De-biasing strategic communication?

-preliminary version-

please do not circulate without author's permission
prepared for the 26th International Conference on Game Theory 2015

Tobias Gesche*

April 17, 2015

Abstract

This paper studies strategic communication with lying costs and hidden conflicts of interest. I present a simple economic mechanism under which the disclosure of conflicts of interest can lead to more biased messages with average receivers following them more closely. Receivers who delegate their choice or who are naive towards the conflict of interest are then hurt by disclosure while non-delegating, rational receivers benefit from it. In consequence, disclosure is often *not* a Pareto-improvement among the set of receivers and can even lead to a *decrease* in efficiency. I find that the correlation between the sender's incentives to bias his message and the true state of the world is decisive for determining i) when mandatory disclosure hurts receivers, ii) when senders would voluntarily commit to disclose their conflicts of interests, and iii) when mandatory disclosure is efficient.

Key words : Strategic communication, Misreporting, Conflict of Interest, (Mandatory) Disclosure, Naive receivers, Delegation, Financial advise

JEL-Classification: D18, D83, G28, L51

*Department of Economics, University of Zurich. Email: tobias.gesche@econ.uzh

I thank Jeffrey Ely, Fabian Herweg, Roman Inderst, Nick Netzer, María Sáez-Martí, Yuval Salant, Adrien Vigier and seminar participants at the University of Zurich and the ZWE 2014 (Solothurn) for helpful comments.

I Introduction

A substantial part of the world's economic activity deals with the elicitation of information by experts and its dissemination to non-experts. Examples include stock analysts, researchers, consultants or journalists. As a direct consequence, the information which such experts communicate to their clients is not verifiable to the latter and prone to a bias if the expert faces a conflict of interests (henceforth COI). Inefficiencies then arise because of two main reasons: First, receivers may ignore the expert's COI and make poor choices by following his biased advise. Second, receivers who are aware of an expert's bias often lack the information about the magnitude and direction of the bias. Without such information, they cannot accurately correct the expert's advise and may then rationally decide to ignore the expert's message, at least partially, such that information is lost.

Disclosure of COIs promises to be a simple remedy to this problem. By the expert's obligation to inform the sender about his bias, a receiver can correct the distortion it causes and make appropriate choices. Moreover, there is the hope that the act of disclosing his COI itself makes the sender behave more honestly. Disclosure is also luring to policy makers as it carries the - as I will show incorrect - intuition that flattening information asymmetries is always desirable. Compared to alternative interventions such as direct regulation and surveillance it is also less paternalistic and less likely to face resistance from affected industries. A prominent example for such a policy is included in the Sarbanes-Oxlay-Act which was enacted as a response to corporate frauds, in particular among financial analysts. Among the regulations it adopted is the requirement for security analysts to "[...] *disclose conflicts of interest that are known or should have been known by the securities analysts*" (United States Congress (2002), Sec. 501b).

Experience however suggests that disclosure is prone to failure. In a recent paper, Malmendier and Shanthikumar (2014) show that financial analysts do *strategically* inflate their stock recommendations when they earn sales commissions and are not just overly optimistic about the return of stock which they recommend. More relevant in the context of this study, their analysis covers behaviour before and after the enactment of the Sarbanes-Oxlay-Act. Their results show that no change in analysts' strategic bias occurred after it was put into action in 2002. Clean evidence that disclosure may even lead to adverse effects and *decrease* the quality expert's advise comes from Cain et al. (2005): In their experiment, subjects in the role of an expert had to give advise to other

subjects who had to estimate the amount of money inside a jar filled with coins. In contrast to these estimators they could examine the jar carefully and take their time so as to obtain superior knowledge of the relevant information. First, they find that when both, experts and estimators, were paid according to an accuracy scheme, experts' advise and the receivers' estimates were much better than when the experts were paid according to how high the receivers' estimates were. Demonstrating the failure of disclosure, they also show that when receivers where made aware of this COI the experts' bias *increased* and receivers - who did not account for this - made worse decisions than with undisclosed COIs.

The objective of this paper is to provide an economic explanation for the failure and unintended consequences of such disclosure. It does so by considering a communication game with and without disclosed COIs and comparing equilibrium behaviour between these two informational settings. The model assumes that some receivers are naive while others are fully rational, in a Bayesian sense. Naivety can occur because receivers either lack the skills and information needed to de-bias a sender's message, or because they are just agnostic about it, e.g. because they trust him. The combination of these factors then unveils a simple economic mechanism which can make disclosure to reach the opposite of what it is expected to achieve, namely information to become more biased and decisions to be less accurate.

To understand the mechanism behind this adverse effect consider a setting of undisclosed COIs: In such a setting, rational receivers face strategic uncertainty regarding the expert's incentive to bias his message. In consequence, their attempt to de-bias the information they get from the sender can result in a correction of the message which mis-estimates the bias' actual size or even its direction. Facing such a risk, rational receivers then also rely on their prior and not just on the expert's imperfectly corrected message. This implies that they do not react to the message as much as they would if they had better information to de-bias it. Also note that this leads to a relatively low bias since the sender, who faces costs of biasing his message, will adjust his bias to how strongly the average receiver reacts to his message. Now consider a situation with disclosed COIs: Being able to correct the sender's message for its bias, rational receivers react more strongly to it. Following this increase in the rational receivers' reaction to his message, the sender then also increases his bias. While rational receiver can correct for this increase in the bias, naive receiver are not capable of doing so and are therefore hurt by the increase in the bias.

Underlying the above reasoning are two main insights: The first is that the reaction to the

sender's message by rational, risk-averse receiver depends on the quality of information they can extract from it. The second is that the sender will determine his bias in proportion to the reaction it induces. Both of these main building blocks are simple in their economic intuition but yet deliver the surprising result that providing information which, in principle, is useful can be a bad idea when it is not usable by everyone. In particular, it shows that the idea that disclosing COIs does at least not hurt any receiver is wrong because it does not take into account the sender's equilibrium reaction. I will also show that there can be situations when the sender will want to commit to disclose his COI in order to exploit the above mechanism but should be prevented from doing so with regards to efficiencies.

By identifying the above mechanism, the model helps to explain why disclosure often works sub-optimally or even counter-productively. It can be used to determine conditions under which this effect manifests and how severe it is. Key to this is the correlation between the COI which triggers the sender's bias and the variable which reflects the information on which the expert has superior knowledge. I find that among the setting in which the above mechanism applies are all environments in which this correlation is at least weakly positive. Further results then use the identified channels to decide when disclosure is beneficial, when it has to be become mandatory, and when less clarity regarding COIs is better. For all of these, the above correlation between the true information and the sender's COI is again decisive.

The next section reviews the related theoretical literature and its impications for modelling choices. Section III outlines the model's structure and assumptions. In Section IV, I derive the equilibrium behaviour of senders and receivers under general information structures and apply it to the case when the incentives to mis-represent the true state of the world are undisclosed. In doing so it sets the stage for section V which examines disclosed COIs. By comparing the result from these two settings, section VI discusses the implications for the game's players and overall efficiency. Section VII concludes by summarizing the main insights and discussing their policy implications.

II Related Literature

By analysing the consequences of disclosing an expert's COI, this paper contributes to the literature on strategic communication. In their seminal work on the topic, Crawford and Sobel (1982)

find that when there is a COI between the sender and the receiver, communication is partitional: In equilibrium, the sender endogenously partitions the state space and (truthfully) announces the partition which contains the actual state of the world such that information is lost and communication is inefficient. Together with no lying costs [1], the assumption of a bounded state space is crucial to the partitioning result. Allowing for an unbounded support for the variable of interest, e.g. when it is normally distributed and/or there no common known *exact* value for state space's bounds, yields different results. In fact, unboundedness is key among the general conditions derived by Kartik et al. (2007) under which the sender's messaging strategy is continuous and biased but yet revealing to rational receivers. Note that a strategy of the partitional form as obtained with a bounded support is not biased, that is the average message by the sender and the average state of the world do not differ. Given the evidence that there *are* such differences, e.g. among financial analyst (see Michaely and Womack (1999)) I assume an unbounded support.

While the above papers assume a publicly known bias, this paper also considers the cases when the bias is the sender's private information. I find that then, the sender's messaging strategy remains continuous but is not revealing anymore. This work complements other work on strategic communication when the sender's bias is unknown: Morgan and Stocken (2003) find that in a compact state space with a binary, independent bias the sender's strategy remains partitional. Similar to that work, Blanes (2003) also considers a binary and independent bias. His model is closer to this work since he assumes a normally distributed, thus unbounded, state of the world. He finds rational receivers never react to the sender's message in every state of the world. This is because senders will only stop to increase their bias when their implicit costs of doing so, as measured by rational receivers' reaction to the message is sufficiently low. A Further consequence of this approach are multiple and on-existing equilibria under non-monotone conditions. Following Kartik (2009), I also allow for direct cost of biasing the message. These cost then imply a unique equilibrium and a simple, montone condition under which every receiver always reacts, though in different degrees, to the sender's message.

Closest to the aim of explicitly comparing the consequences of disclosed and undisclosed COI are two other papers. Li and Madarasz (2008) show that with a binary bias which is uncorrelated to

[1] Kartik (2009) considers the role of lying costs when the state space is compact but and there is a commonly known sender bias. He finds that equilibria are often partially separating of the "LSHP (Low types separate and High types pool)"-form: When the sender is upward-biased, the sender exaggerates his statement by a fixed bias if the state is below a certain threshold. If it is above, senders only announce the partition of this upper subset of the state space in which the true state lies.

an uniformly distributed distributed state of the world, disclosure can be harmful to and undesired by rational receivers. The reason they identify is that disclosure may induce the sender to be more imprecise than when his incentive were only known to him, thus partitioning remains framework with a compact state space. In another paper by Inderst and Ottaviani (2012), the authors explicitly model the origin of the sender's bias as commissions paid by product providers to experts who advise customers on which of the two competing products suits them best. In their model, one of the two available products is better and the state of the world and message are therefore binary. They show that disclosing such commissions reduces their provision but less so in relative terms for the less suitable product. In consequence, the relative bias rises with disclosure and consumer make worse decisions. The negative effect of disclosing commissions in their paper therefore originates from the product providers. The mechanism underlying adverse effects of disclosure identified in this paper is different since it originates from the rational receivers' increased reaction after disclosure and the sender who adapts to this reaction. It is therefore a direct product of this bilateral relationship and not a consequence of an additional third party being involved.

Finally, this paper is different from the above ones since it uses signal extraction techniques for the model's solution. This allows for the sender's private bias to be positive or negative, drawn from a continuous support and to be correlated with the state of the world. This parsimonious framework to analyse different information regimes is then used to trace out a single key parameter - the sender's *endogenously* chosen correlation between his message and the state of the world - which determines equilibrium reactions and their consequences.

III The model

Consider a non-expert who would like to know about the value of a continuous random variable $s \in S$. She has to make a decision denoted by $d \in S$ which is dependent on realization of s. For example, s might represent the forecast for an asset's return and d the non-experts investment in it. I will assume that the non-experts suffer a loss which is the greater, the more her decision and the actual of the state of the world are misaligned. More precisely, their vNM-utility, given the decision d and the value of s is given by

$$u^R(d, s) = L(d - s) \tag{1}$$

where $L : \mathbb{R} \to \mathbb{R}_0^-$ is a loss function which is strictly concave, maximized at $L(0)$, and symmetric around around this maximum (as an important example consider the quadratic loss function).

