*;*
3. *costs less than any other transfer-only scheme that deters bribes and satisfies Finn's voluntary participation and incentive compatibility constraints.*

*Proof.* We have already shown that this scheme satisfies Finn's voluntary participation and incentive compatibility constraints, deters bribes (up to a constant $\epsilon$), and costs $\pi_c \Pi$. It remains to show that no transfer-only scheme costs less. If $f_e - f_0 > r$ then Isla and Finn both strictly benefit by exchanging any bribe $r < b < f_e - f_0$. Therefore any scheme that deters bribes must have $r \geq f_e - f_0$. Any incentive compatible scheme must have $f_e - f_0 \geq \Pi$, therefore any incentive compatible scheme that deters bribes must have $r \geq f_e - f_0 \geq \Pi$. Any scheme that satisfies voluntary participation must have an expected fine of less than 0. Therefore the expected cost of the scheme must be at least $\pi_c r \geq \pi_c \Pi$. □

The intuition is that, without recourse to information design, any bribery proof scheme must destroy all the joint surplus that Isla and Finn generate when Isla stays silent, so we must have $f_e - f_0 - r \leq 0$. But incentive compatibility requires Finn's surplus to be at least $\Pi$, so Isla's reward must be at least $\Pi$.

# 4 Illustrative examples: one sided adverse selection

Paying rewards is necessarily expensive, but information design is free. Akerlof (1970) famously shows how private information in the market for used cars can cause the market to break down. In his model, a buyer is willing to pay a good price for a 'peach' (good car), but he isn't willing to pay anything for a 'lemon' (bad car). A seller is willing to sell a peach for a good price, but he is willing to sell a lemon for any price. The buyer and the seller would be able to generate surplus by trading a peach, but for the fact that only the seller knows whether the car is peach or a lemon. This creates an adverse selection or 'lemons' problem: the buyer knows that the seller will accept a good price whether she has a peach or a lemon, so the buyer will not want to pay a good price if the proportion of lemons in the market is too high, or if he does not value the peach enough to counteract the risk of buying a lemon. But he won't want to pay anything less than a good price either, because the seller will only accept less than a good price if she has a lemon, which the buyer doesn't value at all. Thus the whole market can collapse, even though there are potential gains from trade.

In this section we give two examples that illustrate how information design can be used to engineer lemons problems for Isla and Finn. These lemons problems make it more difficult for Isla and Finn to negotiate a bribe, and hence enables Rose to deter bribes at a lower cost. Section 4.1 presents the simplest scheme with endogenous private information, section 4.2 presents the cheapest scheme with endogenous private information for one player only. Section 5 presents the cheapest scheme with endogenous private information for both players.

## 4.1 The fair coin toss scheme (informed firm)

Before Isla receives any evidence, Rose and Finn flip a fair coin with a lemon ($L$) on one side and a peach ($P$) on the other. Finn observes the outcome of the coin toss, whilst Isla does not, so we refer to the outcome as Finn's 'private message'. If the coin comes up lemons then Finn's fines are $f_e(L) = (1-\pi_c)k$ and $f_0(L) = -\pi_c k$, and Isla's reward is $r(L) = k$, where, as before, $k$ is a constant. In this outcome, *bribery is a lemon for Isla* because she is better off taking her reward of $k$ than any bribe $0 \le b \le k$ that Finn is willing to pay. Similar to Akerlof's model, staying silent generates no joint surplus in the lemon state. However, if Finn has a peach message then Finn's fines are reduced to $f_e(P) = \frac{f_e(L)}{2}$ and $f_0(P) = \frac{f_0(L)}{2}$, and Isla's reward is reduced to $r(P) = 0$. In this outcome, *bribery is a peach for Isla* because any bribe $b > 0$ gives her a strictly higher payoff than reporting evidence. Just like Akerlof's model, there is a strictly positive joint surplus of $f_e(P) - f_0(P) = \frac{k}{2} > 0$ in the peach state. Thus Isla's situation is analogous to Akerlof's used car buyer, the only difference being that Akerlof's buyer faces uncertainty about the value of trading, whereas Isla faces uncertainty about the value of *not* trading.

It is easy to verify that the coin toss scheme satisfies Finn's voluntary participation constraint for any choice of $k$. However, the fact that Finn's fine's are reduced in the peach state mean that his expected incentive is now only

$$\frac{1}{2}(f_e(L) - f_0(L)) + \frac{1}{2}(f_e(P) - f_0(P)) = \frac{1}{2}\left[(1-\pi_c)k + \pi_c k\right] + \frac{1}{2}\left[(1-\pi_c)\frac{k}{2} - \pi_c\frac{k}{2}\right] = \frac{3}{4}k,$$

so we need to choose $k$ so that $\frac{3}{4}k \ge \Pi$ in order to satisfy this incentive compatibility constraint. The expected cost of this coin toss scheme is

$$
\begin{aligned}
&\frac{1}{2}[\pi_c(r(L)-f_e(L))+(1-\pi_c)(-f_0(L))]+\frac{1}{2}[\pi_c(r(P)-f_e(P))+(1-\pi_c)(-f_0(P))]\\
=&\frac{1}{2}\left[\pi_c\{k-(1-\pi_c)k\}+(1-\pi_c)\pi_c k\right]+\frac{1}{2}\left[-\pi_c(1-\pi_c)\frac{k}{2}+(1-\pi_c)\pi_c\frac{k}{2}\right]\\
=&\pi_c\frac{k}{2},
\end{aligned}
$$

which is strictly increasing in $k$. Hence the cost is minimised by choosing $k$ to be as small as possible, namely $\frac{4}{3}\Pi$. Doing so yields an expected cost of $\pi_c\frac{2}{3}\Pi$, which is strictly less than the cost of the third best scheme, $c^{\mathrm{FB}} = \pi_c\Pi$.

Despite being cheaper than the third-best scheme, the coin toss scheme still deters all possible bribes (for any choice of $k$) because, just like Akerlof's seller, Finn adversely selects to offer bribes when he has a lemon message, as we now show. First, consider small bribes $b < \frac{k}{2}$. Finn is always willing to pay small bribes because the size of his incentive is weakly greater than $\frac{k}{2}$ in both the peach and the lemon state; specifically, $f_e(L)-f_0(L) = k > f_e(P)-f_0(P) = \frac{k}{2} > b$. Therefore Isla's expected reward is $\frac{1}{2}r(L)+\frac{1}{2}r(P) = \frac{k}{2}$. But this is greater than the size of the bribe, so she will prefer to report the evidence than to remain silent and take a small bribe. Now consider big bribes $\frac{k}{2} < b < k$. Finn pays big bribes if and only if he has a lemon, because the size of his incentive is greater that the bribe in the lemon state, but lower in the peach state. Specifically, $f_0(L) - f_e(L) = k > b > \frac{k}{2} = f_0(P) - f_e(P)$. Therefore, Isla does not accept big bribes because $r(L) = k > b$, so she is better off staying silent in the lemon state. It is clear that Finn will never pay fines strictly greater than $k$ and Isla will never accept bribes strictly less than 0. Thus the only cases left to consider are the knife edge cases where the bribe exactly equals $k$ or $\frac{k}{2}$, but, no matter what Finn does, Isla is weakly better off rejecting these bribes, so we can make her strictly better off rejecting them by adding some arbitrarily small amount $\epsilon > 0$ to her rewards.

Thus we have proved the following proposition:

**Proposition 3** (Coin toss scheme)**.** *Suppose Finn receives a private message $x_F$ equal to either $L$ (for lemon), or $P$ (for peach). The scheme $(q^{CT}, (f_e^{CT}, f_0^{CT}, r^{CT}))$ defined by*

$$
\begin{aligned}
f_e^{CT}(L) &= (1-\pi_c)\frac{4}{3}\Pi & \qquad f_e^{CT}(P) &= (1-\pi_c)\frac{2}{3}\Pi\\
f_0^{CT}(L) &= -\pi_c\frac{4}{3}\Pi & f_0^{CT}(P) &= -\pi_c\frac{2}{3}\Pi\\
r^{CT}(L) &= \frac{4}{3}\Pi & r^{CT}(P) &= 0\\
q^{CT}(L) &= \frac{1}{2} & q^{CT}(P) &= \frac{1}{2},
\end{aligned}
$$

*where $q^{CT}(x_F)$ denotes the probability of the message $x_F$,*

1. *deters bribes and satisfies Finn's voluntary participation and incentive compatibility constraints;*
2. *costs* $c^{CT} := \frac{2}{3}\pi_c\Pi$.

