UNIVERSITY OF BATH

Citation for published version:
Khalmetski, K, Rockenbach, B & Werner, P 2017, 'Evasive lying in strategic communication', *Journal of Public Economics*, vol. 156, pp. 59-72. https://doi.org/10.1016/j.jpubeco.2017.10.002

DOI:
10.1016/j.jpubeco.2017.10.002

Publication date:
2017

Document Version
Peer reviewed version

Evasive Lying in Strategic Communication*

Kiryl Khalmetski,† Bettina Rockenbach,‡ Peter Werner§

5 October 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. Evidence from our experimental data and a survey conducted with additional participants indicate that the shift towards evasion can indeed be attributed to lower psychological costs. Overall, 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 effective.

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

JEL classification: C91, D82, D83

---

* Accepted for publication in *Journal of Public Economics*. We thank the Editor and two anonymous referees for very helpful comments and suggestions. Financial support of the German Research Foundation (DFG) through the Research Unit "Design & Behavior" (FOR 1371) is gratefully acknowledged.

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

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

§ 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 and may create substantial market inefficiencies. 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.

Due to the often 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. In particular, Malmendier and Shanthikumar (2007) provide evidence that small investors do not seem to discount the inherent upward bias in recommendations, indicating that they place too much trust in analysts. 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 might be especially prone to fraud are credence goods, such as medical service or repair services, where the customer is both unsure about her need ex-ante and unable to assess the value of a good ex-post. 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).

That markets with information asymmetries will generally not lead to Pareto-efficient outcomes calls for regulations to increase efficiency and prevent the exploitation of market participants (Stiglitz 2009). In case of the financial sector, regulators have put particular emphasis on new laws to overcome information problems in recent years, which has been manifested in the US, for example, “by the most important US consumer financial legislation since the New Deal of the 1930s” (Campbell 2016).[1] An obvious question, however, is if and to what extent stricter regulations by the government can indeed lower the incentives of better informed actors to exploit their informational advantage. Stiglitz (2009) suggests tort law as a means to align incentives of participants in the market for financial products, acknowledging, however, that there are potential problems that might lower its effectiveness. For example, damages due to financial fraud are compensated only ex-post, and in many cases, it might even be impossible to receive compensation at all. Also, often it might be difficult for harmed investors to coordinate in order to file a suit. Moreover, Barr et al. (2009) and Campbell (2016) state that in the field of consumer finance, regulations that attempt to remove information asymmetries by imposing ex-ante disclosure of information might not lead to better financial decision-making. The reason is that it is not clear whether customers are sufficiently able to make use of this information. Also, in the face of ex-ante regulation, financial service providers still have the incentive and the means to manipulate clients' decisions and thus to circumvent the regulation.

Indeed, empirical evidence from the finance industry seems to suggest that sanctions for financial misconduct eventually appear to be not very effective in abolishing fraud. In fact, in their study of the US retail finance industry mentioned above, 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).

One reason why implementing appropriate sanctioning schemes for lying might be challenging for regulators is that 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).[2] 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

[1] Notable examples include Pension Protection Act of 2006, the Credit Card Accountability Responsibility and Disclosure Act (CARD Act) of 2009, and the Dodd-Frank Act of 2010.

[2] 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.

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-telling and direct lying. Our experimental treatments introduce the availability of the option of evasive lying, and besides 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, together with a survey study on classifying the severity of different forms of lying, supports 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. Moreover, 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. Our results highlight the necessity to integrate behavioral motivations (in particular, the variation in the intrinsic lying costs depending on the form of the message) into the design of policy regulations of 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, Irlenbusch and Villeval 2015 and Abeler et al. 2016 for surveys of the experimental literature 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 in 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 2017).[3] 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 (Kajackaite and Gneezy 2017), senders' beliefs about the receiver's behavior or beliefs (Sutter 2009, Charness and Dufwenberg 2006, Beck et al. 2013, Gneezy et al. 2016), form of communication (Lundquist et al. 2009), or game experience (Gneezy et al. 2013). Gächter and Schulz (2016) provide evidence from a large cross-country sample that social norms have a crucial role for intrinsic honesty: Experimental measures for honesty are correlated with an index measure for the prevalence of rule violation on the country level.

Our study is mainly related to two groups of studies in this literature: First, several papers have studied the effect of using vague messages or refraining from communication. These studies show that subjects indeed make use of these options to circumvent both explicit lying and truth-telling. Serra-Garcia et al. (2011) conduct a sequential two-person public good game in which only one player is informed about the value of the public good. In situations where the public good has an intermediate value, vague communication helps to induce cooperation and in this sense has a positive impact. In the study by Agranov and Schotter (2012), an "announcer" has to communicate to two agents who have a common interest to coordinate. In this setting, Agranov and Schotter show that coarser communication can improve the ability of agents to coordinate if doing so leads to payoff inequalities. Khalmetski and Tirosh (2012) consider the effect of the form of communication on the rates of both direct and evasive lies in a setting close to the present study and find that, unlike direct lying, the rate of evasive lying is not affected by whether messages are prefabricated or freely chosen by subjects.[4] In a recent paper, Jin et al. (2016) conduct a sender-receiver game in which the sender can either reveal his private information truthfully or remain silent. While in sequential equilibrium, senders should reveal their information in all states (except for the lowest state in which they are indifferent between reporting and remaining silent), the authors find a substantially higher share of senders who remain silent than predicted, which is driven by too optimistic beliefs of the receiver after observing no message.

Second, our study is related to research that investigates the effect of sanctions on lying behavior. In a seminal study, Brandts and Charness (2003) observe that subjects become more willing to induce costly punishment for an unfavorable payoff distribution if this distribution was reached through a deceptive message by another player. Sánchez-Pagés and Vorsatz (2007) conduct a sender-receiver game and provide evidence that receivers are willing to

[3] Kartik (2009) theoretically analyses the effect of lying costs on strategic communication.

[4] Analyzing the same setting, Chater et al. (2010) mainly focus on the effect of conflict of interest between the advisor and the investor.

punish lies at a cost. The study suggests that senders are of consistent "types", as subjects who punish more frequently are also those who send truthful messages more often than predicted by theory. Ohtsubo et al. (2011) conduct trust games in which an external participant decides on the punishment for a trustee depending on the amount returned, while the trustee has an option to send a message to a trustor in advance. Here, the authors find that punishment increases if the amount returned deviates from a promised amount even if the resulting allocation is fair. Moreover, Peeters et al. (2013) let subjects interact in a sender-receiver game 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 the coexistence of groups with and without sanctions, which is driven by individual heterogeneity concerning 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.

Most closely related to our study is the paper by Sánchez-Pagés and Vorsatz (2009, SPV in the following) who conduct an experimental sender-receiver game in which both parties have opposing interests. Similar to our study, senders are allowed to stay silent (however at a cost) instead of sending an explicit message. Furthermore, receivers are endowed with a costly punishment option. The authors find that a share of subjects remains silent. Moreover, lies are punished with a non-negligible probability when the receiver has trusted a message in the first step. Besides, similarly to our results, SPV show that the introduction of punishment does not increase the rate of truth-telling on average while senders switch to staying silent somewhat more often.