Non-experts do not know the value of s and therefore refer to an expert's opinion who knows it. The advise by the expert is conveyed in a public message $m \in M = S$ about s. Given this information transmission context, I will henceforth also refer to experts and non-experts as senders ("him") and receivers ("her"), respectively. The meaning of the message is literal, thus an honest sender would always send $m = s$. Receivers would then just follow his message and implement their optimal choice. In this case, the sender's message would completely determine the receiver's action. However, such influence of the sender on the receivers decisions can be exploited. Third parties can pay the receiver commissions to induce either a high or low decision by the sender, e.g. a high demand for a specific asset a financial analyst is covering. Such commissions then create a conflict of interest and can lead the sender to bias his message. To describe this situation, I assume that commissions are paid in proportion $c \in C$ to the demand. I will denote the sender's (expected) demand by $D(m)$. There might be further variables which determine demand, but in the context of the game analyzed here I will focus on the effect of the message on demand. When demand is differentiable in the message, $D'(m) \neq 0$ implies that the sender can affect the demand with his message and has therefore an incentive to bias his message.

There are are also reputational, legal or moral costs of lying about s [2]. I capture such costs of reporting dishonestly by the loss function $K : M \times S \to \mathbb{R}_0^-$ whose image $K(m, s)$ is uniquely maximized at $m = s$, thus by telling the truth. I also assume that the cost function is concave in m around s, e.g. $K(s, m) = L(m - s)$ By scaling these costs relative to the senders commission with $k > 0$ and assuming additivity of these components, one can then represent the sender's decision problem as choosing his message $m \in M$ to maximize the following utility function:

$$U^S(s, c, m) = cD(m) + kK(m, s) \tag{2}$$

Note that by appropriate re-scaling of c and its distribution, one can always normalize w.l.o.g. the weight of the lying costs, e.g. $k = 1$. Note that the commission is additive and proportional

[2] Regarding potential legal costs, Dubois et al. (2013) provide evidence that the introduction of the Market Abuse Directive for financial markets in the European Union decreased over-optimistic recommendation relative to the directive's legal consequences in the single member countries. For a rationale of reputational costs, see Sobel (1985) and Morris (2001). Evidence that many people have a preference for being honest per se is provided, amongst in others Gneezy (2005), López-Pérez and Spiegelman (2012), and Abeler et al. (2014).

to the demand as for example in Morgan and Stocken (2003) or Blanes (2003). This differs from other approaches which assume that the receiver and sender have the same class of utility functions which only differ in their bliss points [3]. In particular, this implies that sender utility decreases for deviations of the receivers' actions in *either* direction from their bliss point. While this is appropriate in some environments, it is less likely to be so in others. For example, if a sender gets sales commissions, this should result in an incentive to induce a demand as high as possible and not just up to some bliss point.

Given the above utility function for the sender, his optimal message m^* then has to solve the following expression:

$$cD'(m^*) = -\frac{\partial K(m^*, s)}{\partial m} \tag{3}$$

As a direct consequence, there is no truth-telling in the presence of commissions and receivers reacting to the sender message, thus when $cD'(m) \neq 0$. By concavity of $K(m, s)$ in m around s, a sender's incentive to mis-represent the true prospect is also increasing in the receiver's reaction to the signal and the magnitude of the commission and its sign equals the sign of $cD'(m)$. For example, if receivers follow the sender's message, thus $D'(m) > 0$ and there is a commission on generated sales denoted by $c > 0$, it holds that $m^* > s$ and the sender exaggerates the true state.

In the following, I will assume that that $kK(m, s) = -\frac{1}{2}(m - s)^2$. This function captures the above considerations and allows a tractable analysis in a closed form manner, as the above optimality condition for the sender simplifies to the following linear form [4]:

$$m^* = s + cD'(m^*) \tag{4}$$

The sender's message is therefore additive in the true state of the world and a bias which follows from the incentive for manipulating receivers' demand.

Rational and naive receivers: I will now turn to describe the demand side and its reaction to the sender's message in detail. The above results show that commissions induce the sender to not report truthfully. How should receivers then take such a distortion into account and how in turn, should the sender adjust his signal to the receivers' reaction? In general, a rational (Bayesian)

[3] See for example, Crawford and Sobel (1982) or Ottaviani and Squintani (2006)). The sender's utility function therein in the language of this model would be given by $U^S(d, s, b) = L\left(d - (s + b)\right)$ where b is the sender's bias.

[4] This specific form also equals the one laid out by Kartik (2009).

receiver could do so by adjusting his expected investment to it. Her action then maximizes her expected utility, given the information m from the sender. In a recent work, Deimen and Szalay (2014) shows that if s is symmetrically distributed, then the unique maximizer of $L(d-s)$ w.r.t. d is the (conditional) expectation of s. By assuming that s is symmetrically distributed the following then holds: [5]:

$$d_r(m) = \arg\max_{d \in S} \mathrm{E}[u^R(s,d)|m] = \mathrm{E}[s|m] \tag{5}$$

The above optimal decision is that of fully rational, Bayesian receivers who make use of the sender's message while they are aware that it is potentially biased. While some people may acquire and use the skills to act in such a manner, many people are not capable of acting in this (for recent experimental evidence that people see Brocas et al. (2014) and Brown et al. (2012)). In the context of financial decision making and advise, Malmendier and Shanthikumar (2007) show that small investors, e.g. private households, follow analysts' optimistic recommendation more closely than institutional investors such as mutual funds or other investment firms. Another reason for not behaving in a Bayesian manner is that this not only requires skills but also information to form a prior. Just listening to an expert and following him does not require such information. If the money at stake and/or the expected bias are small relative to the cost of conducting the Bayesian inference, receivers can therefore prefer to delegate their decision to the sender or follow him naively.

To capture these concerns, I allow for the possibility that there a naive receivers who take the sender's signal by its face value. That is, his demand is given for any message by $d_n(m) = m$. This is also strategically equivalent to delegating one decision to the sender. I denote the share of naive or delegating receivers by $\mu \in [0,1)$ [6]. The mass of rational receivers is therefore given by $1-\mu$ which yields the following demand function:

$$D(m) = \mu d_n(m) + (1-\mu) d_r(m) = \mu m + (1-\mu)\mathrm{E}[s|m] \tag{6}$$

This also captures a scenario where the sender faces a single receiver but does not know whether

[5] For details, see the proof of lemma 2 in Deimen and Szalay (2014).

[6] Note that by appropriate scaling of μ, one can always account for situation where naive receivers react less than one-to-one, e.g. when $d_n(m) = r \cdot m$ with $c \in (0,1)$. As an example suppose that there is a mass 0.5 of naive receivers who follow the signal on average in proportion $r = 0.6$. From the sender's point of view, this is the same as if there were mass 0.2 of receivers who ignore him, mass 0.3 who follow one-to-one, and a mass 0.5 of rational receivers. Defining $\mu = \frac{0.3}{0.8}$ would then represent the same strategic situation.

this receiver is naive or rational. Denoting the probability for the former case with μ and for the latter with $1-\mu$ would then also be represented by the demand $D(m)$ as specified above.

Information: In order to correct for the bias in the sender's signal, rational receivers would ideally know the value of the commission c as it specifies how much and in which directions the sender wants to push demand. With undisclosed COIs, this information is missing. To describe how sender and receivers deal with this, I assume the common prior about the state (s, c) to be generated by the jointly normal distribution $F(\boldsymbol{\eta}, \boldsymbol{\Sigma})$ with the vector of means $\boldsymbol{\eta}$ and the variance-covariance matrix $\boldsymbol{\Sigma}$ as follows:

$$F(\boldsymbol{\eta}, \boldsymbol{\Sigma}) = \mathcal{N}\left(\begin{bmatrix} \bar{s} \\ \bar{c} \end{bmatrix}, \begin{bmatrix} \sigma_s^2 & \sigma_{sc} \\ \sigma_{sc} & \sigma_c^2 \end{bmatrix} \right)$$

I assume joint normality for two main reasons: First, it implies linearity of the conditional expectations and therefore tractability of the model [7]. The second reason is based on how players, in particular (rational) receivers, arrive at their common prior: The best they can do is to use publicly available, past information about the stock's return and the senders' messages to form an estimate for s and c. For example, when messages concern asset returns, they could regress a sender's message on the actual return and take the residual as a proxy for the senders bias, which is proportional to c. By using such a frequentist approach, they implicitly assume that their estimates $\bar{s}$ and $\bar{c}$, obtained as sample averages, are approximately normally distributed. Their sample (co-)variance would then generate $\boldsymbol{\Sigma}$. When past commissions on average promoted sales, a positive value for $\bar{c}$ would be observed. A positive covariance σ_{sc} would be observed if inferred sale commissions were particularly high when the asset performed good. Among the reasons for such behaviour is that the sender holds the asset himself and in order to increase the value of his holding, he tries to induce a higher demand for the asset. Conversely, if the sender exerts particular effort and bias to sell a junk asset he holds, then a negative σ_{sc} would be observed.

[7] This is a consequence of joint normal being a elliptical distribution. The linearity of conditional expectations and this model's result also applies to any other elliptical distribution with finite mean and variance, e.g. the logistic distribution, which could therefore be used instead of joint normality. For further details on these properties of elliptical distributions see Deimen and Szalay (2014) and references therein.

IV Undisclosed conflicts of interest

Given the assumptions in the previous section, the communication game with undisclosed commissions has the following timing:

1) the sender's type (s, c) is draw from F and privately observed by the sender

2) the sender sends a signal m about s

3) receivers observe m, if rational update their belief about s, and choose their investments d

4) payoffs are realized

I look for a a perfect Bayesian Equilibrium of this game. It consists of a pair of equilibrium strategies $m^* : S \times C \to M$ for the sender and $d_r^* : M \times C \to S$ for the rational receiver such that each player's expected utility is maximized, given the other players' strategy and consistent beliefs formed by Bayes' rule. The key equilibrium belief in this context is the rational receiver's belief about s, denoted $\mathrm{E}[s|m^*] \equiv \mathrm{E}[s|m]|_{m=m^*(s,c)}$. This is the conditional expectation of state of the world given the sender's message m when this message is the realization of a random variable whose distribution is shaped by the sender's equilibrium messaging strategy $m^*(s, c)$ and the joint distribution of the sender's type, e.g. (s, c) [8]. Naive receivers do not require explicit analysis since they have, by assumption, a dominant strategy of implementing the sender's original message.

I will use this equilibrium concept under different settings of common knowledge, henceforth called information structures $\mathcal{I}$. With undisclosed commissions, one can describe the information structure of this game by $\mathcal{I}_U = \{\sigma_s^2, \sigma_c^2, \sigma_{sc}, \mu\}$, it therefore collects the game's parameters. If commissions are disclosed, they are part of the common knowledge of every player and I will denote this information structure $\mathcal{I}_D = \mathcal{I}_U \cup \{c\}$. To avoid unnecessarily complicated notation, I will assume that (conditional) expectations, covariance and variance are always conditional on the information structure the game which is currently analysed [9]. As an important example, $\mathrm{E}[c] = \bar{c}$ holds under $\mathcal{I} = \mathcal{I}_U$. If in contrast c is common knowledge, thus when $\mathcal{I} = \mathcal{I}_D$, it holds that $\mathrm{E}[c] = c$.

