CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Discussion Paper No. 2016-19

Matthias Dahm, Paula González and Nicolás Porteiro

December 2016

The Enforcement of Mandatory Disclosure Rules

CeDEx Discussion Paper Series
ISSN 1749 - 3293

CENTRE FOR DECISION RESEARCH & EXPERIMENTAL ECONOMICS

The Centre for Decision Research and Experimental Economics was founded in 2000, and is based in the School of Economics at the University of Nottingham.

The focus for the Centre is research into individual and strategic decision-making using a combination of theoretical and experimental methods. On the theory side, members of the Centre investigate individual choice under uncertainty, cooperative and non-cooperative game theory, as well as theories of psychology, bounded rationality and evolutionary game theory. Members of the Centre have applied experimental methods in the fields of public economics, individual choice under risk and uncertainty, strategic interaction, and the performance of auctions, markets and other economic institutions. Much of the Centre's research involves collaborative projects with researchers from other departments in the UK and overseas.

Please visit http://www.nottingham.ac.uk/cedex for more information about the Centre or contact

Suzanne Robey
Centre for Decision Research and Experimental Economics
School of Economics
University of Nottingham
University Park
Nottingham
NG7 2RD
Tel: +44 (0)115 95 14763
Fax: +44 (0) 115 95 14159
suzanne.robey@nottingham.ac.uk

The full list of CeDEx Discussion Papers is available at

http://www.nottingham.ac.uk/cedex/publications/discussion-papers/index.aspx

The Enforcement of Mandatory Disclosure Rules*

Matthias Dahm[†], Paula González[‡] and Nicolás Porteiro[§]

December 14, 2016

Abstract

This paper examines the incentives of a firm to invest in information about the quality of its product and to disclose its findings. If the firm holds back information, it might be detected and fined. We show that optimal monitoring is determined by a trade-off. Stricter enforcement reduces the incentives for selective reporting but crowds out information search. Our model implies that (i) the probability of detection and the fine might be complements; (ii) the optimal monitoring policy does not necessarily eliminate selective reporting entirely; (iii) even when there is some selective reporting in equilibrium and more stringent monitoring is costless, increasing the probability of detection might not be beneficial; and (iv) when society values selectively reported information, the optimal fine might not be the largest possible fine.

Keywords: strategic information transmission, distrust effect, confidence effect, monitoring, penalty, fine, sanction, detection probability
JEL Classification No.: D82, L15

*We would like to dedicate this paper to our co-author Nicolás Porteiro who passed away in April 2012. We are truly grateful for his friendship over many years and for all that we learned from him during that time. We are indebted to Inés Macho-Stadler, David Pérez-Castrillo and Daniel Seidmann for their careful reading of this paper. We are also grateful for valuable comments and suggestions from Helmut Bester, Farasat Bokhari, Subhasish M. Chowdhury, Thomas Gall, Paul Heidhues, Roberto Hernán-González, Navin Kartik, Sang-Hyun Kim, David Myatt, Pau Olivella, Carmelo Rodríguez-Álvarez, Marta Ronchetti, Roland Strausz, Marcos Vera-Hernández and Ansgar Wohlschlegel. Financial support from Fundación Ramón Areces is gratefully acknowledged. This work is partially supported by the research projects ECO2012-36480 and ECO2015-65408-R (Ministerio de Ciencia y Tecnología) and FEDER. All errors are our sole responsibility.

[†]Corresponding author: University of Nottingham, School of Economics, University Park, Nottingham, NG7 2RD (UK); ORCID: orcid.org/0000-0002-3146-6802; email: Matthias.Dahm@nottingham.ac.uk.

[‡]Universidad Pablo de Olavide, Department of Economics; email: pgonzalez@upo.es.

[§]Universidad Pablo de Olavide, Department of Economics.

1 Introduction

The US FDA Amendment Act 2007 requires that results must be posted on clinicaltrials.gov within a year of the completion of the trial for all trials with at least one site in the US. The FDA has the power to fine trial sponsors who do not comply but rarely does this . . . The proposed EU Clinical Trials Regulation will require that summary results for every registered trial must be posted within one year of the completion of the trial, and the European Commission is discussing how to enforce this properly. Trial approval bodies in each country should consider expanding their monitoring of reporting, and ensure there is routine and open public audit of compliance for each individual trial.

- The AllTrials Campaign[1]

In September 2004, the pharmaceutical company Merck voluntarily withdrew Vioxx–a pain medication for arthritis–from the world market, because a clinical trial indicated that it increased the risk of heart attacks and strokes when taken for at least 18 months. Later, however, it was discovered that the company had failed to warn of the drug's dangers before the withdrawal. Following several scandals of so-called selective reporting of clinical trial results, the Food and Drug Administration Amendments Act (FDAAA) included the requirement of basic result reporting described in the quote. Mandatory disclosure rules have also been established in other areas in order to counter incentives for selective reporting. For instance, manufacturers of SUVs are required to report rollover risk in the US. This regulation followed an inquiry into a series of deadly accidents during which it was found that the tire manufacturer Bridgestone/Firestone and the auto company Ford had failed to inform the public about the risk of Ford Explorer SUVs rolling over after tires blew out without warning.[2]

The above quote argues that mandatory disclosure rules for clinical trials should be complemented by strict enforcement. Considering monitoring through penalties and appropriate resources to conduct inspections, our research question is to identify the effects of such an enforcement.[3] Our results uncover a trade-off that optimal enforcement must balance and that policy discussions seem to be unaware of.

[1] The AllTrials campaign was launched in 2013 and at the time of writing has been signed by 89582 people and 711 organisations, see www.alltrials.net, accessed on 12/12/2016.

[2] On Vioxx see Berenson (2006), Antman et al. (2007) or Krumholz et al. (2007). On the FDAAA of September 2007 see Wood (2009). A detailed account of the SUV rollover scandal and the development of the Transportation Recall Enhancement, Accountability, and Documentation Act (TREAD) of November 2000 can be found in Fung et al. (2007). This book also discusses 17 other policy areas in which mandatory disclosure rules exist, including corporate financial disclosure, nutritional labeling and restaurant hygiene disclosure. Dranove and Jin (2010) offer further background on disclosure including a brief history.

[3] Fung et al. (2007) discuss (on p. 45-46) in detail the need for appropriate enforcement through monitoring and levying penalties. The FDAAA allows for civil penalties of as much as $10 000 per day but this

We model enforcement of disclosure regulation as a supervising agency that invests resources in order to detect whether a firm held back information, and if so imposes a fine on the firm. Enforcement is hence captured by combination of a probability of detection and a fine. The agency, however, can also be thought of as a surrogate for a more indirect enforcement arising from litigation, whistle-blowing, political activism or journalistic investigations.[4] We model the firm's investment in information as a persuasion game (Milgrom, 1981; Grossman, 1981; Milgrom and Roberts, 1986; Seidmann and Winter, 1997). This captures that a pharmaceutical firm cannot lie and forge the entire evidence of a clinical trial in its favour. Following Dye (1985) and Shin (1994), however, we allow for the possibility that the firm does not become informed, which mitigates the classical unravelling argument. As a result of these assumptions we reproduce that a clinical trial can either be positive, negative or inconclusive (De Angelis et al., 2004), and obtain selective reporting as an equilibrium phenomenon. As with Shin's sanitization strategy negative trials are suppressed, while positive trials are revealed.

Selective reporting is considered to be harmful to society.[5] It might therefore appear that the effects of enforcement in our context are analogous to those discussed in the law enforcement literature. Following Becker (1968) the deterrence of a harmful act depends on the expected fine. Moreover, it is optimal to combine a low probability of detection with the highest possible fine, for, if the fine were not as high as possible, then one could simultaneously increase the fine and decrease the probability of detection, thereby reducing enforcement costs.

In this paper, however, we show that the concealment of information is qualitatively different from other harmful acts. First, the failure of a firm to disclose positive product information makes the public more pessimistic, as it is aware that information might be withheld. We dub this the *distrust effect* of a lack of evidence. If, however, monitoring does not reveal that the firm concealed information, then the public becomes more optimistic;

is considered to be insufficient, see Prayle et al. (2012), Anderson et al. (2015) or Gopal (2015). As a result, there are calls for greater transparency in clinical trials, including Goldacre (2013), Chan et al. (2014), Hudson and Collins (2015), and the aforementioned AllTrials on-line petition.

[4]For example, during a product liability trial it was discovered that the DePuy Orthopaedics division of Johnson & Johnson failed to warn of the risks of artificial hip implants, see Editorial (2013). Also, in the discoveries that Johnson & Johnson and ATK Launch Systems Inc. held back information, a whistle-blower was involved. Under this interpretation, different institutional designs of, for example, liability trials, confidentiality agreements, and legal protection for whistle-blowers might be related to different magnitudes of the probability of detection. For instance, the False Claims Act in the U.S. incentivizes whistle-blowers to bring hidden information to the government's attention. Whistle-blowers are awarded 10 to 30 percent of the sum recovered. Higher probabilities of detection might then be thought of as being induced by higher percentages.

[5]For the case of clinical trials the International Committee of Medical Journal Editors writes "The case against selective reporting is particularly compelling for research that tests interventions that could enter mainstream clinical practice. . . . When research sponsors or investigators conceal the presence of selected trials, these studies cannot influence the thinking of patients, clinicians, other researchers, and experts who write practice guidelines or decide on insurance-coverage policy." See De Angelis et al. (2004, p. 477).

and the higher the probability of detection, the more optimistic the public becomes. We call this the *confidence effect* of monitoring. As a result of this second effect, the probability of detection and the fine do not have to be substitutes as in Becker (1968), but can have a complementary relationship. Second, since in our model the firm's investment in information is endogenous, the enforcement of disclosure regulation has not only an effect on the way information is reported but also on the incentives to invest in information in the first place. We introduce the term *deterrence effect* of disclosure rules on information search to refer to this effect. Consequently, enforcement might deter both honestly and selectively reported information and the strength of this deterrence effect depends on the specific combination of the probability of detection and the fine. Thus, unlike in Becker's model, optimal enforcement does not only depend on the expected fine. Lastly, unlike in the law enforcement setting, the harmful act might be socially valuable, albeit less than honest reporting.

We show that optimal monitoring is determined by a trade-off: Stricter enforcement reduces the information held back but crowds out information search, including the information that is honestly reported. The incentive to invest in information that is selectively reported declines, because stricter enforcement increases the expected fine. The incentive to invest in information that is honestly reported also declines, because a higher probability of detection increases the confidence effect and therefore the opportunity costs of information search. The existence of this trade-off implies that optimal policies might tolerate some selective reporting. Moreover, we show that the deterrence effect on information search can be very strong; even in the presence of some selective reporting and when more stringent monitoring is costless, increasing the probability of detection might be welfare reducing. Lastly, we show that when society values selectively reported information sufficiently, the optimal fine might not be maximal.

Related Literature

The law enforcement literature does not suggest that maximum penalties are always optimal (Garoupa, 1997; Polinsky and Shavell, 2000). Among the different reasons which may advocate for non-maximal penalties are marginal deterrence incentives (Mookherjee and Png, 1992), socially costly sanctions (Kaplow, 1990), differences in wealth among individuals (Polinsky and Shavell, 1991), imperfect information on the probability of apprehension (Bebchuk and Kaplow, 1992), risk-aversion (Polinsky and Shavell, 1979) or the possibility to engage in socially costly activities that reduce the probability of detection (Malik, 1990). In the context of deterrence of collusion, lenient fines can arise when penalties are dynamic and dynamic conditions for cartel stability are considered (Harrington, 2014). Since none of these papers contains a model of information transmission, these rationales for limits on fines are qualitatively different from the rationale we present.

Corts (2014), however, offers a rationale for finite expected penalties based on a model of information transmission. He considers a model of false advertising claims in which a firm signals unverifiable information about product quality to consumers. Optimal expected penalties might be finite, because when information on quality is expensive, society prefers the firm not to eliminate all uncertainty about the quality of its product. Instead, it is desirable that the firm makes ‘speculative claims’, which sometimes turn out to be false. Too severe penalties deter the firm from making these claims. In contrast, we offer a model of selective reporting of verifiable information, rather than of false claims. In our model the drawback to high penalties is that they deter investment in information, rather than that they deter a firm to make claims that are likely, but not certain, to be true. Moreover, we explicitly model monitoring policies as a probability of detection in combination with a fine, rather than as an expected penalty. This allows us to derive the novel result that the fine and the probability of detection can be complements. The possibility of a complementary relationship shows that the forces in our model are conceptionally different from those in all these previously reviewed papers.[6]

The literature on disclosure rules for endogenous information does not suggest that mandatory disclosure rules are always better than voluntary ones. In particular a deterrence effect on information search is also present in Farell (1986), Shavell (1994), Dahm et al. (2009), Henry (2009) and Polinsky and Shavell (2012). In these papers, however, mandatory disclosure is exogenously enforced. In contrast, we show that the deterrence effect appears also as a consequence of imperfect enforcement and how it depends on the specific combination of the monitoring instruments.[7] Polinsky and Shavell (2012) analyse the effects of mandatory versus voluntary disclosure rules and show that–because of the deterrence effect on endogenous information–welfare under the latter might be higher.[8] We go beyond this result by investigating how the deterrence effect in turn shapes the optimal monitoring policy and show that it might imply imperfect enforcement. The optimal effective disclosure rule might thus fall between the benchmarks of mandatory and voluntary

[6]The complementarity might decrease deterrence following an increase in the probability of detection. This is related to the finding in Andreoni (1991) and Feess and Wohlschlegel (2009) that deterrence might decrease in response to an increase in punishment. The mechanism differs, however, as in our model there is no threshold of reasonable doubt that depends on punishment.

[7]Kartik et al. (2015) provide a related comparative statics result. Kartik et al. assume that the agent who sends a message bears a disclosure cost and show that this leads to less disclosure. Their setting, however, differs from ours in important ways. Most substantially, they assume that the agent is exogenously endowed with information.

[8]Relatedly, Di Tillio et al. (2015) show in a different model that it might be beneficial to allow for manipulation of information, because it might lead to more information in the system. In Kartik et al. (2016) mandatory disclosure rules would harm the decision maker. This implication of their model comes from the fact that adding more experts reduces each expert’s incentive to invest in costly information. The reason for this is related to our result that successful monitoring affects the opportunity costs of information search, although the details differ. Dranove and Jin (2010) review additional reasons why mandatory disclosure does not always raise social welfare.

disclosure.

It is also known that different disclosure rules have different effects on reporting and information search. Dahm et al. (2009) analyse the effects of a clinical trial results database, which is a specific voluntary disclosure rule, and how it interacts with clinical trial registries, which make the firm's decision to search for information observable. Dahm et al. show that a voluntary results database leads to unravelling, because it allows the firm to prove that it has no information when it is uninformed. In addition, the specific combination of registry and results database determines the opportunity costs of information search and hence the strength of the deterrence effect on endogenous information. The present paper differs from our earlier paper by focussing on a mandatory disclosure rule that is successful when appropriately enforced, rather than a voluntary disclosure rule that is followed in order to avoid sceptical inference from the public. In addition, we provide a welfare analysis, rather than assuming that regulation is the better the greater the incentives for information search.

Less narrowly related to our paper is a substantial literature on information transmission making the benchmark assumption that the sender can commit to communicate to the receiver everything he knows (Matthews and Postlewaite, 1985; Farell, 1986; Shavell, 1994; Dahm and Porteiro, 2008; Board, 2009; Dahm et al., 2009; Polinsky and Shavell, 2012; Kamenica and Gentzkow, 2011). We relax this assumption by building a model in which the threat of detection induces truth-telling and show that optimal monitoring does not always imply perfect enforcement of mandatory disclosure rules.

This paper is organized as follows. The next section introduces our model. In Section 3, we analyse the equilibria of the game and establish the existence of the trade-off in monitoring. We then impose in Section 4 further assumptions that sharpen this trade-off and offer results on the optimal monitoring policy, including the finding that the optimal fine might not be the largest possible fine. We offer some concluding remarks in Section 5.

2 The model

2.1 Agents

Consider a firm (or seller) that produces a good and interacts with buyers (or the public) through a market. The quality of the good depends on the state of the world $v \in \{0, 1\}$ and is, ex-ante, unknown both to the seller and the public. A pharmaceutical product might, for example, cure a health problem or be ineffective and/or have severe side effects. The likelihood of the good to generate benefits ($v = 1$) is $q \in (0, 1)$; with probability $1 - q$ the good does not provide benefits and might even be harmful ($v = 0$). To be concise, in our interpretations we will refer to q as measuring the (expected) quality of the good, rather than speaking of the ex-ante likelihood of benefits or risk of harm.

2.2 Information search by the seller

Prior to offering the good for sale, the seller can invest in information search in order to provide information about product quality (hereafter we will call this a test). Product quality might be affected, for example, by safety issues or design flaws and a test might reveal that quality is not compromised. Information search can have three possible outcomes, mimicking the outcomes of clinical trials (see De Angeles et al., 2004). First, the test can demonstrate that the seller's product is of high quality ($v = 1$). We will call this outcome a positive test. Second, the test can show that the seller's product is of low quality ($v = 0$), a situation to which we will refer as negative test. Third, the test can be inconclusive (i.e., it does not provide new evidence).

