Discouragement through incentives

Dmitri V. Vinogradov*

Elena V. Shadrina**

December 29, 2018

Abstract

Incentives are usually designed to promote desirable behaviour. In many instances, however, even in the absence of an incentive scheme, people may deliberately choose to act as desired. In such a case, introducing a system of incentives may discourage people from doing this. The discouragement mechanism works through the possibility of errors that may wrongly classify the observed behaviour as undesirable, and hence trigger penalties. The effect is amplified by pessimism, which leads to an overestimation of the error probability, and by the disappointment from errors, which increases the disutility of unfair penalties. This approach is capable of explaining two typical observations for enterprises/industries subject to environmental regulation – overcompliance (excessive investment in compliance) and discretionary inspections by regulators (raised frequency of inspections to enterprises suspected of non-compliance).

Keywords: pro-environmental behaviour, incentives, pessimism, environmental protection, compliance, overcompliance.

JEL classification: Q01, D86, D03.

*University of Glasgow, UK, and National Research University Higher School of Economics, Russian Federation.

**National Research University Higher School of Economics, Russian Federation.

Research has been supported by the RFBR grant No 18-010-01166.

1 Introduction

To illustrate the logic of choosing between compliance and violation, Becker (1993) refers to his own experience of making a decision whether to park in a designated but inconveniently located parking area, or to take on risk of getting a ticket for illegally parking on the street. The choice is governed by the size of the penalty and the likelihood of being caught (the frequency of inspections); the typical theoretical prediction is that the penalty should be high, in order to keep frequency of inspections (and thus cost) low. Becker warned however that "The state should also consider the likelihood of punishing innocent persons." (ibid, p.391), thus pointing at the possibility of inspection errors. In this paper, we demonstrate that if the probability of errors is strictly positive, an incentive-compatible penalty scheme may lead to adverse choices by pushing decision-makers towards alternative options where the penalty scheme does not apply. Such behavioural factors as pessimism and disappointment aversion aggravate the problem.

Theoretically, penalty should be high in order to keep the expected punishment at such a level as to prevent the unwanted behavior even at low probabilities of detection, and the frequency (probability) of inspections is low in order to reduce society's costs for maintaining the desired behavior of all its members (Becker, 1968). With regards to environmental issues, Harford (1978) investigated the impact of fines and subsidies on the scale of environmental pollution by firms. Empirical studies show that, despite the softness of fines, firms still invest disproportionately much in compliance with environmental standards. Harrington et al. (1988) explains this fact with reputational considerations: the regulator builds a kind of reliability rating and simplifies the verification procedure for reliable firms. An important development in this literature is Andreoni (1991) who analyzes the choice of the punishment allowing for an error or doubt by the judge. In his setting, a universally maximal fine may encourage crime rather than deter it, for which reason he advances the idea that the fine should fit the crime. We incorporate the idea of the error in the model of choice, thus stressing that it is not only the state who should consider the possibility of the error, but also the decision-makers who take such a possibility in the account. As a straightforward implication, this might potentially explain the above mentioned overinvestment in compliance. More importantly, however, we will show that anticipated errors of the state may reverse the effect of the regulation by shifting behaviour of the public towards "safe" options, where the question of compliance does not arise. This shift may be suboptimal for the society as a whole.

To give a few examples, endow first the Becker's parking problem with a possibility of an error by an inspector: the car is parked correctly, and the parking ticket was purchased but it slipped away from the view of the inspector, and as a result, the inspector issues a parking charge notice (PCN). From the inspector's (and the authority's) perspective, the PCN is issued on a valid ground of no parking ticket being visible under the windschield, while from the perspective of the driver, the parking was paid for, and hence the correct behaviour is being penalized. In Gary Becker's story, if he took this into account before driving to the university, he might have decided to use public transport instead, and would spend more time unproductively on the transport, and possibly would be late to the exam. Another example, referring to the compliance with environmental standards, either makes firms overinvest in activities that would reduce the likelihood of being erroneously classified as non-compliant, or forces them out of businesses where the likelihood of this error is sizable. Finally, when the UK was preparing for Brexit, an important issue was the right of EU citizens who were already in the country, to maintain their residence in the UK after Brexit. As stressed in the analytical report of The Migration Observatory at the University of Oxford, (MO, 2018), "how many EU citizens and their family members secure settled status will depend not just on how many applicants are rejected [...] A whole host of factors, from lack of awareness to fear of rejection to simple disorganisation mean that some eligible EU citizens will not apply." It is this fear of rejection despite being eligible that is the object of our investigation and that may lead to a suboptimal outcome, for example forcing EU citizens to leave the UK, contrary to their preferences.

To be specific, the leading example of our analysis is a decision of an entrepreneur to invest in a new technology, that is more efficient than the old one, and also potentially more environmentally-friendly.[1] The entrepreneur is required to comply with the environmental protection regulation. With regards to the old technology, the entrepreneur does not need

[1] One could think of Elon Musk deciding whether to develop Tesla (new technology) or invest in a new line of diesel cars (old tecnology).

to do anything, as all permissions have been obtained earlier. The compliance of the new technology needs yet to be proven. Inspections from the environmental agency can establish this, and impose a fine if no compliance is detected. The fine does not need to be pecuniary, for example it may be more work to create a more compelling case for compliance. The problem arises when inspections can result in an error. Errors of Type I (false positive, non-compliance not detected[2]) create incentives to violate regulation; errors of Type II (false negative, compliant entrepreneur is taken for a non-compliant) create disincentives to comply.

Academic literature on environmental protection documents cases that we classify as "errors" in our model. Duflo et al. (2014) investigate regulatory discretion and report for their field experiment region: "While punishments are severe when meted out, the chance of being caught is low ... [h]alf of plants are inspected less often than the prescribed rate, while other, similar plants are inspected many times more ...[this discretion may help] if it allows the regulator to use local information to target more polluting plants." The authors also point out that little is known about why enforcement fails, with possible reasons being lack of resources, corruption, laziness or incompetence, as suggested by Stigler (1971) and Leaver (2009), for example. In Dasgupta et al. (2001) it is complaints by the local residents to the environmental protection agency that triger an increase in the frequency of inspections, however these authors take it for granted that firms that fail to comply with regulatory standards face penalties, while studies like Duflo et al. (2014) report high probability of not being caught, and Almer and Goeschl (2010) stress that reported rates of violations typically understimate the actual levels of non-compliance. The conditioning of inspections on "local information" about the likelihood of high polution is a feature of regulatory policy, usually seen as diescretion. As we will show, this approach may be socially desirable and needs to be clearly communicated to the public.