[8] If undisclosed, the receiver's complete vector of beliefs also includes his conditional expectation $\mathrm{E}[c|m^*] = \mathrm{E}[c|m]|_{m=m^*(s,c)}$ about the commission the receivers gets. Since it does not affect player's actions, I will not analyse it explicitly.

[9] More formally, under information structure $\mathcal{I}$, I re-define $\mathrm{E}[x|y] \equiv \mathrm{E}[x|\{y'\} \cup \mathcal{I}]$ and thereby also $\mathrm{Cov}[x, x'|y] \equiv \mathrm{Cov}[x, x'|\{y'\} \cup \mathcal{I}]$ where $x, x' \in \{m, c, s\}$ and $y = y' \in \{\emptyset, m\}$. Note that this implies that $\mathrm{E}[x|y] = x$ and $\mathrm{Cov}[x, x'|y] = 0$ whenever $x, x' \in \mathcal{I}$.

This game can be solved by backward induction, starting with the receivers' investment decision. By assumption, naive receivers choose $d_n^*(m) = m$. Not knowing s, rational receivers maximize their expected utility. From (5) one gets that their optimal decision $d_r^*(m)$ and $\mathrm{E}[s|m^*]$ have to be identical mappings, thus equilibrium belief and strategy coincide. This determines the total demand $D(m)$ as specified in (6) and yields the following (expected) utility for the sender[10]:

$$U^S(s, c, m) = c\,(\mu m + (1-\mu)\mathrm{E}[s|m]) - \frac{1}{2}(m-s)^2 \tag{7}$$

The sender's message affects the naive receivers' demand since they follow his message and the sender's cost of mis-representing $K(m, s)$ directly. In addition, it also influences the rational receiver's demand via its effect on $\mathrm{E}[s|m]$. I will denote this marginal change of the message on the inferred type by $\theta(m) \equiv \frac{\partial \mathrm{E}[s|m]}{\partial m}$. The sender's optimal message has to balance these effects by providing a solution to the following first-order condition:

$$m = s + c\,(\mu + (1-\mu)\theta(m)) \tag{8}$$

The sender's message is therefore additive in the state of the world plus a bias. This bias equals the marginal effect of his message on the commission he earns, obtained as the product of the marginal effect of his message on average demand and the commission's value. To determine his equilibrium strategy, it has to be taken into account that the *functional form* of $\theta(m)$ is shaped itself by the signalling strategy, such that the sender has to trade off his incentive to bias the signal with the reaction of rational receivers to such an increase in the bias. Therefore, an equilibrium signal m^* has to be a solution to the above first-order differential equation (8) with the last term replaced by $\theta(m^*) \equiv \frac{\partial \mathrm{E}[s|m^*]}{\partial m}$. This is the marginal change in sender's inferred conditional expectation about s caused due to a change in the message when this message is formed by the sender's equilibrium strategy $m^*(s, c)$. If for example the equilibrium strategy involves a large bias, this that means rational receivers will not infer a lot from it. Consequently, they would react little to a change in the message and the value of $\theta(m^*)$ would be comparatively relatively low.

[10] Possible expectations for the sender's utility refers to uncertainty regarding the receiver's type, e.g. when there is one receiver of which the sender does not know whether he is naive or rational (μ denoting the probability for the former case).

The sender then has to weigh these implicit costs of biasing his message and the explicit costs of doing so from $K(m,s)$ against the benefit of manipulating receivers' demand and earning on the commission. For a more simple notation when deriving how the sender behaves optimally given these incentives, I define, analogously to the definition of $\mathrm{E}[s|m^*]$, the "inference coefficient" ρ^* as follows:

$$\rho^* \equiv \frac{\mathrm{Cov}[s,m^*]}{\mathrm{Var}[m^*]} \equiv \frac{\mathrm{Cov}[s,m]|_{m=m^*(s,c)}}{\mathrm{Var}[m]|_{m=m^*(s,c)}} \tag{9}$$

This equilibrium parameter equals the covariance between the sender's equilibrium message about the the state of the world and its actual value, divided by the equilibrium message's variance. It therefore reflects the message's accuracy and it be shown that in equilibrium, its value has to equal the rational receivers' marginal reaction to the message, thus $\theta(m^*) = \rho^*$. Prooving this result constitutes the main building block of the following result which summarizes the above insights:

Proposition 1. *In every equilibrium of the communication game, the sender has the signalling strategy*

$$m^*(s,c) = s + c\,(\mu + (1-\mu)\rho^*) \tag{10}$$

while a rational receiver has equilibrium belief and strategy

$$\mathrm{E}[s|m^*] = d_r^*(m) = (1-\rho^*)\bar{s} + \rho^*\,(m - \mathrm{E}[c](\mu + (1-\mu)\rho^*)) \tag{11}$$

(unless otherwise stated, all proofs can be found in the appendix)

The result shows that the rational receiver's equilibrium inference $\mathrm{E}[s|m^*]$ is additive in two terms: The first, weighted by ρ^*, denotes the rational receiver's prior about the true state of the world, $\bar{s}$. The other part of the receiver's inference, weighted by $1-\rho^*$, is given by the received message which she corrects by $\mathrm{E}[c]\,(\mu + (1-\mu)\rho^*)$. Following the sender's equilibrium strategy (10), the correction therefore equals the expected sender's equilibrium bias given by

$$\beta^*(c) = m^*(s,c) - s = c\,(\mu + (1-\mu)\rho^*) \tag{12}$$

Since with undisclosed COIs, receivers do not know the actual bias, this correction is based on the *expected* commission and can therefore be wrong in both, direction and magnitude. This possible failure in the rational receiver's message correction and her risk-aversion provides the reason why

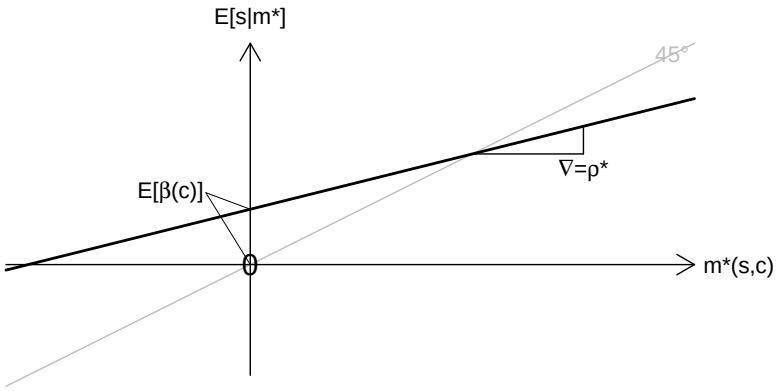

Figure 1: Rational equilibrium receiver's inference as a function of the sender's message

she will often not react fully (e.g. one-to-one) to the corrected message. Whenever $\rho^* \in (0,1)$ she strategically ignores the senders message by degree $1-\rho^*$ and instead puts that weight on her prior about the state of the world, $\bar{s}$. As an example, consider a situation where $\sigma_s^2 \to 0^+$: In this case $\text{Cov}[s, m^*]$ will converge to zero and so does ρ^* implying that a rational receiver act almost entirely according to her prior. The reason is that the actual state of the world s will be very close to its mean and prior $\bar{s}$ about it. Any variation in the signal can then only be due to the sender's bias. Just following $\bar{s}$ is therefore better since it brings her action arbitrarily close to the true state of the world and therefore her optimum.

The sender takes such a moderation of the receiver's reaction through ρ^* - and thereby the bias he chooses - into account: His bias is proportional to the receivers' average marginal demand which is, for rational receivers, is weighted with ρ^*. An important consequence of this, which will be examined in detail in the succeeding sections, is that higher values of ρ^* - thus a higher informativeness of the message - will lead to an increase in the sender's bias. Note the similarity to this important inference coefficient to the coefficient of a linear regression: Both, a regression coefficient and ρ^* describe the marginal change in a conditional expectation due to a marginal change in the conditioning variable. Figure 1 illustrates this by depicting the rational receiver's reaction as a (linear) function of the sender's message with slope ρ^* and an intercept equal to the expected bias of the sender. The crucial difference of the inference coefficient to a normal regression coefficient is that the latter refers to the effect of a change in an exogenous variable. For the inference coefficient, it is the change in the *endogenously* determined equilibrium message $m = m^*(s,c)$.

The results in proposition 1 determine the equilibrium behaviour, independently of the information structure, up to the equilibrium coefficient ρ^*. Determining this parameter's value therefore closes the model's solution. The main insight needed for this is that, given the messaging strategy $m^*(s,c)$, the values of the equilibrium parameter $\rho^* = \frac{\text{Cov}[s,m^*]}{\text{Var}[m^*]}$ can be obtained as a function of the parameters of F which shape the distribution of (s,c). I first do so for the case of undisclosed incentives and obtain the following result:

Proposition 2. *Consider the communication game described above with undisclosed incentives and denote its inference coefficient by ρ_U^*. Then the following holds:*

a) ***Existence:*** *There exists a threshold $\tau < 0$ such that $\rho_U^* > 0$ if and only if $\sigma_{sc} > \tau$.*

b) ***Uniqueness:*** *If it exists, an equilibrium with $\rho_U^* > 0$ is unique.*

c) ***Characterization:*** *The equilibrium inference coefficient ρ_U^* solves*

$$\rho = \frac{\sigma_s^2 + (\mu + (1-\mu)\rho)\sigma_{sc}}{\sigma_s^2 + 2(\mu + (1-\mu)\rho)\sigma_{sc} + (\mu + (1-\mu)\rho)^2\sigma_c^2} \tag{13}$$

The result implies that whenever the correlation of the comission and the state of the world surpasses a negative threshold, then there exists a unique equilibrium in which the sender's message and the actual state of the world are positively correlated. Note that the value of ρ^* is an equilibrium parameter and therefore anticipated by all rational players. I will therefore assume from now on that $\sigma_{sc} > \tau$ holds, such that there are is an unique equilibrium with $\rho_U^* > 0$. Otherwise, players would mutually anticipate that the expert's message about the state of the world and the actual state of the world are not positively correlated which is unlikely to be a feature of any information market with experts. Under this assumption, the solution to (13) characterizes the value of the inference coefficient and thereby the correlation between the message and the state of the world in terms of the model's primitives. While the exact value of ρ_U^* depends on the specific distribution parameters, one can can crucially limit its range for two important cases:

Proposition 3. *Assume that $\rho_U^* > 0$. Then $\rho_U^* < 1$ whenever $\sigma_{sc} \geq 0$ and/or $\sigma_s^2 \leq \sigma_c^2$.*

The above, together with (11) implies that whenever σ_{sc} is non-negative, the rational receiver's demand is a strictly convex combination between his prior $\bar{s}$ and the message, corrected for the average bias. The same applies when the variance of the commission which trigger the COI is larger

than the variance of the state of the world. Non-convex combinations are are however possible when the correlation between the commissions and the state of the world are sufficiently negative $(0 > \sigma_{sc} > \tau))$ *and* the state of the world is relatively more uncertain than these commissions $(\sigma_s^2 > \sigma_c^2)$. In this case, $\rho^* > 1$ can hold and rational receivers "over-react", thus a change in the sender's message can induces a change in rational receiver's demand greater than that original change in the message. In such an equilibrium, the message correlates relatively strongly with the state of the world. The reason is that σ_{sc} is sufficiently low, so that although $\text{Cov}[s, m^*]$ is small, the expected sign difference between the state of the world and the commissions lead to small values of $m = m^*(s, c)$ because the equilibrium messaging strategy combines these opposing effects and therefore even smaller values of $\text{Var}[m^*]$, resulting in $\rho_U^* > 1$. A rational receiver can counteract this bias, which can be expected to be in the other direction than s by over-reacting to the state of the world she extracts from the positively related message. Again, such extraction involves correcting for the *expected* bias and therefore runs the danger of dis-utility for wrong correction. Given the receivers' utility function the harm caused by wrong correct is the greater, the larger the realization of $(\beta^*(c) - \text{E}[\beta^*(c)])^2 = (c - \bar{c})^2$. This explains why when $\sigma_c^2 \geq \sigma_s^2$, thus when the commission which distorts the message about s becomes to unpredictable relative to the actual state of the world, over-reaction cannot occur.

