J Regul Econ (2007) 31:209–233
DOI 10.1007/s11149-006-9020-x

ORIGINAL ARTICLE

Can self regulation work?: a story of corruption, impunity and cover-up

Javier Núñez

Published online: 24 February 2007

Abstract This paper analyzes the reputation-based incentives of a Self-Regulatory Organization (SRO) to detect and expose consumer fraud committed by its members, and the members' incentives to bribe the SRO in exchange for a cover-up to avoid an external punishment. In a corruption-free benchmark, SROs are effective in detecting, exposing and deterring fraud only if exposure yields a reputation gain to the SRO, which depends on consumers inferences about the SRO's type. However, if this case prevails the member can succeed in bribing the SRO in exchange for a cover-up and impunity. Despite this, a bribed SRO yields more vigilance and lower fraud than no self-regulation at all.

Keywords Self-Regulation · Fraud · Corruption · Bribery

JEL Classifications L51 · K20

1 Introduction

Self-regulation of fraud and malpractice is a common regulatory scheme in various industries and institutions worldwide. For example, self-regulation is common in banking and financial services, and in many collegiate professions such as lawyers, doctors, nurses, accountants and auditors throughout many

J. Núñez (✉)
Department of Economics,
Universidad de Chile,
Diagonal Paraguay 257, Santiago, Chile
e-mail: jnunez@econ.uchile.cl

countries.[1] Self-regulation also exists in a variety of public and private organizations, for example in political parties and religious organizations. Even State institutions such as, for example, regulatory agencies and judicial institutions often share some features of self-regulation.[2]

Yet, self-regulation has received little attention from theorists despite its widespread prevalence across different markets and organizations and the public attention often devoted to the scandals that regularly emerge in self-regulated activities. This article develops a theoretical model to investigate the incentives that SROs face to enforce good quality among its members and expose evidence of fraud and malpractice to the public. The model also studies how corruption may emerge endogenously within an SRO, and how it may affect the incentives that SROs and their members face.[3]

As the name suggests, self-regulation is essentially a scheme whereby the enforcement of quality is delegated to the suppliers. The whole rationale for self-regulation rests on the notion that suppliers must somehow form an organization (namely a Self-Regulatory Organization (SRO)) in order to monitor the quality provided by its members and disclose evidence of malpractice and product failure to consumers.[4] Advocates of self-regulation have pointed out an interesting advantage of self-regulation over public regulation; self-regulators are often better informed than public regulators about the provision of quality and the changing conditions of the regulated activity. However, self-regulation also implies *regulatory capture* by definition, as the incentives of the regulator and the regulated party coincide. This raises the fundamental question of whether SROs have the correct *incentives* to deal effectively with fraud and malpractice, and where these incentives may come from. For example, the claims of fraud, negligence, malpractice and corruption by self-regulated auditors and accountants in the Enron-Andersen and Worldcom scandals in 2001 have raised justified scepticism about self-regulation.[5] However, no one really

[1] For example, The Federal Banking Commission and the Swiss Bankers Association jointly draw up binding codes of conduct that define the standards of good industry practice. In the US the American Institute of Certified Public Accountants is a self-regulating body joined by more than three hundred thousand auditors and accountants. In the UK, self-regulation was the fundamental principle in financial regulation since the Financial Services Act of 1986, being replaced since 2001 by the Financial Services Authority, a public body with powers to fight financial fraud and malpractice.

[2] The author is grateful to a referee for having suggested the interpretation of public regulatory agencies as a particular form of self-regulation. For an introduction to the self-regulation of quality, see Scarpa (1999).

[3] There is a good reason to prefer a theoretical to an empirical approach to study self-regulation; an SRO's monitoring and disclosure decisions are usually not observable by an analyst. Hence, an empirical approach may fail to distinguish whether the lack of evidence of fraud reflects adequate or inadequate self-regulation, as fraud might be undetected or underreported by the SRO.

[4] As self-regulation exists in markets as well as in other activities, in what follows we employ the terms "consumers" and "the public" interchangeably.

[5] See for example "The Lessons from Enron", *The Economist*, February 7, 2002. Similar scepticism has risen about other examples of self-regulated activities such as the sex scandals within the Catholic Church worldwide in recent years.

knows whether these are isolated cases within a largely well-functioning self-regulation, or whether they are only the conspicuous tip of widespread fraud and malpractice undetected or concealed by SROs.[6]

Although self-regulation may take different institutional forms, it is characterized by three common features: (i) SROs face a principal-agent problem whereby quality is ultimately determined by the SRO members, (ii) SRO incentives to monitor quality and expose malpractice to the public are mostly reputation-based, and (iii) SROs usually exist in *credence goods* industries, that is, where consumers cannot properly observe product quality either prior to or after purchase. This paper incorporates these essential regularities of self-regulation into a dynamic game of incomplete information, whereby an SRO (the principal) must decide how much to invest in monitoring the conduct of an SRO member (the agent), and whether to expose any evidence of fraud to consumers. The SRO has private information about its vigilance cost. Due to the credence-good assumption, consumers are unable to observe the amount of fraud or the SRO's vigilance effort. Upon observing either exposure or non-exposure, consumers apply Bayes' rule to infer the SRO's type. This provides a reputation-based concern to the SRO to strategically choose its vigilance level and exposure decision.[7] Subsequently, we investigate the role of corruption between the SRO and its member.[8] Upon exposure, the fraudulent member would face an external punishment in addition to the SRO's internal punishment. The member can avoid this external punishment by bribing the SRO to prevent exposure. In this context, this paper studies the reputation-based incentives faced by the SRO to detect fraud and expose it to consumers, the conditions that trigger collusion between an SRO and its members, the consequences of collusion, and whether a corruptible SRO does better or worse than no self-regulation at all. As we shall see, these issues are related. Bribery is necessary only if exposure is a credible threat. This depends on the consequences of exposure to the SRO's reputation, which ultimately depend on how consumers interpret the SRO's exposure decision: exposure would be a good signal of an SROs vigilance if consumers interpreted it as the result of adequate vigilance, but it would be a bad signal if they saw it instead as evidence of widespread fraud *caused* by *little* vigilance.

This article differs from the limited literature on self-regulation on many respects. A large part of this literature does not focus on the performance of self-regulation in preventing and exposing malpractice, but on how it can be

6 However, see Pirrong (1995) for evidence of malfunctioning self-regulation of commodity exchanges.

7 The term "reputation" has different meanings in economics and game theory. In repeated games of complete information players build a reputation for behaving in a certain way, while in games of incomplete information, players have a reputation for having a given characteristic that may affect their behavior. This work employs the term "reputation" in the latter sense. See Fudenberg and Tirole (1991) for a clarifying discussion.

8 The terms "corruption" and "collusion" are employed interchangeably to mean a situation where the member bribes the SRO in exchange for a cover-up.

employed to limit competition, or to promote professional or sectorial interest.[9] Other contributions have focused on how self-regulation emerges as an strategic response to preempt public action or regulation.[10] This work also diverges from the market-mechanism literature of quality provision, which often assumes that quality is observable by consumers either by searching or by experience. This naturally reduces consumers' informational disadvantage and limits the amount of fraud that suppliers can optimally choose. In this work, instead, we focus on credence goods or services where self-regulation is often found. Accordingly, this paper studies whether SROs have incentives to protect consumers against fraud beyond what consumers can do for themselves. In other works quality is typically assumed to be either exogenous (as in a standard signalling model) or a direct choice variable of a black box-type supplier.[11] In this work instead, product quality is determined endogenously from an explicit principal-agent strategic relationship between the SRO and its members. This also allows us to study the possibility of collusion between the SRO and its members, an issue not often addressed by the market-mechanism literature. The rest of the paper is structured as follows. Section 2 presents the model. Section 3 characterizes its equilibria and presents the main results. Finally, Sect. 4 offers some concluding remarks.

2 A model of self-regulation with corruption

In this section, we develop a dynamic model of incomplete information that incorporates the three essential regularities of self-regulation: self-regulation as an explicit principal-agent relationship, a reputation-motivated principal, and the credence-good assumption. We assume that the incentives of self-regulation are mostly reputation-based. Building a reputation is valuable as long as consumers have some degree of uncertainty regarding the features or behavior of the SRO. Consumers might be uncertain about some traits of the SRO, such as its commitment to vigilance, or traits of the SRO members, like members benefit of fraudulent behavior. We focus on the first form of uncertainty, namely the SRO's commitment to vigilance.[12]

The model involves an SRO (the principal), an SRO member (the agent), and consumers of a credence good or service delivered by the SRO, but whose quality is determined by the member. The SRO can monitor the conduct of the

[9] See for example Shaked and Sutton (1981), Bortolotti and Fiorentini (1999) and Van den Bergh (1999).

