De-Biasing Strategic Communication

Tobias Gesche
ETH Zurich*

February 2019

Abstract

This paper studies the effect of disclosing conflicts of interest on strategic communication when the sender has lying costs. I present a simple economic mechanism under which such disclosure often leads to more informative, but at the same time also to more biased messages. This benefits rational receivers but exerts a negative externality from them on naive or delegating receivers; disclosure is thus *not* a Pareto-improvement among receivers. I identify general conditions of the information structure under which this effect manifests and show that whenever it does, full disclosure is socially inefficient. These results hold independently of the degree of receivers' risk-aversion and for an arbitrary precision of the disclosure statement.

Keywords: strategic communication, misreporting, conflict of interest, disclosure

JEL Classification: D82, D83, L51

*Center for Law & Economics, ETH Zurich. Email: tgesche@ethz.ch

I thank Nageeb Ali, Roman Inderst, Navin Kartik, Igor Letina, Ming Li, María Saéz-Martí, David Myatt, Nick Netzer, Deszö Szalay, Adrien Vigier, and Shihong Xiao for helpful discussions and comments on previous versions. I also acknowledge support by the Swiss National Science Foundation (SNSF Grant #P2ZHP1_171900).

1 Introduction

A substantial part of the world's economic activity deals with the elicitation of information by experts and its communication to non-experts. Examples include stock analysts, researchers, consultants or managers reporting to shareholders. Too often, experts face a conflict of interest (henceforth COI) such as sale commissions or affiliations which provide an incentive to bias their reports. Inefficiencies then arise because of two main reasons: First, receivers of such information may ignore the expert's COI and make poor choices by following biased information. Second, receivers lack information about the COI, e.g. its relative magnitude and the direction of the bias it induces. Without such information, they cannot accurately correct the expert's advice. They may then rationally decide to ignore the expert's message, at least partially, such that valuable information is lost. Disclosure of COIs promises to be a simple remedy to this problem. The idea is that information about the expert's COI helps at least those receivers who can use it to correct for potential bias. It is also tempting to policy makers as it carries the, as I will show incorrect, intuition that flattening information asymmetries is always desirable and should at least not hurt anyone. From a regulatory view, disclosure is also an appealing option as it is less paternalistic and less costly to regulators than direct supervision and regulation.[1]

The objective of this paper is to describe an economic mechanism which shows how such disclosure often can lead to consequences opposite to those intended. It does so by considering a communication game in which the sender's private type is two-dimensional. This type consists of the superior information an expert owns and the COI which provides an incentive to communicate this information not truthfully. The sender faces lying costs of doing so, e.g. reputational or expected legal costs which are increasing in the size of the lie – talk is thus not cheap. The model also allows some receivers to be naive towards the sender's COI while others are fully strategic and rational, in a Bayesian sense. Alternatively, naivety in this setup is equivalent to the delegation of decisions to an expert, e.g. to a managed fund. The combination of these factors then unveils a simple economic mechanism through which disclosure can lead to more biased communication which then hurts naive receivers who do not anticipate the strategic effects of disclosure.

To understand the source of this adverse effect, consider an analyst ("he") who knows a share's fundamental value but also benefits from demand for this asset, e.g. via sales commissions. When commenting on the asset, he then faces a COI to overstate its value. The magnitude of this bias is determined by equalizing the marginal costs of lying to the marginal return of doing so. The latter

[1] A prominent example of such a policy is contained in the Sarbanes-Oxlay-Act which was enacted in 2002 as a response to prior corporate frauds, in particular among financial analysts. Among its adopted regulations is the requirement to "[...]*disclose in each research report, as applicable, conflicts of interest that are known or should have been known by the securities analysts*[...]" United States Congress (2002, Sec. 501b). Other cases of disclosure rules can be found in Fung et al. (2007), along with reasons why they seem appealing and examples for their failure.

is given by the average marginal reaction of receivers (i.e., their demand) to the sender's message, weighted with the commission's size. Now regard a client ("she") who receives a message from the sender and is aware of the potential bias. She can try to de-bias it by correcting for the bias' expected value. However, since the COI is the sender's private information, she faces uncertainty regarding the commission's actual size or even its direction. Her de-biasing of the sender's message can thus worsen things when the expected bias differs from the its actual value. Facing such strategic uncertainty, rational receivers will then form an estimate of the actual state of the world. This estimate is a combination between the sender's imperfectly de-biased message and her prior about the state of the world. For this, the relative weight which a rational receiver puts on the de-biased message is inversely related to the strategic uncertainty she faces. Disclosing the sender's COI decreases this strategic uncertainty and increases this weight; disclosure thus translates into a larger marginal reaction to the sender's message. However, as explained above, the marginal reaction of receivers scales the sender's bias which, therefore, increases with disclosure. Delegating or naive receivers who do not account for the strategic effects of biasing and de-biasing communication are then hurt by this increase.

The above reasoning combines two main insights: First, the reaction to the sender's message by rational, risk-averse receivers depends on the quality of information they can extract from it. Second, an expert who faces a COI and has lying costs biases his message in proportion to the reaction it induces. Both of these effects are simple in their economic intuition. Combined, however, they deliver the surprising result that increasing transparency can be a bad idea when the disclosed information is not used by everyone and that lying costs play a crucial role in creating this adverse effect. In particular, it disproves the idea that disclosing COIs is a Pareto-improvement among receivers, except when all of them are fully rational. Accordingly, disclosure can even decrease overall efficiency, depending on the relative share of receiver types.

I model these effects in a framework which allows for arbitrary degrees of risk-aversion as well as arbitrary quality of the disclosure process. General conditions under which this effect manifests and which allow evaluating the welfare consequences of disclosure are identified. The key variable in this regard is the correlation between the expert's COI and the information on which he has superior knowledge. For example, whenever this correlation is weakly positive, disclosure backfires and full disclosure is never optimal for efficiency. I also show that there can be situations in which disclosure is a Pareto-improvement among all receivers and that only in these cases, full disclosure is efficient.

2 Related literature

On the empirical side, the findings Malmendier and Shanthikumar (2014) who examine the communication by financial analysts relate closely to this paper. They show that analysts *strategically* inflate their stock recommendations by tailoring it to the receivers' reactions. This feature is maintained in the following analysis. In their study, they use data which covers a period before and after the Sarbanes-Oxlay-Act which requires financial analysts to disclose COIs. Their analysis shows that the strategic bias did not disappear after the act was put into action in 2001.[2] Similarly, Mullainathan et al. (2012) conducted an audit study and show that after the act came into effect, financial advice remained of poor quality. The present paper show how such effects can arise.

Clean, causal evidence for negative effects of disclosure comes from Cain et al. (2005). In their experiment, subjects in the role of experts could examine a jar filled with coins. These subjects then advised others who had to estimate the amount of money inside the jar but who could not examine it beforehand. Their results first confirm a straightforward intuition: When the pay of the experts is based on the accuracy of final estimates by clients, their advice and the clients' resulting estimates are better then when experts are paid based on how high estimates are. However, they also show that when receivers are made aware of the experts' incentive to induce a high estimate, thus when COIs are disclosed, the experts' bias *increases*, relative to when they are unaware. On average, receivers do not account for this effect and end up making worse decisions than without disclosure. This finding on the adverse effects of disclosure has also been replicated in similar setups (see Koch and Schmidt, 2010; Inderst et al., 2010; Cain et al., 2011).[3] The effects identified here are also in line with these experimental results.

This paper also contributes to the theoretical literature on strategic communication. In their seminal work on the topic, Crawford and Sobel (1982) characterize communication equilibria to be partitional when talk is cheap, that is, when lying costs are absent: An informed sender with a commonly known COI endogenously partitions the state space and just announces the partition which contains the actual state of the world. In consequence, information is lost in communication. This result applies independently of the specific meaning of language, i.e. how exactly states map to messages by the sender and back from messages into actions by receivers, as long as this mapping is common knowledge.[4] Often, however, this meaning of language is determined by the circumstances.

[2]See Malmendier and Shanthikumar (2014), p.1298: They state that their measure of strategic bias remains sizable and positive for affiliated analysts when they split the sample by August 2001, the date when the scandals became public and which contributed to the enactment of the Sarbanes-Oxlay-Act shortly afterward.

[3]For a further review of the failure of disclosure and psychological approaches to it, see Loewenstein et al. (2014). The explanations presented therein are based on the psychological characteristics evolving within relatively close expert-receiver relationships. The results on the negative effects of disclosure presented here do not require such close relationships – they also apply if expert and receiver only very remotely (e.g., in a market setting).

[4]See Sobel (2013) for an overview of the rich literature which has utilized and extended the partitioning result. Also, see the section on pragmatics therein for a further discussion on language and its meaning in the context of

For example, if a financial analyst announces "I expect share X will pay Y this year" many people would understand its meaning to be literal, thus that Y is the share's actual performance or at least the analyst's best estimate. Studies which conclude that financial analysts' statements are often upward biased also adapt this understanding (see Hayes, 1998; Michaely and Womack, 1999; Malmendier and Shanthikumar, 2014). In contrast, the partitioning result in combination with such a literal meaning and understanding of messages implies that, on average, the message and the inferred state of the world should not differ.