Formally, the seller can conduct a test at a constant marginal cost $K_x > 0$. The result of the test is denoted by t. The test reveals with probability $x \in [0, 1]$ the true state of the world, that is, $t = v$. With probability $1 - x$, the test is inconclusive, that is, $t = \emptyset$. The information revealed through the test is hard evidence. For our leading application this assumption captures the fact that a pharmaceutical firm cannot forge the entire evidence of clinical trials and indicate that certain desirable treatment effects exist when they do not. However, in line with the scandals mentioned in the Introduction we allow that the firm reports trial results selectively. We denote the seller's report or message by m. If the test reveals that the seller's product is of low quality, that is $t = 0$, then the seller can hide this evidence. Formally, if $t = v$, the seller can decide to publish the result of the test or conceal it, i.e., $m \in \{v, \emptyset\}$. If the test is inconclusive, that is, $t = \emptyset$, then the seller cannot forge evidence and has to report this fact, that is, $m = \emptyset$. We assume that the public cannot distinguish between a seller who invests in information and hides the result and a seller who does not invest in information; in both situations the public receives the message $m = \emptyset$.

2.3 The seller's payoffs

In our model the interaction between the firm and the public is represented by the seller's profit function $\pi(q)$, which represents the equilibrium profits resulting from the sales process. For our purpose it is sufficient to assume that this function depends directly on the (possibly updated) belief of the public about the quality of the good. This avoids additional notation capturing that the public takes actions based on q, say to buy or not to buy the firm's product, which in turn affects the profit function of the firm. For most of our paper we consider a general setting with a general function $\pi(q)$, rather than assuming a specific interaction that gives rise to a specific functional form. This is so, because in our model the existence of informative equilibria hinges only on the shape of the profit function. The following two conditions define a class of functions that permits a sharp characterization.

We impose them throughout.[9]

Assumption 1 (Monotonicity) *The profit function $\pi(q)$ is weakly increasing in the quality q of the product, with $\pi(0) < \pi(q) < \pi(1)$ for all $q \in (0,1)$.*

Loosely speaking, Assumption 1 requires strict monotonicity at the endpoints of the interval and allows for weak monotonicity for intermediate values. Assuming that the firm's profits are increasing in quality is in line with evidence from the antiulcer-drug market (Azoulay, 2002). Without such a monotonicity the seller has no incentive to search for information and to hold back negative information, precluding to study the problem of selective reporting. Indeed, we will see that $\pi(q) < \pi(1)$ is needed for the possibility of an equilibrium in which the seller invests in selectively reported information. On the other hand, $\pi(0) < \pi(q)$ assures that the firm prefers to report selectively after a negative test.

Assumption 2 (Bounded average quality) *The profit function $\pi(q)$ has bounded average quality if for all $q \in (0,1)$ we have that*

$$\pi(q) < q\pi(1) + (1-q)\pi(0). \tag{1}$$

Bounded average quality requires that $(\pi(q) - \pi(0))/q$, the slope of the chord from $\pi(0)$, is bounded by $\pi(1) - \pi(0)$. Under this condition $\pi(q)$ can have both convex and concave segments but has in some sense a (globally) convex shape. We will see that (1) is a necessary and sufficient condition for the firm to be willing to invest in information when monitoring is successful and all information obtained is honestly reported, independently of what has been learned (Lemma 3). We will later argue that Assumption 2 captures the best case for monitoring. Moreover, it is not unreasonable that the profit function has this shape. In fact, in Section 4 we will relate a stronger version (Assumption 3) to empirical evidence on pharmaceutical market performance.

It is instructive to illustrate Assumptions 1 and 2 considering the following example.

Example 1 *Consider the class of profit functions defined by $\pi(q) = q^\lambda$, with $\lambda > 0$.*[10] *This function fulfils Assumption 1 for any $\lambda > 0$. It fulfils Assumption 2 provided $\lambda > 1$ so that $\pi(q)$ is strictly convex. For later reference we note that the higher λ, the more risk proclivity the seller exhibits, because for this class of profit functions the coefficient of relative risk aversion is $-(\lambda - 1)$.*

[9] In Kartik et al. (2015) and (2016) the payoffs of senders of information also depend directly on the beliefs of the decision maker. Corts (2014) makes a similar assumption to our Assumption 1 in a different model. Assumption 2 is related–but in some sense weaker–than the notion of a star-shaped funtion, which has appeared in the economic literature, see the discussion of Assumption 3 in Section 4. For analytical convenience we also implicitly assume that $\pi(q)$ is differentiable.

[10] The profit function $\pi(q) = q^\lambda$ can be given a micro-foundation based on a monopoly market for a medical treatment. Details are available upon request (and included for the convenience of the referees in Appendix B.1, which is not intended for publication).

2.4 Regulatory options

We assume that the seller is required to disclose quality and safety problems. There is a supervising agency that prior to the sales process aims to detect hidden information and to punish selective reporting. As already mentioned, this probability of being detected can also be interpreted as the probability that selective reporting becomes known through indirect channels, like whistle-blowing.

The monitoring technology is characterized by three assumptions. First, the agency's monitoring might reveal information that the firm is aware of or not. For example, when the Food and Drug Administration controls a production plant, it might find contamination problems of which the seller might be aware or not. In another interpretation a whistle-blower might have specific knowledge of issues related to his daily work, he might have discovered problems, and superiors might have denied acting upon this information. Second, the agency only searches for negative information. This is motivated by our interest in selective reporting and the fact that the seller has no incentive to hide positive information. Third, whenever quality and safety problems are discovered, the agency also learns whether selective reporting has occurred. This is the best-case scenario for monitoring and the conservative assumption to make.[11]

More precisely, we assume that whenever the seller does not reveal the state of the world (that is, $m = \emptyset$) the agency conducts a test. This test detects a flawed product with probability $\rho \in (0,1)$ at a constant marginal cost $K_\rho > 0$ to society.[12] Denoting by $r \in \{0, \emptyset\}$ the outcome of the agency's report, we have that

$$\begin{aligned} \Pr(r=0|v=0) &= \rho = 1-\Pr(r=\emptyset|v=0) \quad \text{and} \\ \Pr(r=0|v=1) &= 0 = 1-\Pr(r=\emptyset|v=1). \end{aligned}$$

Whenever the agency obtains the report $r = 0$ and detects withholding of evidence, it makes this information public and imposes a fine $F > 0$ on the seller. If $r = 0$ but no evidence was concealed, then the quality problems are disclosed but the firm is not fined.

A specific monitoring policy consists of a fine F and a probability of detection ρ and is denoted by (F, ρ). In Section 3 we take the monitoring policy as given, while in Section

[11]Our main result that optimal monitoring might not be as strict as technically feasible is the less surprising, the less powerful the monitoring technology of the agency.

[12]When $m \neq \emptyset$, the agency's test is redundant, as the firm's message is hard evidence. There is, however, an implicit assumption that the agency also searches when the seller is not expected to search (in a non-informative equilibrium). This captures for example the case of the Food and Drug Administration controlling a production plant and finding contamination problems of which the seller is not aware. Because of the confidence effect, this raises the opportunity costs of investment in information and reinforces the seller's behaviour (increases the parameter space for which a non-informative equilibrium exists). Technically, this increases the multiplicity of equilibria but does not matter for our results, as our results do not rely on a particular equilibrium selection. Notice also that assuming that the agency conducts tests with probability smaller than one, would not change our results. If the agency conducted tests with probability β, it would suffice to make a change of variable so that $\rho' = \beta\rho$ and rescale K_ρ.

4 we derive a framework for social welfare and assume that optimal monitoring aims to maximize it.

2.5 Timing

Once a monitoring policy (F, ρ) is in place, the sequence of events is as follows:

1. The firm decides whether to conduct a test (the public does not observe this choice).
2. The seller sends a message m to the buyer which is observed by the agency (if no test has been conducted, $m = \emptyset$).
3. If $m = \emptyset$, the agency conducts its own research and obtains a report r, which is then made public (otherwise, $r = \emptyset$).
4. Depending on the outcomes of m and r the agency decides whether to impose the fine on the firm, and the public updates her beliefs about the expected quality of the seller's product (that is, that the state of the world is $v = 1$) from the prior q to a posterior belief $\tilde{q}$.
5. The sales process captured through Assumptions 1 and 2 unfolds. The seller pays the fine, if the firm was monitored and a fine was imposed.

This game is solved by backward induction. Given that the public does not observe the seller's decision whether to invest in a test or not, she has to base her behaviour on her beliefs about the seller's choice. Moreover, the sales process following each message or report that reveals the state of the world is a proper subgame of the extensive form game. The appropriate equilibrium concept is, hence, a Perfect Bayesian Equilibrium (PBE) in which all agents behave optimally, given their beliefs about the other's action and these beliefs are, at equilibrium, correct.

3 The trade-off in monitoring

We investigate now the equilibria of the game. As usual in this type of models there might exist both an informative and a non-informative equilibrium, depending on how expensive tests are. Clearly, in an informative equilibrium the public expects the firm to invest in information and the beliefs of the public must be consistent with the firm's reporting strategy. Subsection 3.1 derives the beliefs in an informative equilibrium and provides basic results on the interaction between monitoring and the incentives for honest reporting. Building on these results, Subsections 3.2–3.4 establish conditions for informative equilibria that differ in the firms reporting strategy. In Subsection 3.5 we turn to the non-informative equilibrium. Lastly, Subsection 3.6 provides an integrated analysis of the equilibria so far studied in isolation and shows that monitoring policies must balance a trade-off.

3.1 Monitoring and honest reporting

Suppose the firm invests in information search. The incentives to report honestly depend on the monitoring policy. A special case of a monitoring policy is laissez faire. Laissez-faire sets $\rho = 0$ so that the expected fine is zero. Under laissez-faire, given that test results are hard evidence, if the seller reports low quality ($t = 0$), then the public will infer $\tilde{q} = 0$. Under the Monotonicity Assumption, this message strategy is not a best reply. Consequently, the seller only discloses favourable information about product quality.[13] The aim of monitoring is to induce honest reporting of test results. Formally, selective reporting (SR) and honest reporting (HR) are described as

$$m^{SR} = \begin{cases} 1 & \text{if } t = 1 \\ \emptyset & \text{if } t \in \{0, \emptyset\} \end{cases} \quad \text{and} \quad m^{HR} = \begin{cases} v & \text{if } t = v \\ \emptyset & \text{if } t = \emptyset \end{cases}, \tag{2}$$

respectively. The former differs from the latter in that a negative test result ($t = 0$) is not revealed under selective reporting.

In order to derive a general expression for the buyer's posterior beliefs $\tilde{q}$ in an informative equilibrium, suppose that the seller reports honestly with probability y. The posterior beliefs of the buyer are then given by

$$\tilde{q} = \begin{cases} \Pr(v = 1 | m = 0) = 0 & \text{if } m = 0 \\ \Pr(v = 1 | m = \emptyset \wedge r = 0) = 0 & \text{if } m = \emptyset \wedge r = 0 \\ \Pr(v = 1 | m = 1) = 1 & \text{if } m = 1 \\ \Pr(v = 1 | m = \emptyset \wedge r = \emptyset) = \tilde{q}_\emptyset & \text{if } m = \emptyset \wedge r = \emptyset \end{cases}, \tag{3}$$

where the subindex $\emptyset$ in $\tilde{q}_\emptyset$ indicates that the public has not received new information. More precisely, the beliefs of the buyer when neither the firm nor the agency reveal the state of the world are given by

$$\tilde{q}_\emptyset \equiv \frac{\Pr(m = \emptyset | v = 1) \Pr(r = \emptyset | v = 1) \Pr(v = 1)}{\Pr(m = \emptyset \wedge r = \emptyset)} = \frac{q(1-x)}{1 - xq - (1-q)(\rho + xy(1-\rho))}.$$

Notice that $\partial \tilde{q}_\emptyset / \partial y > 0$ holds. The higher the likelihood of selective reporting, the more pessimistic the public. If the firm is suspected to report selectively with a higher probability (and so y declines) and no evidence is published, then the public infers that it becomes more likely that the product is of low quality ($v = 0$) and that information has been withheld. We call this the *distrust effect* of selective reporting. In the extreme case of $x = 1$, it is known that the firm is informed and the classical unravelling argument obtains. Our focus,

[13] As with Shin's (1994) sanitization strategy, unfavourable evidence is concealed

however, is on situations in which the test might be inconclusive and the firm might not be informed, so that $\tilde{q}_\emptyset > 0$. In addition to the distrust effect we have that $\partial \tilde{q}_\emptyset / \partial \rho > 0$ holds. The better the quality of monitoring, the more optimistic the public becomes, when no additional negative information about the good is revealed. We call this the *confidence effect* of monitoring. As we will see, this effect induces some interesting forces to appear in our model, which are absent in the standard law enforcement setting.[14]

Suppose the seller invests in information and the test is negative. Whether a given monitoring policy (F, ρ) induces honest reporting depends on the following comparison. If the seller truthfully reports $m = 0$, then the public becomes as pessimistic as possible (because it is clear that the product cannot be of high quality). But since he is not fined by the regulatory agency, the seller's profits are $\pi(\tilde{q} = 0)$. If, on the other hand, the seller decides to hide the negative evidence, he risks being detected by the agency. In this case it becomes known that the product is of low quality and in addition the fine F is imposed. There is, however, also the chance that he is not detected and the product remains in the market. In equilibrium the public anticipates the frequency of honest reporting y and updates her beliefs about the perceived quality accordingly to $\tilde{q} = \tilde{q}_\emptyset$. Thus, the expected profits from withholding evidence are

$$\rho\left(\pi(\tilde{q} = 0) - F\right) + (1 - \rho)\,\pi\left(\tilde{q} = \tilde{q}_\emptyset\right).$$

This comparison implies the following result.

Lemma 1 *For any probability of detection $\rho > 0$, there exists a minimum penalty*

$$\tilde{F}(q, \rho, y) \equiv \frac{1-\rho}{\rho}\left(\pi\left(\tilde{q} = \tilde{q}_\emptyset\right) - \pi(\tilde{q} = 0)\right) \tag{4}$$

such that for any $F \geq \tilde{F}(q, \rho, y)$, the seller discloses all test results.

This lemma shows that there exist monitoring policies (F, ρ) that can avoid selective reporting. The minimum penalty $\tilde{F}$ is increasing in q, as $\partial \tilde{q}_\emptyset / \partial q > 0$ holds. This reflects the fact that the higher the expected product quality initially is, the higher the stakes (or the more can be lost) when investment in information takes place. In addition, the minimum penalty $\tilde{F}$ is increasing in y, as $\partial \tilde{q}_\emptyset / \partial y > 0$ holds. This is so, because the less the firm is expected to conceal information, the higher the benefit from withholding evidence.

We show now that the relationship between the probability of detection and the fine might be complementary. This implies that starting from a situation in which monitoring provides sufficient incentives to report honestly, an increase in ρ might lead to insufficient incentives. We first provide a lemma specifying a condition for a complementary relationship.

[14] In addition $\tilde{q}_\emptyset$ has the following properties. The higher the quality of the search technology, the more pessimistic the public $\partial \tilde{q}_\emptyset / \partial x < 0$; and the lower the initial product quality estimate, the lower the updated assessment $\partial \tilde{q}_\emptyset / \partial q > 0$.

The discussion following the lemma provides then an example showing that this condition can be satisfied under our assumptions. The possibility of a complementary relationship stands in sharp contrast to the standard law enforcement setting and shows that our setting is qualitatively different.

Lemma 2 *Suppose $F = \tilde{F}(q, \rho, y)$. The probability ($\rho$) and severity ($F$) of punishment are substitutes if and only if*

$$\frac{\left.\frac{d\pi}{dq}\right|_{\tilde{q}=\tilde{q}_{\emptyset}}}{\frac{\pi\left(\tilde{q}=\tilde{q}_{\emptyset}\right)-\pi(\tilde{q}=0)}{\tilde{q}_{\emptyset}}} < \frac{1-xq-(1-q)(\rho+xy(1-\rho))}{\rho(1-\rho)(1-q)(1-xy)}. \tag{5}$$

Proof: See Appendix A.1. *Q.E.D.*

Key for this result is the interplay of two effects. Consider an increase in ρ. First, there is a well-known deterrence effect. Similar to the standard law enforcement setting, the term $(1-\rho)/\rho$ in (4) decreases, so that if no other change occurred the minimum penalty $\tilde{F}$ decreases: it is more likely that selective reporting is detected and the fine is imposed. But there is also a countervailing second effect, which is absent in the law enforcement setting: As ρ increases the monitoring technology becomes more powerful, and when no flaw with the product is detected the public's confidence in the product increases. This confidence effect is reflected in $\pi\left(\tilde{q}=\tilde{q}_{\emptyset}\right)$, which increases with ρ. The overall effect is ambiguous, because the magnitude of the confidence effect depends entirely on the shape of $\pi(q)$. Indeed, the left hand side of (5) is closely related to the elasticity of the profit function with respect to q.[15] Consequently, when the seller exhibits sufficient risk proclivity, the monitoring instruments can be complements. To see this consider Example 1. For this class of profit functions, the left hand side of (5) becomes λ. This is a measure of risk proclivity and Assumption 2 implies that $\lambda > 1$. As a result, whether or not for given parameter values both monitoring policies are complements depends on the seller's attitudes towards risk, which in turn is induced by the nature of the sales process in the last stage. Figure 1 displays Example 1 with $\lambda = 3$ and $x = q = 0.2$. For $y = 0$ the firm is expected to report selectively, while for $y = 1$ it is expected to report honestly. Figure 1 illustrates a complementary relationship for a large interval of intermediate values of ρ; for example when $y = 0$ this is the case for ρ (roughly) in the interval $[0.46, 0.88]$.

Notice that one difference between our model and Becker's law enforcement setting is that in Becker's model the criminal's payoff from committing a harmful act does not depend on a belief of the public that he is liable. This prevents the confidence effect from operating. Under these conditions $\pi\left(\tilde{q}=\tilde{q}_{\emptyset}\right)$ is constant, the right hand side of (4) declines with ρ, and F and ρ are always substitutes.