[10] See for example Stefanadis (2003) and Maxwell,Lyon, and Hackett (2000).

[11] See for example Milgrom (1981),Shapiro (1982), Milgrom and Roberts (1986), Gehrig and Jost (1995) and Emons (1997). In Tirole (1996), a group's reputation is endogenous, but it is not determined in a principal-agent framework, which we argue is a salient feature of self-regulation.

[12] Modeling the two forms of uncertainty in one single model would add significant complexity, as it would require consumers to engage in two separate processes of Bayesian updating in order to decide a course of action. Examining one aspect of reputation helps isolate the various incentives that SROs may face.

member by investing in vigilance at a unit cost c_i, which is high (c_H) with prior probability λ and low (c_L) with probability $1 - \lambda$, with $c_L < c_H$, $i = H, L$. The SRO's unit vigilance cost is known to the SRO and its member, but not to consumers. This informational asymmetry about unit cost c_i represents consumers' uncertainty about the *type* of the SRO regarding its attitude towards vigilance. All things being equal, the low-cost SRO will be more vigilant than the high-cost SRO. SRO type i chooses a level of vigilance $v_i \in [0, \infty)$, which yields a total cost of $c_i v_i$. The member chooses a level of "fraud" or "negligence" $x \in [0, \infty]$, which is beneficial to the member. This benefit might arise, for example, as the direct extraction of resources from consumers, or alternatively as the saving of effort required to deliver a good quality service to consumers. Given vigilance level v_i and fraud level x, SRO type i can detect fraud with probability $p(x, v_i) \leq 1$, which satisfies partial derivatives $p_x > 0, p_v > 0, p_{xx} > 0, p_{vv} < 0$ y $p_{xv} > 0$ and $p_v(x, 0) = \infty$.[13]

If fraud is detected, SRO type i must choose a probability $e_i \in [0, 1]$ of exposing the member to the public, where the pure-strategies "exposure" and "cover-up" correspond to $e_i = 1$ and $e_i = 0$, respectively.[14] Therefore, the probability that consumers observe fraud exposure is $e_i p(x, v_i)$. If the SRO detects fraud, it imposes an internal penalty $T > 0$ on the member regardless of the SRO's exposure decision. This penalty is not observed by consumers. Internal penalty T may be a direct punishment in compliance with the SRO's statutes, or other costs that the SRO can impose on the member, such as unemployment or reducing a discretionary payment or benefit. We assume that the SRO cannot impose an arbitrarily high value for T, which would trivially lead to full fraud deterrence. We justify this with three reasons. First, many forms of SRO punishment have an upper bound beyond the control of the SRO, for example the costs of unemployment. Second, SRO punishments are constrained by what the law allows. Third, T must in practice satisfy a member's participation constraint. This, however, is not addressed here because our focus is on how corruption between a participating member and the SRO can emerge in the presence of an external punishment.

If the SRO exposes fraud to the public, the member faces an *external* punishment $L > 0$ in addition to the internal penalty T. Cost L can be a penalty imposed by the legal system in addition to the internal penalty imposed privately by the SRO, or it may reflect all the costs of facing a trial within the legal system, regardless of the final outcome. Cost L can also reflect the value of a loss in reputation of the member when fraud is exposed to the public. For example, this loss of reputation is likely to be important for professionals exposed and expelled from professional associations.

The possibility of corruption is introduced by assuming that before the SRO commits its exposure decision $e \in [0, 1]$ after detecting a fraudulent member, the member can pay the SRO a bribe $\beta > 0$ in exchange for a cover-up in order to

13 These conditions hold as long as $p(x, v) < 1$.

14 However, only pure strategies survive in equilibrium, as Lemma 1 indicates.

avoid the external penalty L. Instead of modeling such a contract $\{\beta > 0, e = 0\}$ as an explicit bargaining process, we employ all possible Pareto-improving bargaining outcomes.[15] Contracting is costly. Let $\tau > 0$ be the transaction cost of reaching a collusive agreement such that if the member pays bribe β, the net benefit to the SRO is $\beta - \tau$. This transaction cost includes all costs involved in reaching an agreement, any psychological costs of either player associated with bribery, as well as all expected costs associated with a potential public disclosure of the bribe.

We assume that contract $\{\beta > 0, e = 0\}$ is enforceable. This assumption, common in the corruption literature, has been justified in three ways.[16] First, the colluding parties may have a willingness to abide by their promise to cooperate if breaking the "word-of honor" is costly. Second, the parties may be engaged in a long-run relationship, which may have a payoff for developing a reputation for not breaching this kind of contracts. Third, if the benefits from abiding the contract come as a flow, breaking the contract would be costly as it would end the flow. Yet, we offer an additional and perhaps more compelling reason; breaking the contract by exposing fraud increases the risk of disclosing the collusive agreement, which could lead to costly consequences for the SRO.[17]

2.1 The SRO member

The risk-neutral member benefits linearly from the amount of fraud x. This benefit may arise as the direct extraction of resources from consumers or as the saving of the effort required to deliver a good quality service to consumers. The member's optimal choice depends on whether he expects a collusive contract to emerge if the SRO has detected fraud. Let γ be the probability that a collusive contract is established following fraud detection. Then the risk-neutral member chooses an amount of fraud x to maximize,

$$x - p(x, v)T - (1 - \gamma)ep(x, v)L - \gamma p(x, v)\beta \tag{1}$$

and the first-order condition is,[18]

$$1 - p_x(T + (1 - \gamma)eL + \gamma\beta) = 0 \tag{2}$$

and the second-order condition $-p_{xx} < 0$ is satisfied. Equation (2) implicitly provides the member's reaction function $x^*(v)$, from which several implications follow. The slope of the member's reaction function is $\frac{\mathrm{d}x^*}{\mathrm{d}v} = -\frac{p_{xv}}{p_{xx}} < 0$. Hence,

[15] That is, where the SRO and the member are both better-off than under their reservation (non-corruption) payoffs.

[16] See for example Laffont and Tirole (1991).

[17] Note that only the SRO can break the contract; the member is unlikely to succeed in claiming back the bribe once paid.

[18] Recall that an interior solution requires internal penalty T to be sufficiently low.

the member's optimal level of fraud $x^*(v)$ decreases with the level of SRO vigilance. Note also that if $\beta < L$ (which will hold in equilibrium), the optimal level of fraud x^* would be higher under collusion ($\gamma = 1$) than without collusion ($\gamma = 0$) for any given level of vigilance. Equation 2 has yet another implication that is of central importance for this model: the *overall* effect of vigilance v on the probability of fraud detection $p(x(v), v)$ is ambiguous. Indeed, the total effect of vigilance of the probability of fraud discovery along the member's reaction function is given by,

$$\frac{\mathrm{d}p(x^*(v), v)}{\mathrm{d}v} = p_v + p_x \frac{\mathrm{d}x^*(v)}{\mathrm{d}v} = p_v - \frac{p_x p_{xv}}{p_{xx}} \tag{3}$$

Expression (3) has an ambiguous sign because it has two opposite effects; On the one hand, higher vigilance partially increases $p(x, v)$, given assumption $p_v > 0$. On the other hand, higher vigilance reduces the member's optimal fraud level $x^*(v)$, which in turn reduces $p(x, v)$ because $p_x > 0$. Expression (3) indicates that each of these two opposing effects of vigilance can dominate over the other, such that the overall effect of vigilance on the probability of fraud detection can be either positive, zero, or negative, depending on the exogenous derivatives identified in (3).[19] The intuition is that, although vigilance makes fraud detection more likely given a fixed level of fraud, if the optimal level of fraud chosen by the agent is very responsive to vigilance it can eventually overcompensate the positive effect of vigilance on the probability of detection. As we shall see, this feature is an essential aspect of the model because it turns ambiguous the Bayesian posterior beliefs about the SRO's type that consumers should hold after observing either fraud exposure or non-exposure.

2.2 Consumers

Following the credence-good assumption, consumers are unable to observe the SRO's vigilance level, or the level of fraud. Hence, consumers can only observe either evidence of fraud disclosed by the SRO, or lack of it over some time interval. Since our aim is to study the "pure" incentives of self-regulation in a credence-goods context, other sources of information such as the media, a "watchdog" group or a public regulator are ruled out.[20] Nevertheless, introducing these elements into the present analysis remains as an interesting avenue for future research.[21]

[19] For example, functions $p(x, v) = x^2 \mathrm{Ln} v$ and $p(x, v) = (a + e^x)\mathrm{Ln} v$, $(a > 0)$ yield a negative and a positive sign of $\frac{\mathrm{d}p(x^*(v),v)}{\mathrm{d}v}$, respectively.