In order to reconcile a literal understanding and persistently inflated messages, one or both of the two crucial assumptions which underlie the partitioning result need to be changed. Addressing them, Kartik et al. (2007) and Kartik (2009) show that these assumptions are first, the boundedness of the state space and second, cheap talk (i.e., no lying costs).[5] Capturing these insights, this paper allows for an unbounded support, for example when the variable on which the sender has private information is normally distributed. This is similar to Gordon and Nöldeke (2013) who study figures of speech in a related setup. Their work and Kartik (2009) also allow for lying costs which correspond to those used here. These costs increase in the distance of the message to the actual state of the world. They therefore reflect a literal interpretation of language.

In contrast to the above works, I assume that there is strategic uncertainty. Thus, the sender's is type two-dimensional because his COI is not common knowledge. This relates to Morgan and Stocken (2003) who find in a cheap talk framework, absent lying costs and with a compact state space, that the messaging strategy remains partitional when the sender's COI is described by a binary random variable. In a similiar setup, Li and Madarasz (2008) show that when this variable also has an expected value of zero, disclosing its non-zero realization can decrease informational efficiency, as measured by the number of equilibrium partitions. Inderst and Ottaviani (2012) also look on disclosure of COIs. They model them as commissions paid by producers to intermediaries who advise customers on which out of two products to choose. In their model, the state of the world and the expert's message are therefore both binary. They then show that disclosing COIs reduces the provision of commissions but less so, in relative terms, for the inferior product. Consequently, the relative bias rises upon disclosure and consumers make worse decisions. None of these approaches features lying costs. I show that when communication is not partitional or binary, disclosure can be harmful too, but due to another reason for which lying costs are crucial.

The framework in which I establish these findings describes strategic communication when both,

strategic communication.

[5]For one-dimensional sender types Kartik et al. (2007) show that under general conditions, unbounded support is sufficient for the sender's messaging strategy to be continuous and revealing; this also applies when there is a lower bound and lying costs. Kartik (2009) considers a compact state space with lying costs. Equilibria are then of the "LSHP (low types separate and high types pool)"-form: An upward-biased sender exaggerates his statement by a fixed bias if the state is below a certain threshold. If it is above, the sender's messaging strategy becomes partitional.

the state of the world and the sender's COI are represented by continuous, possibly correlated, variables. The model's specific form extends a framework introduced by Fischer and Verrecchia (2000). They study a manager who gets linear utility from influencing his company's share price through his earnings announcement while facing quadratic costs of lying (this setting is, among others, also admissible here). I extend this setup to more general, elliptical distributions as recently used by Deimen and Szalay (2014) who study strategic communication when players cannot agree on the relative importance of different information they hold. This approach has recently been adapted by Kartik and Frankel (2018) as a part of their general treatment of projection-based signaling, i.e. one-dimensional signals sent by two-dimensional sender types.[6] Comparative statics in Fischer and Verrecchia (2000) show that a decrease in strategic uncertainty with regard to the sender's motives leads to an increase in his bias when strategic incentives are uncorrelated to the state of the world. Kartik and Frankel (2018) have a similar result when this correlation is non-negative. This work focuses on this effect and to study it rigorously, I extend the analysis of communication games in this class along three main dimensions:

First and foremost, the following analysis incorporates the presence of naive receivers and their strategic effect on the sender's messaging strategy. This allows to examine the welfare consequences of disclosure with regard to both, the message's informativeness which matters for rational receivers and its bias – the deviation from a honest, literally meant message on which naive receivers rely. Second, this work explicitly studies the role of correlation between the sender's COI and the state of the world, including a negative correlation. Apart from allowing to capture several realistic settings which imply a negative correlation, I show that this can crucially affect comparative statics: Only with a negative correlation, disclosure can decrease the sender's bias and that only then, disclosure is a Pareto-improvement among all receivers, including naive ones. Third and closely related to the preceding point, I explicitly model disclosure through a signal of arbitrary precision. This allows to analyze the effects of disclosure on the whole posterior distribution of beliefs and actions. Just performing comparative statics with respect to a single parameter, such as the variance which describes uncertainty regarding the sender's COI, overlooks the fact that information on one variable also contains information on correlated variables. As the following sections will show, these features allow to comprehensively analyze disclosure and are often crucial in fully evaluating its consequences.

[6]Technically also related, Bénabou and Tirole (2006) study projection of pro-social motivation in a linear-quadratic-normal framework. Kartik and Frankel (2018) use linear benefits for the sender and lying costs which are quadratic in an upward bias but zero for downward-biased messages. In their setup, the sender always wants to induce high beliefs such that this does not restrict their results (see also their footnote 19). I do not allow such free downward deviation as in many settings considered here, both, an upward and a downward-bias can be considered as not telling the truth and has negative consequences for receivers (see the examples discussed above and in subsection 3.2).

3 The model

3.1 General setup

Consider a mass of receivers. Each would like to know the state of the world, denoted by $s \in S \subseteq \mathbb{R}$, because she has to take a decision $d \in S$. The payoff resulting from that decision depends on how well it matches the realization of s. For example, s might represent an asset's return and d the receiver's optimal position into this asset. A receiver then suffers a loss which is the greater, the more d and s are misaligned. This is captured by her ex-post utility

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

where L is a $\mathcal{C}^2$ loss function with argument $d-s$ which is strictly concave and symmetric around its maximum. Without loss of generality, I assume $L(0) = 0$ for its bliss point's value. As the leading example, consider $L(d,s) = -\frac{1}{2}(d-s)^2$, the quadratic loss function.[7]

Receivers do not know s and refer to a risk-neutral sender who knows its value. The sender communicates via a public message $m \in M = S$ which is understood to mean the value of s. I assume that there are two types of receivers, rational and naive ones. Both react differently to the sender's message. This is characterized by two different demand functions $d_r(m)$ and $d_n(m)$ for rational and naive receivers, respectively (they will be described in more detail further below). Denote the share of naive receivers by $\mu \in [0,1)$ so that the mass of rational receivers is given by $1-\mu$. This yields the following expected demand function for the receiver:

$$D(m) = \mu d_n(m) + (1-\mu)d_r(m) \tag{2}$$

Since the sender's message m is supposed to reflect s, there is meaning in his message and stating it falsely creates costs which are measured by $\frac{1}{2}(m-s)^2$. This functional form can capture several possible sources of lying costs such as costs based on social preferences, moral concerns against lying or reputational concerns in a stage version of a repeated game.[8] If minimizing lying costs were the sender's only objective, he would then be honest and always send $m = s$. Receivers would then just follow the message and implement their optimal choice.

[7]This is the canonical example which is put forward by Crawford and Sobel (1982) and used in much of the literature on strategic communication. Ottaviani (2000) shows that this specific function covers the case of a receiver with exponential utility who invests d into a risky asset of which she knows its variance but not its expected value s.

[8] When L is also quadratic, this cost function captures concerns for the utility of a receiver who follow the sender's message at face value. Kartik (2009) uses this specific form of costs as a prominent example, i.e., for intrinsic costs of lying (see Erat and Gneezy, 2012; López-Pérez and Spiegelman, 2012; Abeler et al., 2014). See also Abeler et al. (2018) for a recent meta-study on the determinants of lying costs. Reputational concerns in the spirit of Sobel (1985) or Morris (2001) can also be proxied by it: If the actual value of s became knowledge ex-post, the squared distance of s and m is part of the nominator of the sender's coefficient of determination (R^2) which one obtains by regressing s on the sender's prior messages m; his credulity is thus decreasing in this squared distance.

