Khalmetski, Kiryl; Rockenbach, Bettina; Werner, Peter

Conference Paper

Evasive Lying in Strategic Communication

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Experiments - Games II, No. B18-V2

Provided in Cooperation with:
Verein für Socialpolitik / German Economic Association

Suggested Citation: Khalmetski, Kiryl; Rockenbach, Bettina; Werner, Peter (2017) : Evasive Lying in Strategic Communication, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Experiments - Games II, No. B18-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel, Hamburg

This Version is available at:
https://hdl.handle.net/10419/168119

Evasive Lying in Strategic Communication

Kiryl Khalmetski,[1] Bettina Rockenbach,[2] Peter Werner[3]

18 January 2017

Abstract

Information asymmetries in economic transactions are omnipresent and a regular source of fraudulent behavior. In a theoretical and an experimental analysis of a sender-receiver game we investigate whether sanctions for lying induce more truth-telling. The novel aspect in our model is that senders may not only choose between truth-telling and (explicit) lying, but may also engage in evasive lying by credibly pretending not to know. While we find that sanctions promote truth-telling when senders cannot engage in evasive lying, this is no longer true when evasive lying is possible. Then, explicit lying is largely substituted by evasive lying, which completely eliminates the otherwise positive effect of sanctions on the rate of truth-telling. As outlined in our model, the necessary prerequisite for such an 'erosion' effect is that evasive lying is perceived as sufficiently less *psychologically* costly than direct lying. Our results clearly demonstrate the limitations of sanctioning lying to counteract the exploitation of informational asymmetries and may explain the empirical evidence from the finance industry that sanctions for financial misconduct eventually appear to be not very efficient.

Keywords: lying, sanction, evasion, sender-receiver game, financial fraud

JEL classification: C91, D82, D83

Financial support of the German Research Foundation (DFG) through the Research Unit "Design & Behavior" (FOR 1371) is gratefully acknowledged.

[1] University of Cologne, Albertus-Magnus-Platz, D-50923 Köln, Germany (e-mail: kiryl.khalmetski at uni-koeln.de).

[2] University of Cologne, Albertus-Magnus-Platz, D-50923 Köln, Germany (e-mail: bettina.rockenbach at uni-koeln.de).

[3] Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands (e-mail: p.werner at maastrichtuniversity.nl).

1. INTRODUCTION

Information asymmetries between economic transaction partners may cause serious harm. While it is an immanent characteristic of many economic interactions that one party is better informed than the other, abusing this informational advantage can have severe consequences for the worse informed party. Some of the most impressive cases for tremendous damages caused by the exploitation of informational advantages come from the financial industry. This sector seems to be prone to fraud given the complexity of investment products and the typically large gap between the expertise of financial advisors and customers. Moreover, due to biased incentives for the sale of financial products (for example, set by commissions), financial advisors might be inclined to misrepresent information about the value of a financial product or be reluctant to warn about potential risks, which leads to suboptimal investment decisions. Remarkably, Cohn et al. (2014) show that priming bank employees with their professional identity causes more lying on their side, which suggests that there might be a norm for dishonesty in the banking sector.

Financial fraud seems to occur on all levels: For example, in 2008 the former star investor Bernard Madoff was found to cheat investors by setting up a giant Ponzi scheme in which profits for one group of investors were paid out with new investments of others, thus masking accumulated losses of some 50 billion US-Dollars (Efrati et al. 2008). In an empirical study of the US financial advisors industry, Egan et al. (2016) provide large-scale evidence that fraudulent patterns are a widespread and persistent phenomenon in the US financial advisors industry - 7% of all registered financial advisors in the years 2005 to 2015 have been officially disciplined for financial fraud. Also, the damages caused are substantial, as the median settlement paid in cases of misconduct accounts for $40,000. Furthermore, in October 2016, Ian Nasev, the CEO of Commonwealth Bank in Australia admitted that an independent review concluded that every tenth customer of the bank had received inappropriate advice (Davis 2016).

However, fraudulent behavior based on informational advantages is by no means restricted to financial advice, but also occurs between the managers and the owners of companies, for example, if the former misrepresent information to enrich themselves. In 2003, the biggest Italian food company Parmalat went bankrupt after it was discovered that the management had cheated on the true values of the company's assets and debt, causing thousands of investors to lose their savings (Hooper 2008). Finally, an important daily-life example for economic interactions that bear the risk of fraud are credence goods, such as medical service or repair services, where the customer is able to assess the value of a good only ex-post, but is not sure about her true need ex-ante. This gives rise to incentives for treating the customer in a suboptimal manner, for example through over-charging, under-treatment or over-treatment (see Kerschbamer et al. 2016 for recent field evidence).

Given the considerable risk of being deceived in economic transactions, an obvious question is if formal ex-post sanctions can significantly counteract the decisions of better informed actors to exploit their informational advantage. Empirical evidence from the finance industry seems to suggest that sanctions for financial misconduct eventually appear to be not very efficient. In fact, in their study of the US retail finance industry, Egan et al. (2016) find that almost half of the advisors who took fraudulent actions are set off in the subsequent year, but, at the same time, a substantial fraction of them are again employed in the same industry

in the year afterwards (although facing a substantial pay reduction). Moreover, in August 2016 the Security and Exchange Commission (SEC) settled the last case concerning potential securities fraud due to the misrepresentation of financial risks in subprime loans during the financial crisis with only low settlements (Henning 2016).

In fact, implementing appropriate sanctioning schemes for lying might be challenging for regulators since in many cases lying is not perfectly verifiable. In particular, while it can be possible to verify explicitly stated claims, it is much harder to check whether an advisor *concealed* some of the available evidence by claiming that she/he is uninformed (Okuno-Fujiwara et al. 1990). The previous theoretical literature established that strategic withholding of information can naturally arise under perfectly verifiable information, i.e., when lying is impossible (Dye 1985, Dziuda 2011).[1] Yet, practically no attention was given to the role of concealment of information (or evasion, as termed in our paper) in settings with limited liability of the sender (i.e. when the level of punishment for lying is not fully deterrent), which can be considered as more realistic in light of the aforementioned examples. The specificity of such settings is that the sender decides not only about whether to tell the truth or lie, but in addition faces an explicit tradeoff between lying and evasion. In turn, this can affect the way the sender reacts to a change in the material incentives arising, e.g., due to an introduction of a fine for lying. This is the starting point of our study in which we analyze how the possibility of evasive communication, i.e., credibly pretending to be uninformed about the true state of the world, affects strategic communication and the effectiveness of sanctions against lies. In particular, we study how the tradeoff between lying and evasion is resolved depending on the structure of the intrinsic lying costs, and which implications this has for the efficiency of regulatory sanctions.

As a first step, we consider a model of strategic communication where the rate of lying can be at intermediate levels due to the heterogeneity of intrinsic lying costs in the population, and hence can be effectively altered by changing material incentives for lying, e.g., with corresponding fines. At the same time, the sender has an option to pretend to be uninformed (i.e., to send an evasive message), which is ex-ante credible due to the presence of actually uninformed types, and also cannot be subject to fines (by being unverifiable). In equilibrium, both lying and evasion are chosen by a positive fraction of senders, yet only if the cost of evasion is sufficiently lower than the cost of lying. Besides, we show that under the presence of the evasive option, it might be difficult to achieve an increase in the rate of truth-telling by sanctioning lying behavior. In particular, direct lying can be partially substituted by evasive lying as a result of the sanction, which eventually erodes its positive effect on the rate of truth-telling. Importantly, the necessary prerequisite for this effect is that the (intrinsic) cost of lying is asymmetric between direct and evasive lying. If the sender bears the same psychological cost of lying independently of its form, the fine is fully efficient in substituting lying with truth-telling since evasive lying is never chosen in equilibrium (unless the fine is so large that lying is eliminated completely).

In the next step, we conduct an experimental sender-receiver game that captures the structure of our model. In our benchmark condition, senders have to choose between truth-

[1] While the seminal papers on verifiable disclosure (Grossman 1981, Milgrom 1981) established that the sender eventually discloses all information ("unraveling" result), Dye (1985) and Dziuda (2011) showed that this does not hold if the receiver does not know in advance whether the sender is informed.

telling and direct lying. Our experimental treatments introduce the availability of the option of evasive lying and vary the presence of deterministic punishment for direct lying under both communication regimes. This setting allows us to investigate the impact of sanctions with and without the possibility to evade. In particular, we study the substitution of direct with evasive lying in the presence of external sanctions, which in turn may hinder reaching the welfare objectives of the regulation. We find that the introduction of sanctions induces more truth-telling in the absence of the evasion option for the sender. However, evasive lying is chosen (whenever available) by a non-negligible share of subjects both with and without punishment for (direct) lying. This confirms our hypothesis that the psychological costs of lying depend on the content of the message; costs of evasive messages seem to be substantially lower than these of direct lying. Importantly, we find that deterministic sanctions do not affect the rate of truth-telling when evasive lying is possible. Instead, a substantial number of senders switch to the evasive message in order to circumvent punishment, so that the positive effect of sanctions on the rate of truth-telling is completely eroded. This is in line with the theoretical predictions under asymmetric lying costs, thus highlighting the important role of behavioral motivations in moderating the effect of external regulation on strategic communication.

Related literature

A growing number of laboratory studies have investigated how subjects decide when they have the incentive to be dishonest (see Rosenbaum et al. 2014 for a survey on experiments on honesty and truth-telling). A central result in this literature is that humans are heterogeneous concerning lying behavior - some subjects are found to have intrinsic preferences not to lie, whereas others tend to lie when it is line with their material interest (Gneezy 2005, Sutter 2009, Gibson et al. 2013, Gneezy et al. 2013, Fischbacher and Föllmi-Heusi 2013, Kajackaite and Gneezy 2015).[2] One common interpretation from these studies is that people face (heterogeneous) intrinsic psychological costs of lying. These costs were shown to depend on the monetary consequences of a lie for both the sender and the receiver of the message (Gneezy 2005), senders' beliefs about the receiver (Sutter 2009, Beck et al. 2013), form of communication (Lundquist et al. 2009), or game experience (Gneezy et al. 2013). Several papers so far considered the effect of the possibility of evasive communication (such as staying silent or using vague messages) on the informativeness of communication (see Sánchez-Pagés and Vorsatz 2009, Serra-Garcia et al. 2011, Agranov and Schotter 2012). These studies show that many subjects use such communication patterns to circumvent both explicit lying and truth-telling. This happens even if evasion carries a fixed material cost as in Sánchez-Pagés and Vorsatz (2009).[3]

To the best of our knowledge, only a few studies have investigated the role of sanctions and their effect on lying behavior so far. Mostly relevant for our study, Sánchez-Pagés and Vorsatz (2009) conducted an experimental sender-receiver game with endogenous punishment which could be implemented by receivers at a cost to themselves. At the same time, senders were allowed to stay silent (also at a cost) instead of sending an explicit message. The authors find that those subjects who punished lies with a higher probability are

[2] Kartik (2009) analyses the effect of lying costs on communication that may arise due to either external or psychological factors.

[3] Khalmetski and Tirosh (2012) consider evasive lying in a setting close to the current one, while analyzing the effect of the form of communication on the rates of both direct and evasive lies.

those who exhibited higher rates of truth-telling in case they choose to send a message. Besides, similarly to our results, Sánchez-Pagés and Vorsatz (2009) show that the introduction of punishment does not increase the rate of truth-telling while senders switch to staying silent somewhat more often. At the same time, their setting had important differences from ours. First, staying silent is arguably less psychologically costly than evasive lying considered in our study, which involves a false (explicit) claim of being uninformed while in fact possessing information.[4] Hence, it is less clear ex-ante whether subjects are willing to engage in such type of lying (while simultaneously losing certain control over the receiver's decision) to circumvent punishment. Second, in the experiment of Sánchez-Pagés and Vorsatz (2009) the level of punishment was implemented at the discretion of the receivers and is found to depend on the latter's attitude towards lying. On the contrary, as our study focuses on the role of external (institutionalized) regulation in the presence of information asymmetries, we consider deterministic punishment which depends on the ex-post verifiability of the messages.[5]

The remainder of the paper is organized as follows. Section 2 presents a theoretical model of communication with direct and evasive lying. In Sections 3 and 4 we describe our experimental design and the results. Section 5 discusses our findings and concludes.