At the same time, the study by SPV differs from ours in several important aspects. First, the goal of the study is to explain excessive truth-telling whereas our study rather focuses on the nature of lying behavior. In the formal framework discussed in SPV, the authors compare a sender's preference for truth-telling to lying aversion and conclude that their results are consistent with the latter motivation: Lying-averse senders either choose whether to communicate or to stay silent which avoids lying. On the contrary, in our model senders always have to send an explicit message and thus have to decide whether to lie or to be honest, but may choose messages associated with different psychological costs of lying. This is an important difference as, from a behavioral perspective, it seems reasonable to assume that staying silent (as in SPV) is arguably less psychologically costly than evasive lying in our decision situation which involves a false (explicit) claim of being uninformed while in fact possessing information.[5] In our study, we thus investigate how the exact nature of a lie influences sender behavior. We allow for lies of varying "severity", an issue that has received little attention in the literature so far. In fact, besides the implicit evidence from our experimental data for the conjecture that evasive lies are less psychologically costly than direct lies, we also provide more direct evidence: The results of a follow-up survey clearly reveal that direct lies are judged to be more severe.[6]

[5] Another difference between the setup of SPV and ours is that in SPV the sender is always informed, while in our case this happens only with some probability (which allows for the credibility of the evasive message).

[6] In a more abstract setting of a die-reporting game, Gneezy et al. (2016) find that the pure quantitative size of a lie (i.e., the deviation of the reported number from the true number) does not much affect individual propensity to lie. In contrast, our study highlights the importance of the *qualitative* dimension of a lie (explicit vs. evasive).

Second, SVP's study, as well as all other studies discussed above, focus on the endogenous emergence of sanctions. Here, either the receiver or a third party observes the behavior of the sender and then decides whether or not to impose a sanction. In SPV, punishment is implemented at the discretion of the receiver and is found to depend on the latter's attitude towards lying, with more lying averse subjects exhibiting a higher propensity to punish. Moreover, the receivers could punish both those senders who lied and those who remained silent. On the contrary, as our study focuses on the role of external (institutionalized) regulation, we consider deterministic sanctions imposed by a regulator which depends on the ex-post verifiability of the messages and analyze the strategic adjustments of senders and responders towards these institutional regulations.

On a more general level, our study is related to experimental work that tests the role of hidden information about the agent's actual choice. In particular, after getting the evasive message in our setting, the investor is never able to verify whether the advisor is indeed uninformed or not. Similarly, a number of studies investigate the impact of transparency of decisions on prosocial behavior in bargaining and dictator game environments. An important result here is that subjects tend to become more selfish if they are able to conceal their choices (see, for example, the studies by Mitzkewitz and Nagel 1993, Güth et al. 1996, Güth and Huck 1997, Dana et al. 2007, Andreoni and Bernheim 2009, Ockenfels and Werner 2012, Besancenot et al. 2013, Kriss et al. 2013, Grossman 2015 and the references cited therein). Fischbacher and Föllmi-Heusi (2013) introduced a die-reporting game where subjects had to report the outcome of a die roll to the experimenter, who in turn is unaware about this outcome. They found that a non-negligible amount of subjects tend to conceal their lying behind reporting a number which does not maximize their payoff.

Besides, several studies focus on principal-agent settings with hidden action. These studies implement a different sequence of actions than in our case: the uninformed party moves first and the informed party takes an unobservable action in the next step. Keser and Willinger (2000 and 2007) conduct a principal-agent game in which the principal offers an output-contingent wage contract to the agent who, after acceptance of the contract, chooses (unobservable) effort in turn. In these studies, fairness considerations are found to be important drivers for contract choices; however, the relative importance of fair contract offers is positively related to the net surplus that can be divided between the two parties. At the same time, subjects are found to exploit information asymmetries also in these settings: Vranceanu et al. (2012) find in a trust game that if there is a small probability that the amount sent by the trustor is destroyed by a random unobservable process, trustees become more selfish. Charness and Dufwenberg (2006) show how communication can overcome moral hazard problems. The authors conduct a trust game with hidden action in which a trustworthy choice by the second mover can be overruled by nature with some probability. Here, free-form communication promotes both trust and trustworthiness which is driven by messages that include promises to act in a trustworthy manner. The propensity to make and keep promises is in line with guilt aversion (Battigalli and Dufwenberg 2007) among second movers.

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.

Table 1. Payoff structure of the game.

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

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 σ ex-ante unknown to both players, which can be either good (*G*) or bad (*B*). Both realizations are equally likely ex-ante, i.e., occur with 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.

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. Finally, we assume $P > 0$ and $L < 0$, which implies that the receiver prefers to invest only in the good state. The structure of the decision situation and all payoffs are common knowledge.

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 can be indeed uninformed with some probability (see Dye 1985 and Okuno-Fujiwara et al. 1990 for a discussion).[7]

[7] 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).

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.,

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

where $\eta_X = \Pr[\sigma = G \mid m_X]$, $X \in \{G, N, B\}$, and ϕ is a continuous strictly increasing function with $\phi(0) = 0$ and $\phi(1) = 1$. For instance, if in equilibrium $\Pr[\sigma = G \mid m_G] > \Pr[\sigma = G \mid m_N]$, then the receiver invests with a higher probability after receiving m_G than after receiving m_N. Also, (1) can be considered as a reduced form of modeling heterogeneity of risk aversion in the population of receivers.

It is easy to verify that 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.[8]

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 contents of communication can be associated with different costs of lying. Herewith we distinguish two types of lies 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 σ is then given by

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

where θ is the coefficient measuring individual aversion against lying, and $c(\sigma, m_X)$ is the cost of lying from sending message m_X after observing state σ. In particular, $c(\sigma, m_X)$ 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 lies in that the sender chooses the evasive lie once he is indifferent between the direct and the evasive lie.

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

We also assume that the value of 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.,[9]

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

The receiver does not observe θ, i.e., she does not know ex-ante how trustworthy the sender is. At the same time, the distribution of θ is assumed to be common knowledge.

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.

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}.[10]

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

a) If $\frac{c_{DL}}{c_{EL}} \leq \omega(c_{EL})$, *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* $\theta \in [\theta^{**}, \bar{\theta}]$.

[9] 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).

[10] 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.

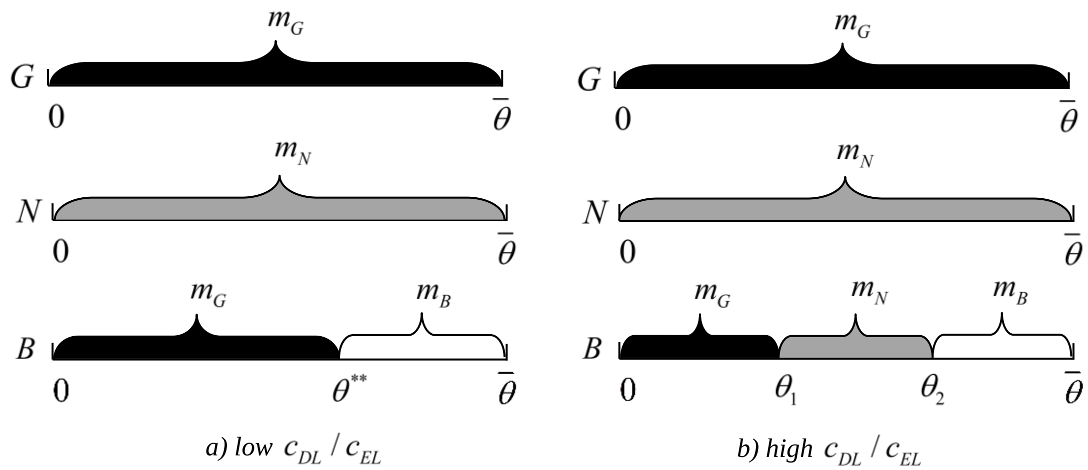

Fig. 1. Structure of equilibrium communication.