The non-detection of violators is our Type I error, which appears to be quite common. Our Type II errors are less frequently documented.[3] A few reasons may be accountable

[2] Konisky and Reenock (2013) refer to a "systematic non- detection of violations" as "compliance bias".

[3] These errors are often in the centre of public attention when it is about wrongful executions; for some analysis, see, for example, the September 2011 Report of the advisory committee on wrongful convictions of the Joint State Government Commission, General Assembly of the Commonwealth of Pennsylvania (available at deathpenaltyinfo.org).

for this. First, entrepreneurs who believe they have been wrongfully classified as non-compliant, may prefer not to appeal and not to report these cases any further, and thus these occurences remain largely unobservable. Second, as we will show, entrepreneurs may choose activities where the probability of such an error is low. Thus the lack of observations does not imply zero probability of such an error. Heyes (2000) stresses that "under monitoring uncertainty Type II errors [where the enforcement process mistakenly penalises a compliant firm] may, of course, occur."[4] Some indirect efidence of Type II errors follows from Almer and Goeschl (2010) who study criminal prosecutions of environmental offences in Germany, and in their data "the clearance rate of reported environmental crime is on average 60%"m which means that 40% cases do not find any positive link between offenders and the crime; furthermore, out of identified offenders about three-quarters result in dismissal due to lack of evidence or insufficient severity, and a quarter of the remaining cases is found not guilty by the court.[5]

By including the two types of errors in the simple model of optimal penalty, we find that both types are costly for the regulator for the following reason: the Type I error makes punishment less likely, to compensate for which the regulator has to increase the frequency of inspections (assuming the penalty is set at the upper bound); the type II error weakens incentives for compliance, which forces the regulator to further tighten the expected punishment. Our main (benchmark) result is that even if the incentive compatible punishment mechanism is feasible, it may lead to a socially suboptimal outcome once an outside option (such as sticking to an old technology) is available to the entrepreneur. We further demonstrate that this is more likely to happen if entrepreneurs's beliefs (perception of error probabilities) are pessimistic. As a final step of analysis, we consider a benevolent entrepreneur who enjoys extra utility from compliance, yet experiences disappointment if unfairly classified as an offender (Type II error). As long as expected

[4] He then goes on to suggest that an implication of the possibility of Type II errors would be that firms overcomply, to reduce the probability of such an error. As we will show, once an alternative option (project) is available to firms, other implications are possible.

[5] Qualitative studies using in-depth interviews may be better suited to reveal Type II errors. For example, Edmond (2010) cites farmers who "know that [inspector's proposal is] wrong, but I cannot say it is wrong because they may look after other things" and "[I] fear that [inspectors] may make things difficult for me during inspections", or "If we do not do what we are asked to do, we will be punished or they may close the company. We accept because it is easier to do it than face the problems that may arise from failure to do it."

disappointment does not outweigh the extra utility gains from compliance, the incentive compatible policy is more likely to be socially optimal than in the benchmark case. However, an increase in disappointment or in the frequency of inspections enlarges the range of parameters under which an incentive compatible policy is not socially optimal.

On the one hand, our results help explain overcompliance and justify discretionary inspections, discussed above. Overcompliance can be seen as the outside option where extra arrangements are made to ensure the project is "bullet proof" to any inspection (including arrangements with lawyers and insurance). When the probability of the Type II error is relatively high, or when these extra arrangements are inexpensive, entrepreneurs strictly prefer overcompliance. To justify discretionary inspections, we need to consider different types of entrepreneurs, which we achieve by considering benevolent types who have a preference for compliance (as compared to the rational bechmark case). By giving a signal to the benevolent entrepreneurs that the inspection rate for them may be lower, the regulator achieves a reduction in the perceived probability of a Type II error by them, and thus makes socially optimal projects more attractive for these entrepreneurs, as this signal reduces the likelihood of disappointment from wrongly classification.

On the other hand, emphasizing the behavioural components of decision-making allows us to suggest that optimal penalty schemes are less likely to promote socially efficient investment in countries with high uncertainty and/or poverty, as both factors may lead to high levels of pessimism and thus to an overestimation of the likelihood of Type II errors by entrepreneurs, which deters them from the socially optimal investment. Similarly, harsh punishment schemes may be unnecessary in countries or industries where pro-environmental behaviour is inherent in entrepreneurs.

The paper is structured as follows. Section 2 presents the benchmark model with inspection errors and derives implications with regards to overcompliance. Section 3 discusses the formation of beliefs about the possibility of inspectors' errors and demonstrates amplification of the benchmark result under pessimism. Section 4 introduces benevolent behaviour in the form of preference for compliance, coupled with disappointment from the Type II error. Section 5 concludes.

2 The model

2.1 Standard case

Consider an entrepreneur who decides to invest a unit endowment in two projects, both risk-free in terms of returns. The first project, in which the entrepreneur invests $1 - a$ of his endowment, uses a familiar, long established technology, yielding return F; as the technology is well-known, it costs nothing to ensure it complies with existing regulation. We assume here the old technology is not outdated. Although old technologies may fail some newly introduced regulatory criteria, in particular in terms of environment protection, it always requires time to get outdated, and for competitors to develop and establish a better technology. The other project, to which the entrepreneur dedicates share a of his endowment, uses a new technology, which outperforms the old one by offering return $R > F$. However as the technology is new, it is costly to prove regulatory compliance: net return after all compliance procedures is r such that the following holds:

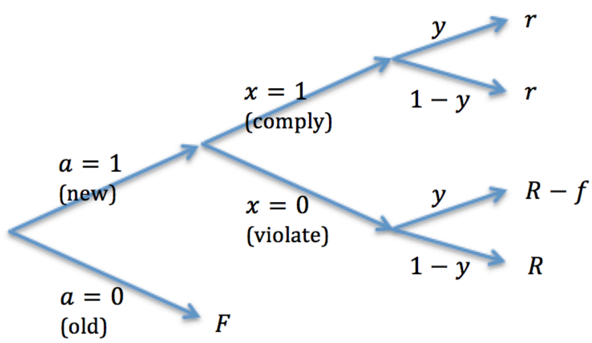
$$F < r < R \tag{1}$$

"Compliance procedures" here refer both to the red tape cost and to any technological adjustments needed to ensure regulatory requirements are met. Main point is that after all these adjustments, the new technology still outperforms the old one, and therefore is strictly preferred by the entrepreneur, hence the amount invested in the new technology is $a = 1$. We shall also assume that the new technology is strictly better for the society as it is more productive.