2. MODEL

2.1. Material game

There are two players: the sender (he) and the receiver (she). Besides, there is a state of the world s ex-ante unknown to both players, which can be either good (*G*) or bad (*B*). Each state is realized with ex-ante probability 1/2. The timing of the game is as follows. First, the sender observes the state of the world with probability $\kappa < 1$. Hence, there can be 3 possible sender types: observing the good state (type G), observing the bad state (type B), and observing no information (type N). If the sender is informed, he can choose between the following 3 messages: m_G, m_B, or m_N to be sent to the receiver (each message being associated in meaning with the corresponding sender's type so that, for example, m_G means that the sender claims to have observed the good state). If the sender is uninformed, he can only send message m_N (the claim to be uninformed). After observing the message, the receiver takes a choice between Invest (*I*) or Abstain (*A*), and the payoffs are realized.

[4] The choice of the message space in Sánchez-Pagés and Vorsatz (2009) was designated to test the hypothesis of whether senders have lying aversion or rather a preference for truth-telling, unlike in our paper where we studied the difference between the costs of direct and evasive lies and the corresponding implications for regulation. Other differences in the theoretical predictions stem from the fact that in Sánchez-Pagés and Vorsatz (2009) the sender was always informed, while in our case this happened only with some probability (which allows for the credibility of the evasive message).

[5] In another study of endogenous punishment, Peeters et al. (2013) let subjects interact both under a sanctioning and a non-sanctioning institution (in later rounds self-selected by the participants), where the sanctioning institution provides subjects with an option to punish lying. The data reveals individual heterogeneity with respect to both psychological costs of lying and the willingness to impose sanctions. The study by Xiao (2013) provides evidence that endogenous punishment may signal norms of truth-telling that substantially improves communication.

Table 1. Payoff structure of the game.

	Good state	**Bad state**
Invest	π , P	π , L
Abstain	0, 0	0, 0

The payoffs structure is given in Table 1. If the receiver chooses Abstain, the payoffs of both players are normalized to 0 (without loss of generality). At the same time, the sender gets a fixed commission π if the receiver invests, so that he strictly prefers investment independently of the state of the world. At the same time, we assume $P > 0$ and $L < 0$, which implies that the receiver prefers to invest only in the good state.

The fact that the truly uninformed sender can only send message m_N creates an asymmetry of the message space between the informed and the uninformed types (as, e.g., in the model of Austen-Smith 1994). Besides simplifying the subsequent analysis, this feature of our model is designated to reflect the conjecture that it is easier to prove the mere fact of being informed for a truly informed sender (e.g., by presenting the part of evidence which can be verified), rather than for an uninformed one (who lacks any evidence to be presented). Thus, it is reasonable to assume that the informed sender can always (credibly) separate from the uninformed one. At the same time, one can well imagine that the informed sender can still pretend to be uninformed by simply concealing his available evidence, with such concealment being in principle not verifiable if the sender is indeed uninformed with some probability (see Dye 1985 and Okuno-Fujiwara et al. 1990 for a discussion).[6]

2.2. Preferences

We assume that the expected probability of investment from the perspective of the sender is an increasing function of the probability of the good state conditional on the message, i.e.,

$$\Pr_s[I \mid m] = \phi(\eta_X), \tag{1}$$

where $\eta_X = \Pr[s = G \mid m_X]$, $X \in \{G, N, B\}$, and ϕ is a continuous strictly increasing function with $\phi(0) = 0$ and $\phi(1) = 1$. Thus, (1) can be considered as a reduced form of modeling heterogeneity of risk aversion in the population of receivers.

Then, it is easy to verify that in case if the sender is driven by mere payoff maximization, in (pure-strategy) equilibrium type B always pools with type G (on any of the 3 messages). This makes the communication completely uninformative independently of the message which is pooled on.[7]

Let us consider how the structure of equilibrium communication is altered if the sender bears a cost of lying (Gneezy 2005, Kartik 2009). We extend the seminal idea that lying is associated with intrinsic psychological costs to also allow for the possibility that different

[6] The fact that non-disclosure of information cannot be verified and punished is one of the key motivations of strategic disclosure literature, starting from Grossman (1981) and Milgrom (1981).

[7] There also exists a continuum of mixed-strategy equilibria (with types B and G adopting the same messaging strategy) which are uninformative as well.

contents of communication can be associated with different costs of lying. Herewith we distinguish two types of lie possible in our game:

- Direct lie: the sender falsely states to have observed a specific state of the world (that is, sending message m_G while being of type B, and vice versa).

- Evasive lie: the sender falsely states to have not observed the state of the world (that is, sending message m_N while being of type B or G).

We denote the behavioral cost of direct (evasive) lie as c_{DL} (c_{EL}). The sender's expected utility from sending message m_X while observing state t is then given by

$$U_t(m_X) = \pi\phi(\eta_X) - \theta c(t, m_X), \tag{2}$$

where θ is the coefficient measuring individual aversion against lying, and $c(t, m_X)$ is the cost of lying from sending message m_X for type *t*. In particular, $c(t, m_X)$ is equal to $c_{DL}(c_{EL})$ in case of direct (evasive) lie, and is 0 in case of truth-telling. Herewith, we naturally assume

$$c_{DL} \geq c_{EL} > 0, \tag{3}$$

i.e., the more explicit direct lie is at least as costly as the less explicit evasive lie. Besides, we assume a lexicographic preference for less explicit lie in that the sender chooses evasive lie once he is indifferent between direct and evasive lie in terms of utility.

We also assume that θ varies in the population of senders being continuously distributed on $[0, \bar{\theta}]$ according to some cumulative distribution function Z (thus, the value of θ is also referred below as the sender's 'type'). Herewith, $\bar{\theta}$ is assumed to be sufficiently large so that at least some types always prefer truth-telling over lying even if the expected probability of investment conditional on lying is 1, i.e.,[8]

$$\bar{\theta} > \frac{\pi}{c_{EL}}. \tag{4}$$

The receiver does not observe θ, i.e. she does not know ex-ante how trustworthy the sender is.

2.3. Equilibrium

We solve for perfect Bayesian equilibrium of this game. First, consider the optimal strategy of type G. Intuitively, while this type has no material conflict of interest with the receiver, he has no incentives to deviate from the truth and hence always sends m_G (which in equilibrium yields the highest probability of investment).

Lemma 1. *In any equilibrium all sender types observing G send message* m_G.

[8] Note that the probability mass on such types can still be arbitrarily small. This assumption allows to not consider equilibria where m_B is not used, thus streamlining the exposition (while being also consistent with subsequent experimental evidence).

Proof. See Appendix A. ■

Given Lemma 1, the sender observing *B* has 3 choices: 1) to pool with type G by sending m_G, which then yields the highest probability of investment at cost c_{DL}, 2) to pool with uninformed types by sending m_N for somewhat lower likelihood of investment at cost c_{EL}, or 3) to tell the truth at no cost eventually getting 0 payoff (as m_B is perfectly informative about state *B* by Lemma 1). The next proposition characterizes how this tradeoff is resolved depending on the ratio between c_{EL} and c_{DL}.[9]

Proposition 1. *There exists a threshold* $\omega(c_{EL})>1$ *such that:*

a) If $\frac{c_{DL}}{c_{EL}} \le \omega(c_{EL})$*, then there exists a unique equilibrium characterized by* $0<\theta^{**}<\bar{\theta}$ *such that the sender observing B never sends* m_N*, sends* m_G *if* $\theta \in [0,\theta^{**})$*, and sends* m_B *if* $\theta \in [\theta^{**},\bar{\theta}]$.

b) If $\frac{c_{DL}}{c_{EL}} > \omega(c_{EL})$*, then there exists a unique equilibrium characterized by* $0<\theta_1<\theta_2<\bar{\theta}$ *such that the sender observing B sends* m_G *if* $\theta \in [0,\theta_1)$*,* m_N *if* $\theta \in [\theta_1,\theta_2)$*, and* m_B *if* $\theta \in [\theta_2,\bar{\theta}]$.

Proof. See Appendix A. ■

The graphical scheme of two types of equilibrium is given on Fig. 1, which shows the distribution of messages over two-dimensional sender types (with the first dimension being the observed information, and the second one being the sensitivity to lying θ). In both cases, senders observing the bad state with sufficiently low θ would prefer to earn the highest possible material payoff by sending m_G, thereby lying directly. In contrast, senders with very high θ prefer to send m_B as no material benefit would compensate their lying cost. Yet, it is not immediate to see whether any sender type observing the bad state would send evasive message m_N. In particular, if the cost of evasive lying is sufficiently close to the cost of direct lying, there is no evasive lying in equilibrium. The reason is that switching to evasive message m_N (from m_G) is costly *per se*, since the likelihood of subsequent investment is smaller. Hence, if there is no large benefit in the form of lower lying costs which could compensate for this loss (i.e., c_{DL} is close to c_{EL}), then no type would find it profitable to choose evasive lying (Fig. 1(a)).[10] Yet, if the difference between c_{EL} and c_{DL} is sufficiently high, there is an intermediate range of lying sensitivities where the sender finds it too costly to lie directly, yet still prefers a (psychologically cheaper) way of evasive lying over truth-telling

[9] Herewith, we make a purely technical assumption of lexicographic preference for truth-telling to pin down the strategy of the marginal cutoff type, which is without loss of generality.

[10] More specifically, although types with sufficiently high θ might get sensitive to a small difference in lying costs between direct and evasive lying, one can show that such types prefer truth-telling in the first place.

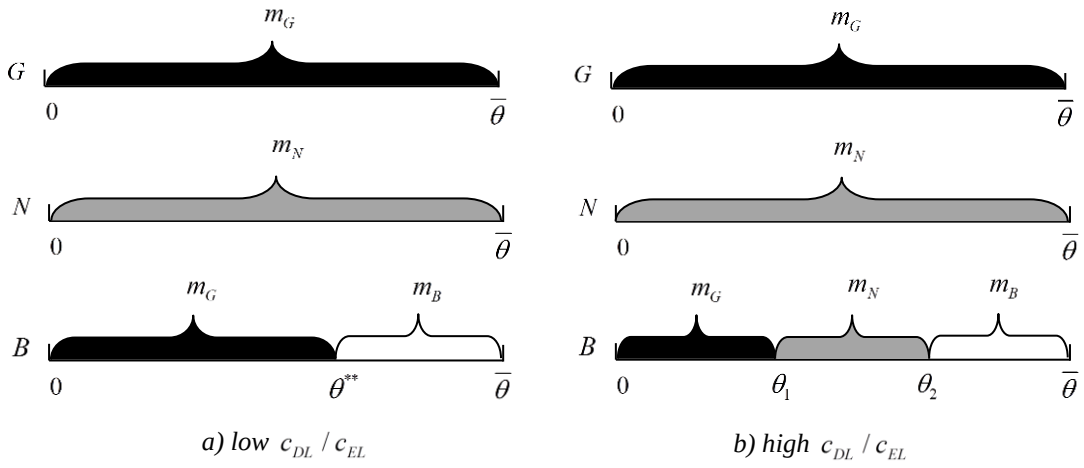

Fig. 1. Structure of equilibrium communication.

(Fig. 1(b)). Thus, the prerequisite for the emergence of evasive lying in equilibrium is that the cost of this type of lying is sufficiently smaller than that of direct lying.

2.4. The effects of policy interventions

As the focus of our study is on the effect of exogenously implemented punishment on the informativeness of communication, we now consider the effect of the policy of imposing a monetary fine for any (verifiable) lying, i.e. if the sender misrepresents his information while this can be verified ex-post. Herewith, we make the following assumptions:

- if the sender claims to be uninformed, this cannot be verified (see section 2.1);
- the state of the world can be verified only if the receiver invests.