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 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 the 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 the 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)).[11] Yet, if the difference between c_{EL} and c_{DL} is sufficiently large, 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 (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.

[11] 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.

2.4. The effects of policy interventions

As the focus of our study is on the effect of policy interventions on the informational content 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.[12] 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, which means that direct lying is not eliminated completely as a result of the fine (consistent with the field evidence outlined in the introduction). The following proposition characterizes the change in the equilibrium structure of communication if a non-deterrent fine is introduced.[13]

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.*

---

[12] 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.

[13] Note that the term $\frac{\pi - f}{\pi}\omega(c_{EL})$ in Proposition 2 may be smaller than 1 if *f* is sufficiently large. One can show that in this case, an equilibrium with a non-deterrent fine can exist only if $c_{EL} < c_{DL}$ (under the assumption on lexicographic preferences).

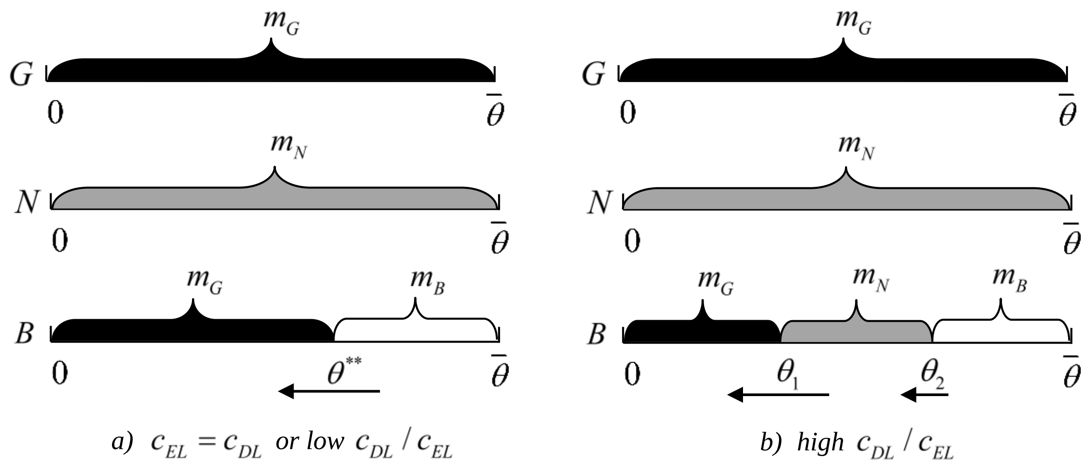

Fig. 2. The effect of the fine for lying on the equilibrium communication.

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 a small fine), no evasive lying emerges in equilibrium also after a non-deterrent 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, if the sender switches from direct to evasive lying, this lowers the likelihood that the receiver invests after receiving the message, hence decreasing the expected payoff of the sender (even taking the fine after direct lying into account).[14] If this decrease in the payoff 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}, the rate of evasion strictly increases after the introduction of the fine. 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, endogenously resulting in a lower probability of investment conditional on this message. Consequently, evasion becomes less attractive than 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 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).

[14] Otherwise, the level of fine would be fully deterrent to direct lying.

3. EXPERIMENTAL DESIGN AND HYPOTHESES

3.1. Experimental design

Our experimental design reflects the interaction between a better and a worse-informed party and follows our model setup described in the previous section. Our study aims to shed light on the effect of sanctions on expert advice, which is particularly relevant in the financial sector. We therefore frame our instructions as an investment decision; participants in the role of senders (receivers) are denominated as advisors (investors). All participants received the same instructions for the experiment.

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 with each other. The timing of the decisions in each round follows the timing of the theoretical game in section 2.1, while the whole structure of the game is common knowledge. 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 the advisor observes the state afterwards, his corresponding decision is implemented.[15]

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 6 ECU relative to abstaining, whereas she loses 3 ECU from the investment if the bad state occurs. The large difference between the gain and the loss

[15] The strategy method was used to increase the number of useful observations. Otherwise, with a direct response method, only 30% of advisor observations could be used for the analysis (50% of the state being bad × 60% of the advisor being informed). See Brandts and Charness (2011) and Fischbacher et al. (2012) for providing evidence in favor of the behavioral validity of the strategy method in various contexts.

Table 2. Experimental payoff structure.

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

is chosen to make investment a sufficiently attractive option for the investor even under sufficiently high risk or uncertainty aversion.[16]

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.[17] Thus, message "Don't know" is never fined. Besides, in line with our model, the fine is deducted only 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. Our hypotheses closely follow the predictions of the model outlined in the previous section. First, we consider the introduction of the evasive communication option in the absence of sanctions. As Proposition 1 and Fig. 1 indicate, the advisor's response to the introduction of the evasive message "Don't know" in the 3M-NP treatment depends on the relative psychological costs

---

[16] Note that the payoff structure of the game provides some room for the effect of inequality aversion (Fehr and Schmidt 1999, Bolton and Ockenfels 2000). For instance, if the advisor is averse to advantageous inequality, this might make him more reluctant to recommend investment after observing the bad state, thus mitigating the conflict of interest in this state. On the other hand, if the investor is averse to disadvantageous inequality, this can make her less prone to invest due to the risk of strongly unequal allocation after investing in the bad state. At the same time, since inequality aversion just alters the value of the ex-post payoffs for the players, adding it does not change the qualitative predictions of the model.

[17] 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 sufficiently low so that truth-telling still remains a strictly dominated option in terms of monetary payoff.

between the direct and the evasive lie. If this ratio is only small, we should see no evasive lying in 3M-NP when the advisor observes the bad state. However, if direct lying leads to sufficiently higher psychological costs than evasive lying, this will lead to a share of advisors choosing the "Don't know" message after observing the bad state even if there is no fine. This leads to our first hypothesis:

Hypothesis 1: *A positive share of advisors sends the "Don't know" message in the 3M-NP treatment.*

In the next step, we consider the effect of the introduction of the fine. The 2-message case is a benchmark where the effect of the fine can be clearly predicted: as the relative benefit from lying drops after the introduction of the fine, this should lead to a lower rate of lying. [18]

Hypothesis 2: *The rate of lying (truth-telling) is lower (higher) in the 2M-P than in the 2M-NP treatment.*

In the 3M treatment, the advisor additionally obtains an option to send an evasive message while being informed. As Proposition 2 implies, under a 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 a non-deterrent fine), because the probability of investing after the "Don't know" message is sufficiently lower than after the "Good" message, while providing no compensation in terms of lower intrinsic lying costs.

However, when evasive lying is intrinsically less costly for the advisor than direct lying, our model predicts that the effect of the fine is qualitatively different in the 3M treatment compared to the 2M treatment. In the 3M treatment, our model predicts that direct lying is substituted by both truth-telling and evasive lying in response to the fine. Concerning the prevalence of evasive communication - the main focus of our study - we can thus state our third hypothesis:

Hypothesis 3: *The introduction of the fine increases the share of evasive messages in the 3M-P treatment relative to the 3M-NP treatment.*

This can be seen as an erosion effect: in the 3M-P treatment a 1-percent drop in the lying rate in response to the fine 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).