The fair coin toss is the simplest scheme with private information, but not the cheapest. Two variations on the coin toss scheme are possible. Firstly, instead of using a fair coin, Rose can use a biased coin which puts a lower weight on the (costly) lemon message. Doing so increases the fraction of peaches in the market at low bribe levels, so Rose must further reduce Finn's incentive in the peach state so as to prevent peaches from entering the market at low bribe levels. The optimal biased coin toss scheme costs strictly less than the fair coin toss scheme. Secondly, Rose can let Isla observe the outcome of the coin toss instead of Finn. This yields an 'informed inspector' coin toss scheme. We develop this further in the following subsection.

## 4.2 The one-sided informed inspector scheme

Coin toss schemes give the informed player a binary signal, which is the minimal amount of private information possible. Here we present a scheme that gives Isla the inspector a whole continuum of possible messages. The reasons for presenting this particular scheme are three-fold. Firstly, it demonstrates that Rose can attain the first best outcome if she can use infinitely large fines. We consider the case of infinitely large fines to be unrealistic, so this fact motivates us to consider cases where fines are bounded. Secondly, it is the optimal one-sided scheme (when the bound on fines is not too small[3]), so the fact that our two-sided scheme costs strictly less than it motivates our interest in two-sided schemes. Thirdly, it provides the best basis for comparing our main result with previous literature (Ortner and Chassang, 2018). We expand more on this latter point in section 6.

Suppose Isla receives a private message $x_I \in [0,1]$ with density $q(x_I)$.

**Proposition 4** (Informed Inspector)**.** *For any constant* $k \geq \Pi$, *the scheme* $(q^{II}, (f_e^{II}, f_0^{II}, r^{II}))$ *defined by*

$$q^{II}(x_I) = 1 \text{ i.e. } q^{II} \text{ is uniform on } [0,1]$$

$$f_e^{II}(x_I) = \begin{cases} 0 & \text{if } x_I \leq \frac{k-\Pi}{k} \\ (1-\pi_c)k & \text{otherwise} \end{cases}$$

$$f_0^{II}(x_I) = \begin{cases} 0 & \text{if } x_I \leq \frac{k-\Pi}{k} \\ -\pi_c k & \text{otherwise} \end{cases}$$

$$r^{II}(x_I) = \begin{cases} 0 & \text{if } x_I \leq \frac{k-\Pi}{k} \\ k - \frac{k-\Pi}{x_I} & \text{otherwise,} \end{cases}$$

[3]If the bound on fines is small enough then the one-sided informed inspector scheme is undercut by a one-sided informed firm scheme.

1. *deters bribes and satisfies Finn's voluntary participation and incentive compatibility constraints;*

2. *costs* $c^{II} := \pi_c\left[\Pi + (k-\Pi)\ln\left(1-\frac{\Pi}{k}\right)\right] \xrightarrow{k\to\infty} 0$;

3. *costs less than any other one-sided, informed inspector scheme that satisfies Finn's voluntary participation and incentive compatibility constraints.*

*Proof.* We prove here that the informed inspector scheme deters bribes. The rest of the proof is given in a appendix A.1.

Consider a bribe $b$. Isla will agree to the bribe if and only if she receives a message for which her reward $r(x_I)$ is less than the bribe $b$. We have $r(x_I) = k - \frac{k-\Pi}{x_I}$ so Isla agrees to the bribe if and only if $b \geq k - \frac{k-\Pi}{x_I}$, or equivalently, $x_I \leq \frac{k-\Pi}{k-b}$. This is an example of a cutoff strategy with cutoff equal to $\frac{k-\Pi}{k-b}$. Finn's expected incentive, conditional on Isla's cutoff strategy, is equal to

$$\begin{aligned}\mathbb{E}\left[f_e^{\mathrm{II}}(x_I) - f_0^{\mathrm{II}}(x_I) \,\middle|\, b \geq k - \frac{k-\Pi}{x_I}\right] &= \mathbb{P}\left[x_I \leq \frac{k-\Pi}{k} \,\middle|\, x_I \leq \frac{k-\Pi}{k-b}\right] 0 + \mathbb{P}\left[x_I \geq \frac{k-\Pi}{k} \,\middle|\, x_I \leq \frac{k-\Pi}{k-b}\right] k \\ &= \min\left\{1, \frac{\frac{1}{k-b} - \frac{1}{k}}{\frac{1}{k-b}}\right\} k \\ &= \min\{k, b\},\end{aligned}$$

so he is indifferent about accepting bribes less than $k$, and strictly prefers to reject bribes greater than $k$. If $b = 0$, then Isla strictly prefers to take her reward if her message is $x_I > 1 - \frac{\Pi}{k}$, otherwise she is indifferent. Therefore, Finn's conditional expected incentive is equal to 0, so he is indifferent as well. In all cases we can deter Isla and Finn from exchanging the zero bribe at arbitrarily small cost (e.g. by adding $\epsilon$ to Isla's reward). □

The key feature of this scheme is that the distribution of rewards is chosen so that Finn's probability of facing a peach conditional on a given bribe increases in proportion to the size of the bribe, so as to keep him indifferent about accepting the bribe. In other words, 'peach inspectors' enter the market at the highest rate possible without giving Finn a strict preference to enter the market. In the limit, Finn's incentive, $k$, becomes arbitrarily large with vanishing probability, which corresponds to the use of extreme incentives in Becker (1968). By contrast, Isla's reward never exceeds $\Pi$. Since Rose only has to pay Isla with the same vanishing probability that she punishes Finn, Isla's expected reward can be made arbitrarily small. This in turn means that the cost of the scheme approaches the first best cost, so no scheme can do better.

However, there is still scope for improvement because transfers are bounded in most practical applications. We show in a appendix A.2 that, when the $k$ is restricted to be small enough, there is an informed firm scheme that costs less than the optimal informed inspector scheme. In section 5 we show that Rose can create a two-sided adverse selection problem by jointly designing private information for both the firm and the inspector, and that doing so deters bribes at strictly lower costs than all one-sided schemes, for any bound on transfers.

# 5 Main result: two-sided lemons

In this section we formally model the regulator's problem with both transfer design and information design. We then describe the players' bribery negotiations and show that the regulator can restrict attention to bribery-proof schemes. Our main result characterises an optimal bribery-proof scheme.

## 5.1 The regulator's problem

Rose the regulator commits to a scheme $\mathcal{S} := (q, (f_e, f_0, r))$, where $q$ is a public distribution over private messages $x = (x_I, x_F) \in [0,1]^2$. Finn observes the message $x_F$ and Isla observes the message $x_I$. The objects $f_e, f_0$ and $r$ are message-contingent transfers. We assume that Isla is protected by limited liability, so Rose must choose $r(x) \geq 0$. We also assume that Rose cannot use extreme incentives for Finn, this means that she must choose $f_e(x)$ and $f_0(x)$ so that $f_e(x) - f_0(x) \leq \kappa$ for some $\kappa \geq \Pi$. If $\kappa < \Pi$ then it will be impossible to provide incentives for Finn to comply; if $\kappa = \infty$ then Rose can achieve the first best with the one-sided scheme described in section 4.2. Like before, Rose's main objective is to incentivise Finn to choose to comply instead of polluting. This requires her to satisfy an incentive compatibility constraint. She also needs to incentivise Finn to stay in business, which requires her to satisfy a voluntary participation constraint. We say that the scheme $\mathcal{S}$ is *feasible* if is satisfies these three constraints: limited liability, incentive compatibility, and voluntary participation. Rose wants to find the cheapest feasible scheme.