The first assumption refers to the idea that when there is uncertainty about the information the sender has, he can always credibly pretend to be uninformed. This implies that the sender can never be punished for evasive lying. The second assumption ensures, in particular, that the sender is never punished if the receiver abstains, in which case she suffers no actual losses.[11] Since the sender still has no incentives to lie after observing G (Lemma 1 remains to hold), he is then punished in equilibrium if and only if he sends m_G after observing B while the receiver invests. Thus, the expected utility function of the sender observing B takes the form

$$U_B(m_X) = (\pi - f(m_X))\phi(\eta_X) - \theta c(B, m_X), \tag{5}$$

where $f(m_X) > 0$ if $X = G$ and is 0 otherwise.

In our subsequent analysis, we focus on the cases where the level of fine is non-deterrent, that is direct lying is not eliminated completely as a result of the fine (consistent with the field

[11] This structure is relevant in many real-life settings. For example, financial advisors or doctors can hardly be punished if the consumer/patient has abstained from taking advice, in which case hypothetical losses resulting from the advice can hardly be verified.

evidence outlined in the introduction). The following proposition characterizes the change in the equilibrium structure of communication if the non-deterrent fine is introduced.[12]

Proposition 2. *If a (non-deterrent) fine for lying is introduced, then:*

a) *If* $c_{EL} = c_{DL}$, *the rate of truth-telling increases, the rate of direct lying decreases and the rate of evasive lying stays the same at 0.*

b) *If* $c_{EL} < c_{DL}$ *:*

i) *If* $\frac{c_{DL}}{c_{EL}} \leq \frac{\pi - f}{\pi}\omega(c_{EL})$, *the rate of truth-telling increases, the rate of direct lying decreases and the rate of evasive lying stays the same at 0.*

ii) *If* $\frac{c_{DL}}{c_{EL}} > \frac{\pi - f}{\pi}\omega(c_{EL})$, *both the rate of truth-telling and the rate of evasive lying increase while the rate of direct lying decreases.*

Proof. See Appendix A. ■

Thus, the relation of the cost of evasive lying to the cost of direct lying has an important implication for the effect of the fine on the communication structure. In particular, if c_{EL} is equal to c_{DL} (or sufficiently close to it under small fine), no evasive lying emerges in equilibrium also after the fine is introduced, so that the fine leads to a substitution of direct lying with truth-telling (see Fig. 2(a)). The mechanism behind this result is the same as described in the previous section. Namely, there is a drop in the investment likelihood after switching from direct to evasive lying. If it is not compensated by the corresponding decrease in the intrinsic lying costs, no type would ever prefer evasion over both truth-telling and direct lying.

In the other case, if c_{EL} is sufficiently small relative to c_{DL}, evasion emerges in equilibrium and the policy has only indirect effect on the rate of truth-telling. In particular, direct lying is substituted by *both* evasive lying and truth-telling, i.e., the reduction in the (direct) lying rate is higher than the corresponding increase in the truth-telling rate (see Fig. 2(b)). Thus, the positive effect of the fine on the rate of truth-telling is partially eroded by the substitution of direct with evasive lying. Intuitively, the mechanism behind this comparative statics is the following. Once a fine is introduced while evasive lying occurs in equilibrium, some marginal types previously involved in direct lying prefer to avoid being fined by switching to unverifiable evasive communication. As a result, message m_N becomes less credible, resulting in a lower probability of investment conditional on this message. Consequently, evasion becomes less attractive as before the fine and some marginal types (in the vicinity of the cutoff θ_2) switch from evasive lying to truth-telling. However, one can

[12] Note that the term $\frac{\pi - f}{\pi}\omega(c_{EL})$ may be smaller than 1 (even if the fine is not fully deterrent). This is the reason why the proposition separately distinguishes cases $c_{EL} = c_{DL}$ and $c_{EL} < c_{DL}$.

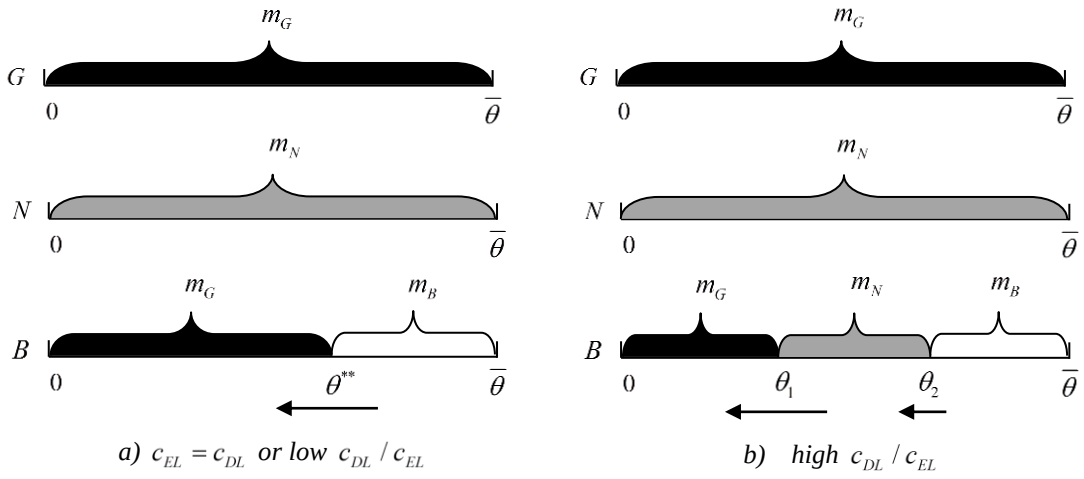

Fig. 2. The effect of the fine for lying on the equilibrium communication.

show that this switch is insufficient to offset the initial switch from direct to evasive lying, so that the total rate of evasion increases (see Appendix A for details).

3. EXPERIMENTAL DESIGN AND HYPOTHESES

3.1. Experimental design

Our experiment design reflects the interaction between a better and a worse-informed party and follows our model setup described in the previous section. Because the main motivation for our study, outlined in the introduction, is the effect of sanctions on the quality of financial advice, our instructions are framed as an investment decision; participants in the role of senders (receivers) are denominated as advisors (investors).

Before the start of our experiment, participants are randomly assigned the roles of advisor and investor and keep the roles during the entire experiment. In each round, one advisor and one investor are randomly matched to each other. The timing of the decisions in each round follows the timing of the theoretical game in section 2.1. First, nature randomly chooses a state of the world (the "investment conditions"), which can be either good (*G*) or bad (*B*), with each state being realized with probability 0.5. The advisor observes the state of the world with probability 0.6 and then has to choose one of several pre-defined messages to be sent to the investor. In particular, in all experimental treatments (see below for a description), the advisor can choose between messages announcing the good and the bad state ("I believe that the investment conditions in this period are good (bad)"; denoted as "Good" and "Bad" in the following). In cases where the advisor does not observe the realization of the investment conditions, the message "I do not know the investment conditions in this period" (the "Don't know" message) is automatically sent. Hence, the advisor is able to manipulate his message only if he in fact learns about the true state. Advisors' choices are elicited with the help of the strategy method (Selten 1967) so that each advisor has to provide his message for each of the possible cases (before learning the state): if he learns that the state is good, and if he learns that the state is bad. In case if the advisor indeed learns the state afterwards, the corresponding decision is implemented.

Table 2. Experimental payoff structure.

	Good state	**Bad state**
Invest	*Investor: 12* *Advisor: 11*	*Investor: 3* *Advisor: 11*
Abstain	*Investor: 6* *Advisor: 8*	*Investor: 6* *Advisor: 8*

In the next step, the message is transmitted to the investor who is ignorant about the state of the world and then has to decide whether to invest or to abstain. Subsequently, payoffs from the investment decision are realized and reported to both players. Table 2 gives an overview of the payoffs for both players, depending on the true investment conditions and the investor's decision.

As Table 2 shows, the conflict of interest between the advisor and the investor arises because the former gains additional 3 ECU if the latter decides to invest, irrespective of the realized state. Yet, for the investor, it is only optimal to invest if the good state has materialized, as here the investor gains 5 ECU relative to abstaining, whereas he loses 3 ECU from the investment if the bad state occurs.

Our experiment is conducted in a 2 X 2 between-subjects design. The first treatment variation refers to the number of messages the advisor can choose in case of learning the state: "Good" and "Bad" for the two message (2M) treatment and "Good", "Bad" and "Don't know" for the three message (3M) treatment. Therefore, our 3M treatment enables the advisor to strategically pretend that he is uninformed in cases he learns about the true state. In contrast, evasive lying is not possible in 2M treatment, which thus serves as a control treatment.

The second treatment dimension refers to the existence of a deterministic sanction after false messages, as modelled in section 2.4. In the punishment treatment (P), a fine of 1 ECU is automatically deducted from the advisor's payoff if the advisor sends a message of having observed a state of the world different from the true state.[13] Thus, message "Don't know" is never fined. Besides, in line with our model, the fine is deducted only in case if the investor has decided to invest. This amount might be interpreted as a direct sanction for (verifiable) fraudulent behavior, for example, by a regulator, in case if the investor suffers losses as a result of bad financial advice. On the contrary, in the "no punishment" treatment (NP), no sanction is implemented. In the remainder of our paper, we use the following abbreviations for our four treatments: 2M-NP, 2M-P, 3M-NP and 3M-P, respectively.

3.2. Experimental hypotheses

Our main research question refers to the effect of the option of evasive lying on the structure of communication, and in particular, its interaction with sanctions for lying.

The 2-message case is a benchmark treatment where the effect of the fine can be clearly predicted: as the relative benefit from lying drops after the introduction of the fine, the rate of

[13] The fine is deducted if the "Good" message is sent in the bad state and if the "Bad" message is sent in the good state. Note that the size of the fine is only low so that truth-telling still remains a strictly dominated option in terms of monetary payoff.

lying (truth-telling) should decrease (increase).[14] In the 3M treatment, the advisor additionally obtains an option to send an evasive message while being informed. As Proposition 2 implies, under non-deterrent fine, this addition does not alter the theoretical predictions relative to the 2M treatment if the advisor's psychological cost of lying does not depend on its type (evasive or direct), i.e., if $c_{EL} = c_{DL}$. In particular, direct lying is still substituted by truth-telling as a result of the fine, while the evasive message is never sent (conditional on being informed) both before and after the fine. As discussed in section 2, in such case evasion yields a strictly lower material benefit than direct lying (also after the fine), while providing no compensation in terms of lower intrinsic lying costs. Hence, our null hypothesis is that the effect of the fine on the structure of communication is the same in the 2M and 3M treatments.

Our alternative hypothesis refers to the case when evasive lying is supposed to be intrinsically less costly for the advisor than direct lying. Then, our model predicts that the effect of the (non-deterrent) fine is qualitatively different in the 3M treatment in that direct lying is substituted by both truth-telling and evasive lying. This can be seen as an erosion effect in that a 1-percent drop in the lying rate should yield a less than 1-percent increase in the truth-telling rate (unlike in the 2M treatment, where the ratio is always 1 to 1 by construction). Moreover, as discussed in section 2.4, the positive effect of the fine on the rate of truth-telling in this case does not work through the direct change in material incentives (as in the case of direct lying). Instead, there is an indirect effect through updated beliefs: advisors are supposed to anticipate the reduced credibility of the evasive message m_N as a result of the fine (when this message is abused more extensively), and consequently switch from evasion to truth-telling in marginal cases. Hence, if subjects have difficulties in assessing the change in the credibility of the messages as a result of the fine, the rate of truth-telling might get even more rigid with respect to the fine in the 3M treatment, thus exacerbating the erosion effect.

3.3. Experimental procedures

Advisors and investors play altogether 10 rounds of the described decision situation. After each round, feedback is given regarding players' payoffs and the investors' investment decisions.

In each session, we divide participants into cohorts of six, with three participants acting in the role of advisors and investors, respectively, so that each cohort forms one statistically independent observation. Prior to the each round, an advisor and an investor are matched to each other, ensuring that no pair of subjects will interact in two consecutive rounds. Before they state their decisions, we ask investors and advisors about their beliefs concerning the behavior and beliefs of others.[15]

---

[14] Formally, one can show that the effect of the fine in case if the advisor cannot send evasive message m_N is equivalent to the case of Proposition 2(a) (the proof proceeds analogously and is available upon request).