At the same time, in the 3M treatment the rate of truth-telling is still increasing as a result of the fine. Yet, this effect works through a different channel than in the 2M treatment. In the 3M treatment, there is an indirect effect through updated beliefs: advisors are supposed to

[18] 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).

anticipate the reduced credibility of the evasive message m_N as a result of the fine (since this message is abused more extensively) which lowers the expected payoff from sending the evasive message. Consequently, some advisors with relatively high values for θ switch from evasion to truth-telling in marginal cases. Hence, we state the last hypothesis as follows.

Hypothesis 4: *The introduction of the fine increases the rate of truth-telling in the 3M-P treatment relative to the 3M-NP treatment.*

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 each round, an advisor and an investor are matched to each other, ensuring that no pair of subjects will interact in two consecutive rounds (which was made common knowledge).[19]

Note that our theoretical analysis, as well as our experimental hypotheses, are based on the premise of fully rational Bayesian belief updating of both advisors and investors. In particular, in order for evasive lying to be less attractive in terms of expected monetary payoff than direct lying (so that the rate of direct lying is non-zero), the advisors should anticipate that investors expect a lower payoff from investing after the "Don't Know" message, and hence invest less frequently conditional on receiving it. Moreover, shifts in the equilibrium distribution of messages as a result of the fine should also be adequately reflected in the first-order beliefs of investors and the second-order beliefs of advisors for the predictions to be valid (in particular, for the validity of Hypothesis 4, see section 3.2). To account for this, we ask investors and advisors about their beliefs concerning the behavior and beliefs of others (before they state their decisions). Beliefs are elicited in an incentivized manner using a quadratic scoring rule (see Schlag et al. 2015 for a survey of belief elicitation techniques).[20]

[19] Six is the smallest cohort size which implies a non-deterministic matching within a cohort. The participants were not aware of the exact cohort size, and they were unlikely to infer it during the experiment given the small number of discrete strategies and relatively low chance for an investor to meet the same advisor in the same contingency (e.g., the latter being informed and observing the good state). Hence, it is highly unlikely that subjects were affected by reputational concerns due to anticipating further interaction with the same subject.

[20] 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 advisors who will choose the "Good" message and the "Don't know" message (the latter guess is only elicited in the 3M treatments) conditional on observing the bad state. Finally, we elicit advisors' beliefs about the average investors' answers to the latter questions, i.e., advisors' second-order beliefs about the credibility of the messages. To mitigate potential hedging between belief-elicitation and investment tasks, we always ask not about the matched partner's behavior/beliefs, but about the average behavior/beliefs of the subjects in the corresponding role. This reduces the correlation between the payoff from belief elicitation and the behavior of the matched partner (see Schlag et al. 2015 for a discussion). Besides, the payoffs from belief elicitation are set to be relatively small constituting at most 1 ECU per question (under the perfectly correct guess).

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. 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.[21] After the experiment, participants had to complete a questionnaire eliciting their demographic characteristics, as well as risk and trust preferences. Then, they 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 and the post-experimental questionnaire can be found in Appendix C.

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.

4.1. Advisors' behavior

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 99.3% 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.[22]

First, we observe a 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 insignificantly different from the 2M-NP treatment ($p = 0.439$, two-sided permutation test on the level of experimental cohorts).[23] At the same time, evasive lying occurs in some 11% of the cases in 3M-NP. This share is significantly larger than zero ($p = 0.001$, one-sided

[21] 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 two cohorts were dropped from the data set. We then collected data for two new cohorts with new subjects to complete our experimental protocol.

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

[23] We do not have a directional hypothesis here, hence a two-sided test is used. In what follows, we use one-sided tests whenever we have a directional hypothesis.

Table 3. Distribution of messages conditional on observing bad state (in percent).

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

permutation test). Note that advisors have a significantly lower expected probability of investment conditional on "Don't know" than conditional on "Good" in 3M-NP, as their elicited beliefs reveal (40.7% vs. 83.4%).[24] Hence, advisors generally expect a lower monetary payoff from the evasive message than from the message "Good". Thus, the finding that some advisors opt for pretending to be ignorant instead of lying directly (under no fine) is in line 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}). This also confirms our Hypothesis 1.

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 only weakly significant (p = 0.090, one-sided permutation test). Most importantly, the sanction has virtually no effect on truth-telling in the 3M case - 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 = 0.577, 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).[25]

When conducting parametric analyses we reach similar conclusions (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).[26] As the independent variables we use the dummies for the experimental treatments while having the 3M-NP treatment as the benchmark. Besides, we control for individual characteristics elicited in the post-experimental questionnaire. Finally, 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.

[24] 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).

[25] Again, the share of evasive messages is significantly larger than zero (p < 0.001, one sided permutation test).

[26] We estimate binary Logit models instead of multinomial Logit to be consistent across estimations for different treatments. Multinomial Logit could only be applied to the 3M-treatments, while otherwise there are only two possible choices.

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	-3.133*	-0.813	2.696**
	(1.825)	(1.739)	(1.189)
2M-NP	1.761	0.485	
	(1.919)	(1.912)	
2M-P	-2.169	4.355**	
	(1.935)	(1.951)	
Period	0.190***	-0.200***	0.012
	(0.064)	(0.074)	(0.075)
Age	-0.269*	0.278*	-0.062
	(0.148)	(0.158)	(0.128)
Female	0.331	-0.629	0.815
	(1.609)	(1.465)	(0.998)
Social sciences	1.454	-0.759	0.217
	(1.669)	(1.702)	(1.302)
Risk aversion	-0.799	0.721	0.559
	(0.777)	(0.716)	(0.609)
Trust	-0.468	0.126	0.930
	(0.719)	(0.664)	(0.574)
Constant	12.15*	-12.89**	-10.27**
	(6.280)	(6.141)	(5.089)
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. We additionally control for whether a subject is female, studies Social Sciences, and for aggregated measures of risk aversion and trust derived from the answers to the post-experimental questionnaire (see Appendix C for details). 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.

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 = 0.0495, 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).

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 = 0.043, 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 the “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.

These findings are summarized in Results 2, 3, and 4, corresponding to the respective Hypotheses in section 3.2. While the results support Hypotheses 2 and 3, we find no support for Hypothesis 4.

Result 2: *The introduction of punishment decreases (increases) the rate of lying (truth-telling) in the 2M-P treatment relative to 2M-NP treatment.*

Result 3: *The introduction of punishment leads to a higher frequency of evasive lying in the 3M-P treatment relative to the 3M-NP treatment.*

Result 4: *The introduction of punishment does not lead to more truth-telling in the 3M-P treatment relative to the 3M-NP treatment.*

Hence, as formulated in Result 4 and in contrast to Hypothesis 4, 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 the "Good" message, as their elicited beliefs in the 3M-P treatment demonstrate (35.5% vs. 81.7%).[27] As we will describe below in section 4.4, both the implicit evidence from our experiment and more direct evidence from a survey conducted after the experiment suggest that the occurrence of evasive lying can at least partly be explained by different psychological costs for direct and evasive lying.