However, since non-compliance offers a higher return $R > r$ to the entrepreneur, the latter has incentives to violate regulatory prescriptions. To ensure compliance, the regulator inspects the project with probability y and if violation is discovered, the entrepreneur pays the penalty f, which results in his net return $R - f$. The decision tree is in Figure 1.

Note that since the regulator is also familiar with the old technology, there is no asymmetric information with regards to which technology is used - the technology choice

Figure 1: Decision tree: standard case



Notes: The decision is between investing in the new ($a = 1$) or old ($a = 0$) technology, and between compliance ($x = 1$) and non-compliance ($x = 0$) with regulatory (e.g. environmental) requirements, given the inspection probability y and fine f that is due if inspection detects non-compliance.

by the entrepreneur is clearly observable. Information is asymmetric only with regards to the decision of the entrepreneur to comply or not to comply with the regulation.

Let x be the binary variable that describes the entrepreneur's choice: $x = 1$ corresponds to compliance, and $x = 0$ – to violation. The entrepreneur is aware of penalty f and probability of inspection y and maximises expected return, which leads to the following choice:

$$x = 1 \Leftrightarrow r > (1 - y) \cdot R + y \cdot (R - f) = R - yf \tag{2}$$

As standard in the literature, compliance is ensured by the expected value of the fine, yf:

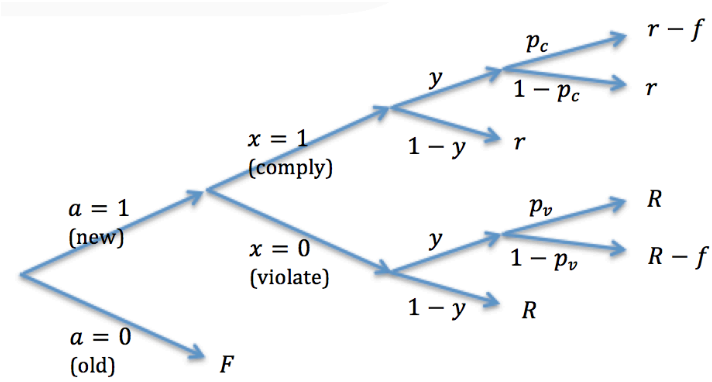
$$yf > R - r, \tag{3}$$

If inspections are costly, y is to be minimised, while f is to be maximised. The latter is achieved at $f = r$, thus yielding $y = \frac{R-r}{r} < 1$. Note that so far $a = 1$ is the only optimal choice of the entrepreneur, independent on the regulatory choice of yf.

2.2 Incentive compatibility with errors

Assume, the inspection can result in an error: with probability p_c a compliant project is classified as non-compliant (Type II error), and with probability p_v the inspection fails to

Figure 2: Decision tree: inspection errors



Notes: The decision is as in Figure 1 except that the inspection result can be erroneous: with probability p_c it classifies a complying project as a non-complying (Type II error) and with probability p_v it fails to detect violation (Type I error).

classify the project that violates regulation as non-compliant (Type I error). The penalty is imposed according to the inspection outcome. The new decision tree is in Figure 2. We will assume both error probabilities are sufficiently small, and in particular holds

$$p_c < 1 - p_v \tag{4}$$

The entrepreneur again compares expected payoffs in the compliance case

$$(1 - y) \cdot r + y \cdot (p_c \cdot (r - f) + (1 - p_c) \cdot r) = r - y \cdot p_c \cdot f,$$

and in the non-compliance case

$$(1 - y) \cdot R + y \cdot (p_v \cdot R + (1 - p_v) \cdot (R - f)) = R - y \cdot (1 - p_v) \cdot f.$$

The choice between compliance and non-compliance now is governed by

$$x = 1 \Leftrightarrow r - y \cdot p_c \cdot f > R - y \cdot (1 - p_v) \cdot f,$$

which implies

$$yf > \frac{R - r}{1 - p_v - p_c}. \tag{5}$$

The right-hand side is strictly positive by assumptions $R > r$ and $p_c < 1-p_v$. Equation (5) is the incentive compatibility constraint the regulator has to meet to ensure compliance. A few observations are due. First, errors make the regulator's job harder, since for any fixed f, a higher probability of inspections y is required than in the case of no errors. This is because Type II errors (which happen with probability $p_c > 0$) lower incentives to comply while Type I errors ($p_v > 0$) raise the expected benefit of non-compliance.

Second, if assumption (4) is violated, and the probability of Type II errors is too high compared to the probability of Type I errors, $1 - p_v \leq p_c$, then no penalty policy is able to ensure incentive compatibility. This is because benfits from non-compliance exceed those from compliance due to the high likelihood of not being penalised. Third, as we are about to demonstrate, errors may discourage the entrepreneur from investing in the new technology, and lead to a socially suboptimal outcome $a = 0$.

2.3 Social optimum with errors

We have assumed that the new technology is socially preferrable to the old one, i.e. $a = 1$ is the socially optimal choice. To be socially oprtimal, the policy (y, f) should be designed in such a way that the expected payoff of the entrepreneur from running the new technology is above that of the old (familiar) technology. This leads to the following social optimality constraint for the regulator:

$$yf < \frac{xr + (1-x) \cdot R - F}{x \cdot p_c + (1-x) \cdot (1-p_v)}, \tag{6}$$

By taking into account that the optimal policy ensures $x = 1$, and combining (5) and (6), we obtain that the socially optimal outcome with incentive compatibility through penalties is feasible if and only if

$$\frac{R-r}{1-p_v-p_c} < yf < \frac{r-F}{p_c}, \tag{7}$$

It immediately follows that an increase in the probability of Type II errors reduces the range of policy variables that ensure both incentive compatibility and social optimality.

The following proposition establishes the critical value of p_c that makes socially optimal penalty policy unfeasible.