[20] However, the former two sources of information are unlikely to be significant in credence-goods industries, as these parties can also have difficulties of observing product quality.

[21] Nunez (2001) combines self-regulation and parallel regulation in a collusion-free context, and shows that Parallel regulation often sharpens SRO vigilance incentives, as the threat of fraud discovery by the public regulator would yield a loss in reputation to the SRO. However, the incentives to cover-up remain.

We assume that consumers place value on the SRO's type, which affects the SRO's payoff. Let w_i be the value to the SRO of being perceived by consumers as an SRO of type i, with $w_L - w_H = W$. We first explore equilibrium for any exogenous $W > 0$. This reduced-form assumption provides an "optimistic" environment for studying SRO vigilance and exposure incentives. Hence, proving non-existence of such incentives under these favorable conditions would constitute a significant negative result about the performance of self-regulation. Alternatively, W could be made endogenous by assuming that consumers are concerned about (unobservable) equilibrium fraud. In this context inferring the SRO's type is valuable to consumers because they know that equilibrium fraud x^* is determined by the level of vigilance, which in turn is determined by the SRO's type. In this case, $w_i(x^*(v_i^*))$, with $\frac{\partial w_i}{\partial x} < 0$.

After observing either fraud exposure or no exposure by the SRO, consumers employ Bayes' Rule to update their prior belief λ on the low-cost SRO type. Let equilibrium fraud discovery probabilities for SRO type i be $p_i = p(x^*(v_i), v_i)$. Then the equilibrium probabilities that consumers observe exposure by SRO type i is $e_i p_i$. Therefore, consumers' Bayesian update of λ conditional upon SRO exposure, b_e, and conditional upon non-exposure, b_n, are given by:

$$\begin{aligned} b_e &= \frac{e_L p_L \lambda}{e_L p_L \lambda + e_H p_H (1-\lambda)} \\ b_n &= \frac{(1-e_L p_L)\lambda}{(1-e_L p_L)\lambda + (1-e_H p_H)(1-\lambda)} \end{aligned} \quad (4)$$

Note that if pooling equilibrium $e_i^* = 0$ occurs, out-of-equilibrium belief b_e would not be constrained by Bayes' rule in (4) because b_e would become indeterminate.[22] As we shall see, neither the dominance criterion nor Cho and Kreps (1987) "intuitive criterion" can be applied here to refine belief b_e because both SRO types face exactly the same exposure incentives.[23] Therefore, if $e_i^* = 0$, out-of-equilibrium belief b_e is allowed to take any value in interval [0,1].[24]

The assumption of Bayesian updating by consumers deserves a discussion. Although Bayesian updating demands a somewhat sophisticated reasoning from consumers, unfortunately there are not yet consensual alternatives to it in the literature on games of incomplete information. However, in the context of this model Bayesian updating has some interesting features. As we shall see, the SROs incentives to expose or cover-up fraud depend on the *sign*, not the magnitude, of the reputation change associated with exposure (term R introduced in the next section). Therefore, if consumers employed a different updating rule than (4) that yielded similar signs of reputation change from exposure, SRO exposure incentives would be similar. However, having consumers depart from the Bayes' rule in (4) could in some cases alter the sign of the reputation change

22 Note that belief b_n, on the other hand, can be directly computed using Bayes' rule in (4) for any $e_i \in [0, 1]$.

23 See Lemma 1.

24 However, later on we shall restrict belief b_e to rule-out some "implausible" equilibria.

associated with exposure. In order to address this possibility, we restrict our attention to "robust" equilibria, that is, those Perfect Bayesian Equilibria that are stable to small perturbations in the value of the reputation change (term R).

It seems also implausible that all consumers share the same prior belief λ. However, as shown later, the sign of the reputation change does not depend on the value of λ, but on the functional properties of the fraud detection probability $p(x, v)$, in particular on the sign of the derivative $\frac{\mathrm{d}p(x^*(v),v)}{\mathrm{d}v}$. Hence, having consumers with different values for λ would yield a similar sign of the reputation change associated with exposure. Note also that consumers do not need to know the *exact* functional form of function $p(x, v)$. Instead, it is required that consumers somehow share a belief on whether the probability of fraud detection is increasing or decreasing in SRO vigilance v, once the effects of vigilance on fraud have been taken into account.

2.3 The Self-Regulatory Organization

As mentioned earlier, the utility of the SRO is determined by its reputation in front of consumers, that is, by the economic value that the SRO obtains from consumers according to their perceptions about the SRO's type. However, the SRO also benefits from potential bribes from fraudulent members in exchange for a cover-up. Suppose that the SRO does not expect a collusive agreement upon discovering fraud. Then the conditional expected utility of SRO type i amounts to,

$$e_i p(x, v_i)[b_e w_L + (1 - b_e)w_H] + [1 - e_i p(x, v_i)][b_n w_L + (1 - b_n)w_H] - c_i v_i \tag{5}$$

which simplifies to,

$$[e_i p(x, v_i)b_e + (1 - e_i p(x, v_i))b_n]W + w_H - c_i v_i \tag{6}$$

On the other hand, if the SRO expects a collusive agreement with the SRO member after discovering fraud, its conditional expected utility, net of vigilance costs, amounts to the expected value of the bribe minus the transaction cost, $p(x, v_i)(\beta - \tau)$, plus the expected value to the SRO associated with the beliefs about the SRO's type that consumers hold conditional on non-exposure, $p(x, v_i)(b_n w_L + (1 - b_n)w_H)$. Therefore, expecting a collusive contract with the member with probability γ if fraud is detected, the risk-neutral SRO types choose e_i and v_i to maximize;

$$\begin{aligned}&\gamma[p(x, v_i)(\beta - \tau) + p(x, v_i)(b_n w_L + (1 - b_n)w_H)]\\&\quad + [1 - \gamma][(e_i p(x, v_i)b_e + (1 - e_i p(x, v_i))b_n)W + w_H] - c_i v_i\end{aligned} \tag{7}$$

or equivalently,

$$(b_n w_L + (1 - b_n) w_H) + p(x, v_i)[(1 - \gamma) e_i R + \gamma(\beta - \tau)] - c_i v_i \tag{8}$$

where $R = (b_e - b_n)W \in [-W, W]$. Term R denotes the value to the SRO of the "reputation change" associated with fraud exposure ($e = 1$); exposing fraud when $R > 0$ yields a reputation *gain* to the SRO, while exposing fraud when $R < 0$ leads to a reputation *loss*.

Term R is determined by two expressions. Expression $(b_e - b_n)$ corresponds to the difference in the posterior Bayesian probability that consumers hold about the SRO being the low-cost (or vigilant) type, given exposure versus non-exposure. This term is multiplied by W, namely, the differential value to the SRO of being perceived by consumers as a low-cost versus a high-cost SRO type. Note that if $W > 0$, the sign of R depends only on how Bayesian consumers update beliefs b_e and b_n in (4) after observing either fraud exposure or non-exposure. The sign of R is a significant feature of this model because it determines the SRO's exposure incentives, as shown next.

Tie-breaking Assumption: Exposure entails an exogenous cost arbitrarily close to zero.

Lemma 1 *Assume there is not a collusive contract between the SRO and the SRO member. Then if* $R > 0$, $e_H^* = e_L^* = 1$. *Otherwise,* $e_H^* = e_L^* = 0$.

Lemma 1 follows from the SRO's objective function in (8); if the SRO discovered fraud but a collusive contract with the member did not emerge, the exposure decision is fully determined by the sign of the reputation change R associated with exposure.[25] Lemma 1 also indicates that there will always be pooling behavior in the exposure decision $e_i \in [0, 1]$ because the value of reputation change R is identical to both SRO types. Lemma 1 shows also that, although SRO types could employ mixed strategies on e_i, only pure strategies $e_i = 0$ or $e_i = 1$ will be employed in equilibrium, depending on the sign of R.

In order to study SRO incentives in a wider context, we analyze both the *simultaneous* and *sequential* choices of SRO vigilance and member's fraud levels. The relevance of each alternative ultimately depends on the SRO's ability to commit credibly to a fixed level of vigilance, an issue that we do not address here. From (8) it follows that the first-order conditions of SRO type i when maximizing with respect to v_i in the simultaneous and sequential cases are, respectively;

$$\frac{\partial p(x, v_i)}{\partial v_i}[(1 - \gamma) e_i R + \gamma(\beta - \tau)] - c_i = 0 \tag{9}$$

$$\frac{\mathrm{d}p(x^*(v_i), v_i)}{\mathrm{d}v_i}[(1 - \gamma) e_i R + \gamma(\beta - \tau)] - c_i = 0 \tag{10}$$

[25] Note that the tie-breaking assumption requires R to be strictly positive to make exposure profitable.