After Rose has sent Isla and Finn their private messages, Isla may or may not obtain evidence. If she does obtain evidence then Isla is able to commit to stay silent in return for a bribe from Finn. For each possible bribe $b$ we define a 'bribery game' in which Isla and Finn choose to accept the bribe with respective probabilities $\sigma_I^b(x_I)$ and $\sigma_F^b(x_F)$. If they both accept the bribe then Finn pays the bribe $b$ to Isla and Isla stays silent: Isla's payoff is then $b$ and Finn's payoff is $f_0(x) - b$. If either of them rejects the bribe then Isla reports the evidence and gets paid $r(x)$; Finn receives the fine $f_e(x)$. We refer to the bribe-strategy pair $(b, (\sigma_F^b, \sigma_I^b))$ as a *bribe contract*, and we denote it by $\mathcal{C} := (b, (\sigma_F^b, \sigma_I^b))$. We refer to the bribe contract in which Isla always reports evidence and Finn never pays bribes as the *null bribe contract*, denoted $\mathcal{C}_0$. If Isla doesn't obtain evidence then she gets 0 and Finn gets $f_0(x)$. A scheme $\mathcal{S}$ is *bribery proof* if Isla always reports evidence in every Bayes-Nash equilibrium of every bribery game. A scheme $\mathcal{S}$ is *$\epsilon$-bribery proof* if the scheme $(q, (r + \epsilon, f_e, f_0))$ is bribery proof for all $\epsilon > 0$. In other words, a scheme is $\epsilon$-bribery proof if it can be made bribery proof at arbitrarily small cost. The reason for introducing $\epsilon$-bribery proofness is purely technical — it ensures that an optimal scheme exists. Without it we would have to add $\epsilon$ to our schemes to destroy any unwanted equilibria. Another potential solution to this problem would be to do partial implementation, but this would give a trivial solution in our model because the null bribe contract is always an equilibrium: rejecting the bribe is always a weak best response if the other player always rejects it.

A natural question arises about how Isla and Finn choose between different possible bribe contracts, but this turns out not to matter because lemma 1 tells us that Rose can restrict attention to $\epsilon$-bribery proof schemes.

**Lemma 1.** *For every feasible scheme, there exists an $\epsilon$-bribery proof feasible scheme with the same cost.*

*Proof.* Let $\mathcal{S}$ be a feasible scheme. If $\mathcal{S}$ is not $\epsilon$-bribery proof then there must exist a non-null, bribery equilibrium. A bribery equilibrium is payoff equivalent to a special case of an incentive compatible collusive side contract.[4] Therefore an interim efficient side contract, $\mathcal{C}$, must exist. Rose can create a new scheme $\mathcal{S}'$ that replicates the ex post payoffs of the bribe contract $\mathcal{C}$ under the original scheme $\mathcal{S}$. This new scheme $\mathcal{S}'$ must be $\epsilon$-bribery proof: otherwise, there would exist a non-null equilibrium side contract $\mathcal{C}'$ that gives either Isla or Finn must get a strictly higher payoff under $\mathcal{S}'$, than $\mathcal{C}_0$ does under $\mathcal{S}'$. The scheme, contract pair $(\mathcal{S}',\mathcal{C}_0)$ is ex post payoff equivalent to $(\mathcal{S},\mathcal{C})$ so $(\mathcal{S}',\mathcal{C}')$ interim dominates $(\mathcal{S},\mathcal{C})$. But then we can construct another side contract $\mathcal{C}''$ that is ex post payoff equivalent under $\mathcal{S}$ to $(\mathcal{S}',\mathcal{C}')$. Since $\mathcal{C}'$ is an equilibrium, $\mathcal{C}''$ must also be an equilibrium, contradicting the fact that $(\mathcal{S},\mathcal{C})$ was assumed to be interim efficient. Therefore $\mathcal{S}'$ must be $\epsilon$-bribery proof. □

The fact that Rose can restrict attention to $\epsilon$-bribery proof schemes simplifies her problem dramatically, because it means that bribes need not feature in Isla and Finn's payoffs. The cost of this simplification is that she has to satisfy an $\epsilon$-bribery proofness constraint. Formally, her problem is

$$\begin{aligned}
\min_{\mathcal{S}}\ & \mathbb{E}[\pi_c(r(x)-f_e(x))+(1-\pi_c)f_0(x)] && \\
\text{s.t. } & \mathbb{E}[\pi_c f_e(x)+(1-\pi_c)f_0(x)] \geq \mathbb{E}[\pi_d f_e(x)+(1-\pi_d)f_0(x)] && \text{(IC)}\\
& \mathbb{E}[\pi_c f_e(x)+(1-\pi_c)f_0(x)] \geq 0 && \text{(VP)}\\
& r(x)\geq 0 \text{ and } f_e(x)-f_0(x)\leq \kappa && \text{(LL)}\\
& \mathcal{S} \text{ is } \epsilon\text{-bribery proof} && \text{(BP)}
\end{aligned}$$

We are now ready to present our main result.

## 5.2 An optimal scheme

Our main result characterises an optimal scheme with two-sided information design. This scheme endogenously creates a two-sided lemons problem for Isla and Finn. In the informed inspector scheme, Isla's willingness to accept bribes was decreasing in her

[4]See appendix A.3.

own message, so she only wanted to agree to bribes when her message was below a certain cutoff. Finn's willingness to accept bribes was increasing in Isla's message, so Isla's choice of cutoff strategy made him unwilling to accept any bribes. In the two-sided scheme, Isla is both informed (about her own message) and uninformed (about Finn's message) so she inherits both of these features. Her willingness to accept bribes is decreasing in her own message, so she continues to adopt a cutoff strategy. But her willingness to accept bribes is increasing in Finn's message, so her optimal choice of cutoff is decreasing in her belief about Finn's message. Her belief about Finn's message depends on his strategy. The distribution of transfers ensures that his best response is also a cutoff strategy, and his optimal cutoff is decreasing in his belief about Isla's message. The distribution of transfers is chosen so that each player wants to choose a cutoff that is slightly below the cutoff that the other player chooses. Therefore there can be no equilibrium in which they both use a positive cutoff, which means that they do not agree to bribes in any equilibrium. Thus, unlike the informed inspector scheme, the two-sided scheme uses contagion in higher order beliefs to amplify the adverse selection problem. This stands in contrast to an earlier result of Carroll (2016), which we describe in more detail in section 6.

The functional form of our optimal two-sided scheme is comparatively simple: messages are independently and uniformly distributed, and, in the special case where $\kappa = 1$ and $\Pi = \frac{3}{4}$, transfers are a linear function of the ratio of the two messages, truncated above at 1:

$$r^*(x) = 1 - \min\left\{1, \frac{x_F}{x_I}\right\}$$
$$f_e^*(x) - f_0^*(x) = \min\left\{1, \frac{x_I}{x_F}\right\}.$$

Figure 1 shows these transfers in the context of the message space. When Isla's message is lower than Finn's, her reward is equal to 0, so she is eager to accept any bribe. However, Finn gets a reduced incentive of $\frac{x_I}{x_F} < 1$, so staying silent becomes less valuable to him and therefore he is not willing to pay such high bribes. Similarly, When Isla's message is higher than Finn's, Finn's incentive is equal to 1, so he is eager to accept any bribe, but Isla gets a reduced reward of $1 - \frac{x_F}{x_I} < 1$, so she is relatively less willing to accept bribes. The fact that messages are independent in this optimal scheme is not surprising: Cremer and McLean (1988) tell us that a third party would be able extract Isla and Finn's private information for free if their messages were correlated.[5] Having extracted their private information, this third part could then choose a bribe that would be mutually agreeable to Isla and Finn.

In the general case where the bound on Finn's incentive takes any large enough to satisfy incentive compatibility (i.e. $\kappa \geq \Pi$), the two sided scheme takes a similar form, described in theorem 1:

[5] See appendix A.3.

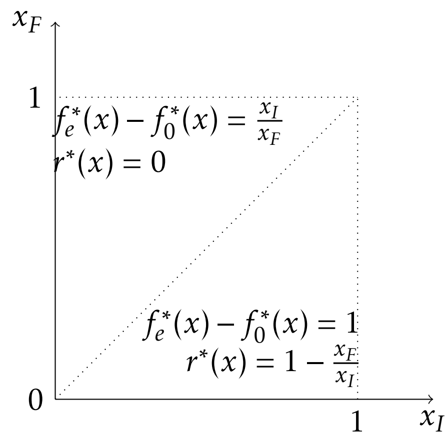


Figure 1: Transfers as a function of messages in the special case $\kappa = 1, \Pi = \frac{3}{4}$.