[15] Notice that $\pi(0) = k$ is not required to be equal to zero and consider the increasing linear transformation $f(q) = \pi(q) - k$. Replacing this expression in the left hand side of (5) yields the elasticity $f'(q)q/f(q)$.

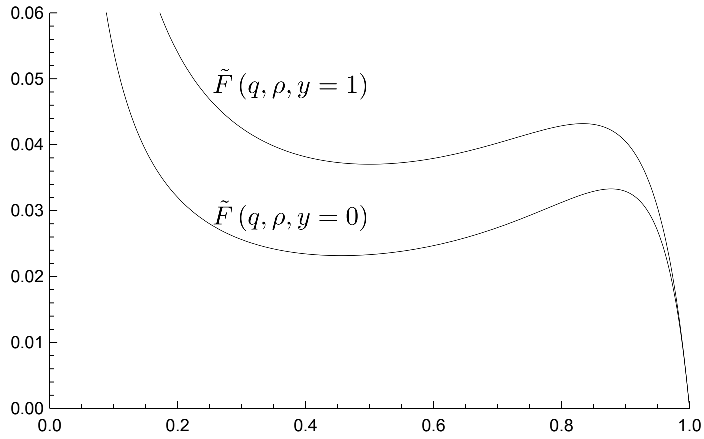


Figure 1: An example for a complementary relationship

3.2 Unsuccessful monitoring

It is instructive to start with the case in which monitoring is unsuccessful, because it is not powerful enough. This includes a situation of laissez-faire, which provides a benchmark for situations with monitoring. In this subsection we analyse when an informative equilibrium exists in which the buyer anticipates correctly that the seller reports selectively, as described in (2). In this case $y = 0$ and $F \leq \tilde{F}(q, \rho, y = 0)$ must hold. In what follows we denote this threshold by $\tilde{F}_0(q, \rho)$. In such a situation we have that

$$\tilde{q}_\emptyset(y = 0) = \tilde{q}_\emptyset^{SR} \equiv \frac{q(1-x)}{1 - xq - \rho(1-q)}. \tag{6}$$

Compared to the initial belief q, the distrust effect makes the public more pessimistic, while the confidence effect has the opposite effect. The distrust effect is stronger than the confidence effect if the agency's search technology is less efficient than the one of the firm: $\tilde{q}_\emptyset^{SR} \leq q$ if and only if $\rho \leq x$.

The seller's expected profits if he searches for information are,

$$\begin{aligned} \Pi_t(q) &= xq\pi(\tilde{q} = 1) + ((1-x)q + (1-q)(1-\rho))\pi\left(\tilde{q} = \tilde{q}_\emptyset^{SR}\right) \\ &\quad + (1-q)\rho(\pi(\tilde{q} = 0) - xF) - xK_x. \end{aligned} \tag{7}$$

With probability xq there will be a positive test and the beliefs of the buyer will be $\tilde{q} = 1$. However, in the remaining cases the test will be negative or inconclusive and the expected quality shifts to $\tilde{q} = \tilde{q}_\emptyset^{SR}$ or $\tilde{q} = 0$, depending on whether the agency detects that the product is flawed or not. If quality problems are detected, the agency learns that information has been concealed and the fine is imposed.

The seller's expected profits from not searching for information are,

$$\Pi_{\neg t}(q) = (1 - (1-q)\rho)\pi\left(\tilde{q} = \tilde{q}_\emptyset^{SR}\right) + (1-q)\rho\pi(\tilde{q} = 0). \tag{8}$$

The seller is expected to invest and the interplay of the confidence and distrust effects affects the seller's position in the market. In addition, there is the risk that the agency's monitoring detects a flaw with the seller's product.

The seller invests in the test if and only if

$$\Pi_t(q) - \Pi_{\neg t}(q) > 0 \Leftrightarrow K_x < \mathbb{K}_t^{SR}(q, \rho, F),$$

where

$$\mathbb{K}_t^{SR}(q, \rho, F) \equiv q\left(\pi\left(\tilde{q}=1\right) - \pi\left(\tilde{q}=\tilde{q}_{\emptyset}^{SR}\right)\right) - (1-q)\,\rho F.$$

We summarize with the following result.

Proposition 1 (Selective reporting equilibrium) *With the monitoring policy (F, ρ), there exists a PBE in which the seller invests in information search and reveals the information obtained selectively, provided*

- *the monitoring technology is not powerful enough, that is, $F \leq \tilde{F}_0(q, \rho)$ and*
- *the test is cheap enough, that is, $K_x \leq \mathbb{K}_t^{SR}(q, \rho, F)$.*

An example for a situation in which such an equilibrium potentially exists is laissez-faire. As $\rho \to 0$, we have that $\tilde{F}_0(q, \rho) \to \infty$. Moreover, the cost threshold $\mathbb{K}_t^{SR}$ is strictly positive, so that the second condition holds, provided information search is cheap enough. The behaviour of the firm described in this informative selective reporting equilibrium might appear cynical. But it rationalizes behaviour in line with a statement by Representative Henry Waxman (D-CA) at a hearing: "The pharmaceutical industry has systematically misled physicians and patients by suppressing information on their drugs...," see Couzin (2004, p. 1695).

3.3 Successful monitoring

We analyse now when an informative equilibrium exists in which monitoring induces the seller to report honestly all test results. In this case $y = 1$ and $F \geq \tilde{F}(q, \rho, y = 1)$ must hold. In what follows we denote this threshold by $\tilde{F}_1(q, \rho)$. In such a situation we have that

$$\tilde{q}_{\emptyset}(y=1) = \tilde{q}_{\emptyset}^{HR} \equiv \frac{q}{1-\rho(1-q)}. \tag{9}$$

Notice that $\tilde{q}_{\emptyset}^{HR} > q$, because on one hand, as the seller is not expected to report selectively, the distrust effect is not present, and on the other, the confidence effect makes the public more optimistic than initially.

The seller's expected profits from information search are,

$$\begin{aligned}\Pi_t(q) \;=\;& xq\pi(\tilde{q}=1) + ((1-x)q + (1-q)(1-x)(1-\rho))\,\pi\left(\tilde{q}=\tilde{q}_{\emptyset}^{HR}\right)\\ &+ (1-q)(\rho + x(1-\rho))\,\pi(\tilde{q}=0) - xK_x.\end{aligned} \tag{10}$$

With probability xq there is a positive test and the beliefs of the buyer are $\tilde{q} = 1$. In the remaining cases, however, the trial is negative or inconclusive and the expected quality shifts to $\tilde{q} = \tilde{q}_{\emptyset}^{HR}$ or to $\tilde{q} = 0$. Profits when the seller does not invest in a test are as in (8) with $\tilde{q}_{\emptyset}^{SR}$ replaced by $\tilde{q}_{\emptyset}^{HR}$. The seller invests in information search if and only if

$$\Pi_t(q) - \Pi_{\neg t}(q) > 0 \Leftrightarrow K_x < \mathbb{K}_t^{HR}(q, \rho),$$

where

$$\begin{aligned}\mathbb{K}_t^{HR}(q,\rho) \;\; \equiv \;\; & q\left(\pi\left(\tilde{q}=1\right) - \pi\left(\tilde{q} = \tilde{q}_{\emptyset}^{HR}\right)\right) \\ & -(1-q)(1-\rho)\left(\pi\left(\tilde{q} = \tilde{q}_{\emptyset}^{HR}\right) - \pi\left(\tilde{q}=0\right)\right).\end{aligned}$$

The preceding yields the following result.

Proposition 2 (Honest reporting equilibrium) *With the monitoring policy (F, ρ), there exists a PBE in which the seller invests in information search and reveals honestly all information obtained, provided*

- *the monitoring technology is powerful enough, that is, $F \geq \tilde{F}_1(q, \rho)$ and*
- *the test is cheap enough, that is, $K_x \leq \mathbb{K}_t^{HR}(q, \rho)$.*

Lemma 1 implies that for any quality q of the firm's product there exist monitoring policies (F, ρ) such that the first condition in Proposition 2 can be fulfilled. The second condition is similar to the one in Proposition 1 and requires that the test must be cheap enough. It remains to investigate in which circumstances this second condition may hold.

Lemma 3 *The cost threshold is strictly positive, that is $\mathbb{K}_t^{HR}(q, \rho) > 0$, if and only if Assumption 2 holds so that the profit function $\pi(q)$ has bounded average quality.*

Proof: See Appendix A.2. *Q.E.D.*

As mentioned before, we can think of bounded average quality as imposing a (globally) convex shape on the profit function $\pi(q)$.[16] The intuition is that in our model investment in information that is honestly reported only makes a difference if the test reveals the state of the world. The firm will only accept this lottery if the profit function induces risk proclivity,

[16] It is well known that when the profit function $\pi(q)$ is linear, the firm does not have incentives to invest in costly information acquisition. As Kartik et al. (2016) observe, "the martingale property of Bayesian updating implies that experts would gain nothing by acquiring information." Related conditions inducing acquisition of information that must be revealed appear in Proposition 2 in Dahm and Porteiro (2008), Proposition 5 in Dahm et al. (2009), and Remark 1 in Kamenica and Gentzkow (2011).

which is reflected in a (globally) convex shape. Indeed, $K_x < \mathbb{K}_t^{HR}$ can be rearranged to yield

$$\pi\left(\tilde{q}=\tilde{q}_{\emptyset}^{HR}\right) \;<\; \tilde{q}_{\emptyset}^{HR}\pi\left(\tilde{q}=1\right)+(1-\tilde{q}_{\emptyset}^{HR})\pi\left(\tilde{q}=0\right)-\frac{K_x}{1-\rho(1-q)}.$$

Since the profit function induces risk proclivity, the firm prefers the gamble on the right hand side to $\tilde{q}_{\emptyset}^{HR}$ with certainty, even if it has to pay (a sufficiently low) risk premium. In contrast, in the selective reporting equilibrium under laissez-faire (described in Proposition 1 with $\rho = 0$) the firm is expected to invest in information and to conceal negative test results. Rather than three there are only two outcomes, there is no gamble, and the risk attitude plays no role. As a result the cost threshold is always strictly positive.

Lastly, note that the degree of risk proclivity is also important when the profit function $\pi(q)$ has bounded average quality. To see this notice that under the assumptions of Example 1 the previous condition can be rewritten as

$$K_x < q\left(1-\left(\tilde{q}_{\emptyset}^{HR}\right)^{\lambda-1}\right).$$

Since the right hand side of this expression is increasing in the degree of risk proclivity λ, we have that the lower λ, the lower the incentives to search for information, until in the extreme Assumption 2 does not hold. Therefore, in this model incentives to search are 'overall continuous' in risk attitude.

3.4 Partially successful monitoring

There is also the possibility that monitoring is only successful to some extent. This is so because $\tilde{F}_0(q,\rho) < \tilde{F}_1(q,\rho)$ holds, as the minimum penalty $\tilde{F}$ is increasing in the belief and $\tilde{q}_{\emptyset}^{SR} < \tilde{q}_{\emptyset}^{HR}$, see again Figure 1. Consequently, there are some monitoring policies (F,ρ) for which a pure strategy equilibrium at the reporting stage does not exist. For these values it is neither an equilibrium to report negative information always honestly, nor is it an equilibrium to report it always selectively. A mixed-strategy, however, in which the seller sometimes conceals and sometimes reveals negative information allows to trade-off the frequency of being caught with the benefits from holding back negative information. More precisely, we investigate now when a mixed-reporting equilibrium exists in which the seller not only invests in information search and reveals positive information but in which there is also a certain probability that negative information is revealed honestly.

Consider a monitoring policy (F,ρ) with $F \in (\tilde{F}_0, \tilde{F}_1)$ and assume that $d\pi/dq > 0$ holds for the relevant values for q. The seller is indifferent between reporting honestly and concealing negative evidence if the buyer holds a belief $\hat{y}$ such that $F = \tilde{F}(q,\rho,\hat{y})$. Notice that, since $\tilde{F}$ is a continuous function of y, we can apply Bolzano's Theorem and conclude that there exists such a belief $\hat{y}$. In what follows suppose the seller chooses $\hat{y}$. For simplicity we denote the belief of the public when no information was revealed by $\tilde{q}_{\emptyset}$, rather than $\tilde{q}_{\emptyset}(\hat{y})$.

The seller's expected profits if he searches for information are hence,

$$\begin{aligned}\Pi_t(q) &= xq\pi(\tilde{q}=1) + ((1-x)q + (1-q)(1-\rho)(1-xy))\pi\left(\tilde{q}=\tilde{q}_\emptyset\right) \\ &\quad + (1-q)((\rho+(1-\rho)xy)\pi(\tilde{q}=0) - x\rho(1-y)F) - xK_x.\end{aligned} \tag{11}$$

This expression reduces to (7) and (10) for $y=0$ and $y=1$, respectively. Profits when the seller does not invest in a test are as in (8) with $\tilde{q}_\emptyset^{SR}$ replaced by $\tilde{q}_\emptyset$. The seller invests in the test if and only if

$$\Pi_t(q) - \Pi_{\neg t}(q) > 0 \Leftrightarrow K_x < \mathbb{K}_t^{Mix}(q,\rho,F),$$

where

$$\begin{aligned}\mathbb{K}_t^{Mix}(q,\rho,F) &\equiv q\left(\pi(\tilde{q}=1) - \pi\left(\tilde{q}=\tilde{q}_\emptyset\right)\right) \\ &\quad -(1-q)(1-\rho)y\left(\pi\left(\tilde{q}=\tilde{q}_\emptyset\right) - \pi(\tilde{q}=0)\right) - (1-q)\rho(1-y)F.\end{aligned}$$

Since in a mixed reporting equilibrium (4) must hold with equality, the previous expression simplifies to

$$\mathbb{K}_t^{Mix}(q,\rho,F) \equiv q\left(\pi(\tilde{q}=1) - \pi\left(\tilde{q}=\tilde{q}_\emptyset\right)\right) - (1-q)\rho F,$$

which generalizes $\mathbb{K}_t^{SR}$ because it uses the more general belief of the public $\tilde{q}_\emptyset$. On the other hand, for $y=1$ the first expression for $\mathbb{K}_t^{Mix}$ coincides with $\mathbb{K}_t^{HR}$, as $\tilde{q}_\emptyset = \tilde{q}_\emptyset^{HR}$. This implies that under our assumptions both conditions in the following Proposition 3 can be fulfilled, provided the test is cheap enough. We summarize with the following result.

Proposition 3 (Mixed-reporting equilibrium) *With the monitoring policy (F,ρ), there exists a PBE in which the seller invests in information search and reveals with probability $\hat{y}$ the information obtained honestly, provided*

- *the monitoring technology is intermediately powerful, that is, $F \in [\tilde{F}_0, \tilde{F}_1]$ and*
- *the test is cheap enough, that is, $K_x \leq \mathbb{K}_t^{Mix}(q,\rho,F)$.*

3.5 Monitoring in the absence of the firm's information search

As usual in this type of models there might exist both an informative and a non-informative equilibrium, depending on how expensive tests are. We now study the latter and show that when the test is expensive enough there exists a PBE in which the firm is correctly expected not to invest in tests.

If the public does not expect the firm to invest in information search, the distrust effect is not present. Since we assume that in this case the agency conducts a test, the beliefs of the buyer are then given by expression (9). Consequently, if the firm sticks to the equilibrium play and does not invest, expected profits are given by (8) with $\tilde{q}_\emptyset^{SR}$ replaced by $\tilde{q}_\emptyset^{HR}$.

Expected profits from search depend on whether the most profitable deviation involves selective or honest reporting. For $F > \tilde{F}_1$, the most profitable deviation involves honest reporting, while selective reporting is optimal for $F < \tilde{F}_1$. Consider $F > \tilde{F}_1$. The seller's expected profits if he searches for information are given in (10). Thus, the seller invests in the test if and only if

$$\Pi_t(q) - \Pi_{\neg t}(q) > 0 \Leftrightarrow K_x < \mathbb{K}_{\neg t}(q,\rho,F) = \mathbb{K}_t^{HR}(q,\rho).$$

Consider $F < \tilde{F}_1$. Profits if the seller searches for information are as in (7) with $\tilde{q}_\emptyset^{SR}$ replaced by $\tilde{q}_\emptyset^{HR}$. Consequently, the seller invests in the test if and only if

$$\Pi_t(q) - \Pi_{\neg t}(q) > 0 \Leftrightarrow K_x < \mathbb{K}_{\neg t}(q,\rho,F),$$

where

$$\mathbb{K}_{\neg t}(q,\rho,F) \equiv q\left(\pi\left(\tilde{q}=1\right) - \pi\left(\tilde{q}=\tilde{q}_\emptyset^{HR}\right)\right) - (1-q)\,\rho F.$$

This implies the following result.