[15] Beliefs are elicited in an incentivized manner using a quadratic scoring rule (see Schlag et al. 2015 for a survey of belief elicitation techniques). In particular, we ask both investors and advisors to provide an estimate about the percentage share of all investors who choose to invest after receiving a "Good" and a "Don't know" message. Moreover, we collect a measure for the expected truthfulness of communication by asking investors to provide a guess about the percentage of senders who will choose the "Good" message and "Don't know" message (the latter guess is only elicited in the 3M treatments) for the case that the realization of the investment

We conducted altogether 10 experimental sessions in the Cologne Laboratory for Economic Research (CLER) from May to August 2015 in which 282 subjects took part. Each session was conducted with 24 or 30 subjects, depending on the number of participants who had registered for the experiment but did not show up. The experiment was programmed with the software z-tree (Fischbacher 2007); participants were recruited with the help of the online recruitment system ORSEE (Greiner 2015). Participants arrived at the laboratory, were seated randomly into cubicles and received written instructions. After reading the instructions, participants answered quiz questions to ensure that they had understood the details of the experimental decision situation.[16] After the experiment, participants privately received the payoffs earned through the decision situation and the belief elicitation, converted at an exchange rate of 12 ECU = 1 Euro, and left the laboratory. Average earnings accounted for 12,39 Euro, while the experiment lasted around 1 hour. A copy of translated instructions can be found in Appendix B.

4. RESULTS

In this section, we start with analyzing the communication patterns of advisors and how they are affected by our experimental treatments. In the next step, we explore the responses of investors towards the messages sent by advisors.

In line with our model, advisors virtually always send message "Good" conditional on observing the good state, when there is no conflict of interest (in 99,1% of the cases). The relevant case for our analysis is how an advisor responds to a conflict of interest between himself and the investor that arises when he observes the bad state. Here, it becomes optimal for the investor to refrain from investing while the advisor has an incentive to misinform her. Table 3 lists the average rates of direct lying, evasive lying and truth-telling which we define as the percentage of all cases in which advisors chose, respectively, "Good", "Don't know" and "Bad" messages conditional on observing the bad state, calculated over all periods and experimental cohorts.[17]

First, we observe generally high (direct) lying rate in the 2M-NP treatment (more than 70%). In the 3M-NP treatment, the rate of direct lying is somewhat lower (62.8%), yet evasive lying occurs in some 11% of the cases. This share is significantly larger than zero, as a one-sided Wilcoxon signed-rank test performed on the level of experimental cohorts suggests ($p = 0.002$). At the same time, note that advisors have significantly lower expected probability of investment conditional on "Don't know" than conditional on "Good" message in 3M-NP, as their elicited beliefs reveal (40.7% vs. 83.4%).[18] Thus, the finding that some advisors opt for pretending to be ignorant instead of lying directly (under no fine) is in line

conditions would be in fact bad. Finally, we elicit advisors' beliefs about the investors' answers to the latter questions, i.e., advisors' second-order beliefs about the credibility of the messages.

[16] The quiz was interrupted for two of the experimental subjects who were not able to reach the end of the quiz in reasonable time due to severe lack of understanding of the instructions, and hence could not be considered as having the equivalent set of instructions relative to other participants. The corresponding cohorts were dropped from the estimation, while the data was resampled with new subjects.

[17] In subsequent analysis, we treat each experimental cohort (i.e., matching group) as one independent observation.

[18] The difference in these beliefs is highly significant, also if one considers only those observations when advisors chose evasive lying ($p < 0.001$, one-sided permutation test).

Table 3. Distribution of messages conditional on observing bad state.

	2M-NP	*2M-P*	*3M-NP*	*3M-P*
Direct lying, %	70.3	50.0	62.8	49.4
Evasive lying, %	0.0*	0.0*	11.4	26.7
Truth-telling, %	29.7	50.0	25.8	23.9

* In the 2M treatments, the evasive message could not be chosen by advisors.

with the interpretation provided by Proposition 1 that an advisor's psychological cost of evasive lying (c_{EL}) is indeed substantially smaller than the psychological cost of direct lying (c_{DL}).

Result 1: *A significant share of advisors choose evasive lying under no sanctions for lying.*

Next, consider the effect of the punishment on the messaging strategies. In the 2M game, the rate of direct lying drops by more than 20 percentage points if punishment is introduced, while the difference is statistically significant (p = 0.019, one-sided permutation test). Concerning the 3M treatments, we observe a drop in the direct lying rate from some 63% to some 49% once the punishment is introduced, which is however insignificant (p = 0.090, one-sided permutation test). Most importantly, the sanction has virtually no effect on truth-telling - in 25.8% of the cases in 3M-NP and 23.9% of the cases in 3M-P advisors decide to send message "Bad" conditional on observing the bad state (p = 0.434, one-sided permutation test). At the same time, in line with our model for the case of sufficiently low psychological cost of evasive lying (see Proposition 2(b)), the rate of evasive lying more than doubles in 3M-P relative to 3M-NP: it accounts for some 27% and is significantly higher than in 3M-NP (p = 0.027, one-sided permutation test).[19]

Our conclusions remain similar in parametric analyses (see Table 4). We calculate Logit models with a dependent variable that takes the value of one if the advisor chooses direct lying (Model 1), truth-telling (Model 2) and evasive lying (Model 3). As the independent variables we use the dummies for the experimental treatments while having the 3M-NP treatment as the benchmark. We include random effects for each experimental advisor and cluster the standard errors on the level of the experimental cohorts to account both for subject-specific heterogeneity and the dependency of observations within cohorts.

Model 1 confirms that the introduction of punishment tends to reduce the propensity to lie in the 2M-P treatment, as the difference between the coefficients on 2M-NP and 2M-P is positive and significant (p = 0.035, two-sided Wald test). Moreover, the probability to lie tends to decrease in the 3M-P treatment compared to the 3M-NP treatment, as indicated by a marginally significant negative coefficient on 3M-P. Notably, subjects tend to lie more with time, which is in line with the results of Gneezy et al. (2013).

[19] Again, the share of evasive messages is significantly larger than zero (p = 0.001, one sided Wilcoxon signed-rank test).

Table 4. Determinants of message choices.

Model No.	(1)	(2)	(3)
Dependent Variable	Probability of direct lying	Probability of truth-telling	Probability of evasive lying
3M-P	-3.032*	-0.682	2.483**
	(1.830)	(1.790)	(1.158)
2M-NP	1.570	0.737	
	(1.929)	(1.874)	
2M-P	-2.673	4.801**	
	(2.091)	(2.040)	
Period	0.190***	-0.200***	0.012
	(0.064)	(0.074)	(0.075)
Constant	1.268	-3.365**	-5.180***
	(1.501)	(1.520)	(0.832)
N	1,410	1,410	720

The table shows the results of Random-Effects Logit models with the dependent variables equal to one if a particular message was used by the advisor. Robust standard errors clustered on the level of experimental cohorts are listed in parentheses. ***, ** and * indicate significance on the 1%, 5% and 10%-level, respectively.

Concerning truth-telling, we again find that the fine has a positive effect in the 2M treatment but not in the 3M treatment: the difference between coefficients on 2M-NP and 2M-P is statistically significant (p = 0.034, two-sided Wald test), while the coefficient on 3M-P is not, with the effect pointing even in the opposite direction. Finally, Model 3 is calculated only for the 3M treatments in which advisors were able to choose "Don't know" message. The positive and significant coefficient on the 3M-P treatment dummy suggests that the rate of evasive lying is higher when the sanction is in place.

Result 2: *The introduction of punishment increases the rate of truth-telling in the 2M-P treatment. It does not lead to more truth-telling in the 3M-P treatment, but instead to higher frequency of evasive lying.*

Hence, the effect of the fine on the rate of truth-telling is completely eroded in the 3M-P treatment, where the fine causes just a substitution of direct with evasive lying. This is a non-trivial result given that advisors do anticipate that investors are much less likely to invest after obtaining the evasive message than after obtaining "Good" message, as their elicited beliefs in the 3M-P treatment demonstrate (35.5% vs. 81.7%).[20] In particular, in the 3M-P treatment, the expected monetary utility from sending "Good" message is still higher than that from sending "Don't Know" message for 81.4% of advisors.[21] This share remains almost the same (79.6%) if one takes only those advisors who chose evasive lying. If the intrinsic lying costs of direct lying were the same as the costs of evasive lying, all of these advisors would then like to

[20] The difference in these beliefs is highly significant, also if one considers only those observations when advisors chose evasive lying (p < 0.001, one-sided permutation test).

[21] Specifically, for these advisors $10\alpha_i + 8(1-\alpha_i) > 11\beta_i + 8(1-\beta_i)$, where α_i and β_i are the elicited beliefs of a given advisor about the probability of investment conditional on receiving "Good" and "Don't know" message, respectively.

Table 5. Investment rates (in %) per message and treatment.

	2M-NP	*2M-P*	*3M-NP*	*3M-P*
"Good"-message	88.8	92.5	88.1	92.5
"Don't know"-message	71.2	74.3	50.6	38.7
"Bad"-message	3.4	14.5	3.4	3.7

forego evasive lying at least in favor of direct lying (which, in turn, can potentially be inferior to truth-telling for some of them). This would substantially limit the scope of evasive communication in the 3M-P treatment (and hence, the erosion effect). Thus, given our theoretical analysis in section 2 (in particular, Proposition 2), our results are indicative of a sufficient difference in the intrinsic costs between evasive and direct lying, to which one can attribute the failure of the fine to raise the rate of truth-telling in the 3M treatment (unlike in the 2M treatment).

Besides, advisors seem to not anticipate a drop in the credibility of the evasive message as a result of the fine, which might have further contributed to the erosion effect by reducing incentives to switch from evasion to truth-telling (see section 3.2). Specifically, the average likelihood of investment conditional on "Don't know" message assessed by advisors was 40.7% in 3M-NP and 35.5% in 3M-P (the difference is insignificant with $p = 0.100$, one-sided permutation test). This is somewhat in contrast to the drop in the actual likelihood of investment after this message from 50.6% to 38.7% (which is, however, only marginally significant as considered below).

In the next step, we focus on how investors' choices respond to the messages sent by advisors. Table 5 lists average conditional investment rates, i.e., the percentage share of cases where the investor chose to invest after seeing a particular message.

First, note that the rate of investment after "Good" message is very high throughout all treatments, despite frequent lies by advisors: in the vast majority of cases (between 88% and 93%), investors invest if the advisors choose "Good" message. Also, in the 2M-treatments the majority of investors (more than 70% of the cases) invest after "Don't know" message.[22] This percentage drops in the 3M-treatments by more than 20 (35) percentage points for the conditions without (with) punishment, and these differences are significant: p-values of one-sided permutation tests account for $p = 0.014$ ($p < 0.001$) concerning the comparison of the 2M-NP and 3M-NP (2M-P and 3M-P) treatments. This is an indication that investors are sophisticated in the sense that they anticipate the deliberate use of "Don't know" message by advisors who observe the bad state. Also, investors seem to foresee the stronger tendency of advisors to use evasive messages in the presence of sanctions as their investment rate after observing "Don't know" message is some 12 percentage points lower in the 3M-P treatment than in the 3M-NP treatment, yet this difference is not significant on conventional levels ($p = 0.063$, one-sided permutation test). Finally, there is only little investment after the "Bad" message in all treatments, which does not significantly vary between them (all one-sided permutation tests yield values of $p > 0.05$).

If we control again for subject-specific characteristics in similar parametric analyses as reported in Table 4, our results for investors' behavior become somewhat stronger

[22] We do not observe sizeable difference in investment rates between 2M-P and 2M-NP after being sent the "Don't know" message which is plausible given that the message in these treatments credibly signals to the investor that the advisor does not know the true state.

Table 6. Determinants of investment choices.