However, such strong influence of the sender on the receivers' decisions can be exploited. The sender can be incentivized to induce either a high or low demand, for example via sales commissions. Such a COI of the sender manifest through an additional payoff $cD(m)$ with $c \in C \subseteq \mathbb{R}$. A value $c > 0$ then implies that the sender has an incentive to generate a high demand, for example in the case of sales commissions. Conversely, $c < 0$ means that low demand is rewarded, for example when the sender wants to temporarily decrease the price of an asset in which he would like to take a position. The magnitude of c can then interpreted as the strength of such incentives, relative to given lying costs. Alternatively, the magnitude of $1/c$ can represent how much lying costs matter for the sender, relative to a given COI. The sender's expected utility is therefore given by

$$\mathrm{E}[u^S(m) \mid s, c] = cD(m) - \frac{1}{2}(m - s)^2. \tag{3}$$

Note that the commission is proportional to the demand. This distinguishes the current work from other approaches which assume that the sender has the same utility function L as the receivers, only with a shifted bliss point (e.g., Crawford and Sobel, 1982; Ottaviani and Squintani, 2006; Kartik et al., 2007; Gordon and Nöldeke, 2013). In particular, such approaches implicitly assume that the sender suffers the more a receiver's action differs in *either* direction from the sender's bliss point. Inducing a too high reaction by receivers is then as bad as inducing a too low reaction.[9] While such an approach can be sometimes valid, getting demand as high or low as possible is often preferred by the sender who faces a COI, such as when he gets a sales commission. Apart from accommodating such reasoning, this specification also yields an intuitive form of the sender's optimal message:

Lemma 1. *Suppose $D(m)$ is a $\mathcal{C}^2$ function. A sender's optimally chosen message m^* then solves the implicit function*

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

and is strictly increasing in s for all $(s, c, m) \in \mathbb{R}^3$. (proof in appendix)

From (4) one gets that the sender's message equals the state of the world plus a bias. The size of the bias and its direction are determined by the COI's value c and the marginal reaction of receivers to the message by the sender. Thus, the bias in the sender's message is restricted by the costs of lying and his loss of credulity. For example, if no-one listens to the sender, i.e. $D'(m) = 0$ holds, there is no point of lying and the bias equals zero. More generally, this reflects that, in the

[9] In such frameworks, where receivers and sender have the same function similiar to L and only differ in their bliss points, naive receivers who follow the sender one-to-one can have the same adverse consequences for the sender as rational receivers who do not react enough to his message. In these frameworks (but not the present one) naive receivers therefore have a similar negative effect as lying costs for the sender (see Ottaviani and Squintani, 2006; Kartik et al., 2007).

presence of lying costs, a lie should be scaled to the reaction it aims to affect. This feature will be crucial in understanding the adverse conflicts of disclosure.[10]

Note also that the sender's message is strictly increasing in s. If c were commonly known, a situation I will later treat as a special case, the sender's strategy would be an univariate, monotone function from M to S. As such, it would be invertible and the sender's message would therefore be revealing. Accordingly, there would be a bias in the sender's message but Bayesian, rational receivers would not be hurt by it.[11] In the following, I will relax two assumptions which ensure that messages can be biased but fully informative to everyone: First, I will allow that c is the sender's private information. The message is then a projection from $S \times C$ to M and rational receivers face strategic uncertainty when they try to infer s from it. Second, I will also allow that some receivers are naive in the sense that they do not account for the sender's bias and follow his message at face value. Apart from the immediate effect that each single of these two features nullifies the result that a bias does not hurt receivers, their combination leads to the main result that decreasing strategic uncertainty can backfire.

3.2 Information structure

Both, the state of the world s and the COI c are the sender's private information. They are assumed to be drawn from an multivariate normal distribution with support $K \times S = \mathbb{R}^2$. I denote this distribution via $F(\boldsymbol{\eta}, \boldsymbol{\Sigma})$ where $\boldsymbol{\eta}$ and $\boldsymbol{\Sigma}$ represent the vector of expected values and the variance-covariance matrix with finite, real-valued elements:

$$\boldsymbol{\eta} = \begin{bmatrix} \bar{s} \\ \bar{c} \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_s^2 & \sigma_{sc} \\ \sigma_{sc} & \sigma_c^2 \end{bmatrix}.$$

When appropriate, I will refer to the correlation of s and c instead of its covariance σ_{sc}. To make things interesting, I also assume $\sigma_s \sigma_c > 0$ and $|Corr[s, c]| = \frac{|\sigma_{sc}|}{\sigma_s \sigma_c} < 1$ as otherwise, the receiver's inference problem would become effectively one-dimensional or vanish entirely. I will refer to σ_s^2 to as "fundamental uncertainty", reflecting that this parameter captures uncertainty regarding s, the variable in which receivers are primarily interested. Similarly, σ_c^2 will be called "strategic uncertainty" as it reflects the uncertainty regarding the sender's motives.

Note that while the results in this main text are stated for the normal distribution, they also apply for more general elliptical distributions. As for normals, which belong to the class of elliptical

[10] This crucial feature is independent of the functional form of the lying costs, as long as the lying costs are differentiable in m and strictly increasing in the distance between m and s.

[11] See Kartik et al. (2007) who establish this insight in a related but different framework. If their language is adapted to this work, that is s corresponds to their x and $D(m)$ to their $\hat{x}$ (the receiver action which the sender wants to influence), then their assumption A.4 corresponds to $\partial^2 \mathrm{E}[u^S(m) \mid s, c]/(\partial s \partial D(m)) < 0$. This is easily verified to be violated here. Nevertheless, messages are biased but yet revealing (w.r.t. s).

distributions, all presently crucial features can be expressed via the first two moments $\boldsymbol{\eta}$ and $\boldsymbol{\Sigma}$ of an elliptical distribution. Examples for this include the heavier-tailed Laplace or Student-t-distribution (which are often used in financial and risk modeling) or the logistic distribution (which is often used to model latent processes underlying discrete outcomes). Some distributions with compact support such as the uniform distribution are also elliptical.[12] For further details and a formal definition of elliptical distributions see the appendix.

This information structure is suited to naturally model how players, in particular, rational receivers, arrive at their prior. First, assuming unbounded support for (s, c) means that no commonly known bounds on the state space are required. In contrast, assuming compact support implies common knowledge of such sharp bounds. Sometimes, this is straightforward and reasonable, e.g. when the sender communicates the share of one's wealth which should be invested in a certain asset. However, once leverages become available or when s reflects asset returns, appropriate bounds are not clear-cut. A solution to this is then to assume that all real values are theoretically possible while unrealistic, extreme realizations receive appropriately low probabilities by choosing the above moment parameters accordingly. Similarly, if the prior is based on historical data (e.g., if (4) forms the basis of a structural model), the resulting estimates of (s, c) are, by the central limit theorem, approximately normally distributed.[13]

Importantly, this framework also allows to handle the case of σ_{sc} being positive or negative. As the reference case, consider a third party who benefits from a higher price because it has to sell a good on which the sender reports, for example an asset. Suppose that supply is temporarily fixed so that the price p is determined by demand: $p(m) = \beta D(m)$ with $\beta > 0$. Normalize the reservation price of the third party to zero and allow it to transfer share τ of the transaction gain as a commission c to the sender. Denote with $q > 0$ the amount which the third party wants to sell. The normalized transaction gain is then given by $qp(m)$ so that, from the sender's point of view, $c = \tau q \beta$ holds. In consequence, $\sigma_{sc} = 0$ applies from the perspective of a receiver (who, for example, may not know τ). Now consider the same situation but assume that the third party has also information about the

[12] Note that by choosing the distribution's parameters appropriately, the probability that realizations of (s, c) are within some compact set can be made arbitrarily high. For an example of how the uniform distribution can be generated from the general definition of elliptical distributions see the survey by Gómez et al. (2003), pp. 359/360. For a related application of elliptical distributions with compact support, see Kartik and Frankel (2018). The additional assumptions which are, in some cases, necessary to maintain all results presented here are essentially to specify suitable out-off-equilibrium beliefs, an issue which is bypassed under unbounded support. Embrechts et al. (2002) describes how elliptical distributions can be used in financial and risk analysis, including the caveats of doings so.

[13] Another scenario worth to be mentioned and which can be easily captured in this framework is scientific fraud: First, empirical research and its assessment often boils down to statistics with elliptical distributions, e.g. normally or student-t-distributed regression coefficients. Second, pressure to publish and publication bias create a COI to inflate these statistics (Simmons et al., 2011). Third, outright cheating such as making up data or more subtle techniques such as selective sampling and data-mining are methods with which these test statistics can be misrepresent or manipulated at the costs of ethical, reputational and legal concerns (Fanelli, 2009; Steen, 2011; Simonsohn, 2014). Finally, disclosure policies are employed in this context, e.g. by many journals. This paper's framework can thus be applied to study their consequences.

asset's fundamentals s, for example, because it is affiliated with the sender. It then wants to sell more if s is low. Thus, q is the image of a decreasing function $q(s)$. In consequence, $c = \tau q(s)\beta$ holds for the sender and, therefore, $\sigma_{sc} < 0$ from the receiver's perspective. Now consider the same situation, but in order to sell, put options are used. Using such an option is only profitable if its strike price $\tilde{p}$ is "in the money", i.e., if $\tilde{p} - p(m) > 0$ holds. This difference then determines the transaction gain of using the options so that a lower market price is beneficial for the sender. In consequence, $c = -\tau q(s)\beta$ then applies for the sender and with it, also $\sigma_{sc} > 0$ from the receiver's perspective.[14] Thus, the present framework allows to analyze different modes of how COIs arise, in particular those which are relevant in the context of strategic information transmission in financial markets.

3.3 Rational and naive receivers

As shown in Lemma 1, COIs can induce the sender to not report truthfully. How should receivers then take such a distortion into account? A receiver who is rational, in a Bayesian sense, should do so by acting on the information she can extract from the sender's message such that it maximizes her expected utility. That is, here action is given by $d_r(m) \equiv \arg\max_{d \in S} \mathrm{E}[L(s,d)|m]$. If L is the quadratic loss function and s and m are jointly normally distributed, this is clearly $\mathrm{E}[s|m]$. The following generalizes this to any $\mathcal{C}^2$, strictly concave and symmetric loss function L:

Lemma 2. *If m is distributed according to F, rational receivers choose $d_r(m) = \mathrm{E}[s|m]$.*

(The proof adapts the proof of part i) of Lemma 2 in Deimen and Szalay (2014) to this paper's framework and can be found in the appendix.)