The first-order conditions in (9) and (10) have an important difference. While in the simultaneous case in (9) the SRO takes the member's fraud choice as parametrical, in the sequential case the SRO internalizes the effect of its vigilance level on the member's optimal fraud level $x^*(v)$, as Eq. 10 indicates. Unless stated otherwise, the propositions that follow apply both to the sequential and simultaneous versions of the model. We finally establish the following fact:

Lemma 2 *Assume* $v_H^* < v_L^*$. *Then* $R^* > 0$ *iff* $\frac{dp}{dv} > 0$.

Proof : See the Appendix.

Lemma 2 indicates that if the low-cost SRO type exerts more vigilance than the high-cost type, then exposure would yield a reputation gain to the SRO, provided that the *overall* effect of vigilance on the probability of fraud discovery is positive. The intuition is that if consumers expect that finding fraud is more likely to happen for the low-cost SRO type, then observing exposure should lead them to assign a higher probability to the low-cost SRO. Table 1 summarizes the key notation of the model described in this section.

3 Equilibrium

Since the model is a dynamic game of incomplete information the equilibrium concept employed is that of a Perfect Bayesian Equilibrium. For the two possible cases—the SRO and the member moving simultaneously or in sequential order—a Perfect Bayesian Equilibrium consists of strategies x^*, v_i^*, e_i^*, an equilibrium probability γ^* of reaching a collusive agreement, a bribe $\beta^* \geq 0$ and consumer beliefs b_e and b_n that satisfy Bayes' Rule in (4). Given the known interdependence between strategies and Bayesian beliefs typical of Perfect

Table 1 Summary of key notation

$i = \{H, L\}$	SRO type, either high-vigilance-cost (H) or low-vigilance-cost (L)
λ	Prior probability that SRO is low-vigilance-cost type
c_i	Unit vigilance cost of SRO of type i, with $c_L < c_H$
v_i	Vigilance level chosen by SRO of type i
x	Fraud level chosen by SRO member
$p(x, v_i) = p_i$	Probability that SRO type i detects fraud, given fraud x and vigilance v_i
$e_i \in [0, 1]$	Exposure decision of SRO type i (exposure = 1, cover-up = 0)
T	Cost to SRO member of internal punishment
L	Cost to SRO member of external punishment
β	Bribe paid to the SRO by the SRO member in exchange for a cover-up
τ	Transaction cost of a collusive contract between SRO and SRO member
γ	Expected probability of reaching a collusive contract upon fraud detection
b_e	Consumers' posterior belief that SRO is low-cost after exposure
b_n	Consumers' posterior belief that SRO is low-cost after non-exposure
w_i	Value to SRO of being perceived by consumers as SRO of type i
$W = (w_L - w_H)$	Differential value to SRO of being perceived as low versus high-cost type
$R = (b_e - b_n)W$	Value to SRO of Reputation gain or loss associated with exposure

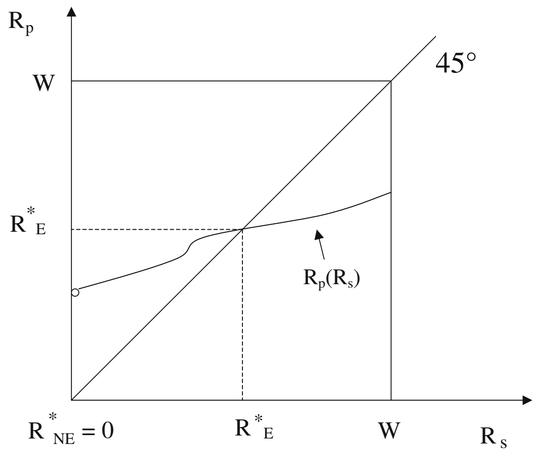

Fig. 1 If $\frac{dp}{dv} > 0$, $R^*_{NE} = 0$ and R^*_E are Perfect Bayesian Equilibria. However, only Exposure Equilibrium R^*_E is stable

Bayesian Equilibria, we study equilibrium by proving the existence of a fixed point for the SRO's reputation change $R^* \in [-W, W]$. To do so, let R_s denote a "starting" (or "initial") level of reputation change, and let R_p denote a "posterior" level of reputation change after Bayesian updating has taken place. A situation $R^* = R_s = R_p$ denotes a fixed point and a Perfect Bayesian Equilibrium of the model. Term R_p is computed from consumers' updated beliefs about the SRO's type conditional upon both SRO types' equilibrium exposure decisions e_i^* and levels of vigilance v_i^*, and the member's optimal amount of fraud x^* for a given level of R_s. Figure 1 shows the features of function $R_p(R_s)$ when $\frac{dp}{dv} > 0$, which yields equilibria $R_E > 0$ and $R_{NE} = 0$. The propositions that follow are derived for both the simultaneous and sequential cases. The proofs are provided in the Appendix separately for each of the two cases.

3.1 Corruption-free benchmark

We first characterize equilibrium when collusive contracts between the member and the SRO are not possible or enforceable. This case amounts to finding the equilibria of the model when either $\gamma = 0$ or $\beta = 0$. Let $x^{e=0}_{\max} = x^*(v = 0, e = 0)$ and $x^{e=1}_{\max} = x^*(v = 0, e = 1)$ denote the optimal levels of fraud when the SRO exerts no vigilance at all, and the member expects cover-up and exposure, respectively. The member's reaction function implicit in (2) guarantees that $x^{e=0}_{\max} > x^{e=1}_{\max}$.[26]

Proposition 1 *If collusive contracts between the SRO and the SRO member are not possible;*

[26] In both cases the optimal amount of fraud is strictly finite because fraud increases the probability of fraud detection due to assumptions $p_x > 0$ and $p_{xx} > 0$.

(i) *Equilibrium exists. Pooling equilibrium with no vigilance, fraud cover-up and impunity ($R^* = v_i^* = e_i^* = 0$) and maximum fraud ($x^* = x_{\max}^{e=0}$) exists for all $W \geq 0$ and any sign of $\frac{dp}{dv}$.*

(ii) *The pooling equilibrium identified in part 1(i) is the unique equilibrium if $\frac{dp}{dv} \leq 0$ for all $W \geq 0$. Moreover, this equilibrium is stable to small perturbations in R if $\frac{dp}{dv} \leq 0$.*

(iii) *If $\frac{dp}{dv} > 0$ there is also a separating equilibrium with $R^* > 0$, voluntary fraud exposure $e_L^* = e_H^* = 1$, positive vigilance $v_L^* > v_H^* > 0$ and fraud deterrence $x^*(v_L^*) < x^*(v_H^*) < x_{\max}^{e=1}$ for any $W > 0$.*

(iv) *If $\frac{dp}{dv} > 0$, the equilibrium identified in part 1(iii) is the only equilibrium that is stable to small perturbations of reputation change R.*

Proof : See the Appendix.

Proposition 1 is essentially the same as in Nunez (2001). Part 1(i) shows that if there is no reputation gain from exposure, then cover-up and no vigilance are optimal pooling strategies. This impedes the separation of types that is necessary to create a reputation gain from exposure. The intuition of part 1(ii) is that if consumers think higher vigilance decreases the chances of fraud detection (i.e. $\frac{\mathrm{d}p}{\mathrm{d}v} < 0$), then Bayesian consumers must place a *lower* probability on the low-cost SRO after observing exposure. Anticipating this, neither SRO type would be willing to choose positive vigilance nor expose fraud. Part 1(ii) also states that if $\frac{\mathrm{d}p}{\mathrm{d}v} \leq 0$, this equilibrium is unique and stable to small perturbations in reputation change R. In what follows, we shall refer to this equilibrium as the *Non-Exposure Equilibrium*, in which exposure yields a null reputation change $R_{NE}^* = 0$.

Part 1(iii) shows that self-regulation *may* yield positive vigilance, fraud deterrence and fraud exposure only if $\frac{\mathrm{d}p}{\mathrm{d}v} > 0$. Intuitively, if the probability of fraud detection increases with vigilance, after observing exposure consumers should think that the low-cost SRO is the more likely type to have detected and exposed fraud. Accordingly, SROs would have incentives to exposure fraud, as exposure would yield a reputation gain. We name this outcome the *Exposure Equilibrium*, in which exposure yields a reputation gain $R_E^* > 0$.