Model No.	(1)	(2)
Dependent variable	Probability of investment after "Good" message	Probability of investment after "Don't know" message
3M-P	0.492	-0.857**
	(0.709)	(0.383)
2M-NP	0.0637	1.129**
	(0.671)	(0.552)
2M-P	0.598	1.297***
	(0.697)	(0.449)
Period	-0.120***	0.091**
	(0.046)	(0.038)
Constant	4.052***	-0.378
	(0.595)	(0.352)
N	667	603

The table shows the results of Random-Effects Logit models with the dependent variables equal to one if a investor decided to invest after a particular message was used by the advisor. Robust standard errors clustered on the level of experimental cohorts are listed in parentheses. ***, ** and * indicate significance on the 1%, 5% and 10%-level, respectively.

pronounced. We calculate Random-Effects Logit models with dummy indicating whether the investor of a message chose to invest after receiving "Good" message (Model 1) and "Don't know" message (Model 2) as the dependent variable. Table 6 reports the results of the estimations.

According to Model 1, there are no treatment effects on investor behavior after receiving "Good" message, which might be attributed to the fact that the investment rate after this message is already quite high in all treatments.[23] The only notable effect is the decrease of investment after "Good" with time, which might be a response to an increasing rate of direct lying (see Table 4).

At the same time, the credibility of the evasive message "Don't know" is significantly affected by both message space and punishment variations. In particular, the significantly positive coefficients on 2M-NP and 2M-P indicate that subjects tend to trust the evasive message more in these treatments relative to the 3M game, where this message can be strategically abused by advisors. Moreover, there is a further drop in the investment rate conditional on "Don't know" message in the 3M treatment as a result of the punishment, manifested in the significantly positive coefficient on 3M-P. This suggests that the credibility of the "Don't know" message further deteriorates in the presence of the fine, as investors foresee the corresponding increase in evasive lying by advisors (i.e., Result 2). Therefore, we can state our next result:

Result 3: *Investor behavior in response to the "Good" message does not differ across the treatments. Investors are sophisticated in the sense that they seem to be able to foresee the strategic use of the evasive message by advisors.*

[23] The difference in the coefficients on 2M-NP and 2M-P is insignificant with p = 0.445 according to two-sided Wald test.

5. CONCLUSION

We have conducted an experimental communication game to test if institutional sanctions can induce more truth-telling when advisors can conceal their true information state (besides lying directly). Importantly, we find that this is not the case. Instead, if sanctions are deterministic and advisors can send evasive messages, they frequently do so, thereby circumventing punishment at a cost of lower likelihood of subsequent investment. In particular, the probability of evasive messages more than doubles after the introduction of a small punishment for direct lying. This completely offsets the positive effect of sanctions on truth-telling that emerges if the advisor has no option to tell an evasive lie.

Our theoretical analysis suggests that such behavioral pattern can only be explained if the advisor's intrinsic cost of evasive lying is lower than that of direct lying, while investors sufficiently trust both direct and evasive messages. Otherwise, evasive lying would always be dominated by direct lying, which yields a larger expected monetary benefit also after the introduction of a non-deterrent fine.[24] Inter alia, our experiment provides a direct evidence in favor of such asymmetry in lying costs, as evasive lying is chosen by a significant fraction of advisors even when there is no fine for direct lying. Overall, we conjecture that the moral wiggle room associated with evasive lying is supposed to cause the erosion of the effect of external sanctions on the rate of truth-telling.

On a more abstract level, our study allows for more thorough insights into the nature of lying behavior, suggesting that the exact character of the message matters for modelling the disutility of lying and the corresponding strategic implications. In our case, pretending to be uninformed seems to be less aversive for senders than explicitly telling a lie. Yet, it remains unclear whether the variation in intrinsic costs associated with direct and evasive lying is driven by the exogenous formulation of the message (as modeled in Kartik 2009), by the beliefs conditional on the message which may trigger guilt aversion (Battigalli et al. 2013, Khalmetski 2014), or by procedural fairness considerations (Bolton et al. 2005) due to equalizing the ex-ante chances of earning high outcomes for both players after the evasive message. Further studies are required to shed more light on this issue.

Overall, while our study considers a very abstract and stylized framework, it offers a number of implications for transactions under incomplete information in real world settings. First, our data suggest that even small institutionalized sanctions may lead to a non-negligible shift in lying behavior. However, this shift is not in the direction of more truthfulness, but rather leads to more evasive communication. Therefore, it seems questionable whether formal sanctions in the field effectively solve the problem of the exploitation of informational asymmetries. The finding from Egan et al. discussed in Section 1 that a non-negligible share of convicted financial advisors in the US remain in the industry (in fact, a third of these financial advisors are repeated offenders) seems to suggest that the threat of punishment is not deterrent.

The question that remains is which measures are suited to prevent uninformed parties to suffer from bad advice. Our study suggests that limiting the scope of evasive communication can help to increase the effectiveness of formal sanctions for lying. One obvious way to do it

[24] As noted in section 4, this does not necessarily mean that the advisors would then eventually choose direct lying, since it can in turn be dominated by truth-telling depending on the individual sensitivity to lying.

is to restrict the form of messages delivered to uninformed parties in that communication should be as explicit as possible, so that its ex-post verifiability can be improved. Another, more indirect but probably more efficient way to counteract evasive communication is to put stricter requirements on the level of expertise of professional advisors. This would make evasive claims of the advisors (like being not able to obtain or process the necessary information) less credible in front of sophisticated receivers, who then would rather attribute such claims to an attempt to conceal information.[25] As a result, the scope of evasive communication might be reduced, in turn, enhancing the ability of regulators to implement external sanctions for (verifiable) lying.

REFERENCES

Agranov, M., and A. Schotter (2012): "Ignorance is bliss: An experimental study of the use of ambiguity and vagueness in the coordination games with asymmetric payoffs," *American Economic Journal: Microeconomics*, 4(2), 77–103.

Austen-Smith, D. (1994): "Transmission of costly information," *Econometrica*, 62(4), 955–963.

Battigalli, P., G. Charness and M. Dufwenberg (2013): "Deception: The role of guilt," *Journal of Economic Behavior & Organization*, *93*, 227–232.

Beck, A., R. Kerschbamer, J. Qiu and M. Sutter (2013): "Shaping beliefs in experimental markets or expert services: Guilt aversion and the impact of promises and money-burning options," *Games and Economic Behavior*, 81, 145–164.

Bolton, G. E., J. Brandts und A. Ockenfels (2005): "Fair procedures: Evidence from games involving lotteries," *The Economic Journal*, 115(506), 1054–1076.

Cohn, A., E. Fehr and M. Marèchal (2014): "Business culture and dishonesty in the banking industry," *Nature*, 516, 86–89.

Davis, A. (2016): Commonwealth Bank chief admits about 800 customers were given the wrong financial advice TWICE," *Daily Mail Australia*, 4 October.

Dye, R. A. (1985): "Disclosure of nonproprietary information," *Journal of Accounting Research*, 23(1), 123–145.

Dziuda, W. (2011): "Strategic argumentation," *Journal of Economic Theory*, Volume 146(4), 1362–1397

Efrati, A., T. Lauricella and D. Searcey (2008). "Top broker accused of $50 billion fraud" *Wall Street Journal*, 12 December.

Egan, M., G. Matvos and A. Seru (2016): "The market for financial adviser misconduct," working paper.

Fehr, E., and K. M. Schmidt (1999): "A theory of fairness, competition and cooperation," *Quarterly Journal of Economics*, 114, 817–868.

[25] In terms of our model, in the limit case when the advisor is known to be perfectly informed ($\kappa = 1$), no evasive communication emerges in equilibrium under any level of fine.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10(2), 171−178.

Fischbacher, U. and F. Föllmi-Heusi (2013): "Lies in disguise - an experimental study on cheating," *Journal of the European Economic Association*, 11(3), 525–547.

Gibson, R., C. Tanner and A. F. Wagner (2013): "Preferences for truthfulness: Heterogeneity among and within individuals," *The American Economic Review*, 103(1), 532–548.

Gneezy, U. (2005): "Deception: The role of consequences," *The American Economic Review*, 95(1), 384–394.

Gneezy, U., B. Rockenbach and M. Serra-Garcia (2013): "Measuring lying aversion," *Journal of Economic Behavior & Organization*, 93 (2013), 293– 300.

Greiner, B. (2015): "Subject pool recruitment procedures: organizing experiments with ORSEE," *Journal of the Economic Science Association*, *1*(1), 114–125.

Grossman, S. J. (1981): "The informational role of warranties and private disclosure about product quality," *Journal of Law and Economics*, 24, 461–489.

Henning, P. J. (2016): "Prosecution of financial crisis fraud ends with a whimper", *New York Times*, 29 August.

Hooper, J. (2008): "Parmalat founds gets 10 years in prison," *The Guardian*, 19 December.

Kajackaite, A. and U. Gneezy (2015): "Lying costs and incentives," working paper.

Kartik, N. (2009): "Strategic communication with lying costs," *The Review of Economic Studies*, 76(4), 1359–1395.

Kerschbamer, R., D. Neururer, and M. Sutter (2016): "Insurance coverage of customers induces dishonesty of sellers in markets for credence goods," *Proceedings of the National Academy of Sciences of the United States of America*, 113(27), 7454–7458.

Khalmetski, K. (2014): "The hidden value of lying: Evasion of guilt in expert advice," working paper.

Khalmetski, K., and G. Tirosh (2012): "Two types of lies under different communication regimes," working paper.

Lundquist, T., T. Ellingsen, E. Gribbe and M. Johannesson (2009): "The aversion to lying," *Journal of Economic Behavior & Organization*, 70(1), 81–92.

Milgrom, P. R. (1981): "Good news and bad news: Representation theorems and applications," *Bell Journal of Economics*, 12, 380–391.

Okuno-Fujiwara, M., A. Postlewaite and K. Suzumura (1990): "Strategic information revelation," *The Review of Economic Studies*, 57(1), 25–47.

Peeters, R., M. Vorsatz and M. Walz (2013): "Truth, trust, and sanctions: On institutional selection in sender-receiver games," *The Scandinavian Journal of Economics*, 115(2), 508–523.

Rosenbaum, S. M., S. Billinger and N. Stieglitz (2014): "Let's be honest: A review of experimental evidence of honesty and truth-telling," *Journal of Economic Psychology*, 45, 181–196.

Schlag, K. H., J. Tremewan and J. J. Van der Weele (2015): "A penny for your thoughts: a survey of methods for eliciting beliefs," *Experimental Economics*, *18*(3), 457–490.

Sánchez-Pagés, S. and M. Vorsatz (2009): " Enjoy the silence: an experiment on truth-telling," *Experimental Econonomics*, 12, 220–241.

Selten, R. (1967): "Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperiments," in *Beiträge zur experimentellen Wirtschaftsforschung*, ed. by H. Sauermann. Tübingen, Germany: J.C.B. Mohr (Paul Siebeck), 136–168.

Serra-Garcia, M., E. van Damme and J. Potters (2011): "Hiding an inconvenient truth: Lies and vagueness," *Games and Economic Behavior*, 73(1), 244–261.

Sutter, M. (2009): "Deception through telling the truth?! Experimental evidence from individuals and teams," *The Economic Journal*, *119*(534), 47–60.

Xiao, E. (2013): "Profit-seeking punishment corrupts norm obedience," *Games and Economic Behavior*, 77, 321–344.

APPENDIX A. Omitted proofs.

Proof of Lemma 1. Let us first show that in any equilibrium $\eta_G \geq \max\{\eta_N, \eta_B\}$ (which would then directly lead to the claim).