The limit to acting upon expectation-based corrections are reached when $\sigma_{sc} \leq \tau$. In this case, the bias can be expected to be in the other direction than the state of the world and to be particularly strong when the magnitude of s is large. Consequently, a wrong correction has particularly strong, negative effects on the receiver's expected utility so that she is not willing to follow the receiver's corrected message anymore. Figure 2 illustrates these findings: It depicts the inference coefficient for possible correlations $\text{Corr}[s, c]$ and different values of σ_s^2. For the parameter combinations, it also portrays the cut-off value $\tilde{\tau} = \frac{\tau}{\sigma_s \sigma_c}$ as a vertical line [11]. If the correlation between between s and c is below it, an equilibrium in which s and m^* are positively correlated and rational receivers follow the corrected message does not exist.

[11] The normalization of $\tilde{\tau}$ follows from the fact that the graph depicts the correlation $\text{Corr}[s, c] = \frac{\sigma_{sc}}{\sigma_s \sigma_c}$ instead of the covariance σ_{sc} directly.

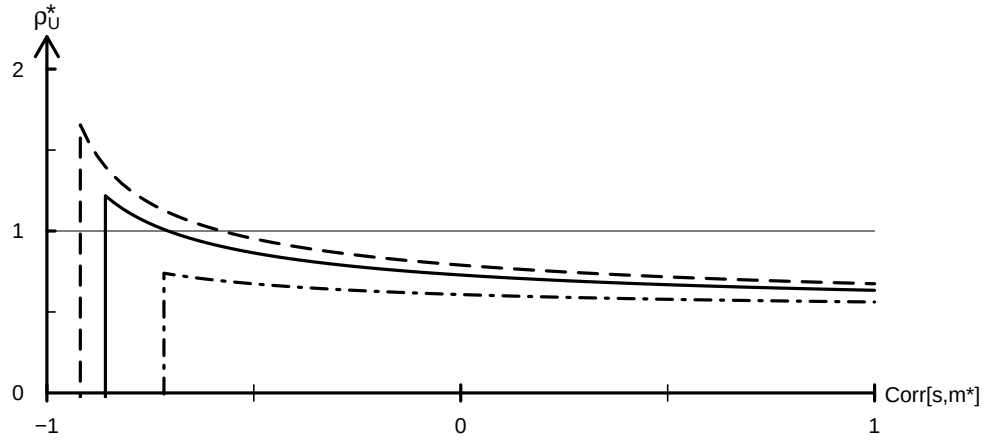

Figure 2: Equilibrium inference coefficients $\rho_U^* > 0$ over possible correlations between s and c. The chosen parameters are $\mu = 0.5$, $\sigma_c^2 = 1$, and $\sigma_s^2 \in \{1, 2, 3\}$ (increasing from bottom to top line).

V Disclosed conflicts of interests

The above discussion shows in a communication game with undisclosed COIs, the state of the world and the message describing it are often positively correlated but also, that this message is biased. Naive receivers who do not account for this bias are deceived by the sender and make wrong decisions. Rational receivers try to correct for the bias but whenever $c \neq \bar{c}$, thus almost surely, their decision is also sub-optimal. A tentative remedy to this consequence of the sender's COI is that he has to disclose it to the receivers. The underlying idea is that at least some receivers can take the sender's incentive for mis-representing the state of the world into account when they make their decision and thereby de-bias the sender's message. A further thought might be that after disclosure, the sender has a smaller incentive to mis-represent the true state of the world since he knows that rational receivers will correct for his bias. The analysis to follow shows that such reasoning can be wrong and that disclosure often results in an *increase* of the sender's bias.

For this, I first examine the game with disclosed COIs. Technically, I assume that a stage is added to the previous information transmission game. In it, a binary decision can be taken whether to disclose the sender's commission - and therefore his COI - after it has been realized. If this decision is negative, the game is that of the previous section, with information structure $\mathcal{I}_U$. If the decision is positive, this means that the sender's COI as captured by c becomes common knowledge and strategic uncertainty in this communication game disappears. The game with disclosed incentives then has the following timing:

1') the sender's type (s, c) is draw from F and observed by the sender; receivers observe c

2') the sender sends a signal m about s

3') receivers observe m, if rational update their belief about s, and choose their investments d

4') payoffs are realized

Accordingly, the information structure of the game with disclosed COIs is given by $\mathcal{I}_D = \{c\} \cup \mathcal{I}_U$. Since all result of proposition 1 apply independently of the information structure, the sender equilibrium messaging strategy in (10) and the rational receiver's given by (11) remain valid. Note however that $\mathrm{E}[c]$ which occurs in the latter expression equals the realization c - not $\bar{c}$ - since expectations are with respect to the information structure. Letting ρ_D^* denote the inference coefficient with disclosed commissions, one then gets the following equilibrium actions and beliefs in this game:

$$m^*(s, c) = s + c\,(\mu + (1 - \mu)\rho_D^*) \tag{10'}$$

$$d_r^*(m) = \mathrm{E}[s|m^*] = (1 - \rho_D^*)\bar{s} + \rho_D^*\,(m - c(\mu + (1 - \mu)\rho_D^*)) \tag{11'}$$

To determine the value of ρ_D^*, consider its nominator which is given by $\mathrm{Cov}[s, m^*]$. Because expectations are taken with respect to the respective information structure, so is this covariance. Also, with disclosed commissions the only random element in the sender's message is s. It follows that $(m - \mathrm{E}[m])|_{m=m^*(s,c)} = (s - \bar{s})$. Taken together, these points imply the following:

$$\mathrm{Cov}[s, m^*] = \mathrm{E}[(s - \bar{s})(m - \mathrm{E}[m])]|_{m=m^*(s,c)} = \sigma_s^2 = \mathrm{Var}[m]|_{m=m^*(s,c)} \tag{14}$$

In consequence, one gets that $\rho_D^* = \frac{\mathrm{Cov}[s,m^*]}{\mathrm{Var}[m^*]} = 1$. Applying this to (10') and (11') then proves the following result:

Proposition 4. *In the communication game with disclosed commissions there exists a unique equilibrium with $\rho_D^* = 1$. The sender has the signalling strategy $m^*(s, c) = s + c$ while a rational receiver has the equilibrium strategy and belief $\mathrm{E}[s|m^*] = d_r^*(m) = m - c$.*

The above result implies that in equilibrium, rational receivers are indeed able to completely de-bias the sender's signal since $\mathrm{E}[s|m^*] = d_r^*(m) = s$ when $m = m^*(s, c) = m + c$. They therefore perfectly extract s from the signal and choose their demand d_r^* accordingly. The main reason for this result is

that with disclosed COIs, the sender type becomes one- dimensional. His signalling function which is strictly increasing in s then becomes invertible such that, given knowledge of c, his complete type profile can be recovered from it. This is a special case of the more general result by Kartik et al. (2007) who show that with an unbounded, one-dimensional state space the sender's message is revealing [12].

Also in line with their result is the above finding that although the message is revealing, it is still biased. Disclosing COI does therefore not de-bias message per se but rather the content which receivers, at least rational ones, can get from it. No that this even holds when $\mu = 0$, thus when all receivers infer the true state of the world and communicating dishonestly is costly. To see why can truthful communication cannot be an equilibrium nevertheless, suppose this were the case and senders sent the truthful message $m = s$. Rational receivers would then implement this message directly by choosing $d_r = m$. Given such a response any $c \neq 0$ would induce the sender to over- or understate s, thus not to report truthfully.

VI Consequences of disclosure

The above analysis shows that disclosure of COIs enables rational receivers to completely learn the actual state of the world from the sender's message and react optimally to it. It however also shows that communication does not become truthful upon disclosure and may become even more biased. The reason is that the sender's bias is proportional to the marginal average reaction of receivers to his signal, given by $\mu + (1 - \mu)\rho^*$. Whenever the inference coefficient increases after disclosure, the bias increases. This hurts naive receivers who do not correct for the sender's bias or an increase therein. By denoting with with $u_j^R\left(\mathcal{I}_{i\in\{D,U\}}\right)$ the utility for receiver type $j \in \{n, r\}$ when equilibrium behaviour under the respective information structure is plugged into (1), this insight can be summarized as follows:

Corollary 1. *Disclosure helps naive receivers* $[u_n^R(\mathcal{I}_D) > u_n^R(\mathcal{I}_U)]$ *if and only if* $\rho_U^* > 1$.

Corollary 2. *Disclosure is a weak Pareto-improving among receivers* $[u_j^R(\mathcal{I}_D) \geq u_j^R(\mathcal{I}_U)$ *with the inequality strict for at least one element of* $j \in \{n, r\}]$ *if and only if* $\rho_U^* > 1$.

By proposition 3, this condition never met when $\sigma_{sc} \geq 0$ and/or $\sigma_s^2 \leq \sigma_c^2$. If for example the latter

[12] "Revealing" here refers to the fact that with disclosed COI the receiver learns the sender's type s in equilibrium.

condition holds, this would mean that the uncertainty about commissions add more to the overall uncertainty than the fundamental uncertainty about the state of the world. In this situation, with such (seemingly) salient causes of the inefficiencies in communication, demand for intervention is likely to be high. The above result however shows that in the presence of naive receivers, disclosure of COIs harms these receivers and leads to an increase in the bias. It is only when the necessary conditions for $\rho_U^* > 1$ are fulfilled, thus when $\sigma_s^2 > \sigma_c$ *and* $\sigma_{sc} < 0$ apply simultaneously, taht disclosure *can* help all receivers. An example for such a situation where disclosure is a Pareto-improvement is when over-reaction by receivers is observed, thus when they react stronger than one-to-one to a change in the sender's message. Since the average receivers marginal reaction to the message is given by $\mu + (1-\mu)\rho_U^*$ they can only over-react when this is greater than one, thus when $\rho_U^* > 1$ holds. The opposite is however the case when there is under-reaction, thus an average receiver reaction less than one-to-one. In particular, the reasoning that an under-reaction indicates an informational inefficiency and disclosing COIs help in improving efficiency without harming anyone than sender is wrong.