The optimal decision $d_r(m)$ is that of fully rational, Bayesian receiver who is capable of adjusting for the effect of the sender's COI on his message and connecting it to the common prior. While some receivers, for example professional ones, can act in such a manner, empirical evidence shows that many people do not anticipate and correct for others' strategic behavior or strategically ignore important information (e.g., Brown et al., 2012; Brocas et al., 2014). In line with such reasoning, Malmendier and Shanthikumar (2007) show that small investors such as private households follow analysts' optimistic recommendations more closely than bigger, institutional ones who are more cautious and adjust for biases. To capture these observations, I allow for the possibility that there are naive receivers who take the sender's signal at face value. Their action is thus given by $d_n(m) = m$.[15]

[14] For the case that the third party wants to buy quantity $q > 0$ (that is selling quantity $q < 0$), analogous arguments can be derived.

[15] By appropriate scaling of μ, one can always account for situations where naive or delegating receivers do not react one-to-one, e.g. when $d_n(m)$ is a positive affine transformation with $d_n'(m) = r > 0$. As an example, suppose that there is a mass 0.5 of naive receivers for whom, on average, $d_n(m) = 0.6m$ holds. From the sender's point of view, this is the same as if there were a mass 0.2 of receivers who ignore him, mass 0.3 of naive receivers who follow one-to-one, and a mass 0.5 of rational receivers. Using $\mu = \frac{0.3}{0.5+0.3}$ would then be strategically equivalent.

In consequence, the total demand function (2) becomes

$$D(m) = \mu m + (1 - \mu)\mathrm{E}[s|m]. \tag{5}$$

The assumption of some receivers naively following the sender's message is also used by others in related settings (see Ottaviani and Squintani, 2006; Kartik et al., 2007; Chen, 2011; Gordon and Nöldeke, 2013). However, this work differs since the sender's objective is to shift demand into the direction of c, not to bring it as close as possible to his bliss point. If this were the case, naive receivers would impose lying cost in itself (see footnote 9). Also note that naive receivers who directly implement the sender's message are strategically equivalent to receivers who delegate their decision d to the sender, for example a managed fund. This can be either because of blind trust (Gennaioli et al., 2015) or because handling the assets for themselves and fully rationally de-biasing a sender's message creates fixed costs which are too high, relative the informational gain from acting rationally (Sims, 2003). Similarly, the current setting also captures a scenario in which a risk-neutral sender faces a single receiver but does not know whether this receiver is naive (with probability μ) or rational (with probability $1 - \mu$). Finally, note that one can also assume different loss functions for naive and rational receivers, for example to measure different levels of risk aversion. As long as these distinct loss functions obey the assumptions made on L, all of the results presented here remain valid.

3.4 Disclosure and timing

In the following, I will consider the communication game described above, appended by a disclosure stage in which receivers get a signal over the sender's COI. Formally, disclosure is captured by a signal $\tilde{c} = c + \epsilon$ where ϵ is an error term which is jointly but independently distributed with (s, c), has an expected value of zero, and variance σ_ϵ^2. Thus, the lower this variance is, the more informative is disclosure. I will refer to the scenario with $\sigma_\epsilon^2 = 0$, where the signal is perfectly informative, as "full disclosure". In contrast, the reference scenario with $\sigma_\epsilon^2 \to \infty$, a completely uninformative or absent signal, is called "no disclosure". Cases in between are referred to as "imperfect disclosure". The communication game with disclosure then has the following timing:

a) The vector (s, c, ϵ) is drawn from F,

b) $\tilde{c} = c + \epsilon$ becomes common knowledge, (s, c) is privately observed by the sender,

c) the sender sends a signal m about s,

d) receivers observe m, if rational update their belief about s, then choose d,

e) payoffs are realized.

4 Communication: Biasing and de-biasing

I look for a Perfect Bayesian Equilibrium of the above game. It consists of a pair of equilibrium strategies $m^* : S \times C \to M$ for the sender and $d_r^* : M \to S$ for a rational receiver such that each player's expected utility is maximized, given the other players' strategies when their beliefs are formed by Bayes' rule. Naive receivers are assumed to have a dominant strategy of following the sender so that their beliefs do not matter. The key equilibrium belief is then the belief of rational receivers about s as, by Lemma 2, they choose $d_r^*(m) = \mathrm{E}[s|m^*] \equiv \mathrm{E}[s|m]|_{m=m^*(s,c)}$.[16] Accordingly, the behavior of rational receivers is governed by their belief about the determinants of the sender's messaging strategy $m^*(s, c)$ and, therefore, by what they learn from disclosure via the signal $\tilde{c}$.

4.1 Updating after disclosure

To desribe the updating process of rational receivers' beliefs, it will be useful to express disclosure via the parameter ψ defined as follows:

$$\psi \equiv \frac{Cov[c, \tilde{c}]}{Var[\tilde{c}]} = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_\epsilon^2} \in [0, 1] \tag{6}$$

This signal-to-noise-ratio reflects how much variation in c can be explained by the signal $\tilde{c}$ and is therefore a measure of its quality. For example, $\psi = 0$ denotes the case of no disclosure ($\sigma_\epsilon^2 \to \infty$). Conversely, $\psi = 1$ captures full disclosure ($\sigma_\epsilon^2 = 0$) while intermediate values of ψ inversely reflect intermediate values of the signal variance. Accordingly, ψ is key for the updated prior:

Lemma 3. *The posterior distribution of $(s, c \mid \tilde{c}\,)$ is given by $F(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}})$ with*

$$\hat{\boldsymbol{\eta}} = \begin{bmatrix} \bar{s} + (\frac{\sigma_{sc}}{\sigma_c^2})(\tilde{c} - \bar{s})\psi \\ \bar{c}(1-\psi) + \tilde{c}\psi \end{bmatrix} \quad \textit{and} \quad \hat{\boldsymbol{\Sigma}} = \begin{bmatrix} \sigma_s^2(1 - Corr[s,c]^2\psi) & \sigma_{sc}(1-\psi) \\ \sigma_{sc}(1-\psi) & \sigma_c^2(1-\psi) \end{bmatrix}$$

(proof in appendix)

First note that in the case of no disclosure ($\psi = 0$), the posterior equals the prior distribution. If there is disclosure ($\psi > 0$), the signal $\tilde{c}$ directly affects the expected value of c which is used to de-bias the received message. In consequence, it also reduces strategic uncertainty σ_c^2 by a factor $1 - \psi$. In addition, the distribution's parameters with regard to s are also affected if $\tilde{c}$ also contains information about it, that is, if s and c are correlated. In this case, disclosure reduces uncertainty in all dimensions (i.e., every element of $\hat{\boldsymbol{\Sigma}}$ decreases with ψ). These effects would be overlooked if disclosure were modeled as an univariate comparative static as, for example, with respect to σ_c^2.

[16] A complete belief profile over the sender's type also requires to specify an analogously-defined belief $\mathrm{E}[c|m]|_{m=m^*(s,c)}$. As it is payoff-irrelevant for either player it is omitted here.

Also note that full disclosure ($\psi = 1$) is a special case of the above. In it, the posterior distribution and with it, the sender's type, becomes effectively a one-dimensional random variable.

4.2 Equilibrium behavior

The sender has to take the aboce-described updating procedure into account when choosing his messaging strategy $m^*(s, c)$. In order for it to be optimal, it has to solve (4). When plugged into the sender's objective function (3), this means that the sender's messaging strategy has to solve the first order condition

$$m = s + c\left(\mu + (1-\mu)d_r^{*\prime}(m)\right) \tag{7}$$

where $d_r^{*\prime}(m) = \frac{\partial \mathrm{E}[s|m]}{\partial m}|_{m=m^*(s,c)}$. Thus, the sender's messaging strategy is not only determined by how the chosen bias shifts the demand of naive receivers and lying costs. It is also based on how rational receiver react to the sender's message which is formed under a messaging strategy with such a bias, as captured by $d_r^{*\prime}(m)$. To derive how rational receivers – and in response also senders – behave optimally, I define the "equilibrium inference coefficient" ρ^*. This parameter captures how well, given a sender's equilibrium messaging strategy $m^*(s, c)$ and rational receivers' posterior $F(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}})$, variations in the resulting equilibrium messages capture variations in the underlying state of the world s:

$$\rho^* \equiv \frac{Cov[s, m^*]}{Var[m^*]} \equiv \frac{Cov[s, m|\tilde{c}]|_{m=m^*(s,c)}}{Var[m|\tilde{c}]|_{m=m^*(s,c)}} \tag{8}$$

Throughout this paper, I focus on the case that in equilibrium, ρ^* is a real, strictly positive constant. This precludes situations where the expert's message is completely uninformative ($\rho^* = 0$). It also does not situations where the message is "reverted" ($\rho^* < 0$), that is, where higher values of m are associated with lower values of s – features which are unlikely to be observed in an information market with experts, especially when this is an equilibrium feature. One can then show that, in equilibrium, the inference coefficient has to be equal to the marginal reaction of rational receivers' to the message, i.e., that $d_r^{*\prime}(m) = \rho^*$ holds. Proving this relationship constitutes the main building block for the following proposition which characterizes the players' equilibrium actions and the relevant equilibrium beliefs:

Proposition 1. *Every pure strategy Perfect Bayesian Equilibrium of the communication game with strategies $m^*(s, c)$ for the sender and $d_r^*(m) \in \mathcal{C}^2$ for rational receivers takes the form of*

$$m^*(s, c) = s + c\left(\mu + (1-\mu)\rho^*\right) \tag{9}$$

$$d_r^*(m) = (1-\rho^*)\mathrm{E}[s] + \rho^*\left(m - \mathrm{E}[c](\mu + (1-\mu)\rho^*)\right). \tag{10}$$

The rational receivers equilibrium belief w.r.t. to s is given by $\mathrm{E}[s|m^*] = d_r^*(m)$.