Proposition 4 (Non-informative equilibrium) *With the monitoring policy (F,ρ), there exists a PBE in which the seller does not invest in information search, provided the test is expensive enough, that is,*

$$K_x \geq \mathbb{K}_{\neg t}(q,\rho,F) \equiv \begin{cases} \mathbb{K}_t^{HR}(q,\rho) & \text{if } F \geq \tilde{F}_1 \\ q\left(\pi\left(\tilde{q}=1\right) - \pi\left(\tilde{q}=\tilde{q}_\emptyset^{HR}\right)\right) - (1-q)\,\rho F & \text{if } F \leq \tilde{F}_1 \end{cases}. \tag{12}$$

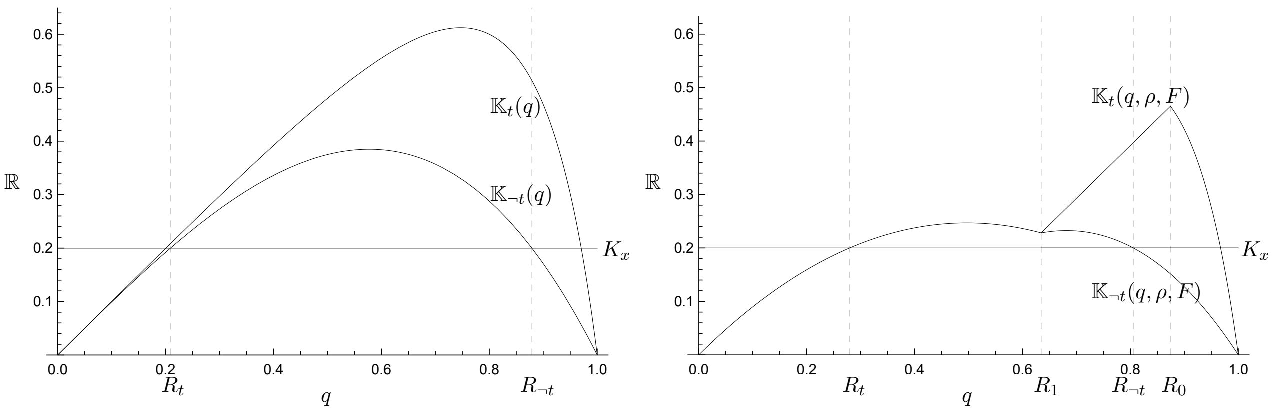
A non-informative equilibrium requires that the test is sufficiently expensive. Using (4) we see that the threshold $\mathbb{K}_{\neg t}(q,\rho,F)$ is continuous at $F = \tilde{F}_1$.

3.6 The trade-off between the quality and the quantity of information

From Lemma 1 we already know that for any quality q of the firm's product there exist monitoring policies (F,ρ) that can avoid selective reporting. An important question is how the introduction of a monitoring policy affects the firm's incentives for information search. In what follows we will refer to the replacement of selective reporting by honest reporting as an increase in the quality of information and to the firm's incentives for information search as the quantity of information. We will see that monitoring policies must balance a trade-off between quality and quantity of information.

We start by defining a counterpart to (12) based on Propositions 1-3 as follows

$$\mathbb{K}_t(q,\rho,F) \equiv \begin{cases} \mathbb{K}_t^{HR}(q,\rho) & \text{if } F \geq \tilde{F}_1 \\ \mathbb{K}_t^{Mix}(q,\rho,F) & \text{if } F \in [\tilde{F}_0,\tilde{F}_1] \\ \mathbb{K}_t^{SR}(q,\rho,F) & \text{if } F \leq \tilde{F}_0 \end{cases}. \tag{13}$$



(a) Laissez-faire

(b) Monitoring

Figure 2: Equilibrium under laissez-faire and monitoring in Example 1

Notice that this function is continuous at $F \in \{\tilde{F}_0, \tilde{F}_1\}$. Comparing (12) and (13), we see that $\mathbb{K}_{\neg t}(q,\rho,F) \leq \mathbb{K}_t(q,\rho,F)$. Thus for any combination of q and K_x, there is a PBE. This equilibrium is unique if the profit function is constant on the relevant interval for q or monitoring induces honest reporting, that is $F \geq \tilde{F}_1$. On the other hand, both an informative and a non-informative equilibrium may exist when monitoring is unsuccessful, that is $F < \tilde{F}_1$, and the profit function is strictly monotone, since $\tilde{q}_{\emptyset}^{SR} < \tilde{q}_{\emptyset}^{HR}$ holds.

Part (a) of Figure 2 illustrates the potential multiplicity of equilibrium. The monitoring policy is laissez-faire. We build on Example 1, assuming in addition that $\lambda = 2$ and $x = 0.75$. The horizontal axis indicates q, while the vertical axis measures the cost of information search K_x. For the level of cost $K_x = 0.2$, the firm does not invest in information search if (roughly) $q \in [0, 0.2]$ and $q \in [0.97, 1]$; if (roughly) $q \in [0.21, 0.88]$ it does invest in information; and the behaviour if $q \in [0.2, 0.21]$ and $q \in [0.88, 0.97]$ depends on the equilibrium selection. Our results do not depend on a particular equilibrium selection, but to fix ideas we focus on the non-informative equilibrium. We denote the interval for q for which the firm invests in information search by $[R_t, R_{\neg t}]$.[17] This is illustrated (based on the non-informative equilibrium) with the interval $[R_t, R_{\neg t}] = [0.21, 0.88]$.

It is important to see that a given a monitoring policy (F, ρ) induces the firm to report honestly, when q, the quality of the firm's product, is relatively low. This is so, because

[17]This notation implicitly assumes that the set of values for q for which the firm invests in information search is an interval. This is, for example, the case if the cost threshold functions are concave, as in part (a) of Figure 2. This seems intuitive, because it implies that the incentives to search for information are the higher, the less informative the initial belief q is. It is, however, not always true when other monitoring policies are considered. Our notation is not ambiguous, because Assumption 4 (below) assures that the set of values for q for which the firm invests in information search is an interval.

the minimum penalty $\tilde{F}$ is increasing in q. For instance, for low quality products (q close to zero), a small fine combined with a low probability of detection is already sufficient to induce honest reporting. Thus the requirement for a honest reporting equilibrium that the monitoring technology must be powerful enough (that is $F \geq \tilde{F}_1$) translates into the condition that q must be smaller than some threshold, say R_1. Notice also that, given a fine $F > 0$, as the probability of detection increases from zero, there is a 'continuous' switch from laissez-faire to monitoring, as R_1 increases from zero, and only firms with low quality products are induced to report honestly. Similarly, $F \leq \tilde{F}_0$ translates into the condition that q must be larger than some threshold, say R_0.[18] A firm with a product of high quality cannot be deterred from reporting selectively, as the stakes are too high. Moreover, $R_1 \leq R_0$ holds and the inequality is strict if the profit function is strictly monotone. In such a case monitoring is sometimes successful in inducing the firm to report honestly. This likelihood $\hat{y}$ of honest reporting is the lower, the higher the initial product quality q, because $\partial \tilde{q}_\emptyset / \partial y > 0$ implies that $\partial \tilde{F} / \partial y > 0$.

Consider now the monitoring policy in part (b) of Figure 2. The monitoring policy has $F = 16$ and $\rho = 0.025$. Again, a firm with either very low or very high product quality does not invest in information (and unlike in laissez-faire, the first interval does not depend on the equilibrium selection, as $F \geq \tilde{F}_1$). Focussing again on the non-informative equilibrium, a seller in $[0, R_t]$ and $[R_{\neg t}, 1]$ does not invest in information, while firms in $[R_t, R_1]$ invest and report honestly. All other firms invest and play a mixed reporting strategy. Since the interval $[R_0, R_{\neg t}]$ is empty, there is no possibility of an informative equilibrium in which all information is reported selectively. Comparing part (b) to part (a) of Figure 2, we see that inducing honest reporting through monitoring comes at a price, because the incentives for information search are reduced. More precisely, focussing on the non-informative equilibrium, the interval $[R_t, R_{\neg t}]$ shrinks from $[0.21, 0.88]$ to $[0.28, 0.81]$. But from the figures we also see that this conclusion does not depend on the selection of equilibrium.

Our next result shows that this trade-off is general and applies to any monitoring policy different from laissez-faire. Such a policy increases the quality of information, because $R_1 > 0$ and hence there exist values for the quality q of the firm's product for which honest reporting is induced. This effect is weak, however, since it is only effective if for such a q the firm invests in information, that is, $R_t < R_1$ holds. Concerning the deterrence effect on the quantity of information, the proof shows the following. First, for any quality q the cost interval $[0, \mathbb{K}_t(q)]$ in which search takes place in an informative equilibrium is larger under laissez-faire than under monitoring. Second, the interval $[\mathbb{K}_{\neg t}(q), \infty)$ in which search does not take place in a non-informative equilibrium is smaller under laissez-faire than under monitoring. Consequently, the incentives for information search are reduced through the

[18]The threshold R_1 is formally defined as the value $q \in [0, 1]$ such that $F = \tilde{F}_1(q, \rho)$. If such a value does not exist, because $F > \tilde{F}_1(q, \rho)$ for all $q \in [0, 1]$, then we set $R_1 = 1$. The threshold R_0 is defined analogously based on $F = \tilde{F}_0(q, \rho)$.

introduction of monitoring and this conclusion does not depend on equilibrium selection.[19]

Proposition 5 *Compared to laissez-faire, any monitoring policy* (F, ρ) *with* $F > 0$ *and* $\rho > 0$

- *increases the quality of information, that is,* $R_1 > 0$ *but*
- *reduces the quantity of information, that is, both* $\mathbb{K}_t$ *and* $\mathbb{K}_{\neg t}$ *decrease strictly.*

Proof: See Appendix A.3. *Q.E.D.*

This shows that optimal monitoring is determined by a trade-off. Stricter enforcement reduces the incentives for selective reporting but crowds out information search. It is worth pointing out that Assumption 2 captures the best case for monitoring. On one hand, we have seen that if it holds, the deterrence effect on information search is (continuously) reduced, as the firm exhibits more risk proclivity and the benefits from monitoring increase. On the other hand, if Assumption 2 does not hold, because the profit function induces risk aversion, then the deterrence effect of monitoring is even stronger. The firm cannot be induced to report honestly, as Lemma 3 does not hold. Laissez-faire outperforms all monitoring policies, whenever selectively reported information has a positive value for society. Therefore, in order to evaluate the deterrence effect on information search, it is crucial to develop a framework to assess the value of information for society. The next section analyses the welfare implications of monitoring in detail. It explores a favourable setting for monitoring, in the form of a stronger regularity condition on profits than Assumption 2.

4 Optimal monitoring

Proposition 5 establishes the existence of a trade-off in monitoring. Investigating the optimal monitoring policy requires to go beyond this (i) by looking at the effects of small changes in the monitoring policy and (ii) by evaluating the value of information to society. The next two subsections provide a framework for (i) and (ii), respectively, while the remaining subsections investigate the optimal monitoring policy.

4.1 A sharper trade-off of monitoring under additional assumptions

In order to look at the effects of small changes in the monitoring policy, we impose in this section two additional assumptions. We first strengthen Assumptions 1 and 2 by imposing more structure on the shape of the profit function $\pi(q)$. The following condition is related

[19]This reasoning only assumes that if there is multiplicity of equilibria before and after the introduction of monitoring, the same type of equilibrium is played. If multiplicity of equilibria only appears before or only after the introduction of monitoring, then the conclusion is unambiguous.

to the notions of concave and convex functions. The difference is that concavity and convexity are based on the monotonicity of the marginal slope, while the following assumption requires monotonicity of the average slope. Monotonicity of the average slope is a generalization of the notion of strict star-shapedness with respect to the origin.[20]

Assumption 3 (Strictly increasing average quality) *The profit function* $\pi(q)$ *has strictly increasing average quality if for each* $\zeta \in (0,1)$ *and all* q *we have that*

$$\zeta\,(\pi(q)-\pi(0)) > \pi(\zeta q)-\pi(0). \tag{14}$$

The inequality in (14) is equivalent to the requirement that the slope of the chord from $\pi(0)$, given by $(\pi(q)-\pi(0))/q$, must be strictly increasing. Thus Assumption 3 implies Assumption 2. It is weaker than strict convexity of $\pi(q)$, because $\pi(q)$ can have both convex and concave segments. The inequality in (14) is equivalent to

$$\frac{d\pi(q)}{dq} > \frac{\pi(q)-\pi(0)}{q}, \tag{15}$$

which implies that $\pi(q)$ is strictly increasing. Thus Assumption 3 implies also Assumption 1. Notice that Example 1 fulfils Assumption 3.

Evidence on pharmaceutical market performance seems to be consistent with Assumption 3. Grabowski et al. (2002) estimated a highly skewed distribution of returns (net present values) for new drug introductions. More precisely, the top decile of most successful new drugs accounted for a 52% of the total present value generated by all new drugs. This seems to suggest that market rewards higher quality at a highly increasing rate. Moreover, it seems that this pattern has not changed over time. A similar analysis conducted for the 1980–1990 period (Grabowski and Vernon, 1994) also found a highly skewed distribution of returns. In this study, the top two deciles accounted for more than a 70% of the total net present value.

In order to introduce the second assumption, notice that the cost threshold functions $\mathbb{K}_{\neg t}(q,\rho,F)$ and $\mathbb{K}_t(q,\rho,F)$ given in (12) and (13) are piecewise defined functions. This implies that at R_1, when the firm starts to mix between selective and honest reporting, the threshold functions might switch from decreasing to increasing. This opens the door to a situation in which the set of values for q for which the firm invests in information is not an interval. In fact this happens in the example displayed in part (b) of Figure 2 for values of

[20] Variations of this property have appeared in the economic literature (see e.g. Landsberger and Meilijon, 1990; Chateauneuf et al., 2004; Armantier and Treich, 2009). The exact relationship between (14) below and strict star-shapedness is as follows. Following Bullen (1998) we say that a continuous and non-negative function f with $f(0)=0$ is strictly star-shaped (with respect to the origin) if the slope of the chord from the origin, given by $f(q)/q$, is strictly increasing. This is equivalent to the requirement that for each $\zeta \in (0,1]$ and all q, we have $f(\zeta q) < \zeta f(q)$. Notice that in our model $\pi(0)$ is not required to be equal to zero. If, however, $\pi(0)=0$, then (14) reduces to $\pi(\zeta q) < \zeta\pi(q)$.

K_x a little higher than the cost threshold function $\mathbb{K}_t(q,\rho,F)$ at R_1. To simplify the exposition we impose now an assumption assuring that this cannot happen.[21] We note, however, that allowing for unconnected sets of values for q for which the firm invests in information would reduce the benefits of monitoring. The reason is that it would increase the deterrence effects of monitoring on information search, because monitoring would reduce the quantity of information not only via R_t and $R_{\neg t}$ but also through other thresholds. For analytical convenience the following assumption also rules out an interval in which the cost threshold function $\mathbb{K}_{\neg t}(q,\rho,F)$ is constant. We also note that we make the following assumption only in Subsection 4.3, but not in Subsection 4.4.

Assumption 4 (Quasiconcavity) *Given a monitoring policy (ρ,F) the cost threshold functions $\mathbb{K}_t(q,\rho,F)$ and $\mathbb{K}_{\neg t}(q,\rho,F)$ are quasiconcave and do not have a linear portion with slope* 0.

We are now in a position to determine the effects of small changes in the monitoring policy. The following result complements Proposition 5.

Proposition 6 *Suppose a monitoring policy (ρ,F) is in place and that there is a (non-degenerate) interval of values for q for which information search takes place.*
An increase in the fine F

- *reduces the quantity of information: $\partial R_t/\partial F \geq 0$ and $\partial R_{\neg t}/\partial F \leq 0$*
- *increases the quality of information: for $y \in \{0,1\}$ we have $\partial R_y/\partial F > 0$ and for $y \in (0,1)$ we have $\partial y/\partial F > 0$.*

An increase in the probability of detection ρ

- *reduces the quantity of information:*[22] *$\partial R_t/\partial\rho > 0$ and $\partial R_{\neg t}/\partial\rho < 0$*
- *affects the quality of information: both $\partial R_y/\partial\rho > 0$ for $y \in \{0,1\}$ and $\partial y/\partial\rho > 0$ for $y \in (0,1)$ if and only if ρ and F are substitutes.*

[21] The profit function $\pi(q)$ underlying part (b) of Figure 2 is a simple quadratic function (Example 1 with $\lambda = 2$). It seems hence that the fact that (12) and (13) are piecewise defined functions is responsible for the switch from decreasing to increasing, rather than the shape of $\pi(q)$. Consequently, it does not seem possible to impose a condition on the profit function $\pi(q)$ preventing this from happening.

[22] For simplicity of the exposition the following statement is slightly incomplete. When in case of multiplicity the non-informative equilibrium is played or the monitoring instruments are substitutes, then the statement is correct. When, however, the non-informative equilibrium is not played and the monitoring instruments are complements, then $\partial R_t/\partial\rho > 0$ and $\partial R_{\neg t}/\partial\rho < 0$ require that the (direct) positive confidence effect of monitoring on the public's belief is stronger than the indirect negative effect via the reduced frequency of honest reporting. The precise condition can be found in the proof of Proposition 6 in Appendix A.4.

Proof: See Appendix A.4. *Q.E.D.*

Proposition 6 adds to Proposition 5 by formalizing the trade-off in monitoring in a different way. On one hand, the quantity of information is reduced, because the interval between R_t and $R_{\neg t}$ shrinks. On the other hand, however, it increases the quality of information, because it induces the firm to replace selective through honest reporting for some values of q. Increasing the fine has similar effects than increasing the probability of detection. There are, however, two subtle differences. First, when a firm with quality q at the threshold R_t or firms at both thresholds R_t and $R_{\neg t}$ report honestly, then the incentives for information search for such firms are not deterred from increasing the fine but there is a deterrence effect on information search if the probability of detection is increased. Second, increasing the fine is more reliable in implementing stricter monitoring than increasing the probability of detection, unless the possibility of a complementary relationship between the instruments can be excluded. This, however, is not the case. To see this consider again Example 1. We know that it fulfils Assumption 3 and in the discussion of Lemma 2 we saw that for this example the possibility of a complementary relationship between the probability of detection and the fine cannot be excluded. This implies that given parameter values ρ and x the firm might be characterized by a value for q such that an increase in the probability of detection ρ might induce this seller to report more often selectively.