Note from part 1(i) that condition $\frac{\mathrm{d}p}{\mathrm{d}v} > 0$ is not a sufficient condition for self-regulation to attain positive vigilance, fraud exposure and fraud deterrence, because the equilibrium identified in parts 1(i) and 1(iii) coexist. However, part 1(iv) provides an equilibrium refinement to solve this multiple-equilibrium situation: only the Exposure Equilibrium $R_E^* > 0$ identified in 1(iii) is stable to small perturbations in R. Accordingly, for the case $\frac{\mathrm{d}p}{\mathrm{d}v} > 0$ we shall restrict out attention only to this equilibrium. Figure 1 portrays the two equilibria, namely the Non-Exposure Equilibrium $R_{NE}^* = 0$ and the stable Exposure Equilibrium $R_E^* > 0$ for the case $\frac{\mathrm{d}p}{\mathrm{d}v} > 0$. Note the discontinuity of function $R_p(R_s)$ near $R = 0$, which turns equilibrium $R_{NE}^* = 0$ unstable to small perturbations in R.

3.2 The possibility of corruption

The central message from Proposition 1 is that self-regulation provides ambiguous reputation-based incentives towards vigilance and exposure. In this section, we examine the consequences of possible collusive agreements between the SRO and the member. In particular, this section investigates the following issues: whether the equilibria identified in the corruption-free benchmark survive the possibility of corruption, whether new equilibria emerge, and the performance of self-regulation in comparison to the corruption-free benchmark.

A Pareto-improving contract between the SRO and the member involving a bribe in exchange for a cover-up $\{\beta > 0, e = 0\}$ will emerge only if $\beta - \tau \in (R, L - \tau)$, where R and L are the SRO's and the member's reservation (no corruption) payoffs, respectively. In other words, bribe β must be higher than the forgone value of the SRO's reputation change R plus the transaction cost τ, but lower than cost L that the member would face if exposed. In order to study equilibrium it is necessary to establish the consequences that follow from actions taken out of the equilibrium path of play. Recall that in an equilibrium involving cover-up ($e_i^* = 0$), out-of-equilibrium belief b_e in (4) is not constrained by Bayes' rule, or by Cho and Kreps' (1987) "intuitive" criterion. In order to rule out implausible equilibria we make the following assumption about belief b_e.[27]

Assumption 2 If $e_i^* = 0$, out-of-equilibrium belief b_e must be derived from Bayes' rule employing any e_L and $e_H \in (0, 1]$ when $\frac{\mathrm{d}p}{\mathrm{d}v} > 0$. However, if $\frac{\mathrm{d}p}{\mathrm{d}v} \leq 0$, belief b_e must be derived from Bayes' rule employing values for e_L and e_H that satisfy $e_L p_L \leq e_H p_H$, where $p_L \leq p_H$.

Assumption 2 does not constrain belief b_e when $\frac{\mathrm{d}p}{\mathrm{d}v} > 0$ because this case does not lead to "implausible" or "unreasonable" equilibria. However, when $\frac{\mathrm{d}p}{\mathrm{d}v} \leq 0$ Assumption 2 establishes that e_L cannot be greater than e_H by an amount large enough as to outweigh the condition $p_L \leq p_H$ that must hold in any equilibrium. Note that Assumption 2 is satisfied if, for example, after observing exposure unexpectedly, consumers believed that the low-cost SRO is at least as likely as the high-cost SRO to have exposed fraud ($e_L \leq e_H$), or if they decided instead to stick to the prior belief such that $b_e = \lambda$.[28]

Proposition 2 *If corruption is possible;*

(i) *Equilibrium exists. The pooling equilibrium $R_{NE}^* = 0$ identified in Proposition 1(i) involving no bribery ($\beta^* = 0$), no vigilance, fraud cover-up,*

[27] The "implausible" equilibria that Assumption 2 rules out is the following: despite that more vigilance decreases the probability of fraud detection ($\frac{\mathrm{d}p}{\mathrm{d}v} < 0$), consumers believe without any Bayesian support that exposure was done by the low-cost SRO type. Exposure would then lead to a reputation gain to the SRO, which would give the member an incentive to offer a bribe in exchange for a cover-up.

[28] This requires that $e_L p_L = e_H p_H$, which implies that $e_L \geq e_H$.

impunity ($v_i^* = e_i^* = 0$) *and maximum fraud* ($x^* = x_{\max}^{e=0}$) *exists for all* $W, \tau, L \geq 0$, *and any sign of* $\frac{dp}{dv}$.

(ii) *Equilibrium* $R_{NE}^* = 0$ *identified in 2(i) is unique and stable to small perturbations in R if* $\frac{dp}{dv} \leq 0$.

Proof : See the Appendix.

The intuition of Proposition 2(i) is the same as that provided earlier for the equilibrium identified in Proposition 1(i): if the SRO does not expect consumers to differentiate types after exposure, both SRO types have incentives to conceal fraud and choose zero vigilance, which reinforces consumers inability to differentiate SRO types. As a consequence, the member optimally chooses maximum fraud and offers no bribe to the SRO. Note that Proposition 2(i) holds regardless of the sign of $\frac{\mathrm{d}p}{\mathrm{d}v}$.

Proposition 2 (ii) shows that if the SRO suffers a loss in reputation from exposing fraud (due to $\frac{\mathrm{d}p}{\mathrm{d}v} \leq 0$), exposure would not be perceived by the member as a credible threat. Hence, the member would not need to bribe the SRO to ensure fraud cover-up and avoid punishment cost L. This implies that if $\frac{\mathrm{d}p}{\mathrm{d}v} \leq 0$ no other equilibria exists apart from the Non-Exposure Equilibrium R_{NE}^* presented in Propositions 1(i) and 2(i). In conclusion, allowing the SRO and the SRO member to collude when the probability of fraud detection is non-increasing in vigilance ($\frac{\mathrm{d}p}{\mathrm{d}v} \leq 0$) adds nothing new to the outcome of self-regulation presented earlier, namely, the Non-exposure Equilibrium $R_{NE}^* = 0$.

However, new incentives and new equilibria emerge when $\frac{\mathrm{d}p}{\mathrm{d}v} > 0$. Recall that in this case fraud exposure yields a reputation gain to the SRO. This makes the threat of fraud exposure by the SRO credible to the member. As a result, the member would have an incentive to offer a bribe $\beta > R + \tau$ in exchange for a cover-up in order to avoid the external punishment L, as the next proposition shows.

Proposition 3 *If collusive contracts between the SRO and the member are possible and if* $\frac{dp}{dv} > 0$;

(i) *If* $L > \tau$, *there exist Collusive Equilibria* $R_C^* > 0$, *which involves bribery* $\beta^* > 0$ *with* $\beta^* - \tau \in (R_C^* > 0, L - \tau)$, *fraud cover-up and impunity* ($e_i^* = 0$). *In these Collusive Equilibria there is positive vigilance* ($v_L^* > v_H^* > 0$) *and some fraud deterrence* ($x^*(v_L^*) < x^*(v_H^*) < x_{\max}^{e=0}$).

(ii) *The Exposure Equilibrium* $R_E^* > 0$ *identified in Proposition 1(iii), which involves voluntary fraud exposure* ($e_L^* = e_H^* = 1$), *positive vigilance* ($v_L^* > v_H^* > 0$) *and fraud deterrence* ($x^*(v_L^*) < x^*(v_H^*) < x_{\max}^{e=1}$) *for any* $W > 0$ *exists if and only if* $L - \tau \leq R_E^*$.

Proof : See the Appendix.

Part 3(i) indicates that the possibility of collusion between the SRO and the member creates collusive outcomes involving bribery, fraud cover-up and impunity. A sufficient condition for such equilibria to exist is that the external punishment L be larger than transaction cost τ, such that Pareto-improving contracts between the SRO and the member would emerge. Part 3(ii) presents a significant finding; the Exposure Equilibrium presented in Proposition 1(iii) (involving exposure, positive vigilance and fraud deterrence) continues to exist only if the external punishment net of the transaction cost is not greater than the reputation gain, that is $L-\tau \leq R^*_E$. Otherwise, only Collusive Equilibria will prevail. Note, however, that Collusive Equilibria exist even if condition $L-\tau \leq R^*_E$ is satisfied, because, as part 3(i) indicates, Collusive Equilibria only requires $L > \tau$.