Proposition 1 *If the probability of Type II errors, p_c, meets*

$$1 - p_v > p_c > \frac{r - F}{R - F} \cdot (1 - p_v)$$

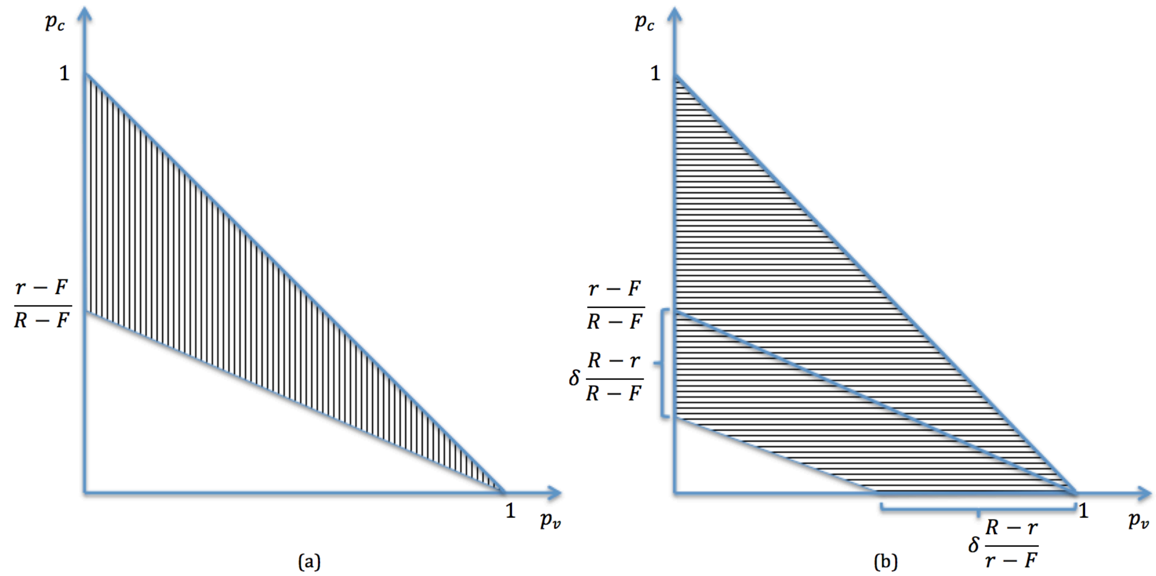
then an incentive compatible policy (y, f) violates social optimality.

The inefficiency result is illustrated in Figure 3, panel (a), where the dashed are represents combinations of error probabilities p_c and p_v under which no penalty policy can establish socially optimal choice. Note that if errors are ruled out, $p_c = 0$ and $p_v = 0$, we are back in the standard case, and the incentive compatibility ensures socially optimal investment. With no Type II errors, there is little problem either, as it may still be possible to choose yf to ensure incentive compatibility. Hovewer even if there are no Type I errors, $p_v = 0$, a large enough probability of a Type II error, $p_c > \frac{r-F}{R-F}$, will make the penalty policy inefficient in the sense of shifting the optimal choice of the entrepreneur towards the socially suboptimal project. This is because on top of regular compliance cost (internalised in r), there is now additional cost of compliance, through the possibility of being erroneously harshly penalised. Importantly, if r is close to F, which may be due to high costs of compliance, even small values of the probability of Type II error, p_c, lead to policy inefficiency. In the next section we show that this inefficiency result is further amplified by pessimism.

2.4 Overcompliance

Overcompliance is a known phenomenon: firms spend on compliance with environmental standards disproportionately much, compared to the minimum required investment, see, for example, Harrington et al. (1988). Heyes (2000) suggests Type II errors may have the potential to explain this phenomenon, as firms might want to overinvest in compliance in order to be sure they pass the inspection, which is a distinct explanation from the reputation build-up argument in Harrington et al. (1988). Our Proposition 1 offers a simple formalisation of the Heyes' point.

Figure 3: Inefficiency of penalty policy due to errors



Notes: Dashed areas correspond to parameter combinations under which no penalty policy leads to a socially optimal outcome. Panel (a) illustrates the bencmark case as in Proposition 1; in panel (b) the effect is amplified by pessimism (overestimation of error probability by δ), as in Proposition 3.

So far we have assumed that a describes the choice between the "old" and the "new" technologies. Now substitute the "old" technology with the same project as the "new" technology, augmented with extra arrangements that ensure compliance. No inspection is ever able to detect non-compliance of this project. Moreover, we can add an insurance policy that covers all expenses that potentially may arise if a stubborn inspector insists on non-compliance. All these arrangements make the project bullet-proof, yet they come at a cost, for which reason the return of this project is below that of the "new" technology, exactly as we assumed, $F < r$.

We can now straightforwardly apply Proposition 1 to this case to obtain that if policy (y, f) is incentive-compatible then as long as the probability of Type II error is relatively high, $p_c > \frac{r-F}{R-F} \cdot (1 - p_v)$, the entrepreneur chooses project F, i.e. strictly prefers overcompliance. Obviously, if overcompliance is inexpensive, i.e. $F \approx r$, the right-hand side in the inequality becomes close to zero and the entrepreneur chooses overcompliance almost surely.

It should be stressed that overcompliance is still an inefficient outcome because it bears unnecessary cost for the entrepreneur. Overcompliance arises purely due to the possibility of inspection errors, which are the source of inefficiency. The inefficiency result in this section is our benchmark case, derived for rational risk-neutral investors only by assuming that inspections may result in an error. To study how departures from rationality affect the result, in the remainder of this paper we introduce two behavioural deviations, pessimism and preferences for compliance coupled with disappointment from errors.

3 Asymmetric beliefs

Above, we have assumed the probability of errors is known to both the regulator and the entrepreneur. However, their beliefs do not need to be aligned, similarly to the case considered in Vinogradov (2012) for bankers' and depositors' beliefs with regards to the probability of a government bailout of failing banks. The key argument for such an asymmetry is the pessimistic view of parties on the outcome of the game. Assume there exists an objectively correct estimate of the probabilities of errors, given by p_c and p_v, as before. The regulator is assumed to perfectly know this estimate and thus faces no ambiguity. The entrepreneur, however, faces ambiguity. A number of models of choice in ambiguity have been suggested in the literature [6], most provide equivalent results for binary choices, which is the case we consider.