(proof in appendix)

The above result characterizes equilibria in which the reaction by rational receivers to the sender's message is smooth in the sense that it is twice continuously differentiable. Note from (5) that assuming a smooth strategy for rational receivers is a necessary condition for the overall demand $D(m)$ to be smooth. I focus on such smooth, pure strategy equilibria because it is sufficient but yet relatively general to demonstrate the main point of this paper, the adverse effects of disclosure. It also captures the idea that arbitrarily small changes in the sender's message should yield no large effect on demand.

In such an equilibrium, the sender's strategy takes an intuitive form. He announces the state of the world and adds a bias, given by $m^*(s,c) - s = c\,(\mu + (1-\mu)\rho^*)$. This bias equals $D'(m)$, the change in expected demand due to a change in the message, scaled by the sender's COI c. Part of this change $D'(m)$ is the change in the best response of rational receivers to the sender's message, given by $d_r^{*\prime}(m) = \rho^*$, weighed by their population share $1-\mu$. This reflects that the best response by rational receivers, as displayed in (10), has two parts: The first is the receivers' prior about the true state of the world $\mathrm{E}[s]$. It is invariant to m and its weight $1-\rho^*$ is inversely related to the message's informativeness. The other part of their best response is given $m - \mathrm{E}[c]\,(\mu + (1-\mu)\rho^*)$, the received message corrected by the expected bias. This is weighted with ρ^*. Therefore, this coefficient measures how much rational receivers react to m (i.e., ρ^* is the weight on $d_r^{*\prime}(m)$). This interplay between the sender's equilibrium message and rational receivers' associated inference through ρ^* is also displayed in Figure 1 and will be crucial in the following.

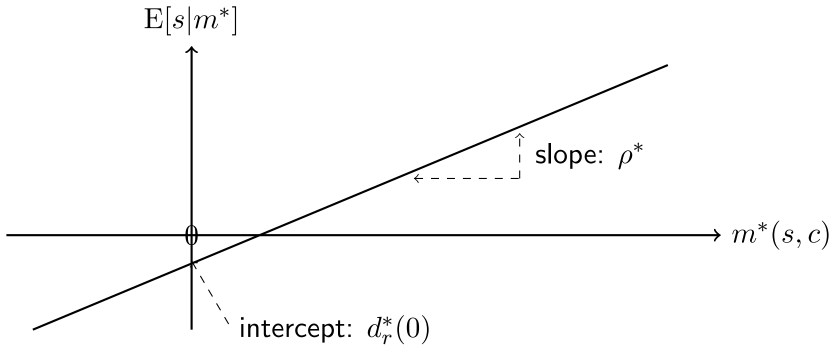

Figure 1: Graphical representation of rational receivers' inference $\mathrm{E}[s|m^*] = d_r^*(m)$ (see Proposition 1).

The underlying reasoning can also be understood by interpreting ρ^* as the coefficient from a linear regression of s on m: Both, a regression coefficient and ρ^*, describe the marginal change in the conditional expectation of a dependent variable due to a marginal change in the independent variable. The crucial difference is that in a regression, this refers to an exogenous change whereas here, it is the change in the sender's endogenously determined equilibrium message.

As a first step in the analysis of ρ^*, note that the correction by rational receivers for the expected bias is based on the *expected* commission. It can thus be wrong in both, direction and magnitude. This possibility of an error in de-biasing the sender's message provides the reason why rational receivers do often not react one-to-one to the corrected message. Whenever $\rho^* \in (0,1)$, they strategically ignore part of the sender's de-biased message and put weight on their prior such that information is left unused. For illustration, consider the case that σ_s^2 is close to zero while σ_c^2 is sufficiently large. In this situation almost all uncertainty is not of fundamental, but of strategic nature. It follows then that $Cov[s, m^*]$ and, through it, also ρ^* are almost zero. Rational receivers then almost completely ignore the signal and act only based on their prior $\mathrm{E}[s]$ because almost all variation in the message m^* can only be due to the sender's bias. In equilibrium, the sender then takes into account this non-reaction of rational receivers and scales down his bias.

Using the functional forms of equilibrium behavior as stated in Proposition 1, one can then determine more generally which values the equilibrium inference coefficient takes and how they are affected by disclosure:

Proposition 2. *Whenever COIs are not fully disclosed* ($\psi < 1$) *the equilibrium inference coefficient* $\rho^* = \frac{Cov[s,m^*]}{Var[m^*]} > 0$ *is a fixed point to*

$$g(\rho) = \frac{\phi\sigma_s^2 + (\mu + (1-\mu)\rho)\sigma_{sc}}{\phi\sigma_s^2 + 2(\mu + (1-\mu)\rho)\sigma_{sc} + (\mu + (1-\mu)\rho)^2\sigma_c^2} \tag{11}$$

with $\phi \equiv \frac{1-\psi(Corr[s,c])^2)}{1-\psi} > 1$*. With full disclosure* ($\psi = 1$), $\rho^* = 1$ *holds. Furthermore, if* $\rho^* \in (0,1]$*, such a fixed point is unique while with* $\rho^* > 1$ *it is either unique or one of three such points.*

(proof in appendix)

This result complements Proposition 1 in describing the game's equilibrium strategies and beliefs. In order to analyze the case that there are multiple equilibria, i.e., that $\rho^* > 1$ holds and there are three such equilibrium values, I will restrict the analysis to "stable equilibria":

Definition. *An equilibrium of the communication game is the collection of strategies and beliefs as specified in Proposition 1 with* ρ^* *as a fixed point to* $g(\rho)$ *as described in Proposition* 2*. An equilibrium is then called an asymptotically "stable equilibrium" if the corresponding fixed point* ρ^* *obeys* $\frac{d}{d\rho}\big(g(\rho) - \rho\big)\big|_{\rho=\rho^*} < 0$.

This stability concept captures the notion of asymptotic stability (Hirsch and Smale, 1974). Although originally a dynamic concept, asymptotic stability has a long history of being used in the analysis of equilibria which originate from one-shot situations, e.g., for tâtonnment processes in (general) equilibrium and recently also in strategic communication settings (see Blume and Board, 2014). In particular, it captures that such stable equilibria converge back to their original value after small perpetuations, are locally unique, and can be found via iteratively. An equilibrium which is not stable does not have these properties.[17]. In the current setup, this applies only in one case:

Lemma 4. *Any unique equilibrium in the communication game is stable. If there are three equilibria (i.e, if $g(\rho)$ has three solutions $\rho_3^* > \rho_2^* > \rho_1^* > 1$), only the equilibrium associated with the intermediate value (i.e., ρ_2^*) is not stable.* (proof in appendix)

4.3 Linking equilibrium behavior to the information structure

The above results show that the equilibrium inference coefficient and, in particular, whether it is larger or smaller than one has a special relevance. As a reference case, consider the case of full disclosure ($\psi = 1$). In this case, the signal precisely indicates the sender's COIs. His message can then be corrected for this bias and inverted; it therefore reveals s and the sender type becomes effectively one-dimensional In consequence, rational receivers react one-to-one to changes in the message as indicated by $d_r^{*\prime}(m) = \rho^* = 1$. For no or imperfect disclosure, this does not hold. Different values of ρ^* then reflect how the strategic interplay of senders and rational receivers to is shaped by informational parameters of this game.