The different aspects of Proposition 3 are illustrated in Figs. 2 and 3, where the set of Collusive Equilibria correspond to the thick line along the 45-degree diagonal between the origin and $R = L-\tau$. The bargaining core of the collusive contract for a given level of reputation gain R corresponds to the horizontal distance between R (on the thick segment of diagonal) and $L-\tau$. In Fig. 2 Exposure Equilibrium $R^*_E > 0$ co-exists with Collusive Equilibria $R^*_C > 0$ because

Fig. 2 If $\frac{dp}{dv} > 0$ and $L - \tau < R^*_E$, Collusive Equilibria $R^*_C > 0$ (thick segment of diagonal) and Exposure Equilibrium R^*_E coexist

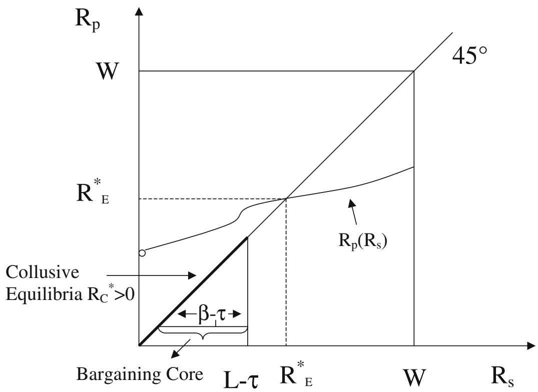

Fig. 3 If $\frac{dp}{dv} > 0$ and $L - \tau < R^*_E$, only Collusive Equilibria $R^*_C > 0$ exist (thick segment of diagonal)

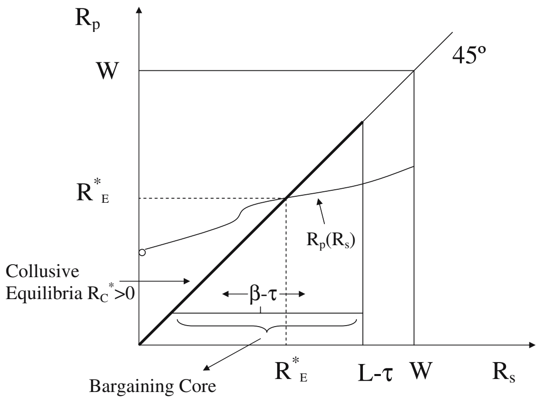

Table 2 Summary of stable perfect Bayesian equilibria in Propositions 1–3

	Fraud detection probability non-increasing in Vigilance $\frac{dp(x^*(v),v)}{dv} \leq 0$	Fraud detection probability increasing in Vigilance $\frac{dp(x^*(v),v)}{dv} > 0$	
Collusion-free Benchmark	Non-exposure equilibrium: Cover-up, impunity No Vigilance Maximum fraud	Exposure equilibrium: Exposure, no impunity Positive Vigilance Fraud deterrence	
		$L - \tau \leq R_E^*$	$L - \tau > R_E^*$
Collusion Allowed	Non-exposure equilibrium: Cover-up, impunity No bribery No vigilance Maximum fraud	Exposure equilibrium: Exposure, no impunity No bribery Positive vigilance fraud deterrence Collusive equilibria: Cover-up, impunity Bribery Positive vigilance Fraud deterrence	Collusive equilibria: Cover-up, impunity Bribery Positive vigilance Fraud deterrence

the reputation gain of exposure $R_E^* > 0$ is greater than the member's maximum willingness to pay for a bribe (net of transaction costs), $L - \tau$. Accordingly, no Pareto-improving contract is possible to make the SRO give up the reputation gain R_E^*. In Fig. 3, however, the member's maximum willingness to pay for a bribe net of the transaction cost, $L - \tau$ is greater than the reputation gain under the Exposure Equilibrium R_E^*. Therefore, the SRO and the member will agree on some bribe β in exchange for a cover-up. As a consequence, in this case only Collusive Equilibria exist.

Table 2 summarizes all the stable equilibria identified in Propositions 1–3. Table 2 provides an overview of the effects that collusion has on exposure incentives and on the levels of vigilance and fraud, as well as the implications of having a fraud detection probability that is increasing versus decreasing in vigilance.

We now investigate the levels of vigilance and fraud under each equilibrium. The next result establishes the effect of corruption on the equilibrium level of vigilance and fraud in comparison to the Exposure Equilibrium.

Proposition 4 *Recall $R_E^* > 0$ denotes the Exposure Equilibrium identified in Propositions 1(iii) and 3(ii). Assume $\frac{dp}{dv} > 0$,*

(i) *For all Collusive Equilibria $R_C^* \leq R_E^*$ equilibrium vigilance is lower and equilibrium fraud is higher than in the Exposure Equilibrium R_E^*.*
(ii) *Recall that if $L - \tau > R_E^*$, Collusive Equilibria $R_C^* > R_E^*$ exist. For all such equilibria, vigilance is higher than in the Exposure Equilibrium R_E^*. However, fraud can be higher or lower than in the Exposure Equilibrium R_E^*.*

(iii) *The level of fraud in a Collusive Equilibrium R_C^* is strictly lower than in the Exposure Equilibrium R_E^* if and only if (a) $L - \tau > R_E^*$ and (b) bribe β^* is sufficiently close to the external punishment cost L.*

Proof : See the Appendix.

Part 4(i) establishes that vigilance is lower and fraud is higher under all Collusive Equilibria where the potential reputation gain is equal or smaller than the reputation gain in the Exposure Equilibrium. The intuition is that the SRO's marginal vigilance incentives are lower in the Collusive Equilibrium because the bribe net of the transaction cost is smaller than the reputation gain in the Exposure Equilibrium. On the other hand, fraud is higher in the Collusive Equilibrium not only due to lower vigilance, but also because the bribe is less costly for the member than facing the external penalty. This increases the member's marginal fraud incentives for a given level of vigilance.

Part 4(ii) shows that if the external punishment minus the transaction cost is sufficiently high, there are Collusive Equilibria where the potential reputation gain is higher than in the Exposure Equilibrium. In these equilibria vigilance is higher than in the Exposure Equilibrium because the bribe net of the transaction cost must exceed the potential reputation gain, which gives the SRO an enhanced incentive to exert more vigilance in order to detect fraud and appropriate the bribe.[29]

However, the higher level of vigilance does not necessarily translate into a lower level of fraud, the reason being that the member's marginal incentives must also be considered. Proposition 4(iii) establishes that fraud would be lower under collusion only if the bribe is sufficiently close to the external punishment. This suggests that the level of fraud ultimately depends on the relative bargaining power of the SRO and the member; if the SRO has more relative bargaining power such that β is close to L, the SRO has enhanced incentives to increase vigilance but the member's marginal fraud incentives for a given level of vigilance remain nearly the same. As a consequence, fraud is reduced. However, if the member has more relative bargaining power such that $\beta - \tau$ approaches R, the SRO's marginal incentives remain nearly constant, while the member faces more incentives to engage in fraudulent behavior because the bribe is now less costly than penalty L. As a result, equilibrium fraud is strictly higher than under the Exposure Equilibrium. While the relevance of each result is an empirical matter, these results do offer a variety of plausible cases where corruption increases the equilibrium level of fraud in comparison to an Exposure Equilibrium.

Despite this, the next proposition highlights a feature that is common to all the equilibria than emerge when collusion is possible and if $\frac{\mathrm{d}p}{\mathrm{d}v} > 0$, which is illustrated also in Table 2.

---

[29] We name this "the corrupt-policeman effect" after the well-known stories in many countries where traffic policemen often undertake their duties with diligence and rigor in order to extract resources from offenders.

Proposition 5 *If* $\frac{dp}{dv} > 0$*, a corruptible SRO always exerts positive vigilance and deters fraud.*

Proof : See the Appendix.

Proposition 5 states that as long as the probability of fraud detection is increasing in vigilance, self-regulation yields positive vigilance and fraud deterrence, regardless of whether an Exposure Equilibrium or a Collusive Equilibrium gets established. Under an Exposure Equilibrium, a corruptible SRO would not be bribed, but it would still yield positive vigilance and fraud deterrence. Under a Collusive Equilibrium, and due to the "corrupt-policeman effect" a corruptible SRO will invest in vigilance in order to detect fraudulent members and extract rents from them, which creates some fraud deterrence. Hence, Proposition 5 indicates that the possibility of corruption between an SRO and its members is not a reason to reject self-regulation and prefer a no-regulation situation if the social goal is to reduce fraud and malpractice.

3.3 A comment on welfare

Although the focus of this article is on the incentives that SROs face to deter and expose fraud, it might be tempting to claim that fraud deterrence is welfare-increasing. Yet, deterring fraud is also costly, and therefore the social benefits and costs of deterrence should be traded off. This section examines the welfare consequences of collusion, impunity and cover-up in self-regulation. We first examine how collusion and cover-up erodes the efficiency of vigilance in deterring fraud. Then, we study the welfare implications of the different possible outcomes of self-regulation by assuming an explicit social welfare function that depends on both fraud and vigilance.