4.2 Social welfare

Analysing the welfare consequences in our setting requires defining the value of information for society. We postulate a social value function $V(q|v)$ that captures the willingness to pay of society for the market interaction between the firm and the public. More precisely, $V(q|v)$ indicates the willingness to pay of society for exchange of the good when probability q is assigned to state 1 conditional on v being the true state. For instance, $V(q = 1|v = 1)$ measures this value when the product is of high quality and the public knows this. We do not impose any specific functional form for $V(q|v)$ but we assume that the willingness to pay is the higher, the more correct the belief is. That is, $q' \leq q''$ implies that

$$V(q''|v=1) \geq V(q'|v=1) \quad \text{and} \quad V(q''|v=0) \leq V(q'|v=0). \tag{16}$$

Notice that such a monotonicity imposes a minimal structure in the sense that

$$H(\tilde{q}) \equiv V(q=0|v=0) - V(\tilde{q}>0|v=0)$$

measures the social harm from selective reporting. If the monotonicity in (16) is violated, then selective reporting might be socially beneficial.[23]

[23] Such a monotonicity can be derived from a micro-foundation based on a monopoly market for a medical treatment. Details are available upon request (and included for the convenience of the referees in Appendix B, which is not intended for publication).

The following expressions describe social welfare in different situations. Welfare depends on the monitoring policy in place, on whether the firm invests in information and, if so, how this information is reported. The following expressions are net of the cost of information xK_x and of the cost of monitoring ρK_ρ, which will both be taken into account later. If no investment in information takes place we have

$$\mathbb{W}_{\neg t}(q,\rho) \equiv qV(\tilde{q}=\tilde{q}_\emptyset^{HR}|v=1)+(1-q)(1-\rho)V(\tilde{q}=\tilde{q}_\emptyset^{HR}|v=0)+(1-q)\rho V(0|v=0). \quad (17)$$

Since the agency always conducts tests, with probability $(1-q)\rho$ it is revealed that the the (expected) quality of the good is low and the state of the world is $v=0$. When the agency's test is unsuccessful, the social values of information assign different values depending on the state of the world.

On the other hand, when investment in information takes place, the three reporting strategies must be distinguished. Successful monitoring obtains

$$\mathbb{W}_t^{HR}(q,\rho) \equiv \begin{cases} xqV(\tilde{q}=1|v=1)+(1-x)qV(\tilde{q}=\tilde{q}_\emptyset^{HR}|v=1) \\ +(1-q)(1-x)(1-\rho)V(\tilde{q}=\tilde{q}_\emptyset^{HR}|v=0) \\ +(1-q)(\rho+x(1-\rho))V(\tilde{q}=0|v=0) \end{cases}, \quad (18)$$

while selective reporting yields

$$\mathbb{W}_t^{SR}(q,\rho) \equiv \begin{cases} xqV(\tilde{q}=1|v=1)+(1-x)qV(\tilde{q}=\tilde{q}_\emptyset^{SR}|v=1) \\ +(1-q)(1-\rho)V(\tilde{q}=\tilde{q}_\emptyset^{SR}|v=0)+(1-q)\rho V(\tilde{q}=0|v=0) \end{cases}. \quad (19)$$

Lastly, in the mixed-strategy equilibrium welfare is given by

$$\begin{aligned} \mathbb{W}_t^{Mix}(q,\rho,F) &\equiv y(F,\rho)W_t^{HR}(q,\rho)+(1-y(F,\rho))W_t^{SR}(q,\rho) \quad (20) \\ &= W_t^{SR}(q,\rho)+y(F,\rho)x(1-q)(1-\rho)H(\tilde{q}_\emptyset), \end{aligned}$$

where again $\tilde{q}_\emptyset$ denotes the belief of the public that makes the seller indifferent between reporting honestly or selectively. We write $y(F,\rho)$ in order to underline that the equilibrium frequency of honest reporting is increasing in F and depends on ρ. Given that a test is carried out, we will also use $\mathbb{W}_t(q,\rho,F)$ in order to indicate welfare from monitoring without specifying which reporting strategy applies. More precisely,

$$\mathbb{W}_t(q,\rho,F) \equiv \begin{cases} \mathbb{W}_t^{HR}(q,\rho) & \text{if } q \leq R_1(F,\rho) \\ \mathbb{W}_t^{Mix}(q,\rho,F) & \text{if } R_1(F,\rho) < q < R_0(F,\rho) \\ \mathbb{W}_t^{SR}(q,\rho) & \text{if } q \geq R_0(F,\rho) \end{cases}.$$

Notice that this function is continuous, since at $R_1(F,\rho)$ the frequency of honest reporting is one, while it is zero at $R_0(F,\rho)$.

4.3 On the optimal monitoring policy

In this subsection we analyse monitoring policies ex-ante, before the quality q of the firm's product is known. At that point in time it is only known that q is distributed following the distribution function $F(q)$, with (continuous) probability density $f(q)$ such that $f > 0$ for all q. We are now in a position to introduce the objective function of the planner. The planner chooses a monitoring policy (ρ, F) in order to maximize

$$\begin{aligned}\Delta(F,\rho) \;&=\; \int_0^{R_t} \mathbb{W}_{\neg t}(q,\rho) f(q) dq + \int_{R_t}^{R_{\neg t}} [\mathbb{W}_t(q,\rho,F) - xK_x] f(q) dq \\ &+ \int_{R_{\neg t}}^{1} \mathbb{W}_{\neg t}(q,\rho) f(q) dq - \rho K_\rho .\end{aligned} \tag{21}$$

It turns out that in general the planner's maximization problem is not well behaved. Consequently, analysis of first-order conditions might not identify the optimal monitoring policy. For this reason we analyse the effects of small changes in the monitoring policy.

We focus first on the welfare effects of adjusting the fine, assuming that the probability of detection is positive. It turns out that increasing the fine has only two effects.

Proposition 7 *The effect of a marginal increase of the fine F on social welfare can be decomposed in two effects:*

(1) The frequency of honest reporting increases, which is socially valuable:

$$\int_{\max\{R_t,\min\{R_1,R_{\neg t}\}\}}^{\max\{R_t,\min\{R_0,R_{\neg t}\}\}} \frac{\partial y(F,\rho)}{\partial F} x(1-q)(1-\rho) H(\tilde{q}_\emptyset) f(q) dq \geq 0.$$

(2) Investment in information is deterred, but the effect on welfare is ambiguous:

$$\begin{aligned}&\frac{\partial R_t}{\partial F}[\mathbb{W}_{\neg t}(R_t,\rho) - \mathbb{W}_t(R_t,\rho,F) + xK_x] f(R_t) \\ &-\frac{\partial R_{\neg t}}{\partial F}[\mathbb{W}_{\neg t}(R_{\neg t},\rho) - \mathbb{W}_t(R_{\neg t},\rho,F) + xK_x] f(R_{\neg t}) \lesseqqgtr 0.\end{aligned}$$

Proof: See Appendix A.5. *Q.E.D.*

Increasing the fine induces more honest reporting. The model captures this through the increased frequency with which in the mixed-reporting equilibrium information is revealed honestly. This is formalized in part (1) of Proposition 7. Using the language of the law enforcement literature, some social harm $H(\tilde{q}_\emptyset)$ from selective reporting is avoided and this occurs with probability $x(1-q)(1-\rho)$. Thus this effect on welfare is positive. In addition, increasing the fine deters investment in information. This is formalized in part (2) of Proposition 7. As shown in Proposition 6, this deterrence effect on information arises, because the interval between the thresholds R_t and $R_{\neg t}$ declines when the fine increases. The effect of deterrence on welfare is negative if and only if investment in information is desirable at the margin.

Proposition 7 decomposes the effect of an increase of the fine on social welfare in two effects. It is important to note that these effects can be different from zero or not (details can be found in the proof). For instance, when the monitoring policy is such that information is always honestly reported whenever investment in information takes place (that is, $R_t < R_{\neg t} < R_1$), then both effects vanish. This is so because in equilibrium the fine is never imposed and does not affect behaviour.

It turns out that the welfare effects of adjusting the probability of detection are more complex than those following an adjustment of the fine. There are two effects, which–although they are analogous to those in Proposition 9–are now both ambiguous. In addition, two new effects appear. In particular, the confidence effect induces consumers to become more optimistic when neither the firm nor the agency reveal the state of the world. This is desirable when the state of nature is good and harmful otherwise. Also, monitoring in itself–similar to the search technology of the firm–might reveal information and reduce social harm but has a marginal cost K_ρ. Unfortunately, given the complexity of the effects and the fact that without further assumptions it is impossible to determine the overall effect, no general conclusion can be drawn.[24]

The following intuition, however, suggests that Becker's conclusion–that the probability of detection should be small and that the fine should be the largest possible fine–should also hold in our model. If the deterrence level of a monitoring policy is higher than needed in order to induce honest reporting, then a small decrease in the probability of detection decreases the deterrence effect, while a decrease in the fine does not affect the incentives to invest in information. This intuition is true if the firm can be induced to report honestly for all realizations of the (expected) quality q. But it leaves out the case in which this is not possible and some selective reporting occurs for some realizations of q.

When information is (at least sometimes) selectively reported (that is, $R_1 < R_{\neg t}$), then the optimal fine should not always be the largest possible fine. Key for this result is that the deterrence effect in part (2) of Proposition 7 is ambiguous. If selectively reported information is socially undesirable, then both effects of raising the fine are aligned and social welfare is increased. But if selectively reported information is socially desirable, then the deterrence effect in part (2) of Proposition 7 is negative and both effects are opposed. In such a case society values the information provided by the firm, although it is (at least sometimes) selectively reported. In other words, society prefers a situation in which the firm conducts a test and reports the result selectively to an alternative policy with a higher

[24] Details are available upon request (and included for the convenience of the referees in Appendix B, which is not intended for publication).

fine, because raising the fine deters the firm's investment in information further. This suggests that the fine should be at its highest level when investment in information that is not always reported honestly is socially undesirable. Conversely, the optimal fine might not be at its highest level when society values the firm's information sufficiently.

In order to show that there are (reasonable) situations in which the optimal fine is not the largest possible fine, we build in the next subsection on Brekke and Kuhn (2006) and provide a micro-founded example of a monopoly market for a medical treatment in which this is the case.

4.4 Optimal monitoring in a simple example

In this subsection we consider a simple example in order to calculate explicitly social welfare for different monitoring policies. We continue Example 1 assuming that $\lambda = 2$. The same micro-foundations based on a monopoly market for a medical treatment that give rise to this example imply the following functional forms for the value of information for society[25]

$$V(q|\nu = 1) = 2q - \frac{1}{2}q^2 \quad \text{and} \quad V(q|\nu = 0) = -\frac{1}{2}q^2. \tag{22}$$

These functions capture the idea that the public decides whether or not to buy the firm's product based on the information it has, which in our model is captured by the (possibly updated) belief q. This is an ex-ante perspective, before the state of the world is revealed. For example, if the public consists of heterogeneous consumers and q is intermediate, then only some consumers might buy the good. If it turns out that the firm's product is of high quality and benefits all types of consumers, then some consumers will have been too pessimistic. Similarly, if it turns out that the firm's product is of low quality and harms all types of consumers, then some consumers will have been too optimistic. Thus, the more correct the belief, the higher the values in (22), as required by (16).

We suppose that it is known that $q = 3/4$. This simplifies the calculations and helps to make the intuitions more visible. We consider two scenarios concerning the efficiency of the firm's search technology. In both scenarios the quality of the firm's test is $x = 3/4$ but the cost of the test varies: $K_x \in \{0.1, 0.19\}$. Lastly, we assume that the cost of monitoring is $K_\rho = 0.3$.[26]

Consider first the laissez-faire benchmark. Since $\mathbb{K}_{\neg t}(q, \rho = 0) = 0.3 < \mathbb{K}_t^{SR}(q, \rho = 0)$, the firm invests in the test for both efficiency levels of its search technology and there is

[25] Details on this and on the calculations on which the following summary is based are available upon request (and included for the convenience of the referees in Appendix B, which is not intended for publication).

[26] A higher value for K_ρ makes monitoring more costly and hence less attractive than in this example. A lower value than 0.28125 makes monitoring such an efficient search technology that the planner prefers to use it even if it has no monitoring value, because the firm does not invest in a test. This follows from evaluating $\mathbb{W}_{\neg t}(q = 3/4, \rho)$ for $\rho = 0$ and $\rho = 1$, which yields 0.84375 and 1.125, respectively. Thus, $0.84375 > 1.125 - K_\rho$ must hold, which is equivalent to $K_\rho > 0.28125$.

ρ	F	type of equilibrium	social welfare
0	any	selective reporting	0.91012
0.025	23	honest reporting	0.97351
0.05	11	honest reporting	0.96736

Table 1: High efficiency of the firm's search technology

no multiplicity of equilibrium. Social welfare is the difference between $\mathbb{W}_t^{SR}(q, \rho = 0) = 0.98512$ and $0.75K_x$, which is reported in the second row of Tables 1 and 2, respectively.[27]

Consider monitoring and suppose first that the efficiency of the firm's search technology is high, that is $K_x = 0.1$. We start by investigating the incentives of the firm to invest in a test, assuming that monitoring is successful in inducing honest reporting. From the proof of Proposition 6 it follows that $\mathbb{K}_t^{HR}(q, \rho)$ is decreasing in ρ. Evaluating the extreme case $\mathbb{K}_t^{HR}(q, \rho = 0) = 0.1875$ suggests that for small detection probabilities there exists an equilibrium in which the firm invests in a test and reports the result honestly. For example, evaluating (4) we see that for a detection probability of 0.025 any fine larger than 23, or for $\rho = 0.05$ any fine larger than 11, is sufficient to induce honest reporting. Indeed, for these detection probabilities there are incentives to invest in a test, as $\mathbb{K}_t^{HR}(q, \rho = 0.025) = 0.18396$ and $\mathbb{K}_t^{HR}(q, \rho = 0.05) = 0.18038$. Notice that there is no multiplicity of equilibria, as (12) and (13) coincide. Social welfare for the two policies $(F, \rho) = (23, 0.025)$ and $(F, \rho) = (11, 0.05)$ is reported in Table 1. Notice that both policies improve over laissez-faire. Moreover, it is beneficial to decrease the detection probability as much as possible, say to $\rho = 0.025$, because it saves resource costs. Thus we see that when the efficiency of the firm's search technology is high, there is no reason to limit the fine, as in Becker's setting.

Now consider monitoring again but suppose that the efficiency of the firm's search technology is low, that is $K_x = 0.19$. The before mentioned fact that $\mathbb{K}_t^{HR}(q, \rho = 0) = 0.1875$ implies that the deterrence effect is now so strong that a honest reporting equilibrium does not exist. There exists no monitoring policy that induces an equilibrium in which the firm invests in a test and always reports the result honestly. A less stringent monitoring policy, however, might induce a mixed-reporting equilibrium that potentially could improve over laissez-faire. Indeed, if the probability of detection remains at $\rho = 0.025$ but the fine is reduced to 21, then there exists such a mixed-reporting equilibrium and there is no multiplicity of equilibria. Row 3 of Table 2 shows that the probability y of reporting honestly is close to one and reports social welfare, which again is higher than in laissez-faire. This

[27]Notice that the case in which the efficiency of the firm's search technology is low is an example for a situation in which the firm has an incentive to conduct tests even though this is inefficient (see Jovanovic, 1982; Milgrom, 2008). This is so because it is welfare improving to prohibit tests, as $\mathbb{W}_{\neg t}(q, \rho = 0) > \mathbb{W}_t^{SR}(q, \rho = 0) - 0.75 \cdot 0.19$.

ρ	F	type of equilibrium	social welfare
0	any	selective reporting	0.84262
0.025	21	mixed, $y = 0.96126$	0.90454
0.025	23	non-informative	0.84933
0.0263	21	non-informative	0.84933

Table 2: Low efficiency of the firm's search technology

reasoning does not imply that $(F, \rho) = (21, 0.025)$ is the optimal monitoring policy, but we can infer that the optimal fine is not the maximal fine. This is so, because we already know that a sufficiently large fine (say $F \geq 23$) induces the firm to report honestly with probability one. This is not an equilibrium, because the firm prefers not to invest in the test. Indeed, a unique non-informative equilibrium exists in which social welfare is low, see row 4. Similarly, since one can show that in this example monitoring instruments are substitutes, a sufficiently large probability of detection induces honest reporting and a similar deterrence effect applies. Again, a unique non-informative equilibrium exists with low social welfare. This is illustrated in row 5 for the best case in which the increase in the probability of detection from $\rho = 0.025$ to $\rho' = 0.0263$ is costless.

Summarizing, this simple example illustrates the trade-off in monitoring between the quality and quantity of information provided. It suggests that when the deterrence effect on information search is not an important concern, because the efficiency of the firm's search technology is high, then choosing the largest possible fine might be beneficial. The reason is that this might allow to reduce the probability of detection, which saves resource costs and reduces the deterrence effect on information. Conversely, when the efficiency of the firm's search technology is low and the deterrence effect on information search is an important concern, then the optimal monitoring policy might not be too stringent. This is so, because the deterrence effect of stricter monitoring on search for information is stronger than the positive effect on the quality of information. Consequently, the optimal monitoring policy might not necessarily eliminate selective reporting entirely. In such a situation increasing the probability of detection beyond the optimal monitoring policy might not be beneficial, even if doing so implies no additional costs, and the optimal fine might not be the largest possible fine. We remark that this example does not rely on equilibrium selection, as there is no multiplicity of equilibria. In addition, since the quality of the firm's product is known, the example does not rely on Assumption 4. Lastly, the example is not built on a complementary relationship of the monitoring instruments.