To quantify this insight, one can derive the receivers expected utilities from an ex-ante perspective, thus before (s, m) is realized. For this, a specific form for the receiver's loss function has to be assumed. Following the seminal example of Crawford and Sobel (1982) and many other papers in the literature on strategic communication, I will assume a quadratic loss function [13]:

$$L(d-s) = -\frac{1}{2}(d-s)^2 \tag{15}$$

Analogously to the use of $\mathcal{I}_i$, I use ρ_i^* for the inference coefficient and let $m^*(\mathcal{I}_i)$ denote the sender's optimal messaging strategic as specified in (10) and (10') for the respective information structure. I then obtain the following:

Proposition 5. *In any communication game with information structure and inference coefficient $\rho_i^* > 0$, the expected utility for naive receivers is given by*

$$\mathrm{E}[u_n^R(\mathcal{I}_i)] = -\frac{1}{2}(\mu + (1-\mu)\rho_i^*)^2(\sigma_c^2 + \bar{c}^2) \tag{16}$$

[13] Ottaviani (2000) considers the case of a receiver who has has mean-variance utility over lotteries (e.g. CRRA-utility). He shows that if the receiver knows the variance of of the lottery's payoff (e.g. returns to a risky asset) but not its expected value s, then the ex-post utility of investing d into the lottery can (up to constant) be represented by this quadratic loss function.

while the expected utility for rational receivers given by

$$\mathrm{E}[u_r^R(\mathcal{I}_i)] = -\frac{1}{2}\left(\sigma_s^2 - \rho_i^{*2}\, Var[m^*(\mathcal{I}_i)]\right) \tag{17}$$

with $\mathrm{E}[u_n^R(\mathcal{I}_i)] \leq \mathrm{E}[u_r^R(\mathcal{I}_i)] \leq 0$.

Expression (16) shows that naive receiver's dis-utility is proportional to the sum of the commission's squared expected value and its variance. The first term reflects the adverse effect of following the sender's biased, whose root-cause is the commission with expected value $\bar{c}$. Due to risk-aversion of receivers there is further dis-utility through strategic uncertainty as measured by the commission's variance σ_c^2. These combined effects are scaled by the squared average marginal reaction of receivers to which the sender adjusts his bias. Note that despite risk-aversion, the naive receiver's expected utility does not depend directly on the underlying fundamental uncertainty, captured by σ_s^2 [14]. The reason is that mis-investment of naive receivers, given by $d_n^* - s$, is just equal to the bias which is independent of the state of the world.

This is different for rational receivers as expression (17) shows: Their strategy rests on partly following their prior for the state $\bar{s}$, such that the (squared) deviations from this prior yield a dis-utility proportional to σ_s^2. However, they do better than just following the prior by also incorporating what they extract about the true state of the world from the sender's message in their strategy. This is reflected in the additional utility proportional to $\rho_i^{*2}\mathrm{Var}[m^*]$. In particular, this expression takes a value of σ_s^2 when commissions are disclosed and thereby induces a maximal utility of zero for rational receivers [15].

Given these insights on the opposing effects of disclosing COIs, I will now consider another criteria to judge its consequences. For this, I define the following welfare measure $W(I_i)$ as a function of the information structure I_i:

$$W(I_i) = \mathrm{E}[(1-\mu)U_r^R(I_i) + (\mu + p)U_n^R(I_i)] \tag{18}$$

This term is a weighted sum of expected utilities for naive and rational receivers. Rational receivers' weight in $W(I_i)$ corresponds to their population share $1-\mu$. For naive receivers, there is an additional weight $p \geq 0$ added to the their population share μ. It represents the costs of mis-

[14] only indirectly through σ_s^2 entering ρ_i^* as specified in (11).
[15] To see this, note that ${\rho_D^*}^2 = 1$ and by (14) that $\mathrm{Var}[m^*] = \sigma_s^2$ when COIs are disclosed.

representing experienced by the sender. Since $d_n(m) = m$, this is equal to adjusting the mass of naive receivers by p. This weight can then be positive when the policy criterion reflects material inefficiencies and lying has pecuniary consequences, e.g. when the sender is a financial analyst who has to invest a share of is own equity according to his advise. Alternatively, if the cost represent expected fines for lying these are effectively transfers and should, when only material efficiency matters, yield a weight of $p = 0$. For the same reason, the sender's earned commission $cD(m)$ does not appear in the above criterion as it is a lump-sum transfer from a third party to the sender. Given this definition, I will call disclosure of COI "efficient" if and only if $W(I_D) \geq W(I_U)$.

The fact that disclosure of COI can lead to an increase in the reaction of rational receivers to the sender's message indicates that he might voluntarily commit to disclose COIs. The underlying motive is that when ρ_U^* is low, a rational receivers will ignore the sender's signal to a large extend and will choose an action close to her prior, independently of the actual state. When there are relative large commissions associated with extreme values of the state of the world, thus when σ_{sc} is relatively large, then the sender however wants receivers to react to the state of the world. By committing to disclose, a sender can achieve just this since rational receivers will extract s from the sender's message and therefore reacts fully to it. To capture such considerations, I will assume that a sender commits to disclose his COI if and only if this increases his expected utility. For voluntary disclosure it therefore has to hold that the sender's expected utility $\mathrm{E}[U^S(\mathcal{I}_i)] = \mathrm{E}[D\left(m^*(\mathcal{I}_i)\right) + u_n^R(\mathcal{I}_i)]$ [16] has to be greater under disclosed than undisclosed commissions.

Applying the above definitions for crucial for efficient and voluntary disclosure then yields the following result:

Proposition 6. *Suppose that $\sigma_{sc} > \tau$ and consider a change from undisclosed to disclosed COIs:*

a) disclosure is efficient $[W(\mathcal{I}_D) > W(\mathcal{I}_U)]$ if and only if

$$\sigma_{sc} \leq \gamma^E \equiv \frac{1-\rho_U^*}{(\mu+(1-\mu)\rho_U^*)\rho_U^*} \cdot \left(\sigma_s^2 - (\mu+p)\left(1+\mu+\rho_U^*(1-\mu)\right)\left(\bar{c}^2+\sigma_c^2\right)\right)$$

b) disclosure is provided voluntarily $[\mathrm{E}[U^S(\mathcal{I}_D)] > \mathrm{E}[U^S(\mathcal{I}_U)]\ ]$ if and only if

$$\sigma_{sc}(1-\rho_U^*) \geq \gamma^V \equiv \frac{1}{2}(1-\mu)(1-\rho_U^*)^2(\bar{c}^2+\sigma_c^2) + \rho_U^*\left(\mu+(1-\mu)\rho_U^*\right)\sigma_c^2 > 0$$

[16] In analogy to the above reasoning for the weight p, the utility for naive receivers appears in the sender's expected utility since $d_n^*(m) = m$ and therefore $u_n^R(\mathcal{I}_i) = K(m^*(\mathcal{I}_i), s)$.

From Corollary 2 one gets that the condition stated under a) is always true when $\rho_U^* > 1$. When this is not the case, the condition becomes more slack the lower $\bar{c}^2 + \sigma_c^2$ which measures the naive receiver's expected additional dis-utility from disclosure. Since this is the same as the sender's expected cost of lying, this term is also scaled by p. The condition also becomes more slack for higher σ_s^2 since this reflects the rational receivers' utility gain from being able to de-bias the message following disclosure.

While the sign of the threshold γ^E for efficient disclosure is not determined, the threshold for voluntary disclosure $\gamma^E/(1-\rho_U^*)$ is strictly positive whenever disclosure is not a Pareto-improvement. This implies that when commissions and the state of the world are un- or negatively correlated, voluntary disclosure will not occur. More generally, the threshold for efficient disclosure is an upper limit on σ_{sc} while in situations when disclosure is not a Pareto-improvement, the threshold for voluntarily disclosure is a lower limit. An important conclusion is therefore that the conditions under which the sender and a regulator (if the latter is concerned with efficiency) would disclose COIs are in principle not aligned. Focussing on the relevant case of no Pareto-improvement among receivers, one relevant case is when disclosure is efficient but not provided voluntarily ($\sigma_{sc} < \min\{\gamma^E, \gamma^V/(1-\rho_U^*)\}$). Such a scenario would then require disclosure to become mandatory. Another case occurs when disclosure is not efficient but provided voluntarily ($\sigma_{sc} > \max\{\gamma^E, \gamma^V/(1-\rho_U^*)\}$). In such a situation a regulator faces the task of preventing a voluntarily commitment to disclosure. If neither of the two conditions applies then either the sender does not disclose voluntarily and this is efficient or he does and disclosure is efficient.

The above reasoning was based on the assumption that there is no Pareto-improvement upon disclosure. If disclosure is an improvement, thus if $\rho_U^* > 1$ holds, the sign of condition in b) changes and senders will then disclose voluntarily whenever $\sigma_{sc} \leq \gamma^V/(1-\rho_U^*) < 0$. The sender then looses the possibility to deceive rational receivers and disclosure will also lead to a decrease in their reaction to his message. The reason why he nevertheless prefers to commit to disclose is that in such situation. σ_{sc} has to have a particular low, negative value. This implies that the sender can expect to have a strong incentive to bias his message in the other direction than the actual state of the world, resulting in high expected lying costs $K(m,s)$. In addition, a high absolute value of the covariance implies that this incentive to bias the message is the stronger, the more extreme the realization of the state of the world which adds again to the expected lying costs. Furthermore, $\rho_U^* > 1$ give high weight to information which, in expectation, has been corrected for the sender's

bias [17]. These then lead to a situation where it is preferable for the sender to commit to disclose his COI and still earn on the commissions.

VII Conclusion and discussion

"*Discounting advise appropriately for a disclosed conflict of interest requieres a mental model of advisor behavior*[...]" is what Loewenstein et al. (2011), p. 424 conclude given in their work on the failure of disclosing COIs. This paper provides such a model in a setting where a biased expert communicates the value of a random variable of interest to uninformed receivers, some of which are naive. Solving this model enables me to compare the implications which the disclosure of COIs has on those giving and those receiving advise.

First, I find that disclosure fulfils the aim of informing *rational* receivers: It transforms the game's non-separating equilibrium into a separating equilibrium which is separating and in which rational receivers can make full use of the expert's information. On the downside however, this paper's core result identifies a channel over which such disclosure can backfire. It does so because disclosure can lead to *both*, increased average reaction to the biased signal and an increased sender bias. Naive receivers who do not account for the higher bias are then hurt by disclosure. In consequence, mandatory disclosure is often not a policy which helps the average receiver. In contrast, it does often constitute a transfer of rents from naive receivers to senders and rational receivers and thereby hurts those which are most vulnerable to strategic biases. The adverse effect on naive receivers originates from an increased reaction of rational receivers and thus scales in proportion to their share in the overall population. This implies that even when the share of naive receivers is relatively low or decreases upon disclosure, the potential negative effect of disclosure on them is particularly strong.

As a second result, I determine more precisely when these adverse effects of disclosure manifest. In terms of economic fundamentals, this is the case when the state of the world and the sender's COI are weakly positively correlated. Another sufficient condition for disclosure to backfire is when the variance of the sender's commissions exceed the variance of the the variable which describes the state of the world. In terms of observed behaviour, a reaction of receivers which is less than one-to-one to the sender's message is a necessary and sufficient condition for negative effects of disclosure on naive receivers. Consequently, the observation that expert markets are not efficient

[17] See also the the explanation on the reasoning underlying equilibria with $\rho_U^* > 1$ on page 15.

because their messages do not have a sufficiently strong impact on receivers' does not imply that disclosure is beneficial policy. On the contrary, I show that in all of the above scenarios, disclosure of COI hurts naive receivers. Only when receivers react on average stronger than one-to-one to the sender's message, then disclosure is a Pareto-improvement among receivers.