As the member's first-order condition in (2) shows, the member can face three different scenarios if detected: exposure, a collusive cover-up and a non-collusive cover-up. Under exposure, the cost to the agent would be the internal plus the external penalties, $T + L$. Under collusion, the cost to the member of being detected would be the internal penalty plus the bribe, $T + \beta$, and under a non-collusive cover-up the cost is just T. Given that under collusion it must be that $\beta < L$, we have that $T + L > T + \beta > T$. From the member's reaction function implicit in (2) it follows that for any level of vigilance $v > 0$,

$$x^*(v, T + L) < x^*(v, T + \beta) < x^*(v, T) \tag{11}$$

Expression (11) implies that for any given level of vigilance the member's optimal level of fraud is lower under exposure than under collusion, the latter in turn lower than under a non-collusive cover-up. The intuition is simple: bribery and cover-up increase the members' marginal incentives to commit fraud. Hence, a given amount of vigilance achieves less deterrence in comparison to an exposure situation. This erodes the efficiency of vigilance in deterring fraud. In fact, denote by $x^*_{\max} - x^*(v)$ the amount of fraud deterrence associated with vigilance level v with respect to the maximum possible level of fraud $x^*_{\max}$, that is, when there is cover-up and no vigilance. If the total cost of vigilance is $c_i v$, then the unit cost of deterring $x^*_{\max} - x^*(v)$ units of fraud is $\frac{c_i v}{x^*_{\max} - x^*(v)}$. Then, it follows that,

$$\frac{c_i v}{x^*_{\max} - x^*(v, T+L)} < \frac{c_i v}{x^*_{\max} - x^*(v, T+\beta)} < \frac{c_i v}{x^*_{\max} - x^*(v, T)} \qquad (12)$$

Expression (12) indicates that for any given level of vigilance the unit cost of fraud deterrence is lower under exposure than under corruption, and it is the highest under non-collusive cover-up.

In order to examine the welfare consequences of collusion and cover-up in a more general setting, let $S(x)$ denote the social welfare associated with x units of fraud, with $S_x < 0$. Evidently, $S(x)$ excludes the benefits of fraud to the member since the preferences for illegal or antisocial behavior should not take part in social welfare evaluations.[30] Let D denote the cost that the member faces if detected, that is, $D = \{T+L,\ T+\beta,\ T\}$. Then, a social planner who knew the SRO's type i would choose x, v and D to maximize,

$$S(x) - c_i v \qquad (13)$$

subject to the member's reaction function $x^*(v, D)$. The first-order condition of this problem with respect to x and v corresponds to $S_x \frac{\partial x^*(v,D)}{\partial v} \geq c_i$, or $S_x \frac{\partial x^*(v,D)}{\partial v} = c_i$ for an interior solution. The left-hand side is positive, as $\frac{\partial x^*(v,D)}{\partial v}$ is the slope of the member's reaction function, and term $\frac{c_i}{S_x}$ corresponds to the slope of the social welfare indifference curve implicit in (13). Note that choosing D is equivalent to choosing the member's reaction function on which the first-order condition above is located. Expression (11) indicates that the most welfare-enhancing value for D (or member's reaction function) is $D = T + L$ (under exposure). This implies that exposure is a necessary condition for a social optimum. Hence, the Collusive Equilibria and the Non-Exposure Equilibrium described earlier are necessarily socially suboptimal. Note, on the other hand, that the Exposure Equilibrium $R^*_E > 0$ in Propositions 1(iii) and 3(ii) is located on the most welfare-enhancing member's reaction function (and so is the social optimum), namely $x^*(v,\ T+L)$. However, how close is the Exposure Equilibria from delivering the socially optimum levels of vigilance and fraud deterrence will depend on how aligned are the SRO's reputation-based incentives and the social welfare function $S(x)$, an issue that we do not explore here. However, the whole rationale for self-regulation rests on the notion that an SRO's concern for reputation should be a sufficiently strong incentive to invest in vigilance an deter fraud to presumably near-optimal levels.

Finally, it must be observed that collusion, cover-up and impunity yield other social costs than those included above. First, establishing collusive agreements

30 See Harsanyi (1982).

demands time and resources from both parties (captured in transaction cost τ). These costs are socially wasteful as their only purpose is to achieve a contract between the SRO and the member. Second, impunity can also be regarded a social cost in itself, as there is agreement that the punishment of illegal or inappropriate behavior is socially desirable, beyond its disciplining effect on potential offenders. Similarly, corruption can be considered also a social cost regardless of its effects on fraud and vigilance levels, as it erodes norms and institutions that are socially valuable.

4 Conclusions

This article has shown a variety of forces that undermine the performance of self-regulation in preventing and disclosing malpractice to the public. This suggests that possibly a significant amount of fraud may in practice go undetected or may be concealed in self-regulated activities. A first reason for concern is that SROs have ambiguous reputation-based incentives to detect and expose fraud committed by their members. This ambiguity originates from consumers' unclear interpretation of exposure; exposure would signal a vigilant attitude if higher vigilance makes it more likely to detect fraud, but it would signal lax vigilance if consumers saw it as indicating the existence of widespread fraud caused by low vigilance.

A second concern is that in the case where exposure is a credible threat because it yields a reputation gain to the SRO, the member has an incentive to bribe the SRO in exchange for a cover-up. This outcome is socially inefficient because it sharpens the members' incentives to engage in fraudulent behavior and reduces the effectiveness of vigilance in comparison to a full exposure situation. Yet, the possibility of corruption is not a reason to reject self-regulation altogether. In fact, a collusive equilibrium is preferred to no regulation at all if the social goal is to reduce fraud because a corrupt SRO would exert some vigilance in order to extract bribes from fraudulent members, which causes fraud deterrence. These results raise a policy dilemma about how fraud and malpractice should be publicly sanctioned: while a high external penalty would discourage fraudulent behavior, it would also increase the incentives to establish collusive agreements within the SRO, which can cause cover-up, impunity and often higher levels of fraud.

The forces at work unveiled in this article can apply also to other less-traditional forms of self-regulation, such as the provision of quality within firms, or the incentives of public regulatory agencies. Both situations can involve a principal concerned by his reputation for being vigilant in charge of monitoring the behavior of potentially fraudulent or negligent agents. In addition, this framework can be extended in several ways, for example, by allowing public regulators in parallel to self-regulation, by introducing other sources of quality information such as consumers' search or experience, the media and "watchdog" groups, and by allowing consumers to have uncertainty about the type of the SRO members.[31] Pursuing these and other possible extensions could add new insights into the incentives and the performance that can be expected from self-regulation.

Acknowledgements I am indebted to John Vickers, Meg Meyer, Tim Besley, Roger Guesnerie, Felipe Zurita, Bernardita Escobar, Juan Dubra, Rodrigo Harrison and Alex Galetovic and two anonymous referees for their valuable comments. All errors are the sole responsibility of the author. This work was supported by FONDECYT—Chile, Project 1010608.

Appendix

Proof of Lemma 2 $R^* > 0$ requires that $b_e > b_n$. From (4) and Lemma 1 it follows that $b_e > b_n$ requires $p_H < p_L$. If $v^*_H < v^*_L$, $p_H < p_L$ holds only if $\frac{\mathrm{d}p}{\mathrm{d}v} > 0$.

Proof of Proposition 1(i) If $R_s = 0$, then in the simultaneous and sequential cases $v^*_i = 0$, $e^*_i = 0$ and $x^* = x_{\max}$. This yields $b_e = b_n = \lambda$ and $R_p = 0$, making $R^* = 0$ a fixed point.

Proof of Proposition 1(ii) Lemma 1 implies that an equilibrium different from $e^*_i = 0$, $v^*_i = 0$ and $x^* = x_{\max}$ requires a fixed point $R > 0$. However, from Lemma 2 for all $R_s > 0$, condition $\frac{\mathrm{d}p}{\mathrm{d}v} \leq 0$ implies that $R_p \leq 0$. Therefore, a fixed point $R^* > 0$ cannot exist. Hence, the equilibrium in Proposition 1(i) is unique if $\frac{\mathrm{d}p}{\mathrm{d}v} \leq 0$. Regarding its stability, if $R_s < 0$, then $e_i = 0$, $v_i = 0$ and $R_p = 0$. If $R_s > 0$, then $e_i = 1$ and $v^*_H < v^*_L$, but $R_p \leq 0$ because $\frac{\mathrm{d}p}{\mathrm{d}v} \leq 0$. By repetition, equilibrium $R^* = 0$ would be reached.

Proof of Proposition 1(iii) *Simultaneous case:* For any $R_s > 0$, $v_L > v_H > 0$ and $p_L > p_H$, which follows from $p_v R \to \infty$ as $v_i \to 0$. This implies that $\lim\limits_{R_s \to 0} p_L/p_H > 1$, $\lim\limits_{R_s \to 0} (1 - p_L)/(1 - p_H) = 1$ and therefore $\lim\limits_{R_s \to 0} R_p > 0$. On the other hand, when $R_s \to W$ then $v_L > v_H > 0$, and therefore $b_e < 1$ and $b_n > 0$, implying that $R_p(R_s = W) < W$. Therefore, given that function $R_p(R_s)$ is continuous, at least one fixed point $R^* > 0$ exists. Any fixed point $R^* > 0$ yields $v^*_i > 0$, $e^*_i = 1$ and $x^*(v^*_i) < x_{\max}$.