Note that Proposition 2 predicts at least some increase in the rate of truth-telling after the introduction of punishment in the 3M-P treatment (underlying Hypothesis 4), which however does not find support in our data. One of the reasons can be that the theoretical mechanism is based on fully rational belief updating. As described in section 2.4, advisors should anticipate a lower investment likelihood after the evasive message once the fine for lying is introduced (due to a lower credibility of this message for investors), and hence partially switch from evasion to truth-telling. At the same time, the aggregate belief data suggest that advisors do not anticipate a drop in the credibility of the evasive message as a result of the fine. The average likelihood of investment conditional on message "Don't know" assessed by advisors is 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 following section). Yet, advisors tend to gradually learn the latter fact over time: while their average expectation over investment rate conditional on message "Don't know" is 38.3% in the first 5 periods of the 3M-P treatment, it is significantly lower at 32.8% in the second half of this treatment ($p = 0.004$, one-sided permutation test). In turn, the rate of evasive lying correspondingly decreases in the second half of the 3M-P treatment from 30.6% to 22.8% ($p = 0.008$, one-sided permutation test). This also implies that the erosion effect is less severe in later periods. In particular, taking only the second halves of the 3M-NP and 3M-P treatments, direct lying is substituted by *both* evasion and truth-telling (see Table 5), as actually predicted by Proposition 2. At the same time, the erosion effect still

[27] 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 5. Distribution of messages conditional on observing the bad state in the first and the second halves of the 3M treatments (in percent).

	Periods 1-5		*Periods 6-10*	
	3M-NP	*3M-P*	*3M-NP*	*3M-P*
Direct lying	58.9	46.7	66.7	52.2
Evasive lying	7.8	30.6	15.0	22.8
Truth-telling	33.3	22.8	18.3	25.0

Table 6. Investment rates per message and treatment (in percent).

	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

cannot be neglected in the second half of these treatments: while the drop in the lying rate from 66.7% to 52.2% is at least weakly significant ($p = 0.083$, one-sided permutation test), the increase in the rate of truth-telling is twice as small and insignificant ($p = 0.216$, one-sided permutation test).

4.2. Investors' behavior

In the next step, we focus on how investors' choices respond to the messages sent by advisors. Table 6 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 the "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 have chosen this message. Also, in the 2M-treatments the majority of investors (more than 70% of the cases) invest after the "Don't know" message.[28] The high share of investment choices after the "Don't know" message suggests that the investors' aversion against uncertainty/risk is not too high in our setting so that it does not overrule the large gain (6 ECU) that can potentially be achieved relative to abstaining from investment. 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-value of one-sided permutation test accounts 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 message "Don't know" 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 message "Don't know" is some 12 percentage points lower in the 3M-P treatment than in the 3M-NP treatment, yet this difference is only weakly significant ($p = 0.063$, one-sided permutation test).[29]

[28] 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.

[29] However, this difference is strongly significant in the first half of the experiment ($p = 0.006$, one-sided permutation test), see Table B.1 in Appendix B. This is consistent with more prominent erosion effect in the first half of the 3M-P treatment (see Table 5).

Table 7. Determinants of investment choices.

Model No.	(1)	(2)
Dependent variable	Probability of investment after message “Good”	Probability of investment after message “Don’t know”
3M-P	0.575 (0.701)	-0.890** (0.375)
2M-NP	0.307 (0.640)	1.115** (0.523)
2M-P	0.829 (0.684)	1.287*** (0.435)
Period	-0.129*** (0.047)	0.089** (0.038)
Age	0.223*** (0.079)	-0.074* (0.039)
Female	0.179 (0.544)	-0.788** (0.350)
Social sciences	0.500 (0.706)	-0.275 (0.390)
Risk aversion	-0.408 (0.275)	-0.369** (0.166)
Trust	-0.459* (0.251)	0.109 (0.138)
Constant	1.700 (2.728)	3.159** (1.324)
N	667	603

The table shows the results of Random-Effects Logit models with the dependent variables equal to one if an investor decided to invest after a particular message was used by the advisor. We additionally control for whether a subject is female, studies Social Sciences, and for aggregated measures of risk aversion and trust derived from the answers to the post-experimental questionnaire (see Appendix C for details). 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.

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 more pronounced. We calculate Random-Effects Logit models with dummies indicating whether the investor chose to invest after receiving the “Good” message (Model 1) and the “Don’t know” message (Model 2) as the dependent variable. Table 7 reports the results of the estimations.

According to Model 1, there are no treatment effects on investor behavior after receiving message “Good”, which might be attributed to the fact that the investment rate after this message is already quite high in all treatments.[30] A notable effect is the decrease of

[30] The difference in the coefficients on 2M-NP and 2M-P is insignificant with $p = 0.372$ according to two-sided Wald test.

investment after "Good" over 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 for 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 message "Don't know" in the 3M treatment as a result of the punishment, manifested in the significantly negative coefficient for 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 3). Therefore, we can state our next result:

Result 5: *Investors' 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.*

4.3. Players' earnings

Let us consider the determinants of advisors' earnings. Table 8 shows advisors' payoffs for the investment stage (i.e., excluding the bonus for belief elicitation) over the experimental treatments and choices.[31] Taking all observations, the introduction of punishment reduces advisors' earnings in both the 2M and 3M treatments, yet only the latter difference is significant ($p = 0.025$, one-sided permutation test). Besides, payoffs are lower in the 3M treatments than in the 2M treatments ($p = 0.076$ for 2M-NP vs. 3M-NP comparison, $p = 0.003$ for 2M-P vs. 3M-P comparison, one-sided permutation tests). This can be attributed to a higher likelihood of investment conditional on the "Don't Know" message in the 2M treatments (see Table 6). Once advisors lie directly conditional on observing the bad state, their earnings are naturally lower in the punishment treatments ($p < 0.01$ for both pairwise comparisons, one-sided permutation tests). Evasive lying also pays much less after introducing punishment in the 3M case ($p = 0.013$, one-sided permutation test), which is again due to a lower credibility of the "Don't know" message for investors and correspondingly lower investment rate hereafter.[32]

Next, let us consider the determinants of investors' earnings (see Table 9). Notably, the introduction of punishment for lying does not lead to a significant change in the investors' earnings if we take all observations. Yet, payoffs are significantly lower in the 3M treatments than in the 2M treatments ($p = 0.013$ for 2M-NP vs. 3M-NP comparison, $p = 0.025$ for 2M-P vs. 3M-P comparison, one-sided permutation tests). As in the case of advisors, this again can be linked to the reduced credibility of the evasive message which distorts incentives to invest after this message, while in fact it is often sent by truly uninformed advisors in which case the ex-ante profitability of investment is high.

[31] Since the advisor has a non-trivial choice only conditional on observing the bad state, the data for separate choices are reported conditional on this case.

[32] The observed patterns of advisors' earnings do not much differ between the first and the second halves of the experiments (see Table B.2 in Appendix B).

Table 8. Advisors' payoffs over treatments and choices.

		2M-NP	*2M-P*	*3M-NP*	*3M-P*
All observations		10.23	10.06	9.96	9.66
Observing the bad state	*Direct lying*	10.74	9.85	10.74	9.78
	Evasive lying	-	-	10.10	8.92
	Truth-telling	8.10	8.44	8.10	8.11

Table 9. Investors' payoffs over treatments and messages.