The third main result considers situations when this is not the case and other criteria have to be applied to decide about disclosure. While clear cut results depend on the specific information structure, the correlation between incentives and the state of the world is again crucial. If this correlation is below a negative threshold then disclosure increases efficiency. I also show that senders may also want to voluntarily commit to disclose his COI. This is the case when the same covariance which determines efficiency of disclosure is above a positive threshold. This result shows that there can be situations in which disclosure is not provided voluntarily though it is efficient, in which case it should be made mandatory. It also show however, that there can be the reverse case where it is provided voluntarily but is not efficient, leaving policy maker with the challenge of preventing disclosure.

These results show that disclosure of conflict of interests is often not a good tool to improve the quality of strategy communication. While it helps some receivers, it often does so on the cost of those who need most help - strategically unsophisticated non-experts. By providing a model which takes account of sender's and receiver's mutually dependent reaction before and after disclosure I identify how and when such a change in the information structure backfires. Besides the presence of some naive or delegating receivers. the model is friction-free. The results therefore show that even when psychological biases and issues of efficient communication and disclosure can be resolved, there are robust economic fundamentals which make disclosure often a, at best, two-sided sword. In consequence, the results support the conclusion that inefficiencies arising from conflicts of interest are best solved by eliminating rather than communicating them.

Appendix

Proof of Proposition 1

The proof is constructive and proceeds in three steps. Step 1 solves the rational receivers problem to extract s from the sender's message. Step 2 determines how this signal extraction manifests in equilibrium when the sender anticipates this process. Step 3 combines these results to obtain equilibrium actions and beliefs.

Step 1:

According to (8) the equilibrium signal m^* has to solve the expression $m - c(1-\mu)\theta(m^*) = s + c\mu = z(s,c)$. The random variable z defined in this way is normally distributed following $\mathcal{N}(\bar{s} + \bar{c}\mu, \sigma_s^2 + \mu^2\sigma_c^2 + 2\mu\sigma_{c,s})$. Suppose now the receiver observes the image z of the function $z(s,c)$. According to standard signal extraction techniques his inference about s given z equals

$$\mathrm{E}[s|z] = \bar{s} + (z - \mathrm{E}[z])\,\frac{\mathrm{Cov}[s,z]}{\mathrm{Var}[z]} = \bar{s} + (z - \bar{s} - \mu\bar{c})\,\frac{\mathrm{Cov}[s,z]}{\mathrm{Var}[z]}$$

However, what the receiver actually observes is m which by (8) is an image of the function $m(s,c) = z + c(1-\mu)\theta(m^*)$. Since the cost of issuing a message is convex in m and receivers' actions are only based on the (conditional) expectation it induces, the sender will never use a mixed strategy when there exists a pure strategy equilibrium. It follows that in any such equilibrium, given (s,c) the associated signalling function $m(s,c)$ is deterministic and so is the marginal inference effect $\theta(m) = \frac{\partial \mathrm{E}[s|m]}{\partial m}$. It follows that the only random element in the term $c(1-\mu)\theta(m)$ is c and the term is normally distributed according to $\mathcal{N}(\bar{c}(1-\mu)\theta(m^*), [(1-\mu)\theta(m^*)]^2\sigma_c^2)$. Since both, z and $c(1-\mu)\theta(m^*)$ are normally distributed, so is their sum which, by (8), equals the senders message $m = z + c(1-\mu)\theta(m^*)$. This implies that the signal extraction can be used again to obtain the expected value of s, given the equilibrium signal m by the following expression:

$$\mathrm{E}[s|m] = \bar{s} + (m - \bar{s} - \mathrm{E}[m])\,\frac{\mathrm{Cov}[s,m]}{\mathrm{Var}[m]} = \bar{s} + (m - \bar{s} - \bar{c}\,(\mu + (1-\mu)\theta(m^*)))\,\frac{\mathrm{Cov}[s,m]}{\mathrm{Var}[m]} \qquad (19)$$

Step 2:

Note that, given the above reasoning both $\mathrm{Cov}[s,m]$ and $\mathrm{Var}[m]$ only depend on μ and the parameters of F. By defining $\rho = \frac{\mathrm{Cov}[s,m]}{\mathrm{Var}[m]}$ and differentiating (19) one obtains

$$\theta(m) = \frac{\partial \mathrm{E}[s|m]|}{\partial m} = \left(1 - \bar{c}(1-\mu)\frac{\partial^2\left(\mathrm{E}[s|m]|\right)}{(\partial m)^2}\right)\rho = \left(1 - \bar{c}(1-\mu)\theta'(m)\right)\rho$$

The solution to this first-order linear differential equation gives the equilibrium reaction of rational receivers to a change in the sender's message when the sender anticipates their inference, $\theta(m^*)$.

When $\rho = 0$ it follows that $\theta^*(m) = \rho = 0$. Similarly, if $\bar{c} = 0$, then $\theta^*(m) = \rho$. Now suppose that $\rho\bar{c} \neq 0$. One then gets $\theta^*(m)$ as the solution to the above general expression given by

$$\theta(m^*) = \rho + \xi \cdot exp\left(-\frac{m}{(1-\mu)\bar{c}\rho}\right)$$

where ξ is an integration factor. To determine its value, I integrate the obtained $\theta(m^*)$ and get

$$\mathrm{E}[s|m] = \int \theta^*(m)dm = m\rho - \xi(1-\mu)\bar{c}\rho \cdot exp\left(-\frac{m}{(1-\mu)\bar{c}\rho}\right) + K \tag{20}$$

where K is a constant of integration. This can be plugged into the sender's expected utility (7) to obtain

$$U^S(s,c,m) = c\mu m + c(1-\mu)\left[m\rho - \xi(1-\mu)\bar{c}\rho \cdot exp\left(-\frac{m}{(1-\mu)\bar{c}\rho}\right) + C\right] - \frac{1}{2}(m-s)^2 \tag{21}$$

To determine ξ, I start with the case that $c > 0$. In this case, $U^S(s,c,m)$ is increasing in $\mathrm{E}^*[s|m]$. If $\rho\bar{c} > 0$, $\mathrm{E}^*[s|m]$ and therefore the sender's expected utility decreases exponentially in m while all other terms involving m are either linear or quadratic. If $\xi < 0$ the sender's maximizes expected utility by choosing $m \to -\infty$ and there is no equilibrium. Therefore, $\xi \geq 0$ has to hold in this case for any equilibrium. For $\xi > 0$ however, $U^S(s,c,m)$ would be lower than with $\xi = 0$. Since ξ is part of the endogenous inference of the sender's signal, he will not send a signal which allows such an inference. With $c > 0$ and $\rho\bar{c} > 0$ only $\xi = 0$ can therefore be the equilibrium integration factor. Continue to suppose that $c > 0$ but now $\rho\bar{c} < 0$ holds. Similar to the above reasoning, $\mathrm{E}^*[s|m]$ now increases exponentially in m which implies a global maximum of the sender's expected utility at $m \to +\infty$ whenever $\xi > 0$. Thus, for an equilibrium $\xi \leq 0$ has to hold. Again, any strictly negative value of ξ would decrease the sender's expected utility. Messaging strategies allowing such inference are therefore not chosen by the sender and $\xi = 0$ holds in any equilibrium with $c > 0$ and $\rho\bar{c} < 0$.

For the case that $c < 0$, $U^S(s,c,m)$ is decreasing in $\mathrm{E}^*[s|m]$. The same reasoning as for the case of $c > 0$ but with reversed signs can then be repeated which rules out any $\xi \neq 0$ in equilibrium when $c < 0$ and $\rho\bar{c} \neq 0$.

Eventually, when $c = 0$ the inference $\mathrm{E}[s|m]$ does not enter $U^S(s,c,m)$ and therefore does neither affect the sender's action nor the receiver's reaction to it and one can assume w.l.o.g. $\xi = 0$. It therefore has to hold in any equilibrium that $\xi = 0$ and therefore $\theta^*(m) = \rho$.

Step 3

Given the above, one can determine the integration constant

$$K = \bar{s} + [\bar{s} - \bar{c}\,(\mu + (1-\mu)\rho)]\rho \tag{22}$$

by combining (19) and (20). Plugging this and $\xi = 0$ into (21) yields after simplifying

$$U^S(s,c,m) = mc\,(\mu + (1-\mu)\rho^*) - \frac{1}{2}(m-s)^2 + c(1-\mu)K$$

which is easily verified to take a unique maximum such that the messaging function is $m(s,c) = s + c\,(\mu + (-\mu)\rho$. Given this functional form, it has to hold in equilibrium that

$$\theta(m^*) \equiv \theta^*(m)|_{m=m^*(s,c)} = \rho^* = \frac{\mathrm{Cov}[s,m^*]}{\mathrm{Var}[m^*]} = \frac{\mathrm{Cov}[s,m]_{m=m^*(s,c)}}{\mathrm{Var}[m]_{m=m^*(s,c)}}$$

with $m^*(s,c) = s + c\,(\mu + (1-\mu)\rho^*$ as stated in (10). Using this, $\xi = 0$, and (22) on (20) then yields the rational receivers inference and strategy as stated in (11). □

Proof of Proposition 2

By the definition of ρ^*, it must be a solution to the expression

$$\rho = \frac{\text{Cov}[s, m^*]}{\text{Var}[m^*]} = \frac{\sigma_s^2 + (\mu + (1-\mu)\rho)\sigma_{sc}}{\sigma_s^2 + 2(\mu + (1-\mu)\rho)\sigma_{sc} + (\mu + (1-\mu)\rho)^2\sigma_c^2} \tag{23}$$

where I used that $m^*(s, c) = s + c(\mu + (1-\mu)\rho)$ from proposition 1. This proves result c).

I will now derive a necessary and sufficient condition for an unique equilibrium with $\rho_U^* \geq 0$. If ρ_U^* is unique, so is the equilibrium since this is the only free parameter in the players' equilibrium strategies.

Necessity

For an equilibrium with $\rho_U^* > 0$ it is necessary that $\text{Cov}[s, m^*] = \sigma_s^2 + (\mu + (1-\mu)\rho_U^*)\sigma_{sc} > 0$ which implies that $\sigma_{sc} > \tau(\rho_U^*)$ where $\tau(\rho) = -\frac{\sigma_s^2}{\mu + (1-\mu)\rho}$. Note that $\tau(\rho)$ is negative and strictly increasing in ρ for all $\rho \geq 0$.

Sufficiency

To see that $\sigma_{sc} > \tau(\rho_U^*)$ is also sufficient for (23) to have a solution $\rho^{U*>0}$ note that this condition implies $\text{Cov}[s, m^*]|_{\rho=0} > 0$. Therefore, the RHS of the expression in (23) is strictly positive at $\rho = 0$. Now consider its derivative w.r.t. ρ:

$$\begin{aligned}\frac{\partial\left(\frac{\text{Cov}[s,m^*]}{\text{Var}[m^*]}\right)}{\partial\rho} &= \frac{(1-\mu)\sigma_{sc}\text{Var}[m^*] - 2(1-\mu)\text{Cov}[s, m^*](\sigma_{sc} + (\mu + (1-\mu)\rho)\sigma_c^2)}{(\text{Var}[m^*])^2} \\ &= (1-\mu) \cdot \frac{\sigma_{sc} - 2\rho(\sigma_{sc} + (\mu + (1-\mu)\rho)\sigma_c^2)}{\text{Var}[m^*]}\end{aligned}$$

The nominator of the above expression decays quadratically in ρ and increases at most linearly in it [18]. while the denominator is strictly positive. Therefore, there exists some $\rho' > 0$ such that the RHS of (23) is monotonically decreasing for all $\rho \geq \rho'$. The LHS of (23) is increasing with a slope of one through the origin. If the value of the RHS of (23) is greater than ρ', then there must be an intersection with the LHS in $[\rho', \infty)$. Conversely, if the value of the RHS of (23) is smaller or equal than ρ', then there must be an intersection with the LHS in $(0, \rho']$. The necessary and sufficient condition then prove result a).