The following two results describe this link in more detail. First, a positive equilibrium inference coefficient and, therefore, a positive correlation between behavior of rational receivers and the sender's message, puts a restriction on some parameters:

Lemma 5. *A fixed point $\rho^* > 0$ to (11) exists if and only if $\sigma_{sc} > \tau^*$ for some $\tau^* < 0$.*

(proof in appendix)

The above links the covariance between s and c, the game's fundamental random variables, to the covariance of s and the sender's message m. As a direct consequence an equilibrium with $\rho^* > 0$ (on which this analysis focuses) requires σ_{sc} to be sufficiently large, i.e., larger than τ^*. The second result presents general conditions under which ρ^* is above and below this threshold of one. This is important with regard to uniqueness and stability of equilbria and how disclosure impacts receivers:

Lemma 6. *Suppose $\rho^* > 0$. Then $\rho^* < \, (=) \, [>] \, 1$ holds if and only if $\sigma_{sc} > \, (=) \, [<] \, -\sigma_c^2$.*

(proof in appendix)

As with the preceding lemma, this one maps the game's economic parameters into equilibrium behavior and back. This represents the following economic intuitions behind the values of ρ^*: Whenever fundamental uncertainty as measured by σ_c^2 is sufficiently high, the above condition and,

[17] Adapting the notation and results of Hirsch and Smale (1974), pp. 185-188 to the notation of this paper, a fixed point is asymptotically stable if $f'(\rho)|_{\rho=\rho_k^*} < 0$ where $f(\rho) \equiv g(\rho) - \rho$ such that $f(\rho)|_{\rho=\rho_k^*} = 0$. This lead to the above definition. Blume and Board (2014) use this to examine endogenously chosen vagueness in a one-shot communication game. They also provide references on how asymptotic stability is relevant in general one-shot situations, in particular to Samuelson's correspondence principle. In a related context, Gordon and Nöldeke (2013) also employ a stability concept which is based on the notion that after perpetuations, equilibria converge back to their original values

with it, $\rho^* < 1$ holds. Thus, for sufficiently high strategic uncertainty, rational receivers put some positive weight, measured by $1 - \rho^*$, on their prior over s and partly ignore the de-biased message; their action is then a strictly convex combination between these two elements. Note that the above condition is equivalent to $Corr[s, c] > -\frac{\sigma_c}{\sigma_s}$. Strategic uncertainty exceeding fundamental uncertainty (as given when $\frac{\sigma_c}{\sigma_s} > 1$) or a weakly positive correlation between s and c are therefore both sufficient conditions for such convex combinations.

However, when strategic uncertainty exceeds fundamental uncertainty, non-convex combinations, as expressed through $\rho^* > 1$, are also possible. By the above lemmas, this happens if first, $\tau^* < \sigma_{sc}$ holds so that the inference coefficient is positive and second, if $\sigma_{sc} < -\sigma_c^2$ also holds so that this coefficient exceeds one. Rational receivers then "over-react" – a change in the sender's message induces a change in rational receivers' demand which is greater than that the change in the sender's message. To understand the economic intuition behind this behavior, note that such a sufficiently negative value of σ_{sc} implies that rational receivers expect the sender to have a strong incentive to push demand into a direction opposite to the actual value of s. But because ρ^* is positive, a higher message m does, in expectation, still reflect a higher value of s. In equilibrium, rational receivers use this feature of the positive correlation between m and s by reacting very strongly, with $\rho^* > 1$, to the de-biased message. However, such extreme correction is based on the expected, not the actual COI. Thus, when the COI is too unpredictable relative to fundamental uncertainty, i.e. when $\frac{\sigma_c}{\sigma_s} \geq 1$ holds, such deliberate over-reaction is to risky and does not occur. The limit to such expectation-based corrections are also reached when $\sigma_{sc} \leq \tau^*$. In this case, the expected bias is so strong and opposed in direction to s that the risk of mis-correction outweighs the benefits of over-reacting. A meaningful communication equilibrium where the sender's message and rational receivers' action covary positively can the not be established.

Figure 2 illustrates these findings. It depicts, for the case of no disclosure, the values of the equilibrium inference coefficient ρ^* over possible correlations between the state of the world s and

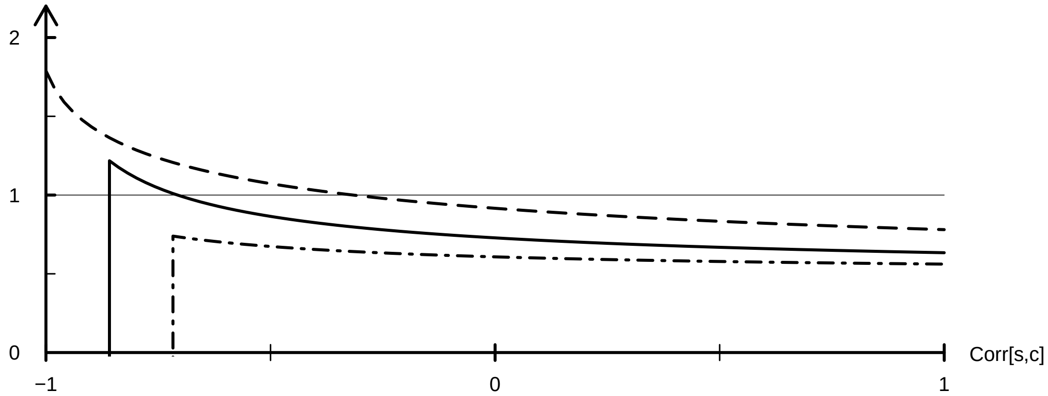

Figure 2: Stable, positive equilibrium inference coefficients ρ^* over $Corr[s, c] = \sigma_{sc}/(\sigma_s\sigma_c)$. Parameters: $\psi = 0, \mu = 0.5$, $\sigma_c^2 = 1$, and $\sigma_s^2 = 1/2/10$ (bottom/middle/top line).

the sender's COI c. The three lines represent different values of σ_s^2, higher ones representing larger variance. This ordering reflects that ceteris paribus, higher variation in s explains more variation in the message and thereby how much rational receiver's can rely on the de-biased message. This is expressed through higher values for ρ^*. Reflecting the previous lemma, the figure also shows that for any positive correlation and whenever $\frac{\sigma_c}{\sigma_s} \geq 1$ holds – as for the lower line – ρ^* is always contained in the unit interval. In contrast, the left region of the upper two lines portrays parameter constellations where the (normalized) covariance $\sigma_{sc}/(\sigma_s \sigma_c)$ is sufficiently low so that $\rho^* > 1$ can hold. The figure also portrays the normalized cutoff value $\tau^*/(\sigma_s \sigma_c)$. For the lower two lines, it is within the range of possible correlations. If $\sigma_{sc}/(\sigma_s \sigma_c)$ is below such a line, an equilibrium with $\rho^* > 0$ does not exist.

5 Consequences of disclosure

Given the preceding analysis of the communication game, I will now look on the effects of disclosing COIs. For this, I will consider changes from a situation of no disclosure ($\psi = 0$) to either imperfect disclosure ($0 < \psi < 1$) or full disclosure ($\psi = 1$). Recall from Proposition 2 and its discussion that under full disclosure, the equilibrium inference coefficient becomes one and the sender's message revealing. The following result captures the notion that this generalizes to imperfect disclosure. It shows that for any level disclosure, the equilibrium inference coefficient moves closer towards the benchmark value of one. Using Lemma 6, one can then relate this back to the game's fundamentals:

Lemma 7. *Upon disclosure, the equilibrium inference coefficient $\rho^* > 0$ of any stable equilibrium increases (remains constant) [decreases] if and only if $\sigma_{sc} >$ $(=)$ $[<]$ $-\sigma_c^2$.*

(proof in appendix)

In the following, I examine how the above-described changes which disclosure brings to the communication game affect welfare. For this, I will take an ex-ante view, thus before a draw of the sender's type takes place. I start with a view on naive receivers. They are agnostic about the sender's bias and do not account for the strategic change as for them, $d_n^*(m) = m$ holds. Recall that the utility of receivers decreases in the distance of their decision and the state of the world. In equilibrium, the sender announces s plus a bias. Since naive receivers follow this message, their (dis-)utility's argument is equal to this bias. Their expected utility is then given by

$$\mathrm{E}[u_n^R] = \mathrm{E}[L(c(\mu + (1-\mu)\rho^*)] = \mathrm{E}[L(|c|(\mu + (1-\mu)\rho^*)] < 0. \tag{12}$$

where the above follows from the fact that L is non-positive and symmetric around its maximum of zero. In consequence, the expected utility of naive receivers strictly decreases in the equilibrium inference coefficient. This reflects that higher values of ρ^* come with a stronger reaction of rational receivers to the sender's message (after correcting it for the expected bias). As this leads to an increased bias for the sender, this then hurts naive receivers:

Corollary 1. *Upon disclosure, the expected utility of naive receivers in any stable equilibrium increases* (*remains constant*) [*decreases*] *if and only if* $\sigma_{sc} < \ (=) \, [>] \ -\sigma_c^2$.

Thus, naive receivers only benefit from disclosure when it leads to a decrease in the equilibrium inference coefficient (i.e., only if $\sigma_{sc} < -\sigma_c^2$ holds).