	2M-NP	*2M-P*	*3M-NP*	*3M-P*
All observations	7.56	7.65	7.22	7.38
Observing message "Good"	7.92	8.81	8.08	8.79
Observing message "Don't Know"	7.48	7.15	6.47	6.29

Splitting the sample depending on which message is received by the investor yields the following pattern of investors' earnings. Conditional on receiving the "Good" message, investors' earnings are significantly larger after the introduction of punishment in the 2M case (p = 0.006, one-sided permutation test). In the 3M case, the effect of punishment is weakly significant (p = 0.085, one-sided permutation test). This can be traced to the lower likelihood of direct lying (i.e., the lower abuse of the "Good" message) once the fine is introduced. Next, exogenous punishment does not increase investors' earnings after the "Don't know" message, which is natural given that this message is sent only by truly uninformed types in the 2M case, and is abused more heavily after punishment in the 3M case. The high likelihood of evasive lying in the 3M treatments also causes a significant decrease in the investors' earnings after the "Don't Know" message in the 3M treatments relative to the 2M treatments (p = 0.001 for both 2M-NP vs. 3M-NP and 2M-P vs. 3M-P comparisons, one-sided permutation tests).[33]

4.4. Rational evasion or varying psychological costs?

Our experimental data has shown that when evasive communication is possible, a share of advisors make use of this possibility. Moreover, when sanctions for explicit lies are introduced, direct lying is substituted by evasive lying. The crucial question, however, is which mechanism drives these patterns. Two competing explanations are possible: First, our model outlined in section 2 would explain the results if the psychological costs for direct lying exceed those of evasive lying. Second, an alternative motivation to switch to evasive communication would be due to mere financial motives. For example, if an advisor would expect a sufficiently high rate of investment after the "Don't know" message, sending this message and avoiding the deterministic fine after explicit lying could be optimal from a strictly rational perspective and irrespective of lying costs. While this alternative explanation is plausible, we provide evidence in the following that corresponds to the interpretation of advisors facing different lying costs depending on the type of their lie.

Our first indication comes from our experimental data. Here, we compare the expected monetary utility for the sender (given his beliefs about the probability of investment) in the

[33] The observed patterns for investors' earnings tend to get more pronounced in the first half of the experiment (see Table B.3 in Appendix B).

3M-P treatment from choosing the direct or the evasive lie. Importantly, in this treatment, the expected monetary utility from sending the "Good" message (and thus, the direct lie) is still higher than that from sending the "Don't Know" message for 81.4% of advisors.[34] This share remains almost the same (79.6%) if one takes only those advisors who chose evasive lying.

The comparison of expected payoffs clearly speaks against the conjecture that the switch to evasive messages is based on rational considerations. 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 forego evasive lying at least in favor of direct lying. 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).

In addition to this indirect evidence, we provide also more direct evidence indicating a difference in the psychological lying costs. As it seems difficult to directly elicit the level of psychological costs of lying at the level of individual advisors, we conducted a classroom survey to elicit the perceived severity of the direct and the evasive lie in our setting. Our survey was conducted in June 2017. We collected data from altogether 160 participants in a large lecture at the University of Cologne, the same institution where we conducted our experimental sessions two years before.[35] Participants were explained that they would be presented the instructions of an economic experiment and that they would have to answer a survey question about this experiment afterwards. They then received the original instructions for our 3M-NP treatment. In the next step, to elicit the perceived severity of the two types of lies in our setting, we described the advisor's choice of which message to send after observing the bad state of the world. Survey participants were confronted with the direct lie (the message "I believe that the investment conditions in this period are GOOD") and with the evasive lie (the message "I do not know the investment conditions in this period") and had to state if one of the messages was either a "much more severe lie" or a "somewhat more severe lie" or if the two messages were "equally severe lies". Moreover, a judgment that neither of the messages was a lie was also possible (see Appendix D for the translated survey).[36]

Overall, the survey provides clear evidence that the direct lie is considered to be more severe than the evasive lie. Out of the 147 participants who consider the messages to be lies,[37] 100 participants (68.0%) state that the direct lie is either a much more severe or a somewhat more severe lie than the evasive lie. Importantly, this pattern is significantly different from a situation in which participants pick randomly between the categories of the survey in which case the respective statement would be chosen with a probability of 2/5 (a binomial test yield a p-value of $p < 0.01$). Only 15.0% of the participants report that the evasive lie is a somewhat

[34] 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.

[35] Since we conducted our survey in an introductory lecture for first-year students in economics it is highly unlikely that a student had already participated in the experimental sessions which took place in 2015.

[36] 20 participants were randomly selected to receive a bonus of 5 Euro after the survey was completed.

[37] 13 subjects or 8.13% of the survey participants agree with the statements that the messages are not lies.

more severe or a much more severe lie than the direct lie.[38] The finding that the majority of participants attribute a higher severity to the direct lie is in line with our interpretation that the use of the evasive message can be at least partly explained by differences in the psychological costs of lying.

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.[39] Inter alia, our experiment is in line with 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.

That said, we acknowledge that our experiment provides only indicative evidence for our central conjecture that the psychological costs of evasive lying are smaller than those of direct lying. An alternative explanation for the choice of the evasive message would be based on financial motives. It would in principle be possible that advisors choose the evasive message because its expected payoff, given advisors' beliefs about the investment probabilities induced by the messages, is higher than the expected payoff of direct lying. However, two empirical observations speak against this interpretation: First, as described in the previous section, for the observed beliefs about investors' propensity to invest, an advisor who faces equal psychological costs for direct and evasive lie would favor the direct lie over the evasive lie. Second, the results of an additional survey in which we show the experimental decision situation to subjects and then let them rate the relative severity of the direct and the evasive lie provide clear evidence that the evasive lie is judged to be less severe.

[38] To control for potential order effects in the presentation of the statements, we conducted two versions of the survey varying which statement (direct or evasive lie) was mentioned first in the instructions. The relative severity of the direct lie is assessed somewhat weaker if the evasive lie is presented first (here some 61.2% of the participants state that the direct lie is either a much more severe or a somewhat more severe lie compared to 73.8% when the direct lie is presented first). However, the pattern is highly significant irrespective of the version.

[39] 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.

All in all, 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. 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), or by the beliefs conditional on the message which may trigger guilt aversion (Battigalli et al. 2013, Khalmetski 2014). Finally, the fact that evasive lying is not verifiable by the investor (while the direct lying is verifiable in case of investment) implies asymmetry in image losses resulting from direct and evasive lies.[40] Further studies are required to shed more light on which channel is the most relevant in determining the difference in the lying costs.

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. As a first answer to this question, 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 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 investors, who then would rather attribute such claims to an attempt to conceal information.[41] 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.

More generally, our study suggests that ex-post sanctions for lying might be ineffective under plausible circumstances, and hence more fundamental interventions, like the reduction of the ex-ante conflict of interest between the advisor and the investor, should be preferred. For example, Barr et al. (2009) propose to ban certain forms of compensation schemes that grant higher commissions for financial brokers if customers pay higher rates. Such schemes induce search for products that are not in the best interest for the customers (so-called "yield-spread premium" in the context of subprime mortgage loans). Inderst and Ottaviani (2010)

---

[40] See Abeler et al. (2016), Gneezy et al. (2016), Dufwenberg and Dufwenberg (2016) and Khalmetski and Sliwka (2017) for theoretical analysis of the effect of image concerns on lying behavior.

[41] 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.

discuss how a regulation that reduces the share of ex-ante commissions and increases the share of commissions paid during the lifetime of the financial product ("trail commissions") helps to align incentives of advisors and investors. Also, strengthening customers' rights to cancel or to terminate a contract might discipline the sellers of financial products and lead to the recommendation of more suitable products. However, as the authors emphasize, the positive impact of these regulations on investment choices depends on the sophistication of customers who have to understand how these interventions change incentives for the sellers of the financial products. Yet, as there is abundant evidence that cognitive biases and limitations may seriously distort financial decisions (see, for example, Inderst and Ottaviani 2010, Campbell 2016, among many others), more research both in the laboratory and the field is needed to evaluate which regulations are indeed suited to reduce consumer exploitation under information asymmetries.