5 Conclusions

The present paper contributed to deepen our understanding of the effects of enforcement of mandatory disclosure regulations. In our model a firm can invest in information about product quality and decide whether or not to disclose the findings. If the firm holds back information, it might be detected and fined. While we showed that such a monitoring policy can improve over laissez-faire, we also saw that optimal monitoring is determined by a trade-off between the quality and quantity of information provided. Monitoring deters not only selective reporting but also information search in the first place. This has several implications. First, optimal monitoring might tolerate some selective reporting and thus it might be optimal to enforce mandatory disclosure rules imperfectly. Second, unlike in Becker's (1968) law enforcement setting, the optimal fine might not be maximal. And lastly, the optimal detection probability might be low, as in Becker's setting–but for a different reason. Indeed, we gave an example for an optimal monitoring policy in which increasing the probability of detection decreases social welfare, even when the additional resources needed are costless.

Our persuasion game with an endogenously informed sender might also shed some light on other situations. Consider the interaction between a firm, a rating agency and investors (see Faure-Grimaud et al., 2009). If the firm asks for a corporate governance rating and has the ownership of it, it might report the rating selectively. Mandatory disclosure, however, deters the firm's incentives to pay the agency for a rating. When there is no obligation to ask for a rating and ratings are valued by investors, our analysis implies that a similar trade-off between the quality and quantity of information to the one in the present paper arises. There might hence be situations in which it is optimal to enforce mandatory disclosure rules for ratings imperfectly.

Our analysis is based on a specific monitoring technology. Of course, depending on the context, other formulations might be reasonable. Relaxing our benchmark assumption that the agency only searches for evidence about quality problems would weaken the confidence effect. In the extreme, when the agency's search technology is equally effective in discovering both states of the world, the confidence effect disappears entirely. While the possibility of a complementary relationship between the two monitoring instruments depends on the existence of the confidence effect, the main forces of our model do not. Even without the confidence effect, eliminating selective reporting makes the public more optimistic. This increases the opportunity costs of information search and deters investment in information. The trade-off between the quality and quantity of information formalized in Proposition 5 hence is likely to be robust.

The aim of the present paper was to provide a formal framework to study enforcement and our results uncover a trade-off that optimal enforcement must balance. We do not offer an empirical assessment of this trade-off and hence our analysis does not imply that the AllTrials Campaign cited in the Introduction is wrong to lobby for strict enforcement. Rather our analysis offers a formal framework to think rigorously about enforcement. We highlighted an implication of enforcement that is currently not taken into account and showed that this side-effect of enforcement can potentially be important enough to imply that mandatory disclosure rules should be imperfectly enforced. We also saw that in other circumstances strict enforcement might be optimal. Future research should quantify the deterrence effect and the welfare consequences through empirical work for specific mandatory disclosure rules, like clinical trials.

There are, however, reasons to believe that the incentives of firms to search for information are an important concern. Indeed, approval of pharmaceutical products often requires the firm to conduct so-called post-marketing clinical trials and it is documented that firms fail to comply with this requirement (US Food and Drug Administration, 2011; Fain et al, 2013). This suggests that we can interpret our model as investigating a benchmark situation in which post-marketing trials are either not required or not enforced. A situation in which firms honour their post-marketing trial commitments has the potential to avoid the trade-off between the quality and quantity of information. But the fact that firms fail to comply points to another enforcement problem when obliging firms to search for information. Thus another interesting avenue for future research considers two-dimensional monitoring policies with an agency that monitors not only selective reporting but also whether investment in information took place in the first place. It appears that the deterrence effect on information search is reduced, as a fine might be imposed when the firm is found out not to have searched. But we leave a rigorous analysis of the overall effects of monitoring and of the optimal two-dimensional monitoring policy for future research.

References

[1] Anderson, M. L.; K. Chiswell; E. D. Peterson; A. Tasneem; J. Topping; and R. M. Califf (2015), "Compliance with results reporting at ClinicalTrials.gov," *New England Journal of Medicine* **372**, p. 1031–1039.

[2] Andreoni, J. (1991), "Reasonable Doubt and the Optimal Magnitude of Fines: Should the Penalty Fit the Crime?," *Rand Journal of Economics* **22** (3), p. 385–395.

[3] Antman, E. M.; J. S. Bennett; A. Daugherty; C. Furberg; H. Roberts; K. A. Taubert (2007), "Use of nonsteroidal antiinflammatory drugs: an update for clinicians: a scientific statement from the American Heart Association," *Circulation* **115** (12), p. 1634–1642.

[4] Armantier, O. and N. Treich (2009), "Star-shaped probability weighting functions and overbidding in first-price auctions," *Economics Letters* **104** (2), p. 83–85.

[5] Azoulay, P. (2002), "Do pharmaceutical sales respond to scientific evidence?" *Journal of Economics and Management Strategy* **11** (4), p. 551–594.

[6] Becker, G.S. (1968), "Crime and punishment: an economic approach," *Journal of Political Economy* **76**, p. 169–217.

[7] Bebchuk, L. A. and L. Kaplow (1992), "Optimal sanctions when individuals are imperfectly informed about the probability of apprehension," *Journal of Legal Studies* **21**, p. 365–370.

[8] Berenson, A. (2006) "A 2nd Loss For Merck Over Vioxx," New York Times, April 6.

[9] Brekke, K.R. and M. Kuhn (2006), "Direct to consumer advertising in pharmaceutical markets," *Journal of Health Economics* **25**, p. 102–130.

[10] Board, O. (2009), "Competition and Disclosure," *Journal of Industrial Economics* **57** (1), p. 197–213.

[11] Bullen, P.S. (1998), "A Dictionary of Inequalities," Pitman Monographs and Surveys in Pure and Applied Mathematics n. 97, CRC Press LLC.

[12] Chan, A.W.; F. Song; A. Vickers; et al. (2014), "Increasing value and reducing waste: addressing inaccessible research," *Lancet* **383**, p. 257–266.

[13] Chateauneuf, A.; M. Cohen; I. Meilijson (2004), "Four notions of mean preserving increase in risk, risk attitudes and applications to the rank-dependent expected utility model," *Journal of Mathematical Economics* **40**, p. 547–571.

[14] Corts, K.S. (2014), "Finite Optimal Penalties for False Advertising," *Journal of Industrial Economics* **62**(4), p. 661–681.

[15] Couzin, J. (2004), "Drug research: legislators propose a registry to track clinical trials from start to finish," *Science* **305**, p. 1695.

[16] Dahm, M.; P. González; and N. Porteiro (2009), "Trials, Tricks and Transparency: How Disclosure Rules Affect Clinical Knowledge," *Journal of Health Economics* **28**, p. 1141–1153.

[17] Dahm, M. and N. Porteiro (2008), "Informational Lobbying under the Shadow of Political Pressure," *Social Choice and Welfare*, **30** (4), p. 531-559.

[18] De Angelis, C.; J.M. Drazen; F. A. Frizelle et al. (2004), "Clinical Trial Registration: A Statement from the International Committee of Medical Journal Editors," *Annals of Internal Medicine*, **141**, p. 477–8.

[19] Dranove, D. and G.Z. Jin (2010), "Quality Disclosure and Certification: Theory and Practice," *Journal of Economic Literature* **48** (4), p. 935–963.

[20] Dye, R.A. (1985), "Disclosure of Nonproprietary Information," *Journal of Accounting Research* **23**, p. 123–145.

[21] Di Tillio, A.; M. Ottaviani; and P.N. Sorensen (2015), "Persuasion Bias in Science: Can Economics Help?," working paper.

[22] Editorial (2013) "What a Company Knew About Its Metal Hips," New York Times, February 10.

[23] Fain, K.; M. Daubresse; and G. Alexander (2013), "The Food and Drug Administration Amendments Act and Postmarketing Commitments" *Journal of the American Medical Association* **310** (2), p. 202–204.

[24] Farell, J. (1986), "Voluntary Disclosure: Robustness of the Unraveling Result, and Comments on Its Importance," in Ronald E. Grieson, ed., *Antitrust and Regulation*, Lexington, MA: Lexington Books, p. 91–103.

[25] Faure-Grimaud, A.; E. Peyrache; and L. Quesada (2009), "The ownership of ratings," *Rand Journal of Economics* **40** (2), p. 234–257.

[26] Feess, E. and A. Wohlschlegel (2009), "Why higher punishment may reduce deterrence," *Economics Letters* **104**, p. 69–71.

[27] Fung, A.; M. Graham; D. Weil (2007), "Full Disclosure: The Perils and Promise of Transparency," New York, Cambridge University Press.

[28] Garoupa, N. (1997), "The theory of optimal law enforcement," *Journal of Economic Surveys* **11** (3), p. 267–295.

[29] Goldacre, B. (2013) "Health Care's Trick Coin," New York Times, February 1.

[30] Gopal, A. D.; N. R. Desai; T. Tse; and J. S. Ross (2015), "Reporting of Noninferiority Trials in ClinicalTrials.gov and Corresponding Publications," *Journal of the American Medical Association* **313** (11), p. 1163–1165.

[31] Grabowski, H. and J. Vernon (1994), "Returns to R&D on new drug introductions in the 1980s," *Journal of Health Economics* **13**, p. 383–406.

[32] Grabowski, H.; J. Vernon; and J.A. DiMasi (2002), "Returns on research and development for 1990s. New Drug Introductions," *Pharmacoeconomics* **20** (3), p. 11–29.

[33] Grossman, S.J. (1981), "The Informational Role of Warranties and Private Disclosure about Product Quality," *Journal of Law and Economics* **24** (3), p. 461–483.

[34] Harrington, J.E. (2014), "Penalties and the Deterrence of Unlawful Collusion," *Economic Letters* **124**, p. 33–36.

[35] Henry, E. (2009), "Strategic disclosure of research results: The cost of proving your honesty," *Economic Journal* **119**, p. 1036–1064.

[36] Hudson, K. L. and F. S. Collins (2015), "Sharing and reporting the results of clinical trials," *Journal of the American Medical Association* **313**, p. 355–356.

[37] Jovanovic, B. (1982), "Truthful Disclosure of Information," *Bell Journal of Economics* **13**(1), p. 36–44.

[38] Kamenica, E. and M. Gentzkow (2011), "Bayesian Persuasion," *American Economic Review* **101**, p. 2590–2615.

[39] Kaplow, L. (1990), "A note on the optimal use of nonmonetary sanctions," *Journal of Public Economics* **42**, p. 245–247.

[40] Kartik, N.; F. X. Lee; and W. Suen (2015), "A Theorem on Bayesian Updating and Applications to Signaling Games," working paper.

[41] Kartik, N.; F. X. Lee; and W. Suen (2016), "Investment in Concealable Information by Biased Experts," *RAND Journal of Economics*, forthcoming.

[42] Krumholz, H. M.; J. S. Ross; A. H. Presler; and D. S. Egilman (2007), "What have we learnt from Vioxx? ," *British Medical Journal* **334**, p.120–123.

[43] Landsberger, M. and I. Meilijson (1990), "Lotteries, insurance and star-shaped utility functions," *Journal of Economic Theory* **52**, p. 1–17.

[44] Malik, A. (1990), "Avoidance, screening and optimum enforcement," *RAND Journal of Economics* **21** (3), p. 341–353.

[45] Matthews, S. and A. Postlewaite (1985), "Quality Testing and Disclosure," *Rand Journal of Economics* **16** (3), p. 328–340.

[46] Milgrom, P. (1981), "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics* **12**, p. 380–391.

[47] Milgrom, P. (2008), "What the seller won't tell you: persuation and disclosure in markets," *Journal of Economic Perspectives* **12**, p. 380–391.

[48] Milgrom, P. and J. Roberts (1986), "Relying on the Information of Interested Parties," *Rand Journal of Economics* **17** (1), p. 18–32.

[49] Mookherjee, D. and I. P. L. Png (1992), "Monitoring vis-a-vis investigation in enforcement of law," *American Economic Review* **82**, p. 556–565.

[50] Polinsky, A. M. and S. Shavell (1979), "The optimal trade-off between the probability and magnitud of fines," *American Economic Review* **69**, p. 880–891.

[51] Polinsky, A. M. and S. Shavell (1991), "A note on optimal fines when wealth varies among individuals," *American Economic Review* **81**, p. 618–621.

[52] Polinsky, A. M. and S. Shavell (2000), "The economic theory of public enforcement of law," *Journal of Economic Literature* **38**, p. 45–76.

[53] Polinsky, A. M. and S. Shavell (2012), "Mandatory Versus Voluntary Disclosure of Product Risks," *Journal of Law, Economic & Organization* **28** (2), 360–379.

[54] Prayle, A. P.; M. N. Hurley; and A. R. Smyth (2012), "Compliance with mandatory reporting of clinical trial results on ClinicalTrials.gov: cross sectional study," *The BMJ* **344**.

[55] Seidmann, D.J. and E. Winter (1997), "Strategic information transmission with verifiable messages," *Econometrica* **65** (1), p. 163–169.

[56] Shavell, S. (1994), "Acquisition and Disclosure of Information Prior to Sale," *RAND Journal of Economics* **25** (1), p. 20–36.

[57] Shin, H. S. (1994), "The burden of proof in a game of persuasion," *Journal of Economic Theory* **64**, p. 253–264.

[58] US Food and Drug Administration (2011), "Guidance for Industry: Postmarketing Studies and Clinical Trials — Implementation of §505(o)(3) of the Federal Food, Drug, and Cosmetic Act," April, available at `http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM172001.pdf`, accessed on 10/11/2016.

[59] Wood, A.J.J. (2009) "Progress and Deficiencies in the Registration of Clinical Trials," Special Report, *New England Journal of Medicine* **360**, p. 824–830.

A Appendix: Proofs

A.1 Proof of Lemma 2

Consider $\tilde{F}$, the minimum value of the penalty preventing selective reporting defined in (4). Both instruments are substitutes provided $\frac{\partial \tilde{F}(\rho)}{\partial \rho} < 0$, for every $\rho \in (0,1)$. Computing the derivative we obtain:

$$\frac{\partial \tilde{F}(\rho)}{\partial \rho} = \frac{1}{\rho}\left[(1-\rho)\left.\frac{d\pi}{dq}\right|_{\tilde{q}=\tilde{q}_\emptyset}\frac{\partial \tilde{q}_\emptyset}{\partial \rho} - \left(\frac{\pi\left(\tilde{q}=\tilde{q}_\emptyset\right)-\pi\left(\tilde{q}=0\right)}{\rho}\right)\right].$$

Using that $\frac{\partial \tilde{q}_\emptyset}{\partial \rho} = \tilde{q}_\emptyset \frac{(1-q)(1-xy)}{1-xq-(1-q)(\rho+xy(1-\rho))}$, we obtain $\frac{\partial \tilde{F}(\rho)}{\partial \rho} < 0$ if and only if (5) holds. *Q.E.D.*

A.2 Proof of Lemma 3

Rewrite the cost threshold as follows

$$\begin{aligned}\mathbb{K}_t^{HR}(q,\rho) &= q\left(\pi\left(\tilde{q}=1\right)-\pi\left(\tilde{q}=0\right)\right)\\ &\quad -\left(1-\rho\left(1-q\right)\right)\left(\pi\left(\tilde{q}=\tilde{q}_\emptyset^{HR}\right)-\pi\left(\tilde{q}=0\right)\right).\end{aligned}$$

This implies that $\mathbb{K}_t^{HR}(q,\rho) > 0$ if and only if

$$\tilde{q}_\emptyset^{HR}\left(\pi\left(\tilde{q}=1\right)-\pi\left(\tilde{q}=0\right)\right)-\left(\pi\left(\tilde{q}=\tilde{q}_\emptyset^{HR}\right)-\pi\left(\tilde{q}=0\right)\right) > 0,$$

which holds if and only if (1) holds. *Q.E.D.*

A.3 Proof of Proposition 5

Let $\rho > 0$ and $F > 0$. We consider first the quality of information and show that $R_1 > 0$. There are two cases.

(a) There exists q such that $F = \tilde{F}_1(q,\rho)$. Notice that we have that $\tilde{F}_1(0,\rho) = 0$ if and only if $q = 0$. Since $F > 0$, we have $R_1 = q > 0$.

(b) There does not exist q such that $F = \tilde{F}_1(q,\rho)$. Since $\tilde{F}_1(0,\rho) = 0$ and since $\tilde{F}_1(q,\rho) = 0$ is weakly increasing and continuous, by the Intermediate Value Theorem we have that $F > \tilde{F}_1(1,\rho)$. By the definition of R_1 we have then that $R_1 = 1$.