The simplest way to capture pessimism of the entrepreneur is to apply a weighting function to these "objective" probabilities of error, as inspired by Kahneman and Tversky (1979) (see also Weber (1994) for the characterisation of weighting functions as optimistic of pessimistic). In particular, pessimism implies overweighting of probabilities of worse

[6] To mention a few most popular, the multiple-priors model by **?** views ambiguity as multiplicity of conceivable probability distributions, and assumes the decision-maker acts as if he faces the worst expected utility among those that can be generated by those distributions. In the neo-additive approach of Chateauneuf et al. (2007) the expected utility is distorted by a weighted average of the worst and the best possible outcomes, where the weights are governed by ambiguity attitudes (pessimism or optimism) of decision-makers. The source function approach by Abdellaoui et al. (2011) is based on the idea that subjects distort the ambiguity-neutral probability by applying a source-specific probability weighting function, the concept known from the prospect theory, Kahneman and Tversky (1979). In the second-order models (Klibanoff et al., 2005; Nau, 2006; Neilson, 2010) subjects form expectations with regards to the value of expected utility; expected utilities are governed by feasible probability distributions, while the "second-order" expectation is taken over the set of these distributions and reflects subjects' ambiguity attitudes.

outcomes, which in our case is probability of the Type II error, and the probability of no Type I error (both assumed strictly positive):

$$w(p_c) > p_c, \tag{8}$$

$$w(1 - p_v) > 1 - p_v. \tag{9}$$

Note that now, even if the objective probabilities of errors are infinitesimally small, but the regulator fails to credibly communicate their values to the entrepreneur, the latter will perceive both probabilities as strictly non-zero. Since weighting functions are usually defined as distortions of cumulative probabilities, and since the event "no error" is a complement to the event "error", with a probability of their union being equal to 1, we can write $1 - w(p_c) < 1 - p_c$ for the probability of no Type II error and $1 - w(1 - p_v) > p_v$ for the probability of Type I error.

We assume the regulator does not account for the entrepreneur's pessimism when designing the optimal policy.[7] As before, the regulator aims to achieve incentive compatibility by setting the policy variables f and y as in (5). The true incentive compatibility for the entrepreneur, however, is governed by expected payoffs as determined by distorted probabilities $w(p_c)$ and $w(p_v)$:

$$(1 - y) \cdot r + y \cdot (w(p_c) \cdot (r - f) + (1 - w(p_c)) \cdot r) = r - y \cdot w(p_c) \cdot f,$$

for the compliance case, and

$$(1 - y) \cdot R + y \cdot (1 - w(1 - p_v) \cdot R + w(1 - p_v) \cdot (R - f)) = R - y \cdot w(1 - p_v) \cdot f$$

for the non-compliance case. It follows that the entrepreneur chooses compliance ($x = 1$) if only if

[7] If the regulator is perfectly aware of the weighting function w, and uses values $w(p_c)$ and $w(1 - p_v)$ when designing the optimal policy, we are in the case of symmetric beliefs, which delivers the same result as Proposition 1, up to notation: one would need to substitute p_c with $w(p_c)$ and $1 - p_v$ with $1 - w(p_v)$. However if there are many entrepreneurs with various probability weighting functions, and one policy is designed for all, the existence of non-pessimistic entrepreneurs would dictate that the regulator chooses to disregard pessimism.

$$yf > \frac{R-r}{w(1-p_v)-w(p_c)}. \tag{10}$$

The right-hand side is strictly positive since assumption $1-p_v > p_c$ and the monotonicity of w imply $w(1-p_v) > w(p_c)$. As the regulator minimises the cost of inspections and disregards pessimism, he sets yf at the lower bound of incentive-compatibility constraint (5), i.e. $yf = \frac{R-r}{1-p_v-p_c}$.[8]

Proposition 2 *If the entrepreneur's beliefs are pessimistic, then policy* $yf = \frac{R-r}{(1-p_v)-(p_c)}$ *is incentive compatible if and only if* $w(p_c) - p_c \leq w(1-p_v) - (1-p_v)$

Proof. If $(1-p_v) - p_c > w(1-p_v) - w(p_c)$ then

$$yf = \frac{R-r}{(1-p_v)-p_c} < \frac{R-r}{w(1-p_v)-w(p_c)}$$

This violates incentive compatibility (10), and the entrepreneur chooses not to comply. Otherwise, if $(1-p_v)-p_c \leq w(1-p_v)-w(p_c)$, policy from (5) suffices to ensure incentive compatibility. ■

To measure the pessimistic distortion of a given probability value p, we may introduce function $\delta(p) = w(p) - p$.[9] With this notation, the condition in the above proposition turns to $\delta(p_c) \leq \delta(1-p_v)$. The result is thus due to the asymmetric overweighting of the small probability of Type II error and the relatively large probability of no error of Type I. Both cases manifest contribution towards a higher likelihood of penalty, and as long as the entrepreneur overemphasises the likelihood of being caught when not complying relative to the likelihood of being erroneously punished when complying, he will choose to comply.

Finally, assume the regulator follows policy dictated by (5), and let this policy be incentive compatibille. Will the entrepreneur still prefer to invest in the new technology?

[8] By using the equality we assume that when the entrepreneur is indifferent between compliance and non-compliance, he prefers the former to the latter.

[9] For most commonly used weighting functions $w(p)$ holds $\delta(0) = \delta(1) = 0$ as there is no uncertainty at $p = 0$ and $p = 1$, yet typically the further away from $p = 0$ and $p = 1$, the higher the distortion. If p is the probability of the "good" outcome, then for pessimists the maximum distortion is achieved at $p < \frac{1}{2}$.

Proposition 3 *If $w(p_c) - p_c \leq w(1-p_v) - (1-p_v)$ and*

$$1 - p_v > p_c > \frac{r-F}{R-F} \cdot (1-p_v) - \delta(p_c) \cdot \frac{R-r}{R-F}, \tag{11}$$

then an incentive-compatible policy (y, f) violates social optimality.