REFERENCES

Abeler, J., Nosenzo, D., and Raymond, C. (2016). "Preferences for truth-telling – a meta study," CESifo Working Paper Series 6087, CESifo Group Munich.

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.

Andreoni, J., and B. D. Bernheim (2009): "Social image and the 50-50 norm: A theoretical and experimental analysis of audience effects," *Econometrica*, 77, 1607–1636.

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

Barr, M. S., S. Mullainathan, and E. Shafir (2009). "The case for behaviorally informed regulation," in: New perspectives on regulation, D. Moss and J. Cisternino, eds., Cambridge, MA: Tobin Project, 2009, 11–23.

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

Battigalli, P., and M. Dufwenberg (2007): "Guilt in games," *American Economic Review*, 97, 170–176.

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.

Besancenot, D., Dubart, D., and R. Vranceanu (2013): "The value of lies in an ultimatum game with imperfext information," *Journal of Economic Behavior & Organization* 93, 239–247.

Bolton, G. E., and A. Ockenfels (2000): "ERC: A theory of equity, reciprocity, and competition," *American Economic Review*, 90, 166–193.

Brandts, J. and G. Charness (2003): "Truth or consequences: An experiment," *Management Science* 49(1), 116–130.

Brandts, J., and G. Charness (2011): “The strategy versus the direct-response method: A first survey of experimental comparisons,” *Experimental Economics*, 14, 375–398.

Campbell J. Y. (2016): “Restoring rational choice: The challenge of consumer financial regulation,” *American Economic Review: Paper & Proceedings*, 106(5), 1–30.

Charness, G. and M. Dufwenberg (2006): “Promises and partnership,” *Econometrica*, 74(6), 1579–1601.

Chater, N., Huck, S., and Inderst, R. (2010): “Consumer decision-making in retail investment services: A behavioural economics perspective,” *Report to the European Commission/SANCO*.

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

Dana, J., R. A. Weber, and J. A. Kuang (2007): “Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness,” *Economic Theory*, 33, 67–80.

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

Dufwenberg, M., Jr. and M. Dufwenberg, Sr. (2016): “Lies in disguise – a theoretical analysis of cheating,” CESifo Working Paper Series 6208, CESifo Group Munich.

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.

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

Fischbacher, U., S. Gächter, and S. Quercia (2012): “The behavioral validity of the strategy method in public good experiments,” *Journal of Economic Psychology*, 33, 897–913.

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.

Gächter, S. and J. Schulz (2016): ”Intrinsic honesty and the prevalence of rule violations across societies,” Nature, 531, 496–499.

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., A. Kajackaite and J. Sobel (2016): "Lying aversion and the size of the lie," Technical report, UCSD.

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.

Grossman, Z., (2015): "Self-signaling and social-signaling in giving," *Journal of Economic Behavior & Organization*, 117, 26–39.

Güth, W., and S. Huck (1997): "From ultimatum bargaining to dictatorship – An experimental study of four games varying in veto power," *Metroeconomica*, 48(3), 262−279.

Güth, W., S. Huck, and P. Ockenfels (1996): "Two-Level ultimatum bargaining with incomplete information: An experimental study," *Economic Journal*, 106, 593–604.

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.

Inderst, R., and M. Ottaviani (2010): "Consumer protection in markets with advice," *Competition Policy International*, 6(1), 47–64.

Irlenbusch, B., and M. C. Villeval (2015): "Behavioral ethics: How psychology influenced economics and how economics might inform psychology?" *Current Opinion in Psychology*, 6, 87–92.

Jin, G. Z., Luca, M. and D. Martin (2016): "Is no news (perceived as) bad news? An experimental investigation of information disclosure," NBER working paper Nr. 21099.

Kajackaite, A. and U. Gneezy (2017): "Incentives and cheating," *Games and Economic Behavior*, 102, 433–444.

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.

Keser, C. and M. Willinger (2000): "Principals' principles when agents' actions are hidden," *International Journal of Industrial Organization*, 18, 163–185.

Keser, C. and M. Willinger (2007): "Theories of behavior in principal-agent relationships with hidden action," *European Economic Review*, 51, 1514–1533.

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

Khalmetski, K. and D. Sliwka (2017): “Disguising lies: Image concerns and partial lying in cheating games,” CESifo Working Paper Series 6347, CESifo Group Munich.

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

Kriss, P., R. Nagel, and R. Weber (2013): „Implicit vs. explicit deception in ultimatum games with incomplete information,” *Journal of Economic Behavior & Organization*, 93, 337–346.

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

Malmendier, U., and D. Shanthikumar (2007): “Are small investors naïve about incentives?” *Journal of Financial Economics*, 85, 457–489.

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

Mitzkewitz, M., and R. Nagel (1993): “Experimental results on ultimatum games with incomplete information,” *International Journal of Game Theory*, 22, 171–198.

Ockenfels, A., and P. Werner (2012): “’Hiding behind a small cake’ in a newspaper dictator game,” *Journal of Economic Behavior and Organization*, 82, 82–85.

Ohtsubo, Y., F. Masuda, E. Watanabe, and A. Masuchi (2011): “Dishonesty invites costly third-party punishment,” *Evolution and Human Behavior*, 31, 259–264.

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.

Sánchez-Pagés, S., and M. Vorsatz (2007): „An experimental study of truth-telling in a sender-receiver game,” *Games and Economic Behavior*, 61(1), 86–112.

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

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.

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.

Stiglitz, J. (2009): "Regulation and failure," in: New perspectives on regulation, D. Moss and J. Cisternino, eds., Cambridge, MA: Tobin Project, 2009, 11–23.

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

Vranceanu, R., A. Sutan, and D. Dubart (2012): "Trust and financial trades: Lessons from an investment game where reciprocators can hide behind probabilities," *The Journal of Socio-Economics*, 41, 72–78.

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.[42]

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

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

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

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

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

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)

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

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

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

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 $\theta_2>\theta_1$).[43] This condition is equivalent to

$$\pi\phi(\eta_N(\theta_1,\theta_1))-\theta_1 c_{EL}>0$$

$$\Leftrightarrow \pi\phi(0.5)-\theta_1 c_{EL}>0$$

$$\Leftrightarrow \theta_1<\pi\frac{\phi(0.5)}{c_{EL}}. \tag{23}$$

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,

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

[43] 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.

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

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

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

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

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

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* $\tilde{f}$ *such that an equilibrium with a non-deterrent level of punishment exists if and only if* $f < \tilde{f}$. *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 $\frac{c_{DL}}{c_{EL}} \le \frac{\pi - f}{\pi}\omega(c_{EL})$, *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

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

$$\Leftrightarrow \pi\phi(\eta_G) - \theta c_{DL} > 0 \tag{39}$$

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

$$(\pi - f)\phi(\eta_G(\theta^{**})) > \pi\phi(0.5)$$

$$\Leftrightarrow (\pi - f)\frac{\theta^{**} c_{DL}}{\pi - f} > \pi\phi(0.5)$$

$$\Leftrightarrow \theta^{**} > \pi\frac{\phi(0.5)}{c_{DL}}, \tag{41}$$

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 $\lambda_4(\overline{\theta}) < 0$ by (4), by the intermediate function theorem, (41) holds if and only if

$$\lambda_4\left(\pi\frac{\phi(0.5)}{c_{DL}}\right) > 0$$

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

$$\Leftrightarrow f < \pi\left(1 - \frac{\phi(0.5)}{\phi\left(\eta_G\left(\pi\frac{\phi(0.5)}{c_{DL}}\right)\right)}\right), \tag{43}$$

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

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

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

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

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

$$\lambda_5(0) > 0, \quad (46)$$

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

The first condition is equivalent to

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

where θ^{***} is implicitly given by

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

(It is easy to verify that $\theta^{***} \in (0, \overline{\theta})$ 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

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

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. Dynamics of behavior and earnings over periods

Table B.1. Investment rates in the first and the second halves of the experiment (in percent).