To evaluate the overall effect of disclosure, one needs to also look at its effects on rational receivers. These receivers de-bias the message based on what they *expect* to be the sender's bias. How much they are hurt when the actual bias differs from this expectation depends on how much they rely on the corrected message and how volatile this bias is. To obtain a tractable measure for rational receivers' expected utility, one can exploit that their decision error $d_r^* - s$ is a linear combination of normally distributed random variables and therefore in itself normally distributed. One can then show that this is sufficient to represent rational receivers' expected utility as mean-variance preferences. As rational receivers' de-biasing is, on average, correct (i.e., the expected error is zero) this admits the following, single-argument representation:

Lemma 8. *The expected utility of rational receivers in equilibrium is given by*

$$\mathrm{E}[u_r^R] = \mathcal{L}\left(\sigma_s^2\Big[1 - \big(Corr[s,m]_{m=m^*(s,c)}\big)^2\Big]\right) \leq 0$$

with $\mathcal{L}'(x) < (=)\, 0$ *for any* $x > (=)\, 0$*. Also,* $\mathrm{E}[u_r^R] = 0$ *holds if and only if there is full disclosure.*

(The proof adapts some techniques from Meyer (1987) and can be found in the appendix.)

The above lemma shows that the expected utility of rational receivers can be expressed as a decreasing function of σ_s^2. This argument is scaled down by the square of one minus the correlation between s and the sender's equilibrium message m^*. This measure of the message's informativeness connects to the previously indicated regression analogy. Its empirical counterpart is the coefficient of determination (the R^2) one would obtain if one regressed past values of s on the corresponding messages by the sender (see footnote 8). This formulation of the expected utility for rational receivers helps in analyzing the opposing effects of disclosure: If $\sigma_{sc} > -\sigma_c^2$ applies and ρ^* increases upon disclosure, rational receivers react more to the message. However, the sender reacts to this and in turn, also increases the bias' magnitude so that the net effect on rational receiver's expected payoff is not clear. Conversely, when ρ^* decreases, so does the bias. But does such a decrease in the inference coefficient then not imply that also the message's informativeness and, with it, also rational receivers' utility decreases? Using the above lemma, the following result shows that in both scenarios, the net effect of disclosure on rational receivers' expected utility is positive:

Proposition 3. *Upon disclosure of COIs, the expected utility of rational receivers increases in every stable equilibrium.* (proof in appendix)

While this is good news from the perspective of rational receivers, naive ones are often hurt by disclosure. The following result captures this notion and follows from the preceding ones:

Corollary 2. *In any stable equilibrium, disclosure is a Pareto-improvement among all receivers if and only if* $\sigma_{sc} \leq -\sigma_c^2$.

The above shows that only when the conditions for the inference coefficient to be at least one are fulfilled (see Lemma 6), then all receivers benefit from disclosure. If this is not the case, naive receivers will suffer from disclosure. Based on a Pareto-criterion, disclosure is then not optimal. Note that similiar arguments can also be made even when naive receivers are absent but the bias in the sender's message and associated lying costs matter. This follows from the fact that the sender's expected lying costs are a special, parametrized case of the naive receivers' expected utility. For both, the sender's bias is the relevant argument. If in either case, the above condition is not fulfilled, a policy maker who can steer disclosure might want to resort to other criteria to examine its consequences. I capture such a criterion by assuming the following welfare function with weights w_n, w_r, and w_k such that $w_r > 0$ and $\max\{w_n, w_k\} > 0$:

$$W = w_n \cdot \mathrm{E}[u_n^R] + w_r \cdot \mathrm{E}[u_r^R] - w_k \cdot \mathrm{E}[(c(\mu + \rho^*(1-\mu)))^2] \tag{13}$$

The first two terms in the above expression capture the expected utility of naive and rational receivers, respectively. The third term captures the sender's expected cost of lying, that is, the squared bias. The sender's COI is not included. This is because it is considered as a transaction gain, either earned by the sender itself or passed through by a third party paying a commission. Consequently, there is a counter-party who makes the corresponding loss so that $cD(m)$ is a welfare-irrelevant transfer (see also the examples in Subsection 3.2). The terms in (13) can therefore capture the dis-utility which arises due to a loss of information and misguided actions by receivers. If $w_k = 0$, this is with regard to receiver welfare only, for example with weights $w_n = \mu$ and $w_r = 1 - \mu$. This then captures receivers' costs of making a wrong decision based on biased or imperfectly de-biased information. Setting $w_k > 0$ includes the sender's lying costs into efficiency considerations, for example because these costs per se matter or because they capture the sender's loss of reputation.[18]

Once such a function is formulated, one can use the described framework to quantify the welfare gains of disclosure. Of course, this requires more assumptions on the utility functions and distribu-

[18] In this case, the above and following results also hold, even if naive receivers are not present ($\mu = 0$) or do not matter for welfare ($w_n = 0$). This is because expected lying costs are a special, parametrized case of the naive receivers' expected utility (12).

tional parameters. However, policy-relevant statements with respect to the effect of disclosure on W can be made even when such exact parameters are unknown:

Proposition 4. *If $\sigma_{sc} \leq (>) - \sigma_c^2$, full disclosure always (never) maximizes welfare W in any stable equilibrium.* (proof in appendix)

The above shows that full disclosure is often inefficient. This happens whenever naive receivers are hurt by disclosure (see Corollary 2). The intuition for this follows from the fact that receivers have smooth, strictly concave utility and that rational receivers can achieve maximum utility only with full disclosure. When they are near to this optimum and some sufficiently low noise is added to a perfect signal $\tilde{c}$, the resulting loss can then be made arbitrarily small. In contrast, the relative gain in expected utility for naive receivers which comes through the associated decrease in the reaction of rational receivers – and therefore the sender's bias – is bigger. Also note that while full disclosure is often not optimal, the reverse reasoning does not work: No disclosure at all can be optimal, in a second-best sense. Determining whether this holds or not and, in the latter case, what the optimal interior level of disclosure is, does however require concrete assumptions on the receivers' loss functions and the game's parameters (an example where no disclosure is optimal is contained in the appendix).[19]

Finally, note that knowledge on whether $\sigma_{sc} \leq \sigma_c^2$ holds, the condition which has been shown to crucially determine the consequences of disclosure, is not always necessary to asses its impact. For an observer who wants to get testable predictions and make informed decisions it can also be sufficient to just observe how the market reacts to new information. To see this, note that since naive receivers follow the sender literally, one can combine (5) and (10) to get that $D'(m) \geq 1$ is equivalent to $\rho^* \geq 1$. Using Lemma 6, one then immediately gets that $\sigma_{sc} \geq -\sigma_c^2$ is equivalent to $D'(m) \geq 1$. In consequence, observing a less than one-to-one reaction implies that upon disclosure, naive receivers will be hurt and that full disclosure is inefficient. Conversely, when receivers react, on average, stronger or equal than one-to-one to a change in the sender's message, it follows that full disclosure is optimal and benefits all receivers. Knowing the elasticity of demand with respect to information can therefore be sufficient to assess the impact of disclosure.

[19]This is because even when assumptions are made which ensure an interior optimal level of disclosure, the comparative statics regarding it are often shaped by two effects: First, there is a direct effect of changes in the game's parameters. They affect the (weighted) marginal utilities of receivers. This then implies a change in the optimal level of disclosure to re-balance these utilities. One can then show that, under assumptions which ensure an interior level of optimal disclosure, this is the only effect for some parameters (e.g., $\bar{s}$, $\bar{c}$, the weights in W) so that for them, clear comparative statics are feasible. However, changes in other parameters (i.e., $\mu, \sigma_s^2, \sigma_c^2, \sigma_{sc}$), do also affect ρ^*, as specified in (11). These indirect effects typically oppose the direct effect on receivers' marginal utilities and prevents generally-valid, clear-cut comparative statics with regard to the optimal interior level of disclosure.

6 Conclusion

This paper describes a setting where a sender communicates the value of a random variable of interest to uninformed receivers. The sender does so while facing a conflict of interest to manipulate the receivers' actions and, at the same time, also facing lying costs. In a parsimonious framework which can accommodate various situations where strategic communication affects market behavior, where receivers can have general levels of risk aversion, I study the effects of disclosing confltICTS of interest via a signal of arbitrary precision.

I find that disclosure fulfills the aim of informing *rational* receivers: Information about the sender's COI helps them to infer more from the sender's biased message and to adjust their actions more closely to the actual state of the world. On the downside, this paper's core result shows that exactly this desired effect of disclosure backfires on naive or delegating receivers. It does so because, in equilibrium, the average reaction to the biased signal and the sender's bias are mutually dependent. Upon disclosure, as rational receivers get helpful information to de-bias the sender's message, their reaction to the sender's message often increases. With this increase, the bias contained in the sender's message also increases. Naive receivers who do not account for the strategic aspect of communication are then hurt by disclosure. Disclosure therefore often amplifies a negative externality which rational receivers exert on their naive peers; it thus hurts those who are most vulnerable to communication biases.