Proof. By Proposition 2 condition $w(p_c) - p_c \leq w(1-p_v) - (1-p_v)$ ensures incentive compatibility of policies that meet (5), and thus $x = 1$. Jointly with social optimality, this requires

$$\frac{R-r}{1-p_v-p_c} < yf < \frac{r-F}{w(p_c)} = \frac{r-F}{p_c + \delta(p_c)}, \tag{12}$$

The system of constraints has no solution for (y, f) if and only if

$$\frac{R-r}{1-p_v-p_c} > \frac{r-F}{p_c + \delta(p_c)}, \tag{13}$$

which is equivalent to

$$p_c > \frac{r-F}{R-F} \cdot (1-p_v) - \delta(p_c) \cdot \frac{R-r}{R-F}. \tag{14}$$

■

If $\delta(p_c) = 0$ the result is identical to that in Proposition 1. The set of feasible policies (y, f) with a pessimistic entrepreneur ($\delta(p_c) > 0$) is strictly smaller than the one in Proposition 1. The new set of parameters under which the penality policy is inefficient is depicted in panel (b) in Figure 3.

Note that now condition $F < r$ does not suffice anymore to ensure the entrepreneur chooses the new technology. From Proposition 3 one straightforwardly obtains that for projects with $r > F > r - \delta(p_c)(R-r)$, no policy leads to the socially optimal choice. In Figure 3 this corresponds to the negative intercept

$$\frac{r-F}{R-F} - \delta(p_c) \cdot \frac{R-r}{R-F} < 0, \tag{15}$$

which makes the whole area $p_c < 1 - p_v$ inefficient as it ensures condition (11) is met for any value $p_v \geq 0$. A socially optimal project, therefore, is abandoned if the entrepreneur believes the likelihood to be punished for the "right" action is too high, as reflected in the distortion of the probability of the Type II error $\delta(p_c) > \frac{r-F}{R-r}$. Thus is likely to hold if cost of compliance makes r too close to F, and/or non-compliance is highly beneficial, i.e. R is high relatively to r. Note that the objective probability of errors may be zero, yet the entrepreneur may still have $\delta(0) > 0$. This is more likely to happen, for example, in poorer societies, or in countries where economic uncertainty is high.[10] In this case the very introduction of the incentive-compatible penalty scheme may lead to a reduction of investment in socially optimal projects.

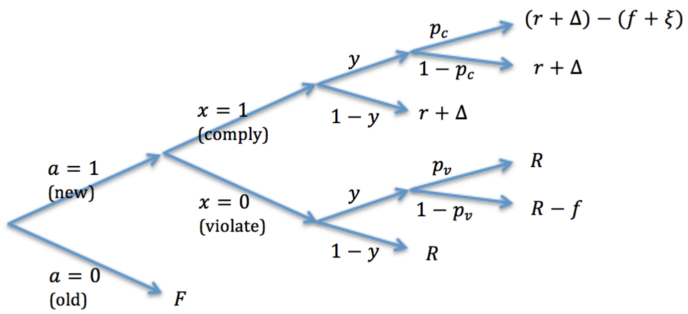
4 Disappointment and preference for compliance

We have so far studied the case of an entrepreneur who prefers to invest in the new technology in the error-free world with an incentive-compatible penalty scheme, but may be deterred from this by the possibility of inspection errors inherent in the incentive mechanism designed to ensure regulatory compliance. As a final step of our analysis, we now assume that compliance has its intrinsic value, denoted with Δ, and thus the return from the new technology is $r + \Delta$ if the entrepreneur complies, and, as before R in the non-compliance case. It might appear, that this "preference for compliance" is well aligned with the penalty policy as the latter is designed to punish non-compliant entrepreneurs. However, with errors, even compliant entrepreneurs may get penalised. If this happens, the entrepreneur's disappointment adds ξ to the fine, f he has to pay. This may turn the entrepreneur away from investing in the new technology, or make him prefer non-compliance.

Disappointment creates an asymmetry in the penalty scheme. If the entrepreneur has low preference for compliance, Δ, he might prefer to violate, if not inspected. If inspection discovers violation, a fine f is due, however we assume no disappointment here. In contrast, if the entrepreneur decides to comply, yet is erroneously classified as an offender, this triggers disappointment. As Gill and Prowse (2012) note, "Disappointment at doing worse than expected can be a powerful emotion." In decision-making disappointment is

[10]For example, Li (2017) provides evidence that ambiguity aversion among the poor and in rural areas is higher than among the rich and in urban areas (in particular, see their Table 3).

Figure 4: Decision tree: disappointment and preferences for compliance



Notes: The decision problem is as in Figure 2 with an addition of preference for compliance, represented by the intrinsic value of compliance, Δ, and of disappointment, as given by extra disutility ξ when an inspection classifies a compliant project as non-compliant.

often represented by a case of "a lottery (or act) result[ing] in a relatively bad outcome" and is seen as a feeling that "can worsen the disutility that the outcome produces directly" (Grant et al., 2001). In our case the complying entrepreneur is evaluated by the inspector, which is our equivalent of "lottery", and becomes disappointed when the inspection disapproves his actions. We model the latter emotional impact on utility as an equivalent to an increase of the fine by ξ. The corresponding decision tree is in Figure 4.

4.1 Discouragement through incentives

The entrepreneur compares expected payoffs in the compliance case [11]

$$(r + \Delta) - y \cdot p_c \cdot (f + \xi),$$

and in the non-compliance case, as before,

$$R - y \cdot (1 - p_v) \cdot f,$$

resulting in the following incentive compatibility condition:

$$(r - y \cdot p_c \cdot f) + (\Delta - y \cdot p_c \cdot \xi) > R - y \cdot (1 - p_v) \cdot f. \tag{16}$$

[11]In this formulation, we assume that when the entrepreneur complies but is falsely accused of non-compliance, elation and disappointment are present at the same time. If only disappoinment is present, the expected payoff changes to $(r + \Delta \cdot (1 - y \cdot p_c)) - y \cdot p_c \cdot (f + \xi) = (r + \Delta) - y \cdot p_c \cdot (f + \xi + \Delta)$. This yields no difference to the result apart from the re-interpretation of ξ.