Periods 1-5:

	2M-NP	*2M-P*	*3M-NP*	*3M-P*
"Good"-message	91.3	91.8	89.3	95.2
"Don't know"-message	69.1	70.8	48.7	27.2
"Bad"-message	5.9	13.3	5.6	6.7

Periods 6-10:

	2M-NP	*2M-P*	*3M-NP*	*3M-P*
"Good"-message	86.5	93.4	87.1	89.5
"Don't know"-message	73.4	77.2	52.6	48.9
"Bad"-message	0.0	16.0	0.0	0.0

Table B.2. Advisors' payoffs in the first and the second halves of the experiment.

Periods 1-5:

		2M-NP	*2M-P*	*3M-NP*	*3M-P*
All observations		10.20	10.03	9.90	9.57
Observing the bad state	*Direct lying*	10.75	9.73	10.88	9.80
	Evasive lying	-	-	9.50	8.64
	Truth-telling	8.18	8.40	8.17	8.20

Periods 6-10:

		2M-NP	*2M-P*	*3M-NP*	*3M-P*
All observations		10.25	10.10	10.02	9.76
Observing the bad state	*Direct lying*	10.74	9.94	10.65	9.76
	Evasive lying	-	-	10.25	9.25
	Truth-telling	8.00	8.48	8.00	8.00

Table B.3. Investors' payoffs in the first and the second halves of the experiment.

Periods 1-5:

	2M-NP	*2M-P*	*3M-NP*	*3M-P*
All observations	7.45	8.17	7.55	7.23
Observing message "Good"	7.76	9.49	8.68	8.82
Observing message "Don't Know"	7.50	7.62	6.73	5.85

Periods 6-10:

	2M-NP	*2M-P*	*3M-NP*	*3M-P*
All observations	7.67	7.13	6.88	7.52
Observing message "Good"	8.06	8.05	7.55	8.76
Observing message "Don't Know"	7.45	6.76	6.20	6.68

APPENDIX C. 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*

<u>Investor:</u>

- 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".

<u>Advisor</u>:

- 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".

Post-experimental questionnaire[44]

1. **Age:** ☐

2. **Gender:** ☐ **m** ☐ **f**

3. **Which department do you study at?**

☐ Social Sciences
☐ Law
☐ Medicine
☐ Philosophy
☐ Mathematics and Natural Sciences
☐ Human Sciences
☐ I am not a student.

[44] The aggregate measure for risk aversion used in the regression analysis is computed as $\frac{(8-a_6)+a_7}{2}$, and for trust as $\frac{a_8+(8-a_9)+(8-a_{10})}{3}$, where a_i is the answer to the question i of the questionnaire.

4. **Native language:** [ ]

5. **Which role did you have in the experiment?**

- ☐ Advisor
- ☐ Investor

6. **How do you personally assess yourself: Are you generally willing to take risks, or do you try to avoid them?**

Not willing to take risks at all						Very willing to take risks
1	2	3	4	5	6	7

7. **Please think what would you do in the following situation:**

Imagine that you have won 100.000 Euro in a lottery. Immediately after receiving the prize you get an offer from a bank of an investment option, which contains the following:

There is a chance TO DOUBLE the money WITHIN TWO YEARS. There is however an EQUALLY HIGH RISK TO LOSE HALF of the invested money. You may invest all your money or a part of it as shown in the following, or you may reject the offer.

Which part of the lottery prize would you put to the on the one hand risky and on the other hand profit-promising investment option?

- ☐ the whole amount of 100.000 €
- ☐ the amount of 80.000 €
- ☐ the amount of 60.000 €
- ☐ the amount of 40.000 €
- ☐ the amount of 20.000 €
- ☐ Nothing at all, I would reject the offer.

How strongly do you agree with the following statements?

8. In general, one can trust people.

Do not agree at all						Fully agree
1	2	3	4	5	6	7

9. Nowadays, one cannot rely on anybody anymore.

Do not agree at all						Fully agree
1	2	3	4	5	6	7

10. When dealing with strangers, it is better to be cautious before trusting them

Do not agree at all						Fully agree
1	2	3	4	5	6	7

11. Finally, please briefly describe according to which criteria you have made your decisions in this experiment.

APPENDIX D. Survey

Below you find instructions for the survey translated from German. Participants received written copies of these instructions prior to the start of the survey. To control for sequence effects in the presentation of the messages in the questionnaire, we used a balanced design, presenting the direct lie as message A in a half of the questionnaires (as shown below) and as message B in the other half of the questionnaires.

General Information on the Survey

Welcome and thank you for your participation in this survey.

From now until the end of the survey, please do not communicate with other participants.

Please read the following pages carefully. If you have any questions after reading this or during the survey, please raise your hand. One of the organizers will come to you and will privately answer your question.

In the following you will get instructions for an experiment, which recently has been conducted at the Cologne Laboratory for Economic Research. Subsequently, we would like to ask you a short question about this experiment.

Participants in this experiment had the opportunity to earn money. How much they earned, depended on their own decisions and the decisions of all other participants. All payments to the participants were made in cash after the experiment. Independent of their decisions within the experiment, participants received 2.50 Euro for their attendance. Payouts and decisions of all participants in the experiment were treated confidentially. Either during or after the experiment, no participant received any information about whom they had interacted with and which payouts other participants received. Thus, all decisions were anonymous.

Questionnaire

Take your time, and make sure you understand the experimental instructions starting on the next page. **Subsequently, please answer the questionnaire on the last page. Your answers will be treated confidentially.**

Submission of questionnaire and the lottery ticket:

- When you have completed the questionnaire, please separate it from the instructions and drop it into the **questionnaire box**.
- At the top of the instruction sheet is a green note with two identical numbers. When submitting the questionnaire, tear **one of these numbers** along the perforation and throw it into the **lottery bin**.
- After submitting all questionnaires, we will randomly draw **20 lottery tickets from the lottery bin** and those of you who have the second sheet of a drawn number on their instruction sheet will receive **5 Euro** each.

The description of the experiment – the following instructions were given to every participant in the Cologne Laboratory before the start of the experiment:

[Instructions for the 3M-NP treatment]

(End of the instructions for the laboratory experiment)

Questionnaire

Consider the situation in which the advisor determines the **message** he would like to send to the investor if he later learns that the investment conditions are "Bad". Now suppose the advisor decides for one of the two messages:

Message A: *"I believe that the investment conditions in this period are GOOD."*

Message B: *"I do not know the investment conditions in this period."*

Please judge both messages by circling **one of the six possible answers**:

A is a **much more severe lie** than **B**	**A** is a **somewhat more severe lie** than **B**	**A** and **B** are **equally severe lies**	**B** is a **somewhat more severe lie** than **A**	**B** is a **much more severe lie** than **A**	**Neither A nor B** is a **lie**