**Theorem 1** (Two-sided adverse selection)**.** *The scheme $\mathcal{S}^* := (q^*, (r^*, f_e^*, f_0^*))$ defined by*

$$\begin{aligned}
q^*(x) &= 1 \text{ \textit{i.e.}, } q^* \text{ \textit{is independent and uniform on} } [0,1]^2 \\
r^*(x) &= \kappa\left(1 - \min\left\{1, x_F^{\lambda}/x_I\right\}\right) \\
f_e^*(x) &= \kappa(1-\pi_c)\min\left\{1, x_I/x_F^{1/\lambda}\right\} \\
f_0^*(x) &= -\kappa\pi_c \min\left\{1, x_I/x_F^{1/\lambda}\right\},
\end{aligned}$$

*where* $\lambda = \sqrt{\frac{\kappa}{\kappa - \Pi}} - 1$,

1. *is feasible;*
2. *costs* $c^* := \pi_c\left(\sqrt{\kappa} - \sqrt{\kappa - \Pi}\right)^2$*;*
3. *costs less than any scheme that can be approximated by feasible schemes with a finite number of messages.*

The two-sided scheme is a substantial improvement on the optimal one-sided scheme. Table 1 compares the cost of the two-sided scheme to the one-sided (informed inspector) scheme (and others) for a range of parameter values. We show in a appendix A.4 that the two-sided scheme costs strictly less than the informed inspector scheme at all parameter values and that the cost of the two-sided scheme converges to *half* the cost of the optimal one-sided scheme as $\kappa$ gets large.

## 5.3 Proof of the main result

We prove each of the numbered points in theorem 1 in turn.

Table 1: The costs of selected schemes for parameters $\pi_p = \frac{2}{3}, \pi_c = \frac{1}{3}, \Pi = 3$.

| Scheme | $\kappa = 3$ | $\kappa = 4$ | $\kappa = 6$ | $\kappa = 30$ |
|---|---|---|---|---|
| third best | 1 | 1 | 1 | 1 |
| fair coin toss (informed firm) | — | 0.667 | 0.667 | 0.667 |
| biased coin toss (informed firm) | 1 | 0.667 | 0.586 | 0.513 |
| biased coin toss (informed inspector) | 1 | 0.75 | 0.5 | 0.1 |
| one-sided (informed inspector) | — | 0.538 | 0.307 | 0.052 |
| two-sided | 1 | 0.333 | 0.172 | 0.026 |

#### 5.3.1 The two-sided scheme is feasible

The proof has three steps: first we show that cutoff strategies are optimal; then we calculate each player's expected returns to bribes when the other player plays a cutoff strategy; finally we show that no bribe is mutually agreeable at any pair of cutoffs. An alternative proof replaces the final step with a demonstration that the best response to any cutoff is a proportionally lower cutoff, with the result that no pair of strictly positive cutoffs can form an equilibrium.

Fix a bribe $b$. The same proof applies to all bribes.

**Step 1.** Isla's opportunity cost of agreeing to the bribe when she receives message $x_I$ is equal to her expected reward conditional on Finn's strategy $\sigma_F$:

$$\mathbb{E}_{x_F}[\sigma_F(x_F) r^*(x)] = \mathbb{E}_{x_F}\left[\sigma_F(x_F)\kappa\left(1 - \min\{1, x_F^\lambda x_I^{-1}\}\right)\right]. \tag{1}$$

This expression is continuous and strictly increasing in her own message $x_I$ whenever $\sigma_F(x_F) > 0$ for some $x_F$ (otherwise it equals 0). Therefore, Isla's best response to any $\sigma_F$ is to agree to a bribe $b$ iff her message is below some cutoff $y_I \in [0,1]$. If equation (1) is greater than $b$ for all $x_I$ then Isla never wants to agree to the bribe, so she chooses a cutoff of $y_I = 0$. If equation (1) is negative $b$ for all $x_I$ then Isla always wants to agree to the bribe, so she chooses a cutoff of $y_I = 1$. In intermediate cases, Isla chooses her cutoff to equal the unique message $x_I$ for which (1) is exactly equal to $b$.

The same argument applies to Finn: his expected incentive conditional on Isla's strategy $\sigma_I$ is

$$\mathbb{E}_{x_I}[\sigma_I(x_I)(f_e^*(x) - f_0^*(x))] = \mathbb{E}_{x_I}\left[\sigma_I(x_I)\kappa \min\{1, x_I/x_F^{1/\lambda}\}\right],$$

which is also continuous and strictly decreasing in his own message $x_F$ whenever $\sigma_I(x_I) > 0$ for some $x_I$. Therefore, Finn's best response to any $\sigma_I$ is to agree iff his message is below some cutoff $y_F \in [0,1]$.

**Step 2.** When Isla receives the message $x_I$ and Finn uses cutoff $y_F$, Isla only agrees to the bribe if it is above her expected opportunity cost

$$\mathbb{E}[r^*(x)|x_I, x_F \leq y_F] = \begin{cases} \left(1 - \frac{1}{\lambda+1}\frac{y_F^\lambda}{x_I}\right)\kappa & \text{if } x_I \geq y_F^\lambda, \\ \frac{\lambda}{\lambda+1}\frac{x_I^{1/\lambda}}{y_F}\kappa & \text{if } x_I < y_F^\lambda. \end{cases} \tag{2}$$

When Finn receives the message $x_F$ and Isla uses cutoff $y_I$, Finn only agrees to the bribe if it is below his expected opportunity cost

$$\mathbb{E}[f_e^*(x) - f_0^*(x)|x_F, x_I \leq y_I] = \begin{cases} \frac{\lambda}{\lambda+1}\frac{y_I^{1/\lambda}}{x_F}\kappa & \text{if } x_F^\lambda > y_I, \\ \left(1 - \frac{1}{\lambda+1}\frac{x_F^\lambda}{y_I}\right)\kappa & \text{if } x_F^\lambda \leq y_I. \end{cases} \tag{3}$$

The full derivation of equations (2) and (3) are shown in appendix A.5.

**Step 3.** To be mutually accepted at the cutuoff outcome $y = (y_F, y_I)$, the bribe $b$ must be greater than Isla's conditional expected reward and below Finn's conditional expected incentive:

$$\mathbb{E}[r^*(x)|x_I = y_I, x_F \leq y_F] \leq b \leq \mathbb{E}[f_e^*(x) - f_0^*(x)|x_F = y_F, x_I \leq y_I].$$

But if $y_F^\lambda \leq y_I$, then

$$\mathbb{E}[r^*(x)|x_I = y_I, x_F \leq y_F] = \frac{\lambda}{\lambda+1}\frac{y_I^{1/\lambda}}{y_F}\kappa = \mathbb{E}[f_e^*(x) - f_0^*(x)|x_I = y_I, x_F \leq y_F].$$

And if $y_F^\lambda > y_I$, then

$$\mathbb{E}[r^*(x)|x_I = y_I, x_F \leq y_F] = \left(1 - \frac{1}{\lambda+1}\frac{y_F^\lambda}{y_I}\right)\kappa = \mathbb{E}[f_e^*(x) - f_0^*(x)|x_I = y_I, x_F \leq y_F].$$

In both cases, Isla and Finn are both indifferent about agreeing to the bribe when they receive their cutoff messages $y_I$ and $y_F$. This means that adding any $\epsilon > 0$ to Isla's reward will make her strictly prefer to reject the bribe at her cutoff and for nearby messages. Therefore, the cutoffs $y_I$ and $y_F$ cannot form an equilibrium. Since this argument applies for any pair of positive cutoffs, the bribe will not be accepted in any equilibrium.

**Incentive compatibility** Evaluating equation (3) at $y_I = 1$ gives $\mathbb{E}[f_e^*(x) - f_0^*(x)|x_F] = \mathbb{E}[f_e^*(x) - f_0^*(x)|x_F, x_I \leq 1] = \left(1 - \frac{1}{\lambda+1}x_F^\lambda\right)\kappa$. Integrating over $x_F$ gives Finn's ex ante incentive:

$$\mathbb{E}[f_e^*(x) - f_0^*(x)] = \int_0^1 \mathbb{E}[f_e^*(x) - f_0^*(x)|x_F]\, dx_F = \kappa \int_0^1 1 - \frac{1}{\lambda+1}x_F^\lambda\, dx_F = \kappa\left(1 - \frac{1}{(\lambda+1)^2}\right).$$

Substituting in $\lambda = \sqrt{\frac{\kappa}{\kappa - \Pi}} - 1$ gives $\mathbb{E}[f_e^*(x) - f_0^*(x)] = \Pi$, as required.