The regulator observes neither ξ nor Δ, and thus designs the penalty scheme as before (see 5). The following result straightforwardly follows from comparing (5) and (16):

Proposition 4 *If $\Delta > 0$ and $\xi > 0$ then penalty policy $yf = \frac{R-r}{1-p_v-p_c}$ is incentive compatible iff $\Delta - y \cdot p_c \cdot \xi > 0$.*

An important implication of the above proposition is that if the extra utiliity from compliance does not outweigh the expected disappointment from being falsely accused of non-compliance, $\Delta < y \cdot p_c \cdot \xi$, the incentive scheme discourages compliance. Note that the entrepreneur with $\Delta > R - r$ would have deliberately chosen to comply if there was no penalties scheme. [12] Thus the introduction of the penalty scheme forces him to switch from compliance to non-compliance if

$$y \cdot p_c \cdot \xi > \Delta > R - r. \tag{17}$$

4.2 Socially optimal choice

From the above it follows that under $y \cdot p_c \cdot \xi > \Delta > 0$ the incentive scheme leads to a socially suboptimal outcome. The remaining question is whether an entrepreneur with $\Delta > y \cdot p_c \cdot \xi > 0$ prefers the new technology to the old one under the existing incentives scheme.

Proposition 5 *Either policy $(y, f) : yf = \frac{R-r}{1-p_v-p_c}$ is not incentive compatible, or it violates social optimality iff*

$$1 - p_v > p_c > \frac{r-F}{R-F} \cdot (1 - p_v) + \frac{\Delta - y \cdot p_c \cdot \xi}{R-F} \cdot (1 - p_v - p_c), \tag{18}$$

where $\frac{\Delta - y \cdot p_c \cdot \xi}{R-F} \cdot (1 - p_v - p_c) > 0$.

Proof. By Proposition 4, policy $yf = \frac{R-r}{1-p_v-p_c}$ is not incentive-compatible iff $y \cdot p_c \cdot \xi > \Delta > 0$, therefore we only need to focus on $\Delta > y \cdot p_c \cdot \xi > 0$, for which incentive compatibility

[12] This can be checked by substituting for $y = 0$ in (16).

ensures compliance, $x = 1$. The entrepreneur's choice in favour of $a = 1$ is then dictated by condition

$$(r - y \cdot p_c \cdot f) + (\Delta - y \cdot p_c \cdot \xi) > F. \tag{19}$$

Substitute for $yf = \frac{R-r}{1-p_v-p_c}$ and rearrange:

$$\Delta - y \cdot p_c \cdot \xi > \cdot p_c \cdot \frac{R - r}{1 - p_v - p_c} - (r - F). \tag{20}$$

If the opposite holds, the policy is not socially optimal:

$$p_c \cdot \frac{R - r}{1 - p_v - p_c} - (r - F) > \Delta - y \cdot p_c \cdot \xi > 0$$

The latter holds iff

$$p_c \cdot (R - r) - (r - F) \cdot (1 - p_v - p_c) > (\Delta - y \cdot p_c \cdot \xi) \cdot (1 - p_v - p_c) > 0. \tag{21}$$

Re-arrange left-hand side to obtain:

$$p_c \cdot (R - r) - (r - F) \cdot (1 - p_v) + (r - F) \cdot p_c = p_c \cdot (R - F) - (r - F) \cdot (1 - p_v)$$

and substitutde in (21)

$$p_c \cdot (R - F) - (r - F) \cdot (1 - p_v) > (\Delta - y \cdot p_c \cdot \xi) \cdot (1 - p_v - p_c) > 0,$$

which yields

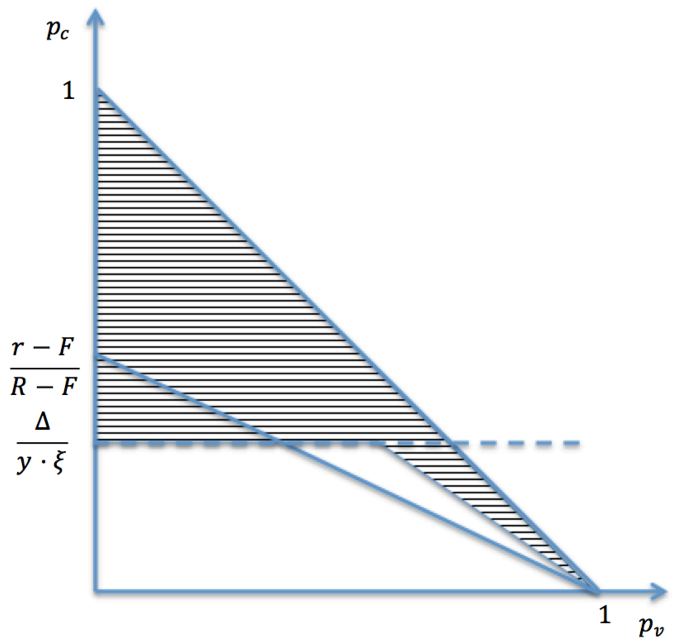
$$p_c > \frac{r - F}{R - F} \cdot (1 - p_v) + \frac{\Delta - y \cdot p_c \cdot \xi}{R - F} \cdot (1 - p_v - p_c).$$

The second term in the right-hand side is strictly positive due to $\Delta > y \cdot p_c \cdot \xi$. ■

Removing behavioural factors, Δ and ξ, from consideration, makes Proposition 5 identical to the benchmark case in Proposition 1, as condition (18) turns into

$$1 - p_v > p_c > \frac{r - F}{R - F} \cdot (1 - p_v).$$

Figure 5: Inefficiency of penalty policy under preferences for compliance and disappointment from errors.



Notes: preferences for compliance are given by Δ, disappointment from erroneous classification of a compliant project as non-compliant is given by ξ.

In the presence of strong preferences for compliance, $\Delta > y \cdot p_c \cdot \xi > 0$, the inefficiency range gets reduced, as compared to the benchmark, see Figure 5. However the requirement of strong preferences for compliance itself implies that the range of parameters is limited by $p_c < \frac{\Delta}{y \cdot \xi}$, which can be a binding constraint if the preference for compliance is rather small, while the disappointment from errors is large.

5 Conclusion

We started with a quote from Gary Becker who warned that when introducing penalty schemes, governments should be careful about the possibility of punishing the innocents. Our objective was to investigate how the mere possibility of an error can distort incentives and thus affect the outcome of decision-making. In a simple model with an entrepreneur who decides between an old and a new (better and possibly environmentally friendlier but requiring more work to comply with regulations) technology, we have introduced two types of errors. Type II errors classify a compliant entrepreneur as a non-compliant and lead to an unjust penalty. Type I errors are examples of an inspection's failure to detect violations of the regulation. We assumed that the new technology is socially optimal, and moreover, the payoff structure is such that the entrepreneur's objectives align with this social optimum. This optimum is always attainable if the regulator introduces an incentive-compatible and error-free penalty scheme to ensure compliance. However with errors not only the frequency of inspections has to go up (because Type II errors disincentivize compliance, and Type I errors create extra incentives for non-compliance) but also the feasibility of the socially optimal outcome is not guaranteed: there is a non-empty range of probabilities of errors under which the entrepreneur strictly prefers the old technology. Pessimism of the entrepreneur amplifies this result.