First, in any equilibrium

$$\eta_G \geq \eta_B. \tag{6}$$

Indeed, assume by contradiction $\eta_G < \eta_B$. Then, no type observing *B* would send m_G which would imply $\eta_G = 1$ (by Bayes rule and the fact that m_G is used on the equilibrium path by assumption (4)) and, thus, a contradiction.[26]

Next, note that whenever some type θ' observing *B* prefers m_G over the other messages, the same type observing *G* should prefer m_G as well (since it is less costly for the latter). Hence, the share of types sending m_G while observing *B* cannot be higher than the share of types sending this message while observing *G*, which implies by Bayes rule

$$\eta_G \geq 1/2. \tag{7}$$

Finally, whenever some type θ' observing *G* prefers m_N, we have

$$U_G(\theta', m_N) \geq U_G(\theta', m_G), \tag{8}$$

[26] In what follows, by "type" we refer to the value of θ.

where U_X is the sender's expected utility while observing state $X \in \{G, B\}$. This is equivalent to

$$\pi\phi(\eta_N) - \theta' c_{EL} \geq \pi\phi(\eta_G) \geq \pi\phi(\eta_B) = U_B(\theta', m_B), \tag{9}$$

where the last inequality is by (6). Thus, type θ' observing *B* would then also prefer m_N over both m_G and m_B, so that the share of types sending m_N cannot be higher in state *G* than in state *B*. Hence,

$$\eta_N \leq 1/2. \tag{10}$$

(6), (7) and (10) together imply

$$\eta_G \geq \max\{\eta_N, \eta_B\}. \tag{11}$$

Consequently, the sender should strictly prefer m_G while observing *G* which then yields a weakly higher monetary payoff than the other messages while having a strictly lower lying cost. ■

Lemma A.1. *In any equilibrium* $\eta_G > 1/2 \geq \eta_N$.

Proof. By assumption (4) a positive share of types observing *B* should send m_B. Consequently, by Bayes rule and Lemma 1 $\eta_G > 1/2$. At the same time, $\eta_N \leq 1/2$ by (10). ■

Proof of Proposition 1. By Lemma A.1, in equilibrium there always exist types with sufficiently low θ who prefer m_G while observing *B*. At the same time, by assumption (4), there should exist sufficiently lying averse types who prefer m_B while observing *B*. Hence, there can be only 2 potential types of equilibria:

- where types observing *B* use all 3 messages (Type 1),
- where types observing *B* use only m_G and m_B (Type 2).

Let us show under which conditions each of these types of equilibria exists.

Type 1.

Since m_N is assumed to be used on the equilibrium path, it must hold that

$$c_{EL} < c_{DL} \tag{12}$$

as otherwise, by Lemma A.1, no type observing *B* would prefer m_N (which then yields a lower likelihood of investment at the same psychological cost). In this case, given that $\eta(m_B) = 0$ by Bayes rule and Lemma 1, the sender observing *B* sends m_N if and only if (given our assumptions on lexicographic preferences)

$$U_B(m_N) \geq U_B(m_G) \ \wedge \ U_B(m_N) > U_B(m_B)$$

$$\Leftrightarrow \pi\phi(\eta_N) - \theta c_{EL} \geq \pi\phi(\eta_G) - \theta c_{DL} \ \wedge \ \pi\phi(\eta_N) - \theta c_{EL} > 0$$

$$\Leftrightarrow \pi \frac{\phi(\eta_G) - \phi(\eta_N)}{c_{DL} - c_{EL}} \leq \theta < \pi \frac{\phi(\eta_N)}{c_{EL}}. \tag{13}$$

Consequently, since m_N is assumed to be sent in equilibrium, we have

$$\pi \frac{\phi(\eta_G) - \phi(\eta_N)}{c_{DL} - c_{EL}} < \pi \frac{\phi(\eta_N)}{c_{EL}}. \tag{14}$$

Analogously, the sender observing *B* prefers m_G over the other two messages if the following incentive constraints are satisfied:

$$U_B(m_G) > U_B(m_N) \ \wedge \ U_B(m_G) > U_B(m_B)$$

$$\Leftrightarrow \pi\phi(\eta_G) - \theta c_{DL} > \pi\phi(\eta_N) - \theta c_{EL} \ \wedge \ \pi\phi(\eta_G) - \theta c_{DL} > 0$$

$$\Leftrightarrow \theta < \min\left\{ \pi \frac{\phi(\eta_G) - \phi(\eta_N)}{c_{DL} - c_{EL}}, \pi \frac{\phi(\eta_G)}{c_{DL}} \right\}. \tag{15}$$

At the same time, inequality (14) implies

$$\pi \frac{\phi(\eta_G) - \phi(\eta_N)}{c_{DL} - c_{EL}} < \pi \frac{\phi(\eta_G)}{c_{DL}}. \tag{16}$$

Consequently, by (13), (15) and (16), the equilibrium where m_N is sent is characterized by the cutoffs implicitly given by

$$\theta_1 = \pi \frac{\phi(\eta_G(\theta_1)) - \phi(\eta_N(\theta_1, \theta_2))}{c_{DL} - c_{EL}}, \tag{17}$$

$$\theta_2 = \pi \frac{\phi(\eta_N(\theta_1, \theta_2))}{c_{EL}}, \tag{18}$$

such that the sender observing *B* sends m_G if $\theta \in [0, \theta_1)$, m_N if $\theta \in [\theta_1, \theta_2)$ and m_B otherwise.

Let us derive the necessary and sufficient conditions for such equilibrium to exist. Denote functions

$$\lambda_1(\theta_1, \theta_2) = \pi\phi(\eta_G(\theta_1)) - \pi\phi(\eta_N(\theta_1, \theta_2)) - \theta_1 (c_{DL} - c_{EL}), \tag{19}$$

$$\lambda_2(\theta_1, \theta_2) = \pi\phi(\eta_N(\theta_1, \theta_2)) - \theta_2 c_{EL}, \tag{20}$$

so that in equilibrium $\lambda_1(\theta_1, \theta_2) = 0$ and $\lambda_2(\theta_1, \theta_2) = 0$ by (17) and (18). These two conditions are equivalent to a single condition $\lambda_3(\theta_1) = 0$ where

$$\lambda_3(\theta_1) = \pi\phi(\eta_G(\theta_1)) - \pi\phi(\eta_N(\theta_1, \theta_2^*(\theta_1))) - \theta_1 (c_{DL} - c_{EL}), \tag{21}$$

with $\theta_2^*(\theta_1)$ implicitly given by $\lambda_2(\theta_1, \theta_2^*) = 0$. Let us show under which conditions the solution to $\lambda_3(\theta_1) = 0$ exists.

First, consider $\theta_2^*(\theta_1)$. By Bayes rule and Lemma 1,

$$\eta_N(\theta_1, \theta_2) = \frac{\Pr[m_N \mid s = G]\Pr[s = G]}{\Pr[m_N]} = \frac{(1-\kappa)0.5}{(1-\kappa) + 0.5\kappa(Z(\theta_2) - Z(\theta_1))}, \tag{22}$$

where Z is the cumulative distribution function of θ. Thus, $\eta(m_N \mid \theta_1, \theta_2)$, and hence $\lambda_2(\theta_1, \theta_2)$, is strictly decreasing in θ_2 for given θ_1. Consequently, by the intermediate value theorem, the unique value of $\theta_2^* > \theta_1$ solving $\lambda_2(\theta_1, \theta_2) = 0$ (for given θ_1) exists if and only if $\lambda_2(\theta_1, \theta_1) > 0$, i.e., $\lambda_2(\theta_1, \theta_2)$ is positive at a minimum possible value of $\theta_2 = \theta_1$ (otherwise, $\lambda_2(\theta_1, \theta_2) < 0$ for any $\theta_2 > \theta_1$).[27] This condition is equivalent to

$$\begin{aligned} &\pi\phi(\eta_N(\theta_1, \theta_1)) - \theta_1 c_{EL} > 0 \\ &\Leftrightarrow \pi\phi(0.5) - \theta_1 c_{EL} > 0 \\ &\Leftrightarrow \theta_1 < \pi\frac{\phi(0.5)}{c_{EL}}. \end{aligned} \tag{23}$$

Next, consider $\lambda_3(\theta_1) = 0$. By the chain rule,

$$\frac{\partial \lambda_3(\theta_1)}{\partial \theta_1} = \pi\phi'(\eta_G)\frac{\partial \eta_G}{\partial \theta_1} - \pi\phi'(\eta_N)\left(\frac{\partial \eta_N}{\partial \theta_1} + \frac{\partial \eta_N}{\partial \theta_2^*}\frac{\partial \theta_2^*}{\partial \theta_1}\right) - (c_{DL} - c_{EL}). \tag{24}$$

Consider each term of the right-hand side of (24). We have

$$\eta_G(\theta_1) = \frac{\Pr[m_G \mid s = G]\Pr[s = G]}{\Pr[m_G]} = \frac{\kappa 0.5}{\kappa(0.5 + 0.5Z(\theta_1)} = \frac{1}{1 + Z(\theta_1).}. \tag{25}$$

Hence, the first term on the right-hand side of (24) is negative. Consider the second term. By the implicit function theorem, $\lambda_2(\theta_1, \theta_2^*) = 0$ yields

$$\frac{\partial \theta_2^*}{\partial \theta_1} = -\frac{\partial \lambda_2 / \partial \theta_1}{\partial \lambda_2 / \partial \theta_2^*} = -\frac{\pi\phi'(\eta_N)\frac{\partial \eta_N}{\partial \theta_1}}{\pi\phi'(\eta_N)\frac{\partial \eta_N}{\partial \theta_2^*} - c_{EL}}. \tag{26}$$

Then,

[27] We need the strict condition $\theta_2^* > \theta_1$ since if $\theta_2^* = \theta_1$, then the cutoff type is indifferent between all three messages, thus choosing m_B by our assumption on lexicographic preferences. In this case, m_N is never sent.

$$\frac{\partial \eta_N}{\partial \theta_1}+\frac{\partial \eta_N}{\partial \theta_2^*}\frac{\partial \theta_2^*}{\partial \theta_1}=\frac{\partial \eta_N}{\partial \theta_1}-\frac{\partial \eta_N}{\partial \theta_2^*}\frac{\pi\phi'(\eta_N)\frac{\partial \eta_N}{\partial \theta_1}}{\pi\phi'(\eta_N)\frac{\partial \eta_N}{\partial \theta_2^*}-c_{EL}}$$

$$=\frac{\partial \eta_N}{\partial \theta_1}\left(1-\frac{\pi\phi'(\eta_N)\frac{\partial \eta_N}{\partial \theta_2^*}}{\pi\phi'(\eta_N)\frac{\partial \eta_N}{\partial \theta_2^*}-c_{EL}}\right)>0, \tag{27}$$

where the inequality follows due to $\frac{\partial \eta_N}{\partial \theta_1}>0$ and $\frac{\partial \eta_N}{\partial \theta_2^*}<0$ by (22). Hence, the second term on the right-hand side of (24) is again negative. Thus, all terms in the right-hand side of (24) are negative so that

$$\frac{\partial \lambda_3(\theta_1)}{\partial \theta_1}<0. \tag{28}$$

At the same time, $\lambda_3(0)=\pi\phi(\eta_G)-\pi\phi(\eta_N)>0$ by Lemma A.1. Consequently, by the intermediate value theorem, there exists a unique equilibrium value of θ_1 solving $\lambda_3(\theta_1)=0$ while satisfying (23) if and only if

$$\lambda_3\left(\pi\frac{\phi(0.5)}{c_{EL}}\right)<0. \tag{29}$$

It is easy to verify that $\theta_2^*\left(\pi\frac{\phi(0.5)}{c_{EL}}\right)=\pi\frac{\phi(0.5)}{c_{EL}}$. Hence,

$$\lambda_3\left(\pi\frac{\phi(0.5)}{c_{EL}}\right)<0$$

$$\Leftrightarrow \pi\phi\left(\eta_G\left(\pi\frac{\phi(0.5)}{c_{EL}}\right)\right)-\pi\phi\left(\eta_N\left(\pi\frac{\phi(0.5)}{c_{EL}},\pi\frac{\phi(0.5)}{c_{EL}}\right)\right)$$
$$-\pi\frac{\phi(0.5)}{c_{EL}}(c_{DL}-c_{EL})<0$$

$$\Leftrightarrow \pi\phi\left(\eta_G\left(\pi\frac{\phi(0.5)}{c_{EL}}\right)\right)-\pi\phi(0.5)-\pi\frac{\phi(0.5)}{c_{EL}}(c_{DL}-c_{EL})<0$$

$$\Leftrightarrow c_{DL}/c_{EL}>\omega(c_{EL}) \tag{30}$$

with

$$\omega(c_{EL})=\phi\left(\eta_G\left(\pi\frac{\phi(0.5)}{c_{EL}}\right)\right)\frac{1}{\phi(0.5)}>1, \tag{31}$$

where the inequality follows by Lemma A.1. Thus, the considered equilibrium exists if and only if the ratio c_{DL}/c_{EL} is sufficiently larger than 1.