Uniqueness

To proof uniqueness, I will make use of the following result:

Theorem. ***(Descarte's rule of signs)*** *Consider a n-degree polynominal $p(x) = \sum_{k=0}^{n} c_k \cdot x^k$ with real coefficients. Order the non-zero coefficients in an descending order of the exponent of x. The number of positive, real roots of the polynomial is less by an even number or equal to the number of sign changes between successive coefficients in this ordering.*

[18] "increases (decreases) at most linearly in ρ" here means that the all terms which increase (decrease) in ρ are bounded above (below) by some linear function of ρ with strictly positive (negative) coefficients.

Solving (23) can be done by finding the roots to the cubic equation $A\rho^3 + B\rho^2 + C\rho + D = 0$ with coefficients

$$A = (1-\mu)^2\sigma_c^2 \quad B = 2(1-\mu)(\sigma_{sc} + \mu\sigma_c^2) \quad C = \sigma_s^2 + \mu^2\sigma_c^2 + (3\mu - 1)\sigma_{sc} \quad D = -\sigma_s^2 - \mu\sigma_{sc}$$

I will now use these coefficients and the sign rule to establish that there is at most one positive, real solution to this polynomial. Together with the above sufficient condition this implies existence of a unique equilibrium.

First note that it always holds that $A > 0$ and $D < 0$. The last inequality is a consequence of $\text{Cov}[s, m^*]|_{\rho=0} > 0$ which has been shown above to follow from the necessary and sufficient condition for equilibria with ρ_D^*. For the signs of B and C consider the following cases:

i) First suppose that $\sigma_{sc} \geq 0$: Then it holds that $B > 0$. Together with $A > 0$ and $D < 0$ the sign rule implies that there is at most one positive real solution, independent of the sign of C.

ii) Now suppose that $\sigma_{sc} \in (\tau(\rho_U^*), 0)$ and $\mu \leq \frac{1}{3}$: It then holds that $C > 0$, together with $A > 0$ and $D < 0$. If in addition $\sigma_{sc} > \tau' \equiv -\sigma_c^2\mu$ holds this is equivalent to $B > 0$. If this condition is fulfilled, thus if $0 > \sigma_{sc} \in (\max\{\tau(\rho), \tau'\}, 0)$, then the sign rule again implies at most one positive real solution.

If in contrast $0 > \sigma_{sc} \in (\tau(\rho), \tau']$ holds, this means that $A > 0$, $B \leq 0$, $C > 0$ and $D < 0$ and there are either one or three positive real solutions. If $B = 0$, there is exactly one sign change and therefore at most one positive solution. Now suppose $B < 0$ and examine the extreme points of the above cubic equation. One obtains them as the roots of the cubic equation's derivative w.r.t. ρ, thus by $\rho_{1/2} = -B/(3A) \pm \sqrt{(B/(3A))^2 - C/(3A)}$. The discriminant of this solution is positive if and only if $B \geq \sqrt{3AC}$ holds. Since $B < 0$ while $A > 0$ and $C > 0$ this can never be true. The cubic function has therefore no extreme points and is monotone in ρ with at most one positive solution.

iii) Finally assume that $\sigma_{sc} \in (\tau(\rho_U^*), 0)$ and $\mu > \frac{1}{3}$: One now has to determine the sign of C. The condition for it to be strictly positive is $\sigma_{sc} > \frac{-\sigma_s^2 - \mu^2\sigma_c^2}{3\mu - 1} \equiv \tau''$. If it holds, then $C > 0$ and one can repeat the same arguments as in case ii). With $\sigma_{sc} \leq \tau''$ one has $C \leq 0$ as well as $A > 0$ and $D < 0$ which, by the sign rule, implies at most one positive real solution, independent of the sign of B.

Taken together, these arguments show that whenever the necessary condition for a reliable equilibrium with ρ_U^* is fulfilled, it has to be unique which establishes result b). □

Proof of Proposition 3

By proposition 2, any equilibrium inference coefficient $\rho_U^* > 0$ is unique. From $\sigma_{sc} \geq 0$ it follows that $\text{Var}[m^*] = \sigma_s^2 + 2(\mu + (1-\mu)\rho_U^*)\sigma_{sc} + (\mu + (1-\mu)\rho_U^*)^2\sigma_c^2 > \text{Var}[s] = \sigma_s^2$ which implies the following condition on this value:

$$0 < \rho_U^* = \frac{\text{Cov}[s, m^*]}{\sqrt{\text{Var}[m^*]}\sqrt{\text{Var}[m^*]}} < \frac{\text{Cov}[s, m^*]}{\sqrt{\text{Var}[s]}\sqrt{\text{Var}[m^*]}} = \text{Corr}[s, m^*] \leq 1$$

To show that $\sigma_s^2 \leq \sigma_c^2$ implies $\rho_U^* < 1$ suppose the opposite, thus $\rho_U^* \geq 1$ and, in particular, that $\rho_U^* > \frac{1}{2}$. Following from (13), this inference coefficient has to be a solution to the following expression:

$$\rho = \frac{\sigma_s^2 + (\mu + (1-\mu)\rho)\sigma_{sc}}{\sigma_s^2 + 2(\mu + (1-\mu)\rho)\sigma_{sc} + (\mu + (1-\mu)\rho)^2\sigma_c^2} > \frac{1}{2}$$

Simplifying the inequality and using the initial assumptions yields $1 > \sigma_s^2/\sigma_c^2 > (\mu + (1-\mu)\rho)^2$. The conditions inherent in this quadratic inequality then require $\rho \in (-\frac{1+\mu}{1-\mu}, 1)$ which contradicts an equilibrium solution $\rho_U^* = \rho \geq 1$.

□

Proof of Proposition 5

By (1), one obtains $\mathrm{E}[u_j^R] = -\frac{1}{2}\mathrm{E}[(d_j^*(m) - s)^2]$ where $j \in \{n, r\}$ is an index which denotes naive and rational receivers, respectively. By assumption it holds that $d_n^*(m) = m$ where, in equilibrium, $m = s + c(\mu + (1-\mu)\rho^*)$. Plugging this into $\mathrm{E}[u_n^R]$ yields $-\frac{1}{2}\mathrm{E}[c(\mu + (1-\mu)\rho^*)^2]$. Since ρ^* is non-random, only the variable c in the argument of the above expectation is random. It follows that

$$\begin{aligned}\mathrm{E}[(c(\mu + (1-\mu)\rho^*))^2] &= \mathrm{Var}[c(\mu + (1-\mu)\rho^*)] + (\mathrm{E}[c(\mu + (1-\mu)\rho^*)])^2 \\ &= (\mu + (1-\mu)\rho^*)^2(\sigma_c^2 + \bar{c}^2) > 0\end{aligned}$$

For the rational receivers' expected utility one obtains from proposition 2 by similar reasoning that

$$\begin{aligned}d_r^*(m) = \mathrm{E}[s|m] &= (1-\rho^*)\bar{s} + \rho^*\left[m - \bar{c}\left(\mu + (1-\mu)\rho^*\right)\right]\rho_U^* \\ &= \bar{s} + \left[m - \bar{s} - \bar{c}\left(\mu + (1-\mu)\rho^*\right)\right]\rho^* \\ &= \bar{s} + \left(m - \mathrm{E}[m^*]\right)\rho^*\end{aligned}$$

By using $\rho^*\mathrm{Var}[m^*] = \mathrm{Cov}[s, m^*]$ one can then establish that the following holds which completes the proof:

$$\begin{aligned}\mathrm{E}[(d_r^*(m) - s)^2] &= \mathrm{E}[(-(s - \bar{s}) + ((m^* - \mathrm{E}[m^*])\rho^*)^2] \\ &= \mathrm{Var}[s] - 2\rho^*\mathrm{Cov}[m^*, s] + (\rho^*)^2\mathrm{Var}[m^*] \\ &= \sigma_s^2 - \rho^{*2}\mathrm{Var}[m^*] > 0\end{aligned}$$

To see why $\mathrm{E}[u_n^R(\mathcal{I}_i)] = -\frac{1}{2}\mathrm{E}[(d_n^*(m) - s)^2] \leq \mathrm{E}[u_r^R(\mathcal{I}_i)] = -\frac{1}{2}\mathrm{E}[(d_r^*(m) - s)^2] \leq 0$ holds suppose that the first inequality does not hold (the validity of the second is immediate). Then $\mathrm{E}[u_n^R(\mathcal{I}_i)] > \mathrm{E}[u_r^R(\mathcal{I}_i)]$ and rational receivers would not be playing a best response whenever $d_r^*(m) \neq m$ which would lead to an immediate contradiction in these cases.

It remains to show $\mathrm{E}[u_n^R(\mathcal{I}_i)] > \mathrm{E}[u_r^R(\mathcal{I}_i)]$ and $d_r^*(m) = m$ cannot be valid simultaneously. To see this, observe that the last statement implies by (11) that $\rho^* = 1$ and $\mathrm{E}[c] = 0$ have to apply simultaneously. To see that this can never be the case, regard the two possible information structures: If $\mathcal{I} = \mathcal{I}_D$, this implies $\mathrm{E}[c] = c = 0$. Form this follows that $\mathrm{E}[u_r^R(\mathcal{I}_D)] = \mathrm{E}[u_n^R(\mathcal{I}_D)] = 0$

which contradicts the initial assumption. If $\mathcal{I} = \mathcal{I}_U$ then $\mathrm{E}[c] = \bar{c} = 0$. Together with $\rho^* = 1$, (13) this implies the following equivalent conditions

$$\sigma_{sc} + \sigma_c^2 = 0 \Leftrightarrow \mathrm{E}[(s - \bar{s})(c - \bar{c}) + (c - \bar{c})^2] = 0 \Leftrightarrow \mathrm{E}[sc] + \mathrm{E}[c^2] = 0$$

where I used $\bar{c} = 0$ to obtain the last statement. This would require that $s = -c$, thus s and c to be perfectly (negatively) correlated. By using this and $\rho^* = 1$ again, the sender would then, according to (10) send $m^*(s, c) = 0$ for any realization (s, c) which implies $\mathrm{Cov}[s, m^*] = 0$ and therefore $\rho^* = 0 \neq 1$. □

Proof of Proposition 6

I start by evaluating the welfare gain of disclosure: (16) indicates that the effect the information structure on $\mathrm{E}[U_n^R(\mathcal{I}_i)]$ is entirely determined by ρ_U^* and $\rho_D^* = 1$ which yields the following:

$$2\left(\mathrm{E}[U_n^R(\mathcal{I}_D)] - \mathrm{E}[U_n^R(\mathcal{I}_U)]\right) = -\left[1 - (\mu + (1-\mu)\rho_U^*)^2\right](\bar{c}^2 + \sigma_c^2)$$