We consider now the quantity of information and show that for both thresholds $\mathbb{K}(q,\rho = 0,F) - \mathbb{K}(q,\rho > 0,F) > 0$ holds. We start with $\mathbb{K}_t$. Notice that when $\rho > 0$ the firm reports selectively. Thus $\mathbb{K}_t(q,\rho = 0,F) - \mathbb{K}_t(q,\rho > 0,F)$ is equal to

$$\begin{cases} q\left(\pi\left(\tilde{q}=\tilde{q}_\emptyset^{HR}\right)-\pi\left(\tilde{q}=\tilde{q}_\emptyset^{SR}(\rho=0)\right)\right)+\phi & \text{if } q \le R_1\\ q\left(\pi\left(\tilde{q}=\tilde{q}_\emptyset(\hat{y})\right)-\pi\left(\tilde{q}=\tilde{q}_\emptyset^{SR}(\rho=0)\right)\right)+\varphi & \text{if } q \in (R_1,R_0)\\ q\left(\pi\left(\tilde{q}=\tilde{q}_\emptyset^{SR}\right)-\pi\left(\tilde{q}=\tilde{q}_\emptyset^{SR}(\rho=0)\right)\right)+(1-q)\rho F & \text{if } q \ge R_0\end{cases},$$

where

$$\begin{aligned} \phi &\equiv (1-q)(1-\rho)\left(\pi\left(\tilde{q}=\tilde{q}_{\emptyset}^{HR}\right)-\pi\left(\tilde{q}=0\right)\right)>0 \quad \text{and} \\ \varphi &\equiv (1-q)(1-\rho)y\left(\pi\left(\tilde{q}=\tilde{q}_{\emptyset}\right)-\pi\left(\tilde{q}=0\right)\right)+(1-q)\rho(1-y)F>0. \end{aligned}$$

Because of the properties of $\tilde{q}_{\emptyset}$, we have that $\tilde{q}_{\emptyset}^{HR} \geq \tilde{q}_{\emptyset}(\hat{y}) > \tilde{q}_{\emptyset}^{SR} > \tilde{q}_{\emptyset}^{SR}(\rho = 0)$. Thus $\mathbb{K}_t(q,\rho = 0,F) - \mathbb{K}_t(q,\rho > 0,F) > 0$ for any monitoring policy (F,ρ). By analogous reasoning it can be shown that $\mathbb{K}_{\neg t}(q,\rho = 0,F) - \mathbb{K}_{\neg t}(q,\rho,F) > 0$, because $\mathbb{K}_{\neg t}(q,\rho = 0,F) - \mathbb{K}_{\neg t}(q,\rho > 0,F)$ is equal to

$$\begin{cases} q\left(\pi\left(\tilde{q}=\tilde{q}_{\emptyset}^{HR}\right)-\pi\left(\tilde{q}=q\right)\right)+\phi & \text{if } q \leq R_1 \\ q\left(\pi\left(\tilde{q}=\tilde{q}_{\emptyset}^{HR}\right)-\pi\left(\tilde{q}=q\right)\right)+(1-q)\rho F & \text{if } q \geq R_1 \end{cases}.$$

Q.E.D.

A.4 Proof of Proposition 6

Consider first the quality of information, that is, R_1, R_0 and $y\,(F,\rho)$. Let

$$\Gamma \equiv \frac{1-\rho}{\rho}\left(\pi\left(\tilde{q}=\tilde{q}_{\emptyset}\right)-\pi\left(\tilde{q}=0\right)\right)-F.$$

Notice that under Assumption 3 $\partial\Gamma/\partial q > 0$ for any q and $\partial\Gamma/\partial y > 0$ for any y. Hence, we have that

$$\frac{\partial R_y}{\partial\rho} = -\frac{\frac{\partial\Gamma}{\partial\rho}}{\frac{\partial\Gamma}{\partial q}}, \qquad \frac{\partial R_y}{\partial F} = -\frac{\frac{\partial\Gamma}{\partial F}}{\frac{\partial\Gamma}{\partial q}}, \qquad \frac{\partial y}{\partial\rho} = -\frac{\frac{\partial\Gamma}{\partial\rho}}{\frac{\partial\Gamma}{\partial y}}, \qquad \text{and} \qquad \frac{\partial y}{\partial F} = -\frac{\frac{\partial\Gamma}{\partial F}}{\frac{\partial\Gamma}{\partial y}}.$$

Since $\partial\Gamma/\partial\rho < 0$ if and only if the monitoring instruments are substitutes and since $\partial\Gamma/\partial F = -1$, the statement follows.

Consider now the quantity of information, that is, R_t and $R_{\neg t}$. Let $R_t < R_1$ so that $\mathbb{K}_t^{HR}(q,\rho)$ is relevant. Given Assumption 4 there are at most two values a and b such that $\mathbb{K}_t^{HR}(q,\rho) = xK_x$. Suppose that two such values a and b with $a < b$ exist and consider $\gamma \in \{a,b\}$. We have that

$$\frac{\partial\gamma}{\partial\rho} = -\frac{\frac{\partial\mathbb{K}_t^{HR}(q,\rho)}{\partial\rho}}{\frac{\partial\mathbb{K}_t^{HR}(q,\rho)}{\partial q}} \qquad \text{and} \qquad \frac{\partial\gamma}{\partial F} = 0.$$

Moreover,

$$\begin{aligned} \frac{\partial\mathbb{K}_t^{HR}(q,\rho)}{\partial\rho} &= (1-q)\left(\pi\left(\tilde{q}=\tilde{q}_{\emptyset}^{HR}\right)-\pi\left(\tilde{q}=0\right)\right) \\ &\quad -(1-\rho\,(1-q))\left.\frac{d\pi}{dq}\right|_{q=\tilde{q}_{\emptyset}^{HR}}\frac{\partial\tilde{q}_{\emptyset}^{HR}}{\partial\rho} < 0. \end{aligned}$$

This last inequality holds, since

$$\frac{\partial \tilde{q}_{\emptyset}^{HR}}{\partial \rho} = \tilde{q}_{\emptyset}^{HR} \frac{1-q}{1-\rho\,(1-q)},$$

implies that it is equivalent to (15) holding at $q = \tilde{q}_{\emptyset}^{HR}$. Thus, under Assumption 3 $\partial \mathbb{K}_t^{HR}(q,\rho)/\partial\rho < 0$ for any q. On the other hand, given that $a < b$ exist and Assumption 4 holds, we have that $\partial \mathbb{K}_t^{HR}(q,\rho)/\partial q \neq 0$. In fact we have that

$$\left.\frac{\partial \mathbb{K}_t^{HR}(q,\rho)}{\partial q}\right|_{q=a} > 0 \qquad \text{and} \qquad \left.\frac{\partial \mathbb{K}_t^{HR}(q,\rho)}{\partial q}\right|_{q=b} < 0.$$

so that

$$\frac{\partial a}{\partial \rho} > 0 \qquad \text{and} \qquad \frac{\partial b}{\partial \rho} < 0.$$

Now let $R_1 < R_{\neg t}$ so that

$$\Upsilon(q,\rho,F) \equiv q\left(\pi\,(\tilde{q}=1) - \pi\left(\tilde{q}=\tilde{q}_{\emptyset}\right)\right) - (1-q)\,\rho F$$

is relevant. In the previous expression $\tilde{q}_{\emptyset}$ denotes the belief of the public in the mixed-reporting equilibrium if $\mathbb{K}_t^{Mix} = \Upsilon$ applies, $\tilde{q}_{\emptyset} = \tilde{q}_{\emptyset}^{SR}$ if $\mathbb{K}_t^{SR} = \Upsilon$ is relevant, and $\tilde{q}_{\emptyset} = \tilde{q}_{\emptyset}^{HR}$ otherwise. Again, given Assumption 4 there are at most two values c and d such that $\Upsilon(q,\rho,F) = xK_x$. Suppose that two such values c and d with $c < d$ exist and consider $\gamma \in \{c,d\}$. We have that

$$\frac{\partial \gamma}{\partial \rho} = -\frac{\frac{\partial \Upsilon(q,\rho,F)}{\partial \rho}}{\frac{\partial \Upsilon(q,\rho,F)}{\partial q}} \qquad \text{and} \qquad \frac{\partial \gamma}{\partial F} = -\frac{\frac{\partial \Upsilon(q,\rho,F)}{\partial F}}{\frac{\partial \Upsilon(q,\rho,F)}{\partial q}}.$$

Also,

$$\frac{\partial \Upsilon(q,\rho,F)}{\partial \rho} \;=\; -q\left.\frac{d\pi}{dq}\right|_{q=\tilde{q}_{\emptyset}}\left[\frac{\partial \tilde{q}_{\emptyset}}{\partial \rho} + \frac{\partial \tilde{q}_{\emptyset}}{\partial y}\frac{\partial y}{\partial \rho}\right] - (1-q)F.$$

Notice that we can be sure that the previous expression is strictly negative, unless $\mathbb{K}_t^{Mix} = \Upsilon$ applies (as only in this case $\partial y/\partial\rho \neq 0$). In such a case a strictly positive sign requires the monitoring instruments to be complements and that the term in square brackets represents a sufficiently small negative number. As explained in footnote 22, we assume for simplicity of the exposition that this is not the case. Moreover,

$$\frac{\partial \Upsilon(q,\rho,F)}{\partial F} = -q\left.\frac{d\pi}{dq}\right|_{q=\tilde{q}_{\emptyset}}\frac{\partial \tilde{q}_{\emptyset}}{\partial y}\frac{\partial y}{\partial F} - (1-q)\rho < 0,$$

where again we have that $\partial y/\partial F = 0$ unless $\mathbb{K}_t^{Mix} = \Upsilon$ applies. Again, given that $c < d$ exist and Assumption 4 holds, we have that $\partial \Upsilon(q,\rho,F)/\partial q \neq 0$, which similarly to the case before is increasing in c and decreasing in d. Thus,

$$\frac{\partial c}{\partial \rho} > 0, \quad \frac{\partial c}{\partial F} > 0, \quad \frac{\partial d}{\partial \rho} < 0, \quad \text{and} \quad \frac{\partial d}{\partial F} < 0.$$

Lastly, observe that $R_t \in \{a,c\}$ implies that $\partial R_t/\partial \rho > 0$ and $\partial R_t/\partial F \geq 0$. Similarly, $R_{\neg t} \in \{b,d\}$ implies that $\partial R_{\neg t}/\partial \rho < 0$ and $\partial R_{\neg t}/\partial F \leq 0$.

A.5 Proof of Proposition 7

We start by rewriting (21). Let

$a \equiv \min\{R_t, R_1\} \quad b \equiv \min\{R_1, R_{\neg t}\} \quad c \equiv \max\{R_t, \min\{R_1, R_{\neg t}\}\} \quad d \equiv \max\{R_t, \min\{R_0, R_{\neg t}\}\}.$

Remember that $R_1 < R_0$ follows, because Assumption 3 implies that the profit function is strictly monotone. Moreover, by definition we have that $R_t \leq R_{\neg t}$ holds.[28] The planner chooses a monitoring policy (ρ, F) in order to maximize

$$\begin{aligned}
\Delta(F,\rho) &= \int_0^{R_t} \mathbb{W}_{\neg t}(q,\rho) f(q)dq + \int_a^b \left[\mathbb{W}_t^{HR}(q,\rho) - xK_x\right] f(q)dq \\
&+ \int_c^d \left[\mathbb{W}_t^{Mix}(q,\rho,F) - xK_x\right] f(q)dq + \int_d^{R_{\neg t}} \left[\mathbb{W}_t^{SR}(q,\rho) - xK_x\right] f(q)dq \\
&+ \int_{R_{\neg t}}^1 \mathbb{W}_{\neg t}(q,\rho) f(q)dq - \rho K_\rho
\end{aligned} \tag{23}$$

Since only $\mathbb{W}_t^{Mix}(q,\rho,F)$ depends on F, the derivative of $\Delta(F,\rho)$ with respect to F is

$$\begin{aligned}
\frac{\partial}{\partial F}\Delta(F,\rho) &= \frac{\partial R_t}{\partial F}\mathbb{W}_{\neg t}(R_t,\rho) f(R_t) + \frac{\partial b}{\partial F}\left[\mathbb{W}_t^{HR}(b,\rho) - xK_x\right] f(b) \\
&- \frac{\partial a}{\partial F}\left[\mathbb{W}_t^{HR}(a,\rho) - xK_x\right] f(a) + \frac{\partial d}{\partial F}\left[\mathbb{W}_t^{Mix}(d,\rho,F) - xK_x\right] f(d) \\
&- \frac{\partial c}{\partial F}\left[\mathbb{W}_t^{Mix}(c,\rho,F) - xK_x\right] f(c) + \int_c^d \frac{\partial}{\partial F}\mathbb{W}_t^{Mix}(q,\rho,F) f(q)dq \\
&+ \frac{\partial R_{\neg t}}{\partial F}\left[\mathbb{W}_t^{SR}(R_{\neg t},\rho) - xK_x\right] f(R_{\neg t}) - \frac{\partial d}{\partial F}\left[\mathbb{W}_t^{SR}(d,\rho) - xK_x\right] f(d) \\
&- \frac{\partial R_{\neg t}}{\partial F}\mathbb{W}_{\neg t}(R_{\neg t},\rho) f(R_{\neg t}).
\end{aligned}$$

[28]To be fully precise, R_t and $R_{\neg t}$ denote the intersections between K_x and $\mathbb{K}_{\neg t}(q,\rho,F)$, when the two functions intersect. If such an intersection does not exists, we set $R_t = R_{\neg t} = 1$. Assumption 4 guarantees that there are at most two intersections.

Consider the following six cases, which are exhaustive. Remember that R_t and $R_{\neg t}$ only depend on F if they are larger than $R_1(F,\rho)$. (We disregard equalities, which would eliminate further terms.)

1. Let $R_t < R_1 < R_0 < R_{\neg t}$. Then $(a,b,c,d) = (R_t, R_1, R_1, R_0)$. Because of the continuity of $\mathbb{W}_t(q,\rho,F)$, the second and fifth term and the fourth and eighth term cancel. This implies the statement (with $\partial R_t/\partial F = 0$).

2. Let $R_1 < R_t < R_0 < R_{\neg t}$. Then $(a,b,c,d) = (R_1, R_1, R_t, R_0)$. The second and third term and the fourth and eighth term cancel. This implies the statement.

3. Let $R_1 < R_0 < R_t < R_{\neg t}$. Then $(a,b,c,d) = (R_1, R_1, R_t, R_t)$. The second and third term and the fourth and eighth term cancel. This implies the statement (with the effect in part (1) being zero).

4. Let $R_t < R_1 < R_{\neg t} < R_0$. Then $(a,b,c,d) = (R_t, R_1, R_1, R_{\neg t})$. The second and fifth term and the seventh and eighth term cancel. This implies the statement (with $\partial R_t/\partial F = 0$).

5. Let $R_t < R_{\neg t} < R_1 < R_0$. Then $(a,b,c,d) = (R_t, R_{\neg t}, R_{\neg t}, R_{\neg t})$. The fourth and fifth term and the seventh and eighth term cancel. This implies the statement (but all effects are zero).

6. Let $R_1 < R_t < R_{\neg t} < R_0$. Then $(a,b,c,d) = (R_1, R_1, R_t, R_{\neg t})$. The second and third term and the seventh and eighth term cancel. This implies the statement.

Lastly, we remark that calculating the derivative of the sixth term yields the statement in part (1) of the Proposition. *Q.E.D.*

B Appendix: Not intended for publication

This appendix is for the convenience of the referees and not intended for publication. We start by substantiating the claims made in footnotes 10, 23 and 25 of the main text. After this we provide the complete calculations for the example in Subsection 4.4. Lastly, we decompose the welfare effects of adjusting the probability of detection, as mentioned in footnote 24.

B.1 A micro-foundation for Example 1 and Subsection 4.4

In this subsection we provide a micro-foundation for Example 1, the monotonicity property mentioned in Subsection 4.2, and the social values postulated in expression (22) in Subsection 4.4. We simplify the Hotelling model of product differentiation considered by Brekke and Kuhn (2006) to a monopoly market for a medical treatment. We start by describing the model and derive then Example 1 and the social values postulated in (22).

In a particular therapeutic market a continuum of patients requires medical treatment. Patients are distributed uniformly on the line segment $[0, 1]$ with mass 1. A patient's location $l \in [0, 1]$ represents disease type and/or personal characteristics. For simplicity, we consider a monopolist that is located at endpoint 0 and sells a drug at a price p. The unit cost of production is assumed to be zero. A patient derives the following expected utility from one unit of the drug

$$U(l, p) = E(v) - sl - \tau p,$$

where $E(v) > 0$, $s > 0$, and $\tau \in [0, 1]$. The parameter τ represents either the co-payment rate or the extent to which physicians take prices into account when they make prescription decisions. The term sl reflects side-effects and other factors like contraindications that reduce the effectiveness of the drug. More precisely, there is the utility loss s that multiplies the mismatch cost $|l - 0|$. Again, for simplicity we suppose that $\tau = s = 1$. Finally, $E(v)$ represents the expected 'quality' or gross effectiveness of the drug. There is also an outside treatment (e.g. surgery, physical exercise, or no treatment at all), the benefit of which is normalized to zero. It is assumed that patients cannot observe their condition nor the effectiveness of the two drugs. As a result, all patients consult a physician. This physician is a perfect agent for patients and has the skill to identify the condition of patients, that is, the location $l \in [0, 1]$.

Standard calculations yield the demand function

$$D(p) = \begin{cases} 1 & \text{if} \quad p \leq E(v) - 1 \\ E(v) - p & \text{if} \quad E(v) - 1 \leq p \leq E(v) \\ 0 & \text{if} \quad E(v) \leq p \end{cases} .$$

The monopoly price is $E(v)/2$ if $E(v) \leq 2$ and $E(v) - 1$ otherwise. The profits of the monopolist are

$$\pi\left(E(v)\right)=\begin{cases}0 & \text{if} \quad E(v)\leq 0\\ \left(E(v)\right)^2/4 & \text{if} \quad 0\leq E(v)\leq 2\\ E(v)-1 & \text{if} \quad 2\leq E(v)\end{cases}.$$

Total (expected) surplus for society is thus

$$TS\left(E(v)\right)=\begin{cases}0 & \text{if} \quad E(v)\leq 0\\ 3\left(E(v)\right)^2/8 & \text{if} \quad 0\leq E(v)\leq 2\\ E(v)-1/2 & \text{if} \quad 2\leq E(v)\end{cases}.$$

Notice that this is different from ex-post total surplus for society, once the quality of the drug is revealed.