Type 2.

Assume that no type observing *B* sends m_N in equilibrium. Then, the sender sends m_G while observing *B* if and only if

$$U_B(m_G) > U_B(m_B)$$

$$\Leftrightarrow \pi\phi(\eta_G) - \theta c_{DL} > 0 \tag{32}$$

while sending m_B otherwise. This implies that there is a cutoff θ^{**} separating two cases implicitly given by

$$\pi\phi(\eta_G(\theta^{**})) - \theta^{**} c_{DL} = 0. \tag{33}$$

It is easy to verify that a cutoff in $(0, \bar{\theta})$ solving (33) always exists (by the intermediate value theorem). At the same time, in the considered equilibrium the sender observing *B* should also never prefer m_N over both m_G and m_B. Clearly, this is always the case if $c_{DL} = c_{EL}$ (when the sender always prefers m_G to m_N by Lemma A.1). If $c_{DL} > c_{EL}$, we must have that incentive constraint (13) implying the existence of types preferring m_N is never satisfied. Given that in the considered equilibrium $\eta_N = 0.5$, this is equivalent to

$$\pi \frac{\phi(\eta_G(\theta^{**})) - \phi(0.5)}{c_{DL} - c_{EL}} \geq \pi \frac{\phi(\eta_G(\theta^{**}))}{c_{DL}}$$

$$\Leftrightarrow \phi(\eta_G(\theta^{**}))c_{EL} \geq \phi(0.5)c_{DL}. \tag{34}$$

Substituting for $\phi(\eta_G(\theta^{**}))$ from (33) we obtain

$$\frac{\theta^{**} c_{DL}}{\pi} c_{EL} \geq \phi(0.5)c_{DL}$$

$$\Leftrightarrow \theta^{**} \geq \pi \frac{\phi(0.5)}{c_{EL}}. \tag{35}$$

Let us show when this is the case. Denote

$$\lambda_4(\theta^{**}) = \pi\phi(\eta_G(\theta^{**})) - \theta^{**} c_{DL} \tag{36}$$

so that in equilibrium $\lambda_4(\theta^{**}) = 0$ by (33). Given that λ_4 is strictly decreasing in θ^{**} while $\lambda_4(\bar{\theta}) < 0$ by (4), by the intermediate function theorem, (35) holds if and only if

$$\lambda_4\left(\pi \frac{\phi(0.5)}{c_{EL}}\right) \geq 0$$

$$\Leftrightarrow \pi\phi\left(\eta_G\left(\pi\frac{\phi(0.5)}{c_{EL}}\right)\right)-\pi\frac{\phi(0.5)}{c_{EL}}c_{DL}\geq 0$$

$$\Leftrightarrow \frac{c_{DL}}{c_{EL}}\leq \omega(c_{EL}), \tag{37}$$

which is then a necessary and sufficient condition for the existence of the considered equilibrium. ■

Proposition A.1. *For any given parameter values, there exists a threshold level of punishment* f *such that an equilibrium with a non-deterrent level of punishment exists if and only if* $f < f$. *Such equilibrium is unique. Hereby:*

a) If $c_{EL}=c_{DL}$, *then the equilibrium is characterized by* $0<\theta^{**}<\bar{\theta}$ *such that the sender observing B never sends* m_N, *sends* m_G *if* $\theta\in[0,\theta^{**})$, *and* m_B *if* $\theta\in[\theta^{**},\bar{\theta}]$.

b) If $c_{EL}<c_{DL}$:

i) If $\frac{c_{DL}}{c_{EL}}>\frac{\pi-f}{\pi}\omega(c_{EL})$, *then the equilibrium is characterized by* $0<\theta_1<\theta_2<\bar{\theta}$ *such that the sender observing B sends* m_G *if* $\theta\in[0,\theta_1)$, m_N *if* $\theta\in[\theta_1,\theta_2)$, *and* m_B *if* $\theta\in[\theta_2,\bar{\theta}]$.

ii) If $\frac{c_{DL}}{c_{EL}}\leq\frac{\pi-f}{\pi}\omega(c_{EL})$, *then the equilibrium is characterized by* $0<\theta^{**}<\bar{\theta}$ *such that the sender observing B never sends* m_N, *sends* m_G *if* $\theta\in[0,\theta^{**})$, *and* m_B *if* $\theta\in[\theta^{**},\bar{\theta}]$.

Proof. Let us characterize all possible equilibria with a non-deterrent level of fine (i.e., with a positive equilibrium rate of direct lying). As in the case without the fine, it applies that types with sufficiently high θ observing *B* should always tell the truth. Hence, there can be only two possible equilibria with a positive rate of direct lying:

- where types observing *B* use all 3 messages (Type 1),
- where types observing *B* use only m_G and m_B (Type 2).

Let us consider the necessary and sufficient conditions for the existence of each type of equilibrium depending on whether $c_{EL}=c_{DL}$ (Case 1) or $c_{EL}<c_{DL}$ (Case 2).

Case 1: $c_{EL}=c_{DL}$.

In this case, for any sender type

$$U_B(m_G)-U_B(m_N)=(\pi-f)\phi(\eta_G)-\pi\phi(\eta_N). \tag{38}$$

Hence, all types observing *B* should prefer either m_G over m_N (if $(\pi - f)\phi(\eta_G) > \pi\phi(\eta_N)$) or m_N over m_G (otherwise, by the assumption on lexicographic preferences). Consequently, the equilibrium with a non-deterrent fine in Case 1 can only be of Type 2. Then, the sender sends m_G while observing *B* if and only if

$$\begin{aligned} & U_B(m_G) > U_B(m_B) \\ & \Leftrightarrow \pi\phi(\eta_G) - \theta c_{DL} > 0 \end{aligned} \tag{39}$$

while sending m_B otherwise. This implies that there is a cutoff θ^{**} separating two cases implicitly given by

$$(\pi - f)\phi(\eta_G(\theta^{**})) - \theta^{**} c_{DL} = 0. \tag{40}$$

At the same time, no type observing *B* should have incentive to send m_N. Given that in the considered equilibrium $\eta_N = 0.5$, this is equivalent to

$$\begin{aligned} & (\pi - f)\phi(\eta_G(\theta^{**})) > \pi\phi(0.5) \\ & \Leftrightarrow (\pi - f)\frac{\theta^{**} c_{DL}}{\pi - f} > \pi\phi(0.5) \\ & \Leftrightarrow \theta^{**} > \pi\frac{\phi(0.5)}{c_{DL}}, \end{aligned} \tag{41}$$

where the second inequality follows from (40). Let us show when this is the case. Denote

$$\lambda_4(\theta^{**}) = (\pi - f)\phi(\eta_G(\theta^{**})) - \theta^{**} c_{DL} \tag{42}$$

so that in equilibrium $\lambda_4(\theta^{**}) = 0$ by (40). Given that λ_4 is strictly decreasing in θ^{**} while $\lambda_4(\bar{\theta}) < 0$ by (4), by the intermediate function theorem, (41) holds if and only if

$$\begin{aligned} & \lambda_4\left(\pi\frac{\phi(0.5)}{c_{DL}}\right) > 0 \\ & \Leftrightarrow (\pi - f)\phi\left(\eta_G\left(\pi\frac{\phi(0.5)}{c_{DL}}\right)\right) - \pi\frac{\phi(0.5)}{c_{DL}} c_{DL} > 0 \\ & \Leftrightarrow f < \pi\left(1 - \frac{\phi(0.5)}{\phi\left(\eta_G\left(\pi\frac{\phi(0.5)}{c_{DL}}\right)\right)}\right), \end{aligned} \tag{43}$$

which is then a necessary and sufficient condition for the existence of the considered equilibrium (note that the right-hand side of (43) is always positive). Thus, if $c_{DL} = c_{EL}$, a unique equilibrium (of Type 2) exists if and only if f is below a certain threshold.

Case 2. $c_{EL} < c_{DL}$.

In this case, one can show that both types of equilibrium are possible. Let us derive the necessary and sufficient conditions for the existence of each type of equilibrium in the considered case.

Type 1.

By the same arguments as in the proof of Proposition 1 (for the equilibrium of Type 1) the considered equilibrium is characterized by cutoffs θ_1 and θ_2 such that θ_1 solves $\lambda_5(\theta_1) = 0$ where

$$\lambda_5(\theta_1) = (\pi - f)\phi(\eta_G(\theta_1)) - \pi\phi(\eta_N(\theta_1, \theta_2^*(\theta_1))) - \theta_1 (c_{DL} - c_{EL}), \tag{44}$$

with $\theta_2 = \theta_2^*(\theta_1)$ implicitly given by

$$\pi\phi(\eta_N(\theta_1, \theta_2^*(\theta_1))) - \theta_2^*(\theta_1) c_{EL} = 0. \tag{45}$$

By the same arguments as in the proof of Proposition 1, we have that the equilibrium exists if and only if

$$\lambda_5(0) > 0, \tag{46}$$

$$\lambda_5\left(\pi \frac{\phi(0.5)}{c_{EL}}\right) < 0. \tag{47}$$

The first condition is equivalent to

$$\begin{aligned} &(\pi - f)\phi(\eta_G(0)) - \pi\phi(\eta_N(0, \theta_2^*(0))) > 0 \\ &\Leftrightarrow (\pi - f) - \pi\phi(\eta_N(0, \theta_2^*(0))) > 0 \\ &\Leftrightarrow f < \pi(1 - \phi(\eta_N(0, \theta^{***}))), \end{aligned} \tag{48}$$

where θ^{***} is implicitly given by

$$\pi\phi(\eta_N(0, \theta^{***})) - \theta^{***} c_{EL} = 0. \tag{49}$$

(It is easy to verify that $\theta^{***} \in (0, \bar{\theta})$ solving (49) always exists by the intermediate value theorem.)

At the same time, analogously as with (30), the second equilibrium condition (47) is fulfilled if and only if

$$\frac{c_{DL}}{c_{EL}} > \frac{\pi - f}{\pi} \omega(c_{EL}). \tag{50}$$

Thus, the necessary and sufficient conditions for the considered type of equilibrium (under $c_{EL} < c_{DL}$) are (48) (i.e., f is below a certain threshold) and (50).

Type 2.

Analogously to the proof of Proposition 1 (for the equilibrium of Type 2), one can show that the corresponding equilibrium exists if and only if

$$\frac{c_{DL}}{c_{EL}} \le \frac{\pi - f}{\pi}\omega(c_{EL}). \tag{51}$$

Summing up the necessary and sufficient conditions for the existence of each type of equilibrium in Cases 1 and 2 leads to the claim of the proposition. ■

Proof of Proposition 2. Consider 3 separate cases depending on the possible parameter values:

- $c_{EL} = c_{DL}$ *or* $\frac{c_{DL}}{c_{EL}} \le \frac{\pi - f}{\pi}\omega(c_{EL})$ (Case 1).
- $c_{EL} < c_{DL}$ *and* $\frac{\pi - f}{\pi}\omega(c_{EL}) < \frac{c_{DL}}{c_{EL}} \le \omega(c_{EL})$ (Case 2).
- $c_{EL} < c_{DL}$ *and* $\frac{c_{DL}}{c_{EL}} > \omega(c_{EL})$ (Case 3).

Case 1: $c_{EL} = c_{DL}$ *or* $\frac{c_{DL}}{c_{EL}} \le \frac{\pi - f}{\pi}\omega(c_{EL})$.