**Voluntary participation**

$$\mathbb{E}[\pi_c f_e^*(x) + (1-\pi_c) f_0^*(x)] = \pi_c \kappa (1-\pi_c) \min\{1, x_I/x_F^{1/\lambda}\} - (1-\pi_c)\kappa\pi_c \min\{1, x_I/x_F^{1/\lambda}\} = 0$$

### 5.3.2 The two-sided scheme costs $\pi_c(\sqrt{\kappa} - \sqrt{\kappa - \Pi})^2$

Evaluating equation (2) at $y_F = 1$ gives $\mathbb{E}[r^*(x)|x_I] = \mathbb{E}[r^*(x)|x_I, x_F \le 1] = \frac{\lambda}{\lambda+1} x_I^{1/\lambda}\kappa$. Integrating over $x_I$ gives Isla's expected reward:

$$\begin{aligned}\mathbb{E}[r^*(x)] &= \int_0^1 \frac{\lambda}{\lambda+1} x_I^{1/\lambda} \kappa \, dx_I \\ &= \left(\frac{\lambda}{\lambda+1}\right)^2 \kappa.\end{aligned}$$

Substituting in $\lambda = \sqrt{\frac{\kappa}{\kappa - \Pi}} - 1$ gives $\mathbb{E}[r^*(x)] = (\sqrt{\kappa} - \sqrt{\kappa - \Pi})^2$, so the cost of the scheme is is $\pi_c(\sqrt{\kappa} - \sqrt{\kappa - \Pi})^2$.

### 5.3.3 Every feasible finite scheme costs at least $\pi_c(\sqrt{\kappa} - \sqrt{\kappa - \Pi})^2$

Rose's problem involves deterring all bribes. We obtain a lower bound on the cost of deterring all bribes by solving a relaxed problem in which Rose only has to deter one specific bribe, $b$. We show that the cost of deterring the 'worst case' bribe $b^* = \sqrt{\kappa}\left(\sqrt{\kappa} - \sqrt{\kappa - \Pi}\right)$, is equal to the cost of the two-sided scheme. Since deterring the specific bribe $b^*$ is easier than deterring all bribes simultaneously, the cost of deterring $b^*$ gives a lower bound on the cost of deterring all bribes. Therefore no feasible scheme can cost strictly less than the two-sided scheme.

It remains to show that any scheme that deters the bribe $b^*$ costs at least $\pi_c(\sqrt{\kappa} - \sqrt{\kappa - \Pi})^2$. This problem is dramatically simplified by a result in Carroll (2016) which shows that we can restrict attention to public schemes. This result is stated in lemma 2.

**Lemma 2** (Corollary of Carroll (2016))**.** *Suppose a finite scheme $\mathcal{S}$ deters a bribe $b$. Then there exists a public, finite scheme, $\mathcal{S}^p$ (i.e. one in which $x_I = x_F$ with probability 1) which deters the bribe $b$ and costs the same as $\mathcal{S}$.*

*Proof.* We first translate our problem into the model of Carroll (2016). The result is then corollary of his propositions 3.1 and 3.2. The details are given in appendix A.6. □

A public scheme deters bribe $b^*$ if and only if, for all $x \in \text{supp}(q)$, either $r(x) > b^*$ or $f_e(x) - f_0(x) < b^*$. An optimal public scheme with incentive equal to $\Pi$ has $r(x) = b^*$ and $f_e(x) - f_0(x) = \kappa$ with probability $\frac{\Pi - b^*}{\kappa - b^*}$; and $r(x) = 0$ and $f_e(x) - f_0(x) = b^*$ with probability $1 - \frac{\Pi - b^*}{\kappa - b^*}$. This scheme has expected reward $\frac{\Pi - b^*}{k - b^*} b^* = (\sqrt{\kappa} - \sqrt{\kappa - \Pi})^2$, hence its cost is equal to $c^*$. Therefore, the two-sided scheme costs less than every finite feasible scheme. It follows from continuity that it costs less than every scheme that can be approximated by a sequence of finite feasible schemes.

# 6 Literature

We contribute to an extensive literature on corruption (Tirole, 1986; Laffont and Martimort, 1997; Strausz, 1997; Baliga and Sjöström, 1998). The closest paper to ours is Ortner and Chassang (2018). They are the first (to the best of our knowledge) to study the use of endogenous asymmetric information to deter bribes. They show that a principal (Rose) can benefit from paying the monitor (Isla) a random wage (privately observed by the monitor) according to a public distribution, known to the agent (Finn). Doing so endows the monitor with private information about their outside option, and thereby creates an informational-friction in subsequent collusive negotiations between the monitor and a would-be criminal agent. In their model, the agent chooses between being criminal and bribing the monitor on the one hand, or being innocent on the other. Paying the monitor a random wage creates a trade-off for the agent: he can either offer a high bribe which guarantees a high probability of successfully corrupting the monitor, or he can offer a low bribe which guarantees a low probability of successfully corrupting the monitor. The principal saves money by paying random wages because low wage monitors can mimic the high wage monitors and demand high bribes.

Our model differs from theirs in two important respects. Firstly, Ortner and Chassang (2018) assume perfect monitoring so they can rule out bribes on the equilibrium path (when no incriminating evidence arises) without needing to rule them out off the equilibrium path (when the monitor receives incriminating evidence with certainty). By contrast, we allow for monitoring mistakes so we have to consider the impact of bribes both on and off the equilibrium path. If, like Ortner and Chassang, we pay rewards according to a distribution which pays rewards that are smaller than the agent's punishment, then we inevitably get on-path bribery because the agent is always weakly better off accepting bribes smaller than the punishment, and there will be a strictly positive probability that the monitor is willing to accept such bribes. This difficulty motivates our second main departure from their model, which is to endogenise the agent's fines. Doing so allows us to replicate the agent's trade-off in their model, because we can use the changes in the agent's fine to imitate his choice to commit crime or not. This gives our result a qualitatively different interpretation from theirs: our agent faces a lemons problem because his fine depends on the monitor's private information. Despite these differences, our informed inspector scheme (proposition 4), which is the closest to theirs conceptually, produces the same distribution of rewards and has the same cost.[6] We showed in section 5.2 that the two-sided scheme costs strictly less than the informed inspector scheme, and costs half as much in the limit as the size of the maximal punishment increases.

Another closely related paper to ours is von Negenborn and Pollrich (2020). They also find that engineering a lemons problem is an optimal solution to a mechanism design problem. Our main contribution relative to theirs is that we impose bounds on all transfers, whereas their proposed mechanisms attain the first best by using large rewards

[6] Garrett et al. (2021) obtain the same distribution as a solution to a similar problem in which an agent chooses their distribution of costs to maximise their information rent.

and/or punishments. Therefore, they do not need to engineer an optimal lemons problem – any lemons problem would suffice.

Our results also speak to the literature on robustness of equilibria to contagion.[7] Carroll (2016) obtains an upper bound on the amount of surplus lost due to contagion in a game with two agents either accepting or rejecting a proposed deal, where both agents have private information about the payoff outcomes of the deal. Surprisingly, Carroll finds that contagion does not prevent the agents from realising joint surplus, so long as they have common knowledge that their ex-post joint surplus from the deal is weakly positive. He concludes by asking "What change[s] if we consider ... mechanism[s] that determine not only whether a deal takes place but which deal is chosen? ... Is it possible to describe the worst-case information structure?" (pp. 355–356). Our two-sided scheme entails common knowledge that the ex-post joint surplus from bribery is weakly positive, and yet we find that contagion does play an important role in this scheme. We conclude that contagion does become problematic when the players are trying to negotiate the terms of a deal, because the players' types adversely select which terms to accept. The two-sided scheme has a worst case information structure which leads to all deals being rejected for a particular distribution of payoffs (payoffs are endogenous in our setting, but exogenous in his). This worst case information structure has independent and uniform signals that quantify the severity of the lemons problem faced by the recipient.

Our problem fits into a larger class of general mechanism design problems in which the designer chooses both transfers and information.[8] A particularly relevant and recent paper is Halac et al. (2021)'s *Ranking Uncertainty in Organisations*. They show how 'ranking schemes' can create strategic uncertainty and thereby induce a team of workers to exert complementary effort on a project. Ranking schemes are superficially similar to ours in two respects. Firstly, all the players receive a private message. Secondly, the distribution of payoffs is chosen so that work is dominant strategy for the players with the highest possible message realisation, and each player finds it optimal to work conditional on the belief that all players with the same message or higher will work. Thus, like ours, their scheme produces an inductive chain that causes working to be a higher order best response for all other workers. However, the mechanism underlying their ranking scheme is qualitatively different from ours. Endogenous private information benefits their designer because each worker's incentives to work are strictly concave in their belief that other workers will work. Therefore, a given incentive is created more cheaply by randomising over beliefs. There is no lemons problem in their scheme because workers have complete information about their own payoffs — other workers' types only affect them indirectly through the other workers' decisions to exert effort. By contrast, asymmetric information only benefits us because it inhibits our the players from negotiating bribes (which are not considered in Halac et al. (2021)). We engineer a lemons problem by designing a scheme in which each player's payoffs depend directly on the message received by the other other player.