Sequential Case: For $R_s > 0$ close to 0, $e^*_i = 1$ but $v^*_i = 0$ because $c_i > 0$. This yields $R_p = 0$ as $R_s \to 0$. As $R_s > 0$ increases, there is a point from which $v^*_L > v^*_H \geq 0$, and therefore $R_p > 0$. A fixed point $R^* > 0$ exists if c_i is small enough relative to $\frac{\mathrm{d}p}{\mathrm{d}v}$.

Proof of Proposition 1(iv) Consider first equilibrium $R = 0$. Suppose a small upward perturbation in R such that $R_s > 0$. In this case there would be

[31] SROs can also be made more complex, for example allowing SROs to manipulate evidence of fraud or frame innocent members in order to extort money from them, as in Polinsky and Shavell (2001), or allow competition for promotion among several "whistle-blowing" principals.

fraud exposure and positive vigilance $v_H^* < v_L^*$, $p_L > p_H$ and accordingly, $R_p > 0$.

Consider now equilibrium $R^* > 0$. Suppose that $R_s > R^*$. In this case there would be fraud exposure and positive vigilance $v_H^* < v_L^*$, $p_L > p_H$ and R_p would be positive but below the 45-degree diagonal, that is $R_p < R_s$, as function $R_p(R_s)$ crosses the 45-degree diagonal from above, as shown in proof of Proposition 1(iii). By successive iterations, fixed-point $R_s = R_p = R^* > 0$ would be established. The same holds for the case of a downward perturbation in R, that is, $R_s < R^* > 0$. Hence, equilibrium $R^* > 0$ is the only equilibrium stable to small perturbations in R if $\frac{\mathrm{d}p}{\mathrm{d}v} > 0$.

Proof of Proposition 2(i) This proof is identical to proof of Proposition 1(i) except that $R^* = 0$ and $e^* = 0$ imply $\beta^* = 0$.

Proof of Proposition 2(ii) An equilibrium different from $\beta^* = 0$, $e_i = v_i = 0$ and $x^* = x_{\max}^{e=0}$ requires a fixed point $R^* > 0$. $R_s > 0$ implies that $v_L^* \geq v_H^* \geq 0$ regardless of whether a collusive agreement is reached. If $\frac{\mathrm{d}p}{\mathrm{d}v} \leq 0$, this implies that $p_L \leq p_H$. Given Assumption 2, it follows that $R_p \leq 0$. Therefore, a fixed point $R^* > 0$ cannot exist if $\frac{\mathrm{d}p}{\mathrm{d}v} \leq 0$. The proof for stability is identical to that provided for Proposition 1(ii).

Proof of Proposition 3(i) $L > \tau$ implies that there exist fixed points $R^* = (b_e - \lambda)W > 0$ such that $R^* < L - \tau$. The latter is guaranteed by choosing appropriately out-of-equilibrium belief b_e, which is not constrained by Bayes' Rule. $R^* > 0$ implies that there exists a contract $\{(\beta^* - \tau) \in (R^*, L - \tau), e^* = 0\}$. From the first-order condition of the member and the SRO types it follows that $v_L^* > v_H^* > 0$ and $x^*(v_L^*) < x^*(v_H^*) < x_{\max}^{e=0}$.

Proof of Proposition 3(ii) If $L - \tau > R_E^*$, R_E^* is not an equilibrium because there would be a collusive contract $\{(\beta^* - \tau) \in (R_E^*, L - \tau), e_i^* = 0\}$.

Proof of Proposition 4(i) $L - \tau \leq R_E^*$ and $\beta - \tau < L - \tau$ implies that $\beta - \tau < R_E^*$. Given that in a Collusive Equilibrium $\beta^* - \tau$ replaces R_E^* in the SRO types' first-order conditions, the SRO's marginal vigilance incentives are lower in a Collusive Equilibrium. On the other hand, in a Collusive Equilibrium β^* replaces L in the member's reaction function. Given that collusion requires $\beta^* < L$, in a Collusive Equilibrium the member's marginal fraud incentives and the equilibrium level of fraud are higher than in Exposure Equilibrium R_E^*.

Proof of Proposition 4(ii) If $R_C^* > R_E^*$ and $(\beta^* - \tau) \in (R_C^*, L - \tau)$, then $\beta^* - \tau > R_E^*$, which implies that the SRO's marginal vigilance incentives are higher. However, $\beta^* < L$ implies that the member's marginal fraud incentives are also higher. While the equilibrium level of vigilance is unambiguously higher, the net effect on the equilibrium level of fraud remains ambiguous due to the counteracting effect of higher vigilance on fraud.

Proof of Proposition 4(iii) Condition (a) is required due to result 4(i). Condition (b) is needed because $\beta^* \to L$ implies that the member's marginal fraud incentives in the Collusive and Exposure Equilibria converge, while $\beta^* - \tau > R_E^*$ raises SRO's marginal vigilance incentives. This implies that only in this situation the equilibrium level of fraud would be unambiguously lower in the Collusive Equilibrium than in the Exposure Equilibrium.

Proof of Proposition 5 Proposition 1(iii) has shown that an Exposure Equilibrium—which can exist under a corruptible SRO—yields positive vigilance and fraud deterrence. In a Collusive Equilibrium, exposure must be credible, which requires that $R^* > 0$. From the SRO's FOCs and assumption $p_v(x, 0) = \infty$, it follows that $v_i^* > 0$. In turn, the latter implies that $x^*(v_i^*) < x_{\max}^{e=0}$, that is, there is fraud deterrence.

References

Bortolotti, B., & Fiorentini, G. (1999). Barriers to entry and the self-regulating professions: Evidence from the market of Italian accountants. In *Organized interests and self-regulation; an economic approach* (pp. 131–157). Oxford: Oxford University Press.
Cho, I.-K., & Kreps, D. (1987). Signalling games and stable equilibria. *Quarterly Journal of Economics, 102*, 179–222.
Emons, W. (1997). Credence goods and fraudulent experts. *RAND Journal of Economics, 28*(1), 107–119.
Fudenberg, D., & Tirole, J. (1991). *Game Theory*. Cambridge, Mass: MIT Press.
Gehrig, T., & Jost, P.-J. (1995). Quacks, lemons, and self regulation: A welfare analysis. *Journal of Regulatory Economics, 7*, 309–325.
Harsanyi, J. (1982). Morality and the theory of rational behaviour. In A. Sen & B. Williams (Eds.), *Utilitarianism and beyond*. Cambridge: Cambridge University Press.
Laffont, J. J., & Tirole, J. (1991). The politics of government decision-making: A theory of regulatory capture. *Quarterly Journal of Economics, 106*(4), 1089–1127.
Maxwell, J., Lyon, T., & Hackett, S. (2000). Self-regulation and social welfare: The political economy of corporate environmentalism. *Journal of Law and Economics, 43*(2), 583–617.
Milgrom, P. (1981). Good news and bad news: Representation theorems and applications. *Bell Journal of Economics, 12*, 380–391.
Milgrom, P., & Roberts, J. (1986). Price and advertising signals of product quality. *Journal of Political Economy, 94*, 796–821.
Nunez, J. (2001). A model of self regulation. *Economics Letters, 74*, 91–97.
Pirrong, S. (1995). The self-regulation of commodity exchanges: The case of market manipulation. *Journal of Law and Economics, 38*, 141–206.
Polinsky, A. M., & Shavell, S. (2001). Corruption and optimal law enforcement. *Journal of Public Economics, 81*, 1–24.
Scarpa, C. (1999). The theory of quality regulation and self-regulation: Towards an application to financial markets. In B. Bortolotti & G. Fiorentini (Eds.), *Organized interest and self-regulation* (pp. 236–260). Oxford: Oxford University Press.
Shaked, A., & Sutton, J. (1981). The self-regulating profession. *Review of Economic Studies, 48*, 217–234.
Shapiro, C. (1982). Consumer information, product quality and seller reputation. *Bell Journal of Economics, 13*, 20–35.
Stefanadis, C. (2003). Self-regulation, innovation, and the financial industry. *Journal of Regulatory Economics, 23*(1), 5–26.

Tirole, J. (1996). A theory of collective reputations with applications to the persistence of corruption and to quality. *Review of Economic Studies, 63*(1), 1–22.

Van den Bergh, R. (1999). Self regulation of the medical and legal professions: Remaining barriers to competition and EC law. In B. Bortolotti & G. Fiorentini (Eds.), *Organized interest and self-regulation* (pp. 89–130). Oxford: Oxford University Press.