By proposition 4, rational receivers infer s from the sender's message and thus $\mathrm{E}[U_r^R(\mathcal{I}_D)] = 0$. By applying (17) to get $\mathrm{E}[U_r^R(\mathcal{I})]$ and using that $\rho^* \mathrm{Var}[m^*] = \mathrm{Cov}[s, m^*] = \sigma_s^2 + (\mu + (1-\mu)\rho^*)\,\sigma_{sc}$ from proposition 1 yields

$$2\left(\mathrm{E}[U_r^R(\mathcal{I}_D)] - \mathrm{E}[U_r^R(\mathcal{I}_U)]\right) = \sigma_s^2 - \rho^* \mathrm{Cov}[s, m^*] = (1-\rho^*)\sigma_s^2 - \rho^*\left(\mu + (1-\mu)\rho^*\right)\sigma_{sc}$$

Weighting these expressions with $\mu + p$ and $1 - \mu$ respectively, I get

$$\begin{aligned}
2(W(\mathcal{I}_D) - W(\mathcal{I})) &= 2(\mu + p)(\mathrm{E}[U_n^R(\mathcal{I}_D)] - \mathrm{E}[U_n^R(\mathcal{I})]) + 2(1-\mu)\left(\mathrm{E}[U_r^R(\mathcal{I}_D)] - \mathrm{E}[U_r^R(\mathcal{I})]\right) \\
&= -(\mu + p)\left[1 - (\mu + (1-\mu)\rho_U^*)^2\right](\bar{c}^2 + \sigma_c^2) \\
&\quad + (1-\mu)\left[(1-\rho^*)\sigma_s^2 - \rho^*\left(\mu + (1-\mu)\rho^*\right)\sigma_{sc}\right]
\end{aligned}$$

Since $(1-\mu)(\mu + (1-\mu)\rho_U^*)\rho_U^* > 0$, $W(\mathcal{I}_D) - W(\mathcal{I}_U)) \geq 0$ applies if and only

$$\begin{aligned}
\sigma_{sc} &\leq \frac{-(\mu + p)\left[1 - (\mu + (1-\mu)\rho_U^*)^2\right](\bar{c}^2 + \sigma_c^2) + (1-\mu)(1-\rho^*)\sigma_s^2}{(1-\mu)(\mu + (1-\mu)\rho_U^*)\rho_U^*} \\
&= \frac{-(\mu + p)\left[1 - \mu^2 - 2\mu(1-\mu)\rho_U^* - (1-\mu)^2\rho^2\right](\bar{c}^2 + \sigma_c^2) + (1-\mu)(1-\rho^*)\sigma_s^2}{(1-\mu)(\mu + (1-\mu)\rho_U^*)\rho_U^*} \\
&= \frac{-(\mu + p)\left[1 + \mu - 2\mu\rho_U^* - (1-\mu)\rho^2\right](\bar{c}^2 + \sigma_c^2) + (1-\rho^*)\sigma_s^2}{(\mu + (1-\mu)\rho_U^*)\rho_U^*} \\
&= \frac{-(\mu + p)\left[1 - \rho^{*2} + \mu(1-\rho_U^*)^2\right](\bar{c}^2 + \sigma_c^2) + (1-\rho^*)\sigma_s^2}{(\mu + (1-\mu)\rho_U^*)\rho_U^*} \\
&= \frac{1-\rho^*}{(\mu + (1-\mu)\rho_U^*)\rho_U^*} \cdot \left(\sigma_s^2 - (\mu + p)\left(1 + \mu + \rho^*(1-\mu)\right)(\bar{c}^2 + \sigma_c^2)\right)
\end{aligned}$$

For result b), I look on the change in the sender's expected utility from disclosure:

$$\begin{aligned}\mathrm{E}\left[U^S(\mathcal{I}_D)\right] - U^S(\mathcal{I}_U)\right] = \mu\mathrm{E}\left[c\left(m^*(\mathcal{I}_D) - m^*(\mathcal{I}_U)\right)\right] + (1-\mu)\mathrm{E}[c\left(d_r^*(m^*(\mathcal{I}_D)) - d_r^*(m^*(\mathcal{I}_U))\right)] \\ + \mathrm{E}[U_n^R(m(\mathcal{I}_D))] - \mathrm{E}[U_n^R(m(\mathcal{I}_U))]\end{aligned}$$

It will be useful to denote the sender under different information structures following (12) by $\beta_i^*(c) = c(\mu + \rho_i^*(1-\mu)$ with $i \in \{D, U\}$. Using (10) and $\rho_D^* = 1$ one obtains for the first element of the first term

$$\begin{aligned}\mu\mathrm{E}[c(m^*(\mathcal{I}_D) - m^*(\mathcal{I}_U))] &= \mu\mathrm{E}[c\left(\beta_D^*(c) - \beta_U^*(c)\right)] \\ &= \mathrm{E}[c^2]\mu(1-\mu)(1-\rho_U^*) \\ &= \left(\bar{c}^2 + \sigma_c^2\right)\mu(1-\mu)(1-\rho_U^*)\end{aligned}$$

For the second element of the first term one can use $d_r^*(m^*(\mathcal{I}_D)) = s$ and get

$$\begin{aligned}(1-\mu)\mathrm{E}[c(s - d_r^*(m^*(\mathcal{I}_U)))] &= (1-\mu)\mathrm{E}[c(s - ((1-\rho_U^*)\bar{s} + \rho_U^*(s + \beta_U^*(c) - \mathrm{E}[\beta_U^*(c)]))] \\ &= (1-\mu)(1-\rho_U^*)\mathrm{E}[c(s-\bar{s})] - (1-\mu)\rho_U^*\mathrm{E}\left[c\left(\beta_U^*(c) - \mathrm{E}[\beta_U^*(c)]\right]\right) \\ &= (1-\mu)(1-\rho_U^*)\mathrm{E}[c(s-\bar{s})] - (1-\mu)\rho_U^*(\mu + (1-\mu)\rho_U^*)\mathrm{E}[c^2 - c\bar{c}] \\ &= (1-\mu)(1-\rho_U^*)\sigma_{sc} - (1-\mu)\rho_U^*(\mu + (1-\mu)\rho_U^*)\sigma_c^2\end{aligned}$$

Using that

$$\begin{aligned}\mathrm{E}[U_n^R(\mathcal{I}_D)] - \mathrm{E}[U_n^R(\mathcal{I}_U)] &= -\frac{1}{2}\left[1 - (\mu + (1-\mu)\rho_U^*)^2\right](\bar{c}^2 + \sigma_c^2) \\ &= -\frac{1}{2}(1-\mu)\left[1 - {\rho_U^*}^2 + \mu(1-\rho_U^*)^2\right](\bar{c}^2 + \sigma_c^2) \\ &= -\frac{1}{2}(1-\mu)(1-\rho_U^*)\left[1 + \rho_U^* + \mu(1-\rho_U^*)^2\right](\bar{c}^2 + \sigma_c^2)\end{aligned}$$

and collecting the above terms yields that $\mathrm{E}\left[U^S(\mathcal{I}_D) - U^S(\mathcal{I}_U)\right] \geq 0$ if and only if

$$\begin{aligned}\sigma_{sc}(1-\rho_U^*) &\geq \left[-\mu + \frac{1}{2}\left(1 + \rho_U^* + \mu + (1-\rho_U^*)\right)\right](1-\rho_U^*)(\bar{c}^2 + \sigma_c^2) + \rho_U^*\left(\mu + (1-\mu)\rho_U^*\right)\sigma_c^2 \\ &= \frac{1 + \rho_U^* - \mu - \mu\rho_U^*}{2}(1-\rho_U^*)(\bar{c}^2 + \sigma_c^2) + \rho_U^*\left(\mu + (1-\mu)\rho_U^*\right)\sigma_c^2 \\ &= \frac{1}{2}(1-\mu)(1-\rho_U^*)^2(\bar{c}^2 + \sigma_c^2) + \rho_U^*\left(\mu + (1-\mu)\rho_U^*\right)\sigma_c^2\end{aligned}$$

References

Abeler, J., A. Becker, and A. Falk (2014). Representative evidence on lying costs. *Journal of Public Economics 113*, 96–104.

Blanes, J. (2003). Credibility and Cheap Talk of Securities Analysts : Theory and Evidence. *mimeo*.

Brocas, I., J. D. Carrillo, S. Wang, and C. F. Camerer (2014). Imperfect choice or imperfect attention? Understanding strategic thinking in private information games. *Review of Economic Studies 81*, 944–970.

Brown, A. L., C. F. Camerer, and D. Lovallo (2012). To Review or Not to Review? Limited Strategic Thinking at the Box Office. *American Economic Journal: Microeconomics 4*(2), 1–26.

Cain, D. M., G. Loewenstein, and D. a. Moore (2005). The Dirt on Coming Clean: Perverse Effects of Disclosing Conflicts of Interest. *The Journal of Legal Studies 34*(1), 1–25.

Crawford, V. and J. Sobel (1982). Strategic information transmission. *Econometrica 50*(6), 1431–1451.

Deimen, I. and D. Szalay (2014). Smooth, Strategic Communication. *mimeo*.

Dubois, M., L. Fresard, and P. Dumontier (2013). Regulating Conflicts of Interest: The Effect of Sanctions and Enforcement. *Review of Finance 18*(2), 489–526.

Gneezy, U. (2005). Deception. The Role of Consequences. *American Economic Review 95*(1), 384–394.

Inderst, R. and M. Ottaviani (2012). Competition through Commissions and Kickbacks. *The American Economic Review 102*(2), 780–809.

Kartik, N. (2009). Strategic Communication with Lying Costs. *Review of Economic Studies 76*(4), 1359–1395.

Kartik, N., M. Ottaviani, and F. Squintani (2007). Credulity, lies, and costly talk. *Journal of Economic Theory 134*(1), 93–116.

Li, M. and K. Madarasz (2008). When mandatory disclosure hurts: Expert advice and conflicting interests. *Journal of Economic Theory 139*, 47–74.

Loewenstein, G., D. M. Cain, and S. Sah (2011). The Limits of Transparancy: Pitfalls and Potential of Disclosing Conflict of Interest. *American Economic Review: Papers Proceedings 101*(3), 423–428.

López-Pérez, R. and E. Spiegelman (2012). Why do people tell the truth? Experimental evidence for pure lie aversion. *Experimental Economics 16*(3), 233–247.

Malmendier, U. and D. Shanthikumar (2007). Are small investors naive about incentives? *Journal of Financial Economics 85*(2), 457–489.

Malmendier, U. and D. Shanthikumar (2014). Do security analysts speak in two tongues? *Review of Financial Studies 27*(5), 1287–1322.

Michaely, R. and K. Womack (1999). Conflict of interest and the credibility of underwriter analyst recommendations. *Review of Financial Studies 12*(4), 653–686.

Morgan, J. and P. C. Stocken (2003). An analysis of stock recommendations. *RAND Journal of Economics 34*(1), 183–203.

Morris, S. (2001). Political correctness. *Journal of Political Economy 109*(2), 231–265.

Ottaviani, M. (2000). The Economics of Advice. *mimeo*.

Ottaviani, M. and F. Squintani (2006). Naive audience and communication bias. *International Journal of Game Theory 35*(1), 129–150.

Sobel, J. (1985). A theory of credibility. *The Review of Economic Studies 52*(4), 557–573.

United States Congress (2002). Sarbanes-Oxley Act of 2002.