[7] See e.g. Kajii and Morris (1997); Morris and Ui (2005).

[8] See Bergemann and Morris (2019); Mathevet et al. (2020); Taneva (2019).

## 7 Conclusion

We show how information design can be used together with transfer design to deter bribes by engineering a lemons problem. The optimal scheme characterised in our main result, theorem 1, accommodates monitoring errors, costs strictly less than other schemes in the literature, and is relatively simple to implement. This scheme also gives insights into 'worst case' information structures and gives an upper bound on the amount of surplus lost to contagion in bargaining games.

We have shown that the two-sided scheme deters bribes, but bribery contracts are only a special case of an incentive compatible side-contract. Side contracts additionally allow for the possibility of message dependent bribes, $b(x)$, and correlated agreement strategies that depend on both messages, $\sigma(x)$. We conjecture that the two-sided scheme deters all side contracts. One possibility for future research is to prove this by showing that the core of the cooperative game with incomplete information (Myerson, 2007; Forge and Serrano, 2013) induced by the scheme is empty.

Studying the core of the cooperative game induced by the two-sided scheme would also be valuable for extending our results to more than two players. We see this as a particularly promising avenue for future research because it could help us to utilise the information held by potential whistleblowers. In particular, if the lemons problem can be made disproportionately worse by spreading information across an even larger number of of 'inspectors', then we expect to find that the costs of implementing compliance can be further reduced by offering stochastic rewards to whistleblowers. This stands in contrast to the case without information design, where hiring multiple monitors does not help to deter bribery (Stapenhurst, 2019).

Finally, the use of endogenous lemons to deter collusion may have applications beyond monitoring. For instance, it can used used to deter illegal trades, such as weapons, drugs, and human tracking. We also speculate that it could also be used to break up cartels or to deter sub-coalitions of would-be signatories from undermining international agreements.

## References

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics 84*(3), 488–500.

Baliga, S. and T. Sjöström (1998). Decentralization and collusion. *Journal of Economic Theory 83*(2), 196–232.

Becker, G. S. (1968). Crime and punishment: An economic approach. *Journal of Political Economy 76*(2), 169–217.

Bergemann, D. and S. Morris (2019, March). Information design: A unified perspective. *Journal of Economic Literature 57*(1), 44–95.

Carroll, G. (2016). Informationally robust trade and limits to contagion. *Journal of Economic Theory 166*(C), 334–361.

Cremer, J. and R. P. McLean (1988, November). Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions. *Econometrica 56*(6), 1247–1257.

Duflo, E., M. Greenstone, R. Pande, and N. Ryan (2013, 09). Truth-telling by Third-party Auditors and the Response of Polluting Firms: Experimental Evidence from India*. *The Quarterly Journal of Economics 128*(4), 1499–1545.

Forge, F. and R. Serrano (2013). Cooperative games with incomplete information: Some open problems. *International Game Theory Review 15*(02), 1340009.

Garrett, D., G. Georgiadis, A. Smolin, and B. Szentes (2021). Optimal technology design.

Halac, M., E. Lipnowski, and D. Rappoport (2021, March). Rank uncertainty in organizations. *American Economic Review 111*(3), 757–86.

Kajii, A. and S. Morris (1997). The robustness of equilibria to incomplete information. *Econometrica 65*(6), 1283–1309.

Laffont, J.-J. and D. Martimort (1997). Collusion under asymmetric information. *Econometrica 65*(4), 875–912.

Mathevet, L., J. Perego, and I. Taneva (2020). On information design in games. *Journal of Political Economy 128*(4), 1370–1404.

Morris, S. and T. Ui (2005). Generalized potentials and robust sets of equilibria. *Journal of Economic Theory 124*(1), 45–78.

Myerson, R. B. (2007). Virtual utility and the core for games with incomplete information. *Journal of Economic Theory 136*(1), 260–285.

Ortner, J. and S. Chassang (2018). Making corruption harder: Asymmetric information, collusion, and crime. *Journal of Political Economy 126*(5), 2108–2133.

Stapenhurst, C. (2019). *How Many Corruptible Monitors does it take to Implement an Action?* Ph. D. thesis, University of Edinburgh.

Strausz, R. (1997). Delegation of monitoring in a principal-agent relationship. *Review of Economic Studies 64*(3), 337–357.

Taneva, I. (2019, November). Information design. *American Economic Journal: Microeconomics 11*(4), 151–85.

Tirole, J. (1986). Hierarchies and bureaucracies: On the role of collusion in organizations. *Journal of Law, Economics, & Organization 2*(2), 181–214.

Topsøe, F. (2004). Some bounds for the logarithmic function. *RGMIA Res. Rep. Collection 7*(2), 1–20.

United Nations (2018). Global cost of corruption at least 5 per cent of world gross domestic product, secretary-general tells security council, citing world economic forum data.

von Negenborn, C. and M. Pollrich (2020). Sweet lemons: Mitigating collusion in organizations. *Journal of Economic Theory 189*, 105074.

# Appendices

## A Mathematical appendix

### A.1 Proof of proposition 4

The cost of the scheme $\mathcal{S}^{II} = (q^{\mathrm{II}}, (f_e^{\mathrm{II}}, f_0^{\mathrm{II}}, r^{\mathrm{II}}))$ is

$$\begin{aligned}
\mathbb{E}[\pi_c(r^{\mathrm{II}}(x_I) - f_e^{\mathrm{II}}(x_I)) + (1-\pi_c)(-f_0^{\mathrm{II}}(x_I))] &= \int_{\frac{k-\Pi}{k}}^{1} \pi_c r^{\mathrm{II}}(x_I) - (\pi_c f_e^{\mathrm{II}}(x_I) + (1-\pi_c) f_0^{\mathrm{II}}(x_I)) dx_I \\
&= \int_{\frac{k-\Pi}{k}}^{1} \pi_c \left(k - \frac{k-\Pi}{x_I}\right) - (\pi_c(1-\pi_c)k - (1-\pi_c)\pi_c k) dx_I \\
&= \pi_c \int_{\frac{k-\Pi}{k}}^{1} k - \frac{k-\Pi}{x_I} dx_I \\
&= \pi_c \left(k\left(1 - \frac{k-\Pi}{k}\right) - (k-\Pi)\Big[\ln x_I\Big]_{\frac{k-\Pi}{k}}^{1}\right) \\
&= \pi_c \left[\Pi + (k-\Pi)\ln\left(1 - \frac{\Pi}{k}\right)\right] \\
&\le \pi_c \left[\Pi - (k-\Pi)\frac{\Pi}{k}\right] \\
&= \pi_c \frac{\Pi^2}{k} \xrightarrow{k\to\infty} 0,
\end{aligned}$$

where the inequality comes from the fact that $\ln(1+x) \le x$ for all $x > -1$ (Topsøe, 2004).

We now show that the scheme is optimal by showing that any feasible solution to Rose's problem costs weakly more than $\mathcal{S}^{II}$. In any scheme where Isla has full information about the transfers, she will accept the bribe $b$ whenever she receives a message $x_I$ such that $r(x_I) < b$. Finn anticipates this, so he accepts bribe $b$ if $\mathbb{E}[f_e(x_I) - f_0(x_I) | r(x_I) < b] >$

$b$. Therefore, any informed inspector scheme which deters bribes must at least satisfy $\mathbb{E}[f_e(x_I) - f_0(x_I) | r(x_I) < b] \leq b$ for all bribes $b \geq 0$.