Apart from pessimism, we considered a benevolent entrepreneur by assuming he is intrinsically motivated to comply. As this intrinsic motivation is unobservable to the regulator, the penalty scheme is still justified from the perspective of the latter. Punishing a compliant benevolent entrepreneur leads to disappointment. If, hypothetically, the benevolent entrepreneur experiences very little disappointment when unjustly penalised, we observe an improvement to the benchmark case: social optimum can be achieved under a wider range of parameters. However the very introduction of the penalty scheme in this case is unnecessary, as the entrepreneur would comply anyway. If however, more realistically, the disappointment from an unjust punishment is strong, it creates further disinsentives to comply, which pushes the inspection rate up. The resulting range of probabilities of errors that still admit the socially optimal outcome becomes smaller.

What we call Type I errors is widely documented in the literature. There are many explanations of why inspectors exercise discretion, or follow the policy of forebearance, or simply fail to detect violations. Type II errors, sadly, exist and become a focus of public attention if it is about death sentences that years after execution are found erroneous (e.g. when the actual murderer is found). In our entrepreneurial context, Type II errors are less frequently observable. One explanation could be that entrepreneurs fear to appeal against incorrect judgments of inspectors. Our analysis suggests a different explanation: the possibility of this error diverts entrepreneurs from activities where this possibility is high to activities where such a possibility is small or nil.

The driving example for our analysis was environmental protection, where inspections are a common tool to enforce compliance. Literature suggests regulators deliberately condition the frequency of inspections on the past observed behaviour of entrepreneurs, sending inspectors more often to enterprises with higher historical pollution levels. This is justified from the perspective of our model as it makes sense to reduce the frequency of inspections of benevolent entrepreneurs. A straightforward implication of our analysis is that this practice should be publicised as a formal rule, thus affecting the beliefs of entrepreneurs with regards to the probability of an inspection error. Formally, this corresponds to a reduction of y in our model, and hence an increase in $\frac{\Delta}{y \cdot \xi}$ in the analysis of the benevolent entrepreneur. Making this practice a rule, therefore, makes switching to the old technology less likely, and thus promotes policy efficiency.

References

Abdellaoui, M., A. Baillon, L. Placido, and P. P. Wakker (2011). The rich domain of uncertainty: Source functions and their experimental implementation. *The American Economic Review 101*(2), 695–723.

Almer, C. and T. Goeschl (2010). Environmental crime and punishment: empirical evidence from the german penal code. *Land Economics 86*(4), 707–726.

Andreoni, J. (1991). Reasonable doubt and the optimal magnitude of fines: should the penalty fit the crime? *The RAND Journal of Economics*, 385–395.

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime*, pp. 13–68. Springer.

Becker, G. S. (1993). Nobel lecture: The economic way of looking at behavior. *Journal of political economy 101*(3), 385–409.

Chateauneuf, A., J. Eichberger, and S. Grant (2007). Choice under unvertainty with the best and worst in mind: Neo-additive capacities. *Journal of Economic Theory 137*(1), 537–567.

Dasgupta, S., B. Laplante, N. Mamingi, and H. Wang (2001). Inspections, pollution prices, and environmental performance: evidence from china. *Ecological Economics 36*(3), 487–498.

Duflo, E., M. Greenstone, R. Pande, and N. Ryan (2014). The value of regulatory discretion: Estimates from environmental inspections in india. Working paper no. w20590, National Bureau of Economic Research.

Edmond, Y. (2010). Environmental management: Communication and conflict during environmental inspection: A case study of uppsala municipality (sweden). Master thesis, Faculty of Natural Resources and Agricultural Sciences Department of Urban and Rural Development, Swedish University of Agricultural Sciences, Uppsala.

Gill, D. and V. Prowse (2012). A structural analysis of disappointment aversion in a real effort competition. *American Economic Review 102*(1), 469–503.

Grant, S., A. Kajii, and B. Polak (2001). Different notions of disappointment aversion. *Economics Letters 70*(2), 203–208.

Harford, J. D. (1978). Firm behavior under imperfectly enforceable pollution standards and taxes. *Journal of environmental economics and management 5*(1), 26–43.

Harrington, W. et al. (1988). Enforcement leverage when penalties are restricted. *Journal of Public Economics 37*(1), 29–53.

Heyes, A. (2000). Implementing environmental regulation: enforcement and compliance. *Journal of regulatory economics 17*(2), 107–129.

Kahneman, D. and A. Tversky (1979). Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the econometric society*, 263–291.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005). A smooth model of decision making under ambiguity. *Econometrica 73*(6), 1849–1892.

Konisky, D. M. and C. Reenock (2013). Compliance bias and environmental (in) justice. *The Journal of Politics 75*(2), 506–519.

Leaver, C. (2009). Bureaucratic minimal squawk behavior: Theory and evidence from regulatory agencies. *American Economic Review 99*(3), 572–607.

Li, C. (2017). Are the poor worse at dealing with ambiguity? *Journal of Risk and Uncertainty 54*(3), 239–268.

MO (2018, 07). Measuring success: Will we ever know how many eligible eu citizens did not apply for settled status? Analytical report, The Migration Observatory, at the University of Oxford COMPAS (Centre on Migration, Policy and Society).

Nau, R. F. (2006). Uncertainty aversion with second-order utilities and probabilities. *Management science 52*(1), 136–145.

Neilson, W. S. (2010). A simplified axiomatic approach to ambiguity aversion. *Journal of Risk and uncertainty 41*(2), 113–124.

Stigler, G. J. (1971). The theory of economic regulation. *The Bell journal of economics and management science*, 3–21.

Vinogradov, D. (2012). Destructive effects of constructive ambiguity in risky times. *Journal of International Money and Finance 31*(6), 1459–1481.

Weber, E. U. (1994). From subjective probabilities to decision weights: The effect of asymmetric loss functions on the evaluation of uncertain outcomes and events. *Psychological Bulletin 115*(2), 228.