We obtain Example 1 with $\lambda=2$ in the following way. Consider the gross effectiveness of the drug, given by

$$E(v)=qu(v=1)+(1-q)u(v=0). \tag{24}$$

One interpretation is that in state $v=0$, the drug has gross effectiveness $u(v=0)$, while in state $v=1$ effectiveness is $u(v=1)$. Assume $u(v=1)=2$ and $u(v=0)=0$. This implies, among other things, that for any quality q the firm has positive demand; no patient benefits from the drug in the bad state but all patients benefit in the good state; and the (expected) quality is initially low enough so that a positive clinical trial can increase demand (and not just the price). Under this assumption we have that $E(v)=2q$ and hence

$$\pi\left(q\right)=q^2 \text{ and } TS\left(q\right)=\frac{3}{2}q^2.$$

Consider first $V(q|v=1)$. Since $E(v)=2q$, the social value is

$$V(q|v=1)=\left(\frac{E(v)}{2}\right)^2\frac{1}{2}+\left(2-\frac{E(v)}{2}\right)\frac{E(v)}{2}=2q-\frac{1}{2}q^2.$$

Notice that this expression is increasing in q. The reason is that all consumers would benefit from buying the good but this does not occur because consumers are too pessimistic. Similarly,

$$V(q|v=0)=-\left(\frac{E(v)}{2}\right)^2\frac{1}{2}=-\frac{1}{2}q^2,$$

which declines with q, as no consumer benefits from buying the good but consumers are too optimistic and suffer the mismatch costs. Thus, both function fulfil the monotonicity property (16).

We obtain Example 1 for any $\lambda>0$ by generalizing (24) to

$$E(v)=w(q)u(v=1)+w(1-q)u(v=0),$$

where $w(.)$ is a probability weighting function (see Kahneman and Tversky (1979); Prelec (1998)). Assuming that $w(.)$ takes the form of a power function q^{γ} with $\gamma > 0$, while maintaining that $u(v=1)=2$ and $u(v=0)=0$, we obtain $E(v)=2q^{\gamma}$ and hence $\pi(q)=q^{2\gamma}$. Setting $\gamma=\lambda/2$ yields Example 1 for any $\lambda>0$.

Additional References

[32] Kahneman, D. and A. Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* **47** (2), p. 263–292.

[33] Prelec, D. (1998), "The Probability Weighting Function," *Econometrica* **66** (3), p. 497–527.

B.2 Derivations of the example in Subsection 4.4

Consider the values of information for the firm. In laissez-faire we have $\rho=0$ and obtain

$$\mathbb{K}_t(q) \equiv q\left(1-\left(\tilde{q}_{\emptyset}^{SR}\right)^2\right) \text{ and } \mathbb{K}_{\neg t}(q) \equiv q\left(1-q^2\right).$$

The minimum penalty becomes

$$\tilde{F}(q,\rho,y) \equiv \frac{1-\rho}{\rho}\left(\tilde{q}_{\emptyset}\right)^2,$$

and $\mathbb{K}_t(q,\rho,F)$ equals

$$\begin{cases} \mathbb{K}_t^{HR}(q,\rho) = q\left(1-\left(\tilde{q}_{\emptyset}^{HR}\right)^2\right)-(1-q)(1-\rho)\left(\tilde{q}_{\emptyset}^{HR}\right)^2 & \text{if } q \le R_1 \\ \mathbb{K}_t^{Mix}(q,\rho,F) = q\left(1-\left(\tilde{q}_{\emptyset}\right)^2\right)-(1-q)(1-\rho)y\left(\tilde{q}_{\emptyset}\right)^2-(1-q)\rho(1-y)F & \text{if } q \in (R_1,R_0) \\ \mathbb{K}_t^{SR}(q,\rho,F) = q\left(1-\left(\tilde{q}_{\emptyset}^{SR}\right)^2\right)-(1-q)\rho F & \text{if } q \ge R_0 \end{cases},$$

while

$$\mathbb{K}_{\neg t}(q,\rho,F) \equiv \begin{cases} q\left(1-\left(\tilde{q}_{\emptyset}^{HR}\right)^2\right)-(1-q)(1-\rho)\left(\tilde{q}_{\emptyset}^{HR}\right)^2 & \text{if } q \le R_1 \\ q\left(1-\left(\tilde{q}_{\emptyset}^{HR}\right)^2\right)-(1-q)\rho F & \text{if } q \ge R_1 \end{cases}.$$

Consider now the values of information for society. When there is no search we obtain

$$\begin{aligned} \mathbb{W}_{\neg t}(q,\rho) &= qV(q|v=1)+(1-q)(1-\rho)V(q|v=0) \\ &= q\left(2\tilde{q}_{\emptyset}-\frac{1}{2}(\tilde{q}_{\emptyset})^2\right)-(1-q)(1-\rho)\frac{1}{2}(\tilde{q}_{\emptyset})^2. \end{aligned}$$

When monitoring deters selective reporting we obtain

$$\mathbb{W}_t^{HR}(q,\rho) = xq\frac{3}{2}+(1-x)q\tilde{q}_{\emptyset}^{HR}\left(2-\frac{1}{2}\tilde{q}_{\emptyset}^{HR}\right)-(1-q)(1-x)(1-\rho)\frac{1}{2}(\tilde{q}_{\emptyset}^{HR})^2.$$

When monitoring is unsuccessful welfare is given by

$$\mathbb{W}_t^{SR}(q,\rho) = xq\frac{3}{2} + (1-x)\,q\tilde{q}_\emptyset^{SR}\left(2 - \frac{1}{2}\tilde{q}_\emptyset^{SR}\right) - (1-q)(1-\rho)\frac{1}{2}(\tilde{q}_\emptyset^{SR})^2.$$

Lastly, the mixed-strategy equilibrium yields

$$\mathbb{W}_t^{Mix}(q,\rho,F) = W_t^{SR}(q,\rho) + y(F,\rho)x(1-q)(1-\rho)\frac{1}{2}(\tilde{q}_\emptyset)^2.$$

B.3 The welfare effects of adjusting the probability of detection

We decompose now the effects of adjusting the probability of detection, assuming that the fine is positive. For notational simplicity we provide a result for the case in which all three types of informative equilibria (honest, selective and mixed-reporting) can occur, depending on the realised quality q of the firm's product. More precisely, suppose that $R_t < R_1 < R_0 < R_{\neg t}$. This case seems particularly relevant, because the monitoring policy has some success in eradicating selective reporting but it falls short from eliminating it completely. It is straightforward to adjust the argument to the other cases.

It turns out that increasing the probability of detection has complex effects. Some of the consequences are similar to the effects of raising the fine described in Proposition 7 but further effects emerge. In order to offer a compact representation of the different effects, the following statement uses the notation $\hat{x} \in \{0, x\}$ and $\hat{y} \in \{0, y\}$, where $\hat{x}$ and $\hat{y}$ are equal to x and y, respectively, if (for a given value of q) the firm invests in information and zero otherwise. Remember also that y—the probability that the firm reports honestly—is equal to zero in a selective reporting equilibrium.

Proposition 8 *The effect of a marginal increase of the probability of detection ρ on social welfare can be decomposed in four effects, each of which is ambiguous:*

(1) The frequency of honest reporting changes:

$$\int_{R_1}^{R_0} \frac{\partial y(F,\rho)}{\partial \rho} x(1-q)(1-\rho)H(\tilde{q}_\emptyset)f(q)dq.$$

(2) Investment in information is deterred:

$$\frac{\partial R_t}{\partial \rho}\left[\mathbb{W}_{\neg t}(R_t,\rho) - \mathbb{W}_t^{HR}(R_t,\rho) + xK_x\right]f(R_t)$$
$$-\frac{\partial R_{\neg t}}{\partial \rho}\left[\mathbb{W}_{\neg t}(R_{\neg t},\rho) - \mathbb{W}_t^{SR}(R_{\neg t},\rho) + xK_x\right]f(R_{\neg t}).$$

(3) The confidence effect induces the public to be more optimistic:

$$\int_0^1 \frac{\partial \tilde{q}_\emptyset}{\partial \rho} \Bigg[(1-\hat{x})q \left. \frac{\partial V(\tilde{q}=\tilde{q}_\emptyset | v=1)}{\partial q} \right|_{q=\tilde{q}_\emptyset}$$
$$+ (1-\hat{x}\hat{y}(F,\rho))(1-q)(1-\rho) \left. \frac{\partial V(\tilde{q}=\tilde{q}_\emptyset | v=0)}{\partial q} \right|_{q=\tilde{q}_\emptyset} \Bigg] f(q)dq$$

(4) Monitoring in itself reveals information and reduces social harm but causes a resource cost:

$$\int_0^1 (1-q)(1-\hat{x}\hat{y}(F,\rho))H(\tilde{q}_\emptyset)f(q)dq - K_\rho.$$

We next compare this result to Proposition 7 and offer then a proof. The first two effects are analogous to the corresponding ones in Proposition 7. There are, however, two additional twists that mirror the discussion after Proposition 6. On the one hand, the frequency of honest reporting might decline, so that the first effect might be negative. On the other hand, the deterrence effect on information is stronger in the sense that the second term is unambiguously negative (and never zero).[29] An increase in the probability of detection *always* crowds out honestly reported information.

There are two additional effects that are not present in Proposition 7. A third effect arises, because the confidence effect induces consumers to become more optimistic when neither the firm nor the agency reveal the state of the world. This is desirable when the state of nature is good and harmful otherwise. The latter negative effect applies less often, the higher the frequency of honest reporting (that reveals the state of the world). Lastly, monitoring—as the search technology of the firm—has a marginal resource cost K_ρ and might reveal the state of the world when the firm does not reveal it. Again, this happens less often, the higher the frequency of honest reporting.

[29]The conclusion that the deterrence effect is negative assumes that honestly reported information is socially desirable at the margin, so that the incentives for strict monitoring are highest. If honestly reported information is not desirable, then the deterrence effect is positive (but never zero).

Proof: The derivative of $\Delta(F,\rho)$ with respect to ρ is

$$\begin{aligned}
\frac{\partial}{\partial\rho}\Delta(F,\rho) &= \frac{\partial R_t}{\partial\rho}\mathbb{W}_{\neg t}(R_t,\rho)f(R_t)+\int_0^{R_t}\frac{\partial}{\partial\rho}\mathbb{W}_{\neg t}(q,\rho)f(q)dq \\
&+\frac{\partial R_1}{\partial\rho}\left[\mathbb{W}_t^{HR}(R_1,\rho)-xK_x\right]f(R_1)-\frac{\partial R_t}{\partial\rho}\left[\mathbb{W}_t^{HR}(R_t,\rho)-xK_x\right]f(R_t) \\
&+\int_{R_t}^{R_1}\frac{\partial}{\partial\rho}\mathbb{W}_t^{HR}(q,\rho)f(q)dq+\frac{\partial R_0}{\partial\rho}\left[\mathbb{W}_t^{Mix}(R_0,\rho,F)-xK_x\right]f(R_0) \\
&-\frac{\partial R_1}{\partial\rho}\left[\mathbb{W}_t^{Mix}(R_1,\rho,F)-xK_x\right]f(R_1)+\int_{R_1}^{R_0}\frac{\partial}{\partial\rho}\mathbb{W}_t^{Mix}(q,\rho,F)f(q)dq \\
&+\frac{\partial R_{\neg t}}{\partial\rho}\left[\mathbb{W}_t^{SR}(R_{\neg t},\rho)-xK_x\right]f(R_{\neg t})-\frac{\partial R_0}{\partial\rho}\left[\mathbb{W}_t^{SR}(R_0,\rho)-xK_x\right]f(R_0) \\
&+\int_{R_0}^{R_{\neg t}}\frac{\partial}{\partial\rho}\mathbb{W}_t^{SR}(q,\rho)f(q)dq+\frac{\partial R_{\neg t}}{\partial\rho}\mathbb{W}_{\neg t}(R_{\neg t},\rho)f(R_{\neg t}) \\
&+\int_{R_{\neg t}}^{1}\frac{\partial}{\partial\rho}\mathbb{W}_{\neg t}(q,\rho)f(q)dq-K_\rho.
\end{aligned}$$

Because of the continuity of $\mathbb{W}(q,\rho,F)$, the third and seventh term cancel; and similarly for the sixth and tenth term. Rearranging we are left with

$$\begin{aligned}
\frac{\partial}{\partial\rho}\Delta(F,\rho) &= \frac{\partial R_t}{\partial\rho}\mathbb{W}_{\neg t}(R_t,\rho)f(R_t)+\int_0^{R_t}\frac{\partial}{\partial\rho}\mathbb{W}_{\neg t}(q,\rho)f(q)dq \\
&+\frac{\partial R_{\neg t}}{\partial\rho}\left[\mathbb{W}_t^{SR}(R_{\neg t},\rho)-xK_x\right]f(R_{\neg t})-\frac{\partial R_t}{\partial\rho}\left[\mathbb{W}_t^{HR}(R_t,\rho)-xK_x\right]f(R_t) \\
&+\int_{R_t}^{R_1}\frac{\partial}{\partial\rho}\mathbb{W}_t^{HR}(q,\rho)f(q)dq+\int_{R_1}^{R_0}\frac{\partial}{\partial\rho}\mathbb{W}_t^{Mix}(q,\rho,F)f(q)dq \\
&+\int_{R_0}^{R_{\neg t}}\frac{\partial}{\partial\rho}\mathbb{W}_t^{SR}(q,\rho)f(q)dq+\frac{\partial R_{\neg t}}{\partial\rho}\mathbb{W}_{\neg t}(R_{\neg t},\rho)f(R_{\neg t}) \\
&+\int_{R_{\neg t}}^{1}\frac{\partial}{\partial\rho}\mathbb{W}_{\neg t}(q,\rho)f(q)dq-K_\rho.
\end{aligned}$$

The first, third, fourth and eighth term of the previous expression prove part (2) of the proposition. Consider the second and penultimate term. We have that

$$\begin{aligned}
\frac{\partial}{\partial\rho}\mathbb{W}_{\neg t}(q,\rho) &= q\left.\frac{\partial V(\tilde{q}=\tilde{q}_\emptyset^{HR}|\nu=1)}{\partial q}\right|_{q=\tilde{q}_\emptyset^{HR}}\frac{\partial\tilde{q}_\emptyset^{HR}}{\partial\rho} \\
&+(1-q)(1-\rho)\left.\frac{\partial V(\tilde{q}=\tilde{q}_\emptyset^{HR}|\nu=0)}{\partial q}\right|_{q=\tilde{q}_\emptyset^{HR}}\frac{\partial\tilde{q}_\emptyset^{HR}}{\partial\rho} \\
&+(1-q)H(\tilde{q}_\emptyset^{HR}).
\end{aligned}$$

Consider the fifth term. We have that

$$\begin{aligned}\frac{\partial}{\partial\rho}\mathbb{W}_t^{HR}(q,\rho) &= (1-x)q\,\frac{\partial V(\tilde{q}=\tilde{q}_\emptyset^{HR}|v=1)}{\partial q}\bigg|_{q=\tilde{q}_\emptyset^{HR}}\frac{\partial\tilde{q}_\emptyset^{HR}}{\partial\rho}\\ &\quad+(1-x)(1-q)(1-\rho)\,\frac{\partial V(\tilde{q}=\tilde{q}_\emptyset^{HR}|v=0)}{\partial q}\bigg|_{q=\tilde{q}_\emptyset^{HR}}\frac{\partial\tilde{q}_\emptyset^{HR}}{\partial\rho}\\ &\quad+(1-x)(1-q)H(\tilde{q}_\emptyset^{HR}).\end{aligned}$$

Consider the seventh term. This yields

$$\begin{aligned}\frac{\partial}{\partial\rho}\mathbb{W}_t^{SR}(q,\rho) &= (1-x)q\,\frac{\partial V(\tilde{q}=\tilde{q}_\emptyset^{SR}|v=1)}{\partial q}\bigg|_{q=\tilde{q}_\emptyset^{SR}}\frac{\partial\tilde{q}_\emptyset^{SR}}{\partial\rho}\\ &\quad+(1-q)(1-\rho)\,\frac{\partial V(\tilde{q}=\tilde{q}_\emptyset^{SR}|v=0)}{\partial q}\bigg|_{q=\tilde{q}_\emptyset^{SR}}\frac{\partial\tilde{q}_\emptyset^{SR}}{\partial\rho}\\ &\quad+(1-q)H(\tilde{q}_\emptyset^{SR}).\end{aligned}$$

Consider the sixth term. We obtain

$$\begin{aligned}\frac{\partial}{\partial\rho}\mathbb{W}_t^{Mix}(q,\rho,F) &= \frac{\partial}{\partial\rho}\mathbb{W}_t^{SR}(q,\rho)+\frac{\partial y(F,\rho)}{\partial\rho}x(1-q)(1-\rho)H(\tilde{q}_\emptyset)\\ &\quad-x(1-q)(1-\rho)y(F,\rho)\,\frac{\partial V(\tilde{q}=\tilde{q}_\emptyset|v=0)}{\partial q}\bigg|_{q=\tilde{q}_\emptyset}\frac{\partial\tilde{q}_\emptyset}{\partial\rho}\\ &\quad-x(1-q)y(F,\rho)H(\tilde{q}_\emptyset).\end{aligned}$$

Notice that the second term on the right hand side of this last expression proves part (1) of the proposition, while collecting and rearranging the remaining terms using the notation $\hat{x}$ and $\hat{y}$ completes the proof of parts (3) and (4). *Q.E.D.*