Then, by Propositions 1 and A.1, both before and after the fine the equilibrium is characterized by single cutoff θ^{**} separating direct lying and truth-telling after observing *B*. Analogously to (33), this cutoff is given by

$$(\pi - f)\phi(\eta_G(\theta^{**})) - \theta^{**}c_{DL} = 0. \tag{52}$$

Then, by the implicit function theorem,

$$\frac{\partial\theta^{**}}{\partial f} = \frac{\phi(\eta_G(\theta^{**})}{(\pi - f)\phi'(\eta_G)\eta_G'(\theta^{**}) - c_{DL}}. \tag{53}$$

Since $\eta_G'(\theta^{**}) < 0$ by (25), we obtain that $\frac{\partial\theta^{**}}{\partial f} < 0$, i.e., the sender switches from direct lying to truth-telling once *f* increases, while the rate of evasion stays at 0.

Case 2: $c_{EL} < c_{DL}$ *and* $\frac{\pi - f}{\pi}\omega(c_{EL}) < \frac{c_{DL}}{c_{EL}} \le \omega(c_{EL})$.

Then, by Propositions 1 and A.1, the equilibrium is characterized by single cutoff θ^{**} (separating m_G from m_B in state *B*) before the fine, and by cutoffs θ_1 and θ_2 after the fine. Then, the claim for the rate of evasive lying follows directly, as the equilibrium is characterized by a positive rate of evasion after the fine and by 0 rate of evasion before the fine.

Consider the change in the lying rate as a result of the fine. We have

$$\pi\phi(\eta_G(\theta_1)) - \theta_1 c_{DL} > (\pi - f)\phi(\eta_G(\theta_1)) - \theta_1 c_{DL} = \pi\phi(\eta_N(\theta_1, \theta_2)) - \theta_1 c_{EL}$$

$$> \pi\phi(\eta_N(\theta_1, \theta_2)) - \theta_2 c_{EL} = 0 = \pi\phi(\eta_G(\theta^{**})) - \theta^{**} c_{DL}$$

$$\Leftrightarrow \theta_1 < \theta^{**}, \tag{54}$$

where the equalities follow from the corresponding indifference conditions which should hold at the cutoffs in equilibrium, and the last inequality follows due to the fact that function $\pi\phi(\eta_G(\theta)) - \theta c_{DL}$ is strictly decreasing in θ. Hence, the rate of lying is strictly lower after the fine.

Finally, let us show that the rate of truth-telling strictly increases as a result of the fine. Note that this rate is characterized by θ^{**} before the fine, and by θ_2 after the fine, where these cutoffs are implicitly given by

$$\theta^{**} = \pi \frac{\phi(\eta_G(\theta^{**}))}{c_{DL}}, \tag{55}$$

$$\theta_2 = \pi \frac{\phi(\eta_N(\theta_1, \theta_2))}{c_{EL}}. \tag{56}$$

At the same time,

$$\pi\phi(\eta_G(\theta^{**})) - \theta^{**} c_{DL} = 0 \geq U_B(\theta^{**}, m_N) = \pi\phi(0.5) - \theta^{**} c_{EL}, \tag{57}$$

where the first equality is by (55), and the inequality is by incentive compatibility, as otherwise the sender would strictly prefer to send m_N at least at θ^{**}, which is a contradiction by the initial assumption of the case and Proposition 1. Hence,

$$\theta^{**} \geq \pi \frac{\phi(0.5)}{c_{EL}} > \pi \frac{\phi(\eta_N(\theta_1, \theta_2))}{c_{EL}} = \theta_2, \tag{58}$$

where the first inequality is by (57), the second inequality is by (22), and the equality is by (56). Given that the rate of truth-telling is $1 - Z(\theta^{**})$ before the fine and $1 - Z(\theta_2)$ after the fine, it increases as a result of the fine.

Case 3: $c_{EL} < c_{DL}$ *and* $\frac{c_{DL}}{c_{EL}} > \omega(c_{EL})$.

By Propositions 1 and A.1, both before and after the introduction of the fine the equilibrium is given by cutoffs θ_1 and θ_2, characterizing the messaging strategy of the sender observing *B*. These cutoffs are implicitly given by $\lambda_5(\theta_1) = 0$ and $\theta_2 = \theta_2^*(\theta_1)$ where

$$\lambda_5(\theta_1) = (\pi - f)\phi(\eta_G(\theta_1) - \pi\phi(\eta_N(\theta_1, \theta_2^*(\theta_1))) - \theta_1(c_{DL} - c_{EL}), \tag{59}$$

$$\theta_2^*(\theta_1) = \frac{\pi\phi(\eta_N(\theta_1, \theta_2^*(\theta_1)))}{c_{EL}} \tag{60}$$

with $f = 0$ before the fine is introduced (see (21) and (44)). Then, the rates of direct lying, truth-telling and evasive lying are given by, respectively,

$$L = Z(\theta_1), \tag{61}$$

$$T = 1 - Z(\theta_2), \tag{62}$$

$$E = Z(\theta_2) - Z(\theta_1). \tag{63}$$

We have

$$\frac{\partial L}{\partial f} = Z'(\theta_1)\frac{\partial \theta_1}{\partial f} = -Z'(\theta_1)\frac{\partial \lambda_5 / \partial f}{\partial \lambda_5 / \partial \theta_1}, \tag{64}$$

where the last equality is by the implicit function theorem. One can show that $\frac{\partial \lambda_5}{\partial \theta_1} < 0$ (by the same arguments as the proof of (28)) while

$$\frac{\partial \lambda_5}{\partial f} = -\phi(\eta_G(\theta_1)) < 0. \tag{65}$$

Substituting this into (64), we obtain that the rate of direct lying strictly decreases with *f*.

Next, consider the rate of truth-telling. We have

$$\frac{\partial T}{\partial f} = -Z'(\theta_2)\frac{\partial \theta_2}{\partial f} = -Z'(\theta_2)\frac{\partial \theta_2^*}{\partial \theta_1}\frac{\partial \theta_1}{\partial f}. \tag{66}$$

Since $\frac{\partial \theta_1}{\partial f} < 0$ as shown in (64) and $\frac{\partial \theta_2^*}{\partial \theta_1} > 0$ by (26) and (22), we obtain that the rate of truth-telling strictly increases with the fine.

Finally, consider the rate of evasive lying. By (60),

$$\phi(\eta_N(\theta_1, \theta_2)) = \theta_2 \frac{c_{EL}}{\pi}. \tag{67}$$

Consequently, since $\frac{\partial \theta_2}{\partial f} < 0$ as shown above, we obtain that $\phi(\eta_N(\theta_1, \theta_2))$ and hence $\eta_N(\theta_1, \theta_2)$ are strictly decreasing with *f*. Then, by (22) the term $Z(\theta_2) - Z(\theta_1)$, i.e., the rate of evasive lying, is strictly increasing with *f*.

Combining the results of Cases 1-3 leads to the claim of the proposition. ■

APPENDIX B. Experimental instructions

Below you find experimental instructions translated from German. Participants received written copies of these instructions prior to the start of the experiment.

General Information

Welcome and thank you for your participation in this experiment.

Please do not communicate with the other participants from now until the end of the experiment. We also ask you to switch off your mobile phone during the experiment. If you do not comply with these rules, we have to exclude you from the experiment and all payoffs.

Please read the instructions carefully. If you have questions after reading them or during the experiment, please raise your hand. One of the experimenters will come to you and answer your questions individually.

Your payoff and your decisions will be treated confidentially. None of the participants will get to know during or after the experiment with whom he interacted and which payoffs other participants receive. Your decisions are hence anonymous.

You can earn money in this experiment. How much you earn depends on your decisions as well as on the decisions of the other participants. Your payoff will be paid to you in cash after the end of the experiment. You receive 2,50 Euro for your participation independently from the decisions in the experiment.

All participants receive identical instructions in this experiment.

Instructions

Procedures of the experiment

You get assigned a fixed role in the experiment: Advisor or Investor. The role assignment will be done randomly and persists for the whole experiment.

The experiment consists of 10 periods in which you have to make decisions. At the end of the experiment your payoffs in ECU from all rounds will be summed up, converted into Euro and paid out to you. The conversion rate here is 12 ECU = 1 Euro. The earnings from the experiment will be paid to you together with the 2,50 Euro for your participation.

Each period proceeds as follows:

At the beginning of each period one advisor and one investor will be matched to one another. It is ensured in the matching that there the same participants will never interact in two consecutive periods.

The investor decides if he “invests” or “does not invest”. The payoffs that the investor and advisor receive in this period depend on this decision.

If the investor „invests“, his payoff additionally depends on the investment conditions („Good“ or „Bad“) that apply in this period. The investment conditions are determined randomly in each period; the probability that the investment conditions are „Good“ or „Bad“ are 50% respectively.

If the investor decides whether he invests or not, he however does not know which investment conditions apply in this period.

Payoffs

The payoffs for the investor and advisor in one period are determined as follows:

		Investment conditions	
		"Good"	"Bad"
Decision of the investor	Do not invest	*Payoff Investor = 6* *Payoff Advisor = 8*	*Payoff Investor = 6* *Payoff Advisor = 8*
	Invest	*Payoff Investor = 12* *Payoff Advisor = 11*	*Payoff Investor = 3* *Payoff Advisor = 11*

Investor:

- If the investor does not invest, his payoff in this period is 6 ECU.
- If the investor invests, his payoff in this period is 12 ECU, if the investment conditions are „Good“.
- If the investor invests, his payoff in this period 3 ECU, if the investment conditions are „Bad“.

Advisor:

- If the investor does not invest, the payoff of the advisor is 8 ECU.
- If the investor does invest, the payoff of the advisor is 11 ECU, independently from whether the investment conditions are „Good“ or „Bad“.

Messages

The advisor gets informed in each period with the probability of 60% whether the investment conditions are „Good“ or „Bad“ in this period. This means that the advisor knows in approximately 6 out of 10 cases which state („Good“ or „Bad“) actually occurred in this period. With the probability 40% (or in approximately 4 out of 10 cases) the advisor does not know which state (“Good” or “Bad”) actually occurred in this period.

Before the investor decides whether he invests or not he receives a message from the advisor. First, the advisor determines for both of the following cases a message to the investor:

- The message that he wants to send if he later gets to know that the investment conditions are “Good”.
- The message that he wants to send if he later gets to know that the investment-conditions are “Bad”.

He can choose freely out of two [*in the 3M treatments*: three] messages

- *"I believe that the investment conditions in this period are GOOD."*
- *"I believe that the investment conditions in this period are BAD."*
- [*in the 3M treatments:*] *"I do not know the investment conditions in this period."*

Then the investment conditions are drawn randomly. If the advisor gets informed about the investment conditions (with a probability of 60 %), the investor receives the respective message that was chosen by the advisor.

If the advisor *does not* get informed about the investment conditions (with a probability of 40%), the investor *automatically* (i.e. independent of the decision of the advisor) receives the following message from the advisor:

- *"I do not know the investment conditions in this period."*

Fines [*only in the 2M-P and 3M-P treatments*]

If the investment conditions that the advisor communicated to the investor differ from the actual investment conditions and the investor invested after receiving the message, 1 ECU will be subtracted automatically from the payoff of the advisor after the period. This means that the amount gets subtracted from the advisor, if the investor invested and

- …either the investment conditions in the period were „Bad“ and the investor received the message „*I believe that the investment conditions in this period are GOOD*“ from the advisor
- …or the investment conditions in this period were „GOOD“ and the investor received the message *„I believe that the investment conditions in this period are BAD“* from the advisor

If the investor receives the message „*I do not know the investment-conditions in this period“*, the advisor gets no subtraction in any case [*in the 3M-P treatment:* even if he knows the investment conditions of this period].

Estimations

Before the investor and the advisor make their decisions, they are asked to estimate the behavior of other participants. The more accurate the estimation is, the higher is the payoff

they can achieve with the estimation. Details will be explained to you at the screen. The payoff possibilities of investors and advisors related to the estimation questions are identical.

End of the period

After each period both the advisor and the investor get informed about their own payoff as well as about the payoff of the matched participant that result from the investment decision of the investor. [*in the 3M treatments:* However, the investor does not get informed whether or not the advisor knew the investment conditions.]

This is the end of the instructions of the experiment. If you have questions, please raise your hand. If you understood the instructions entirely and have no further questions, please press the button "Ready".