Define $E(b) := \underset{x \sim q}{\mathbb{E}}[f_e(x_I) - f_0(x_I) | r(x_I) \geq b]$, and $F_r(b) = \underset{x \sim q}{\mathbb{P}}[r(x_I) < b]$. We use the fact that $E(0) = \mathbb{E}[f_e(x_I) - f_0(x_I)] = \mathbb{E}[f_e(x_I) - f_0(x_I) | r(x_I) < b] F_r(b) + E(b)(1 - F_r(b))$ to rewrite the no-bribery constraint as $\frac{E(0) - E(b)(1 - F_r(b))}{F_r(b)} \leq b$. This rearranges to give $F_r(b) \leq \frac{E(b) - E(0)}{E(b) - b}$. If $E(0) < b$ then $\frac{E(b) - E(0)}{E(b) - b} > 1$, so the constraint is slack. Otherwise, if $E(0) \geq b$, then $\frac{E(b) - E(0)}{E(b) - b}$ is increasing in $E(b)$ and decreasing in $E(0)$. Finn's limited liability constraint requires that $E(b) \leq k$ and his incentive compatibility constrain requires that $E(0) \geq \Pi$. It follows that every bribery-proof informed inspector scheme satisfies

$$F_r(b) \leq \frac{E(b) - E(0)}{E(b) - b} \leq \frac{k - \Pi}{k - b}.$$

In $\mathcal{S}^{II}$, the distribution of rewards is

$$F_r^{II}(b) = \underset{x_I \sim q^{II}}{\mathbb{P}}[r^{II}(x) < b] = \underset{x_I \sim q^{II}}{\mathbb{P}}[x_I < \frac{k - \Pi}{k - b}] = \frac{k - \Pi}{k - b}.$$

Hence the distribution of rewards in any bribery-proof solution must first order stochastically dominate the distribution of rewards in the scheme $\mathcal{S}^{II}$. This implies that every bribery-proof solution has a weakly higher expected reward than does $\mathcal{S}^{II}$. At the same time, $\mathcal{S}^{II}$ exactly satisfies Finn's (VP) constraint, so every feasible scheme must have a weakly lower expected fine than $\mathcal{S}^{II}$. The cost of a scheme is given by Isla's expected reward minus Finn's expected fine, so it follows that every feasible scheme must cost weakly more than $\mathcal{S}^{II}$.

## A.2 Informed firm schemes

Consider the biased coin toss (informed firm) scheme described in table 2. Similar techniques to those used in section 4.1 show that this scheme satisfies voluntary participation, incentive compatibility and deters bribes. The cost of the scheme is $\pi_c \sqrt{\kappa}(\sqrt{\kappa} - \sqrt{\kappa - \Pi})$. When $\kappa = 4$ and $\Pi = 3.9$ we get that this informed firm scheme costs $3.4\pi_c$, whereas the cheapest informed inspector scheme costs $\pi_c \left[\Pi + (k - \Pi) \ln\left(1 - \frac{\Pi}{k}\right)\right] = 3.5\pi_c$. In general, informed firm schemes are cheaper when $\kappa$ is small enough relative to $\Pi$.

Table 2: The biased coin toss (informed firm) scheme.

| | Lemon | Peach |
|---|---|---|
| Probability | $1 - \sqrt{1 - \Pi/\kappa}$ | $\sqrt{1 - \Pi/\kappa}$ |
| Fine Finn | 4 | $\left(1 - \sqrt{1 - \Pi/\kappa}\right)\kappa$ |
| Reward Ina | 4 | 0 |

The optimal informed firm scheme takes the following form,

$$q^{\mathrm{IF}} \text{ is uniform on } [0,1]$$

$$f_e^{\mathrm{IF}}(x_F) = (1-\pi_c)\begin{cases} k & \text{if } x_F \leq \tilde{x}^{IF} \\ \frac{\tilde{x}^{IF}}{x_F} & \text{otherwise} \end{cases}$$

$$f_0^{\mathrm{IF}}(x_F) = -\pi_c\begin{cases} k & \text{if } x_F \leq \tilde{x}^{IF} \\ \frac{\tilde{x}^{IF}}{x_F} & \text{otherwise} \end{cases}$$

$$r^{\mathrm{IF}}(x_F) = \begin{cases} k & \text{if } x_F \leq \tilde{x}^{IF} \\ 0 & \text{otherwise,} \end{cases}$$

where $\tilde{x}^{IF}$ solves $\tilde{x}^{IF}\ln\left(\frac{e}{\tilde{x}^{IF}}\right) = \frac{\Pi}{\kappa}$. Similar techniques to those used in appendix A.1 show that this scheme is feasible and cheaper than any other informed firm scheme. However, it is not easy to work with because no analytical solution for $\tilde{x}^{IF}$ exists.

## A.3 Collusion proofness

In the main body of the paper, we restrict attention to bribery contracts. The revelation principle tells us that for every equilibrium of every bribery game, there exists an incentive compatible collusive side contract. An incentive compatible collusive side contract is a generalisation of a bribery contract in which the players can play correlated strategies and in which side of the bribe can depend on the players' types. We use this more general notion of collusion to prove that the regulator can restrict attention to collusion-proof contracts, and therefore to bribery-proof contracts.

Our notion of collusive side contracts is inspired by Laffont and Martimort (1997). Suppose that a fourth player, Marta the Mafia, offers to enforce collusive side contracts. In a direct collusive side contract, Isla and Finn report their private messages $x = (x_I, x_F)$ to Marta; Marta tells Isla to be silent with probability $s(x)$; and Marta makes (potentially negative) transfers $b_F(x)$ to Finn and $b_I(x)$ to Isla. We denote a collusive side contract by $\mathcal{C} = (s, b_F, b_I)$. A side contract $\mathcal{C}$ is budget balanced if Marta makes does not lose money on average, i.e. $\mathbb{E}_x[b_F(x) + b_I(x)] \leq 0$. It is incentive compatible if it satisfies the usual incentive compatibility constraints that ensure it is in Isla and Finn's best interest to report their message truthfully. Finally, Marta cannot force Isla and Finn to participate, so $\mathcal{C}$ must satisfy the usual voluntary participation constraints. A balanced budget, incentive compatible, voluntary side contract is *interim efficient* if it delivers a strictly higher payoff to at least one type of one player, and a weakly higher payoff to all types of both players, than every other balanced budget, incentive compatible, voluntary side contract. A scheme $(q,(f_e,f_0,w))$ is *weakly collusion-proof* if no interim efficient side contract exists. Hence, weak collusion-proofness implies bribery-proofness, but the reverse does not necessarily hold. In the main paper, we show that our scheme is bribery-proof. We conjecture that it is also weakly collusion-proof.

## A.4 The two sided scheme costs strictly less than the one sided schemes.

The informed inspector scheme costs strictly more than the two-sided scheme at all parameter values:

$$
\begin{aligned}
c^{\mathrm{II}}/\pi_c &= \Pi + (\kappa - \Pi)\ln\left(1 - \frac{\Pi}{\kappa}\right) \\
&\geq \Pi - (\kappa - \Pi)\frac{\frac{\Pi}{\kappa}}{\sqrt{1 - \frac{\Pi}{\kappa}}} \\
&= \Pi - (\kappa - \Pi)\frac{\frac{\Pi}{\sqrt{\kappa}}}{\sqrt{\kappa - \Pi}} \\
&= \Pi - \sqrt{\kappa - \Pi}\frac{\Pi}{\sqrt{\kappa}} \\
&= \frac{\Pi}{\sqrt{\kappa}}\left(\sqrt{\kappa} - \sqrt{\kappa - \Pi}\right) \\
&> (\sqrt{k} - \sqrt{\kappa - \Pi})^2 \\
&= c^*/\pi_c,
\end{aligned}
$$

where the first inequality is a property of the logarithm function, and the second comes from the fact that

$$
\begin{aligned}
\sqrt{\kappa} &> \sqrt{\kappa - \Pi} \\
\sqrt{\kappa}\sqrt{\kappa - \Pi} &> \kappa - \Pi \\
\Pi &> \kappa - \sqrt{\kappa}\sqrt{\kappa - \Pi} \\
\frac{\Pi}{\sqrt{\kappa}} &> \sqrt{\kappa} - \sqrt{\kappa - \Pi}.
\end{aligned}
$$

Now we show that the two-sided scheme costs half as much as the informed inspector scheme in the limit.

$$
\begin{aligned}
\frac{c^*}{c^{\mathrm{II}}} &\leq \frac{\sqrt{\kappa}}{\Pi}(\sqrt{\kappa}-\sqrt{\kappa-\Pi}) \\
&= \frac{\kappa-\sqrt{\kappa(\kappa-\Pi)}}{\Pi} \\
&= \frac{\frac{\kappa-\sqrt{\kappa(\kappa-\Pi)}}{\Pi}(\kappa+\sqrt{\kappa(\kappa-\Pi)})}{\kappa+\sqrt{\kappa(\kappa-\Pi)}} \\
&= \frac{\frac{\kappa^2-\kappa(\kappa-\Pi)}{\Pi}}{\kappa+\sqrt{\kappa(\kappa-\Pi)}} \\
&= \frac{\kappa}{\kappa+\sqrt{\kappa(\kappa-\Pi)}} \xrightarrow{\kappa\to\infty} \frac{1}{2},
\end{aligned}
$$

where the first inequality comes from the previous calculations.

## A.5 Derivation of Conditional Expectations

Here we derive equation (2). The derivation of equation (3) is completely analogous.

$$
\begin{aligned}
\mathbb{E}\left[\min\{1, x_F^\lambda x_I^{-1}\} \mid x_I, x_F \leq y_F\right] &= \begin{cases} \frac{1}{y_F}\int_0^{y_F} \frac{x_F^\lambda}{x_I}\, dx_F & \text{if } x_I \geq y_F^\lambda, \\ \frac{1}{y_F}\int_0^{x_I^{1/\lambda}} \frac{x_F^\lambda}{x_I}\, dx_F + \frac{1}{y_F}\int_{x_I^{1/\lambda}}^{y_F} 1\, dx_F & \text{if } x_I < y_F^\lambda. \end{cases} \\
&= \begin{cases} \frac{1}{y_F x_I}\left[\frac{1}{\lambda+1} x_F^{\lambda+1}\right]_0^{y_F} & \text{if } x_I \geq y_F^\lambda, \\ \frac{1}{\lambda+1}\frac{x_I^{\frac{\lambda+1}{\lambda}}}{y_F x_I} + 1 - \frac{x_I^{1/\lambda}}{y_F} & \text{if } x_I < y_F^\lambda. \end{cases} \\
&= \begin{cases} \frac{1}{\lambda+1}\frac{y_F^\lambda}{x_I} & \text{if } x_I \geq y_F^\lambda, \\ 1 - \frac{\lambda}{\lambda+1}\frac{x_I^{1/\lambda}}{y_F} & \text{if } x_I < y_F^\lambda. \end{cases}
\end{aligned}
$$

Then the fact that $r^*(x) = 1 - \min\{1, x_F^\lambda x_I^{-1}\}$ gives equation 2.

## A.6 Proof of lemma 2

*Proof.* Let $\mathcal{S} = (q, (w, f_e, f_0))$ be any scheme with finite support that deters bribe $b^*$. $\mathcal{S}$ and $b^*$ induce a distribution $p$ over payoffs defined by:

$$p(v_F, v_I) := \mathbb{P}_q[v_F = b^* - r(x), v_I = f_e(x) - f_0(x) - b^*].$$

The payoff distribution $p$ satisfies *Condition A* if $\max\{v_M, v_I\} \geq 0$ with probability 1, i.e. if at least one player benefits from the bribe ex post. Claim 1 shows that it is without loss of generality to restrict attention to such schemes.

**Claim 1.** *We can assume without loss of generality that either $r(x) \le b^*$ or $f_e(x) - f_0(x) \ge b^*$ for all messages $x$.*

*Proof.* Let $\mathcal{S}$ be a feasible scheme with an outcome $x$ in which $r(x) \ge b^*$ and $f_e(x) - f_0(x) \le b^*$. Define a new scheme $\mathcal{S}'$ by

$$
\begin{array}{c|ccc}
\mathcal{S} & \dots & x_I & \dots \\
\hline
\vdots & \ddots & \vdots & \ddots \\
x_F & \dots & (\Delta(x), r(x)) & \dots \\
\vdots & \ddots & \vdots & \ddots
\end{array}
\quad \longmapsto \quad
\begin{array}{c|cccc}
\mathcal{S}' & \dots & x_I^a & x_I^b & \dots \\
\hline
\vdots & \ddots & \vdots & \ddots & \\
x_F^a & \dots & (2\Delta(x), 0) & (0, 2r(x)) & \dots \\
x_F^b & \dots & (0, 2r(x)) & (2\Delta(x), 0) & \dots \\
\vdots & \ddots & \vdots & \ddots &
\end{array}
$$

Suppose $\mathcal{S}'$ has an equilibrium $(\sigma_F', \sigma_I')$ for some bribe $b$. Consider the following cases:

1. If $\sigma_F'(y_F^a) = \sigma_F'(y_F^b)$ and $\sigma_I'(y_I^a) = \sigma_I'(y_I^b)$ then $V_i'(y_i^a; \sigma_{-i}') = V_i'(y_i^b; \sigma_{-i}')$ for $i = F, I$. Define a new strategy pair $(\sigma_F, \sigma_I)$ with $\sigma_i(x_i) = \sigma_i'(x_i)$ for all $x_i \ne y_i$, and $\sigma_i(y_i) = \sigma_i'(y_i^a)$ for $i = F, I$. The fact that $v_i(x_i, y_{-i}) = \frac{1}{2}v_i'(x_i, y_{-i}^a) + \frac{1}{2}v_i'(x_i, y_{-i}^b)$ for all $x_i \ne y_i$, and
$$
\begin{aligned}
v_i(y_i, y_{-i}) &= \frac{1}{2}v_i'(y_i^a, y_{-i}^a) + \frac{1}{2}v_i'(y_i^a, y_{-i}^b) \\
&= \frac{1}{2}v_i'(y_i^b, y_{-i}^a) + \frac{1}{2}v_i'(y_i^b, y_{-i}^b).
\end{aligned}
$$
then implies that $V_i(x_i; \sigma_{-i}) = V_i'(x_i; \sigma_{-i}')$ for all $x_i \ne y_i$, and $V_i(y_i; \sigma_{-i}) = V_i'(y_i^a; \sigma_{-i}') = V_i'(y_i^a; \sigma_{-i}')$. Therefore $(\sigma_F, \sigma_I)$ is an equilibrium under $\mathcal{S}$. The fact that $\mathcal{S}$ deters bribes implies that $\sigma_i = 0$ for either $i = I$ or $i = F$, which in turn implies that implies $\sigma_i' = 0$.

2. If $\sigma_F'(y_F^a) > \sigma_F'(y_F^b)$ then $V_I(y_I^a; \sigma_F') > V_I(y_I^b; \sigma_F')$, which implies that $\sigma_I'(y_I^a) > \sigma_I'(y_I^b)$. This in turn implies that $V_F(y_F^b; \sigma_I') > V_F(y_F^a; \sigma_I')$, but then $\sigma_F'(y_F^a) > \sigma_F'(y_F^b)$ is not a best response for Finn, giving a contradiction.

3. All other cases are symmetric to case 2.

Therefore at least one agent plays the null strategy in every equilibrium of $\mathcal{S}'$. This means that $\mathcal{S}'$ is feasible. Moreover, $c(\mathcal{S}') = c(\mathcal{S})$. □

The *no-deal measure* of an equilibrium $(\sigma_F, \sigma_I)$ gives the probability that no bribery occurs. Formally, it is by $p_{\mathrm{ND}}(E) = \int_E 1 - \sigma_F(x_F)\sigma_I(x_I)\, dp(x)$ for all $E \subseteq \operatorname{supp} p$. A scheme $\mathcal{S}$ deters bribe $b^* \iff$ the 'no-deal' measure of every equilibrium is equal to the prior $p$.

Given a prior distribution $p$ over payoffs, a distribution $\tilde{p}$ has a *disjoint negative decomposition bound* if there exists distributions $p_F$ and $p_I$ with the properties that

1. $\tilde{p} \le p_F + p_I$,

2. $p_F + p_I \le p$ and

3. $\int_{\mathbb{R}} v_F(x)\, dp_M(x) < 0$ and $\int_{\mathbb{R}} v_I(x)\, dp_I(x) < 0$.

**Proposition 5** (3.1, Carroll 2016)**.** *If $p$ satisfies condition A then for every information structure there exists an equilibrium whose no-deal measure has a 'disjoint negative decomposition bound' (DNDB).*

The fact that $p$ deters bribes implies that $p$ has a DNDB.

**Proposition 6** (3.1, Carroll 2016)**.** *If $p$ has a DNDB then there exists a public information structure such that, in any equilibrium, the no-deal measure is equal to $p$.*

Thus there exists a public information structure that deters bribe $b$. The cost of the scheme is unaffected by the information structure (it is determined exclusively by the payoff distribution), so Rose can restrict attention to public schemes. □