I also determine more precisely when and how these adverse effects of disclosure manifest. In terms of economic fundamentals, this is always the case when the state of the world and the sender's conflict of interest are weakly positively correlated. Another sufficient condition for disclosure to backfire is when strategic uncertainty regarding the sender's bias exceeds fundamental uncertainty regarding the state of the world. In terms of observed behavior, this happens when an expert's message does not induce at least a one-to-one average reaction among receivers. Only when they react stronger than one-to-one to changes in the sender's message, then disclosure is an improvement among all, rational *and* naive, receivers. This is also the only case when full disclosure is optimal from an efficiency point of view. In all other cases, a less than perfect signal about the sender's COI, potentially even an uninformative one, is optimal for maximizing efficiency.

The results presented here show that when some people do not have the ability or the information and time to act in a completely Bayesian and rational manner, disclosure often hurts. Merely communicating an expert's conflict of interest does often *not* solve the problems which arise – it rather increases its negative effects. In consequence, disclosure is not the regulatory panacea it promises to be. This suggests that eliminating conflicts of interest promises much better outcomes than just announcing them.

References

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

Abeler, J., C. Raymond, and D. Nosenzo (2018). Preferences for Truth-Telling. *Econometrica*, forthcoming.

Bénabou, R. and J. Tirole (2006). Incentives and prosocial behavior. *American Economic Review 96*(5), 1652–1678.

Blume, A. and O. Board (2014). Intentional Vagueness. *Erkenntnis 79*(S4), 855–899.

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. *Journal of Legal Studies 34*(1), 1–25.

Cain, D. M., G. Loewenstein, and D. A. Moore (2011). When Sunlight Fails to Disinfect: Understanding the Perverse Effects of Disclosing Conflicts of Interest. *Journal of Consumer Research 37*(5), 836–857.

Chen, Y. (2011). Perturbed communication games with honest senders and naive receivers. *Journal of Economic Theory 146*(2), 401–424.

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

Deimen, I. and D. Szalay (2014). Smooth, Strategic Communication. *CEPR Discussion Paper 10190*.

Embrechts, P., A. McNeil, and D. Straumann (2002). Correlation and dependence in risk management: properties and pitfalls. In M. Dempster (Ed.), *Risk Management: Value at Risk and Beyond.*, pp. 176–223. Cambridge University Press.

Erat, S. and U. Gneezy (2012). White Lies. *Management Science 58*(4), 723–733.

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE 4*(5).

Fischer, P. E. and R. E. Verrecchia (2000). Reporting Bias. *The Accounting Review 75*(2), 229–245.

Fung, A., M. Graham, and D. Weil (2007). *Full disclosure: The perils and promise of transparency*. Cambridge University Press.

Gennaioli, N., A. Shleifer, and R. Vishny (2015). Money doctors. *Journal of Finance 70*(1), 91–114.

Gómez, E., M. A. Gómez-Villegas, and J. M. Marin (2003). A survey on continuous elliptical vector distributions. *Revista Matemática Complutense 16*(1), 345–361.

Gordon, S. and G. Nöldeke (2013). Figures of Speech in Strategic Communication. *mimeo*.

Hayes, R. M. (1998). The Impact of Trading Commission Incentives on Analysts' Stock Coverage Decisions and Earnings Forecasts. *Journal of Accounting Research 36*(2), 299–320.

Hirsch, M. W. and S. Smale (1974). *Differential Equations, Dynamical Systeman, and Linear Algebra*. Academic Press.

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

Inderst, R., A. Rajko, and A. Ockenfels (2010). Transparency and Disclosing Conflicts of Interest: An Experimental Investigation. *GEABA Discussion Paper Series in Economic and Management 10*(20).

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

Kartik, N. and A. Frankel (2018). Muddled Information. *Journal of Political Economy*, forthcoming.

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

Koch, C. and C. Schmidt (2010). Disclosing conflicts of interest - Do experience and reputation matter? *Accounting, Organizations and Society 35*(1), 95–107.

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

Loewenstein, G., C. R. Sunstein, and R. Golman (2014). Disclosure: Psychology Changes Everything. *Annual Review of Economics 6*, 391–419.

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.

Meyer, J. (1987). Two-moment decision models and expected utility maximization. *American Economic Review 77*(3), 421–430.

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.

Mullainathan, S., M. Noeth, and A. Schoar (2012). The Market For Financial Advise: An Audit Study. *NBER working paper 17929*.

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.

Simmons, J. P., L. D. Nelson, and U. Simonsohn (2011). False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science 22*, 1359–1366.

Simonsohn, U. (2014). P-curve: A Key To The File Drawer. *Journal of Experimental Psychology: General 143*(2), 534–547.

Sims, C. (2003). Implications of rational inattention. *Journal of Monetary Economics 50*, 665–690.

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

Sobel, J. (2013). Giving and Receiving Advice. In D. Acemoglu, M. Arellano, and E. Dekel (Eds.), *Advances in Economics and Econometrics: Theory and Applications, Tenth World Congress of the Econometric Society*, Volume 2. Cambridge University Press.

Steen, R. G. (2011). Retractions in the scientific literature: is the incidence of research fraud increasing? *Journal of Medical Ethics 37*, 249–253.

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

Appendix

Features of elliptically distributed random variables

While in the main text, F was assumed to be a multivariate normal distribution, all proofs in this appendix will treat F to be of the more general class of elliptical distributions. The purpose of this section is to provide some information on them. It starts with the following definition, obtained from Definition 1 and Theorem 4 ii) in the survey on elliptical distributions by Gómez et al. (2003):

Definition. *A random vector $\mathbf{x} \in \mathbb{R}^n$ is elliptically distributed with expected value $\boldsymbol{\eta} \in \mathbb{R}^n$, positive definite variance-covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$, and the Lebesque-measurable function $g : [0, \infty) \to [0, \infty)$ s.t. $\int_0^\infty t^{\frac{n}{2}-1} g(t)dt < \infty$ as parameters if it has the density function*

$$f(\mathbf{x}; \boldsymbol{\eta}, \tilde{\boldsymbol{\Sigma}}, g) = c_n |\tilde{\boldsymbol{\Sigma}}|^{-\frac{1}{2}} g\left((\mathbf{x} - \boldsymbol{\eta})^T \tilde{\boldsymbol{\Sigma}}^{-1} (\mathbf{x} - \boldsymbol{\eta})\right)$$

where $c_n = \Gamma(\frac{n}{2}) / \left(\phi^{\frac{n}{2}} \int_0^\infty t^{\frac{n}{2}-1} g(t)dt\right)$ and $\tilde{\boldsymbol{\Sigma}} \propto \boldsymbol{\Sigma}$.

Therefore, the exact form of the distribution depends on the density generator g. In the context of this paper, it is assumed to be implicitly defined by the specific elliptical distribution F and its dimensionality. The generic example is when F denotes the normal distribution which would imply that $g(t) = \exp\left(-\frac{1}{2}t\right)$ for $n \geq 1$. Other examples, with different density generators, include the mutivariate logistic, student-t or power exponential families of distributions.

The results in this paper do not depend on the specific distribution F as long as it is elliptical, but just on its first two moments, $\boldsymbol{\eta}$ and $\boldsymbol{\Sigma}$. To illustrate them, consider a random vector $\mathbf{x} \in \mathbb{R}^n$ with $n \geq 2$ which is elliptically distributed according to $F(\boldsymbol{\eta}, \boldsymbol{\Sigma})$. Also consider two non-empty partitions $[\mathbf{x}_1, \mathbf{x}_2]$ of this vector. Partition analogously the corresponding vector of expected values as $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ and the variance-covariance matrix $\boldsymbol{\Sigma}$ into blocks $(\boldsymbol{\Sigma}_{11}, \boldsymbol{\Sigma}_{12}, \boldsymbol{\Sigma}_{21}, \boldsymbol{\Sigma}_{22})$.

A random vector $\mathbf{x}$ from an elliptical distribution F has the following three properties which will be used in the following proofs:

E1: *linear combinations of elements of $\mathbf{x}$ are distributed according to F*

E2: $(\mathbf{x}_2|\mathbf{x}_1)$ *is distributed according to*

$$F(\boldsymbol{\eta}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\eta}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$$

E3: $\mathbf{x}$ *is symmetrically distributed around* $\boldsymbol{\eta}$

E3 follows from the above density function. Properties E1 and E2 are consequences of Theorem 5 and Theorem 8, respectively in Gómez et al. (2003) which also contains further references on the original research establishing these properties for elliptical distributions. It will be useful to note that for the special case that $(\mathbf{x}_2|\mathbf{x}_1) \in \mathbb{R}^2$, E2 implies that $(x_2|x_1)$ is distributed according to

$$F\left(\mathrm{E}[x_2] + (x_1 - \mathrm{E}[x_1])\frac{Cov[x_1, x_2]}{Var[x_1]}, Var[x_2]\left(1 - (Corr[x_1, x_2])^2\right)\right).$$

Proof of Lemma 1

The implicit function $R(s, m, c) = cD'(m) + s - m = 0$, stated as (4), follows directly from the first-order condition for maximizing (3). Note that the associated second-order-condition for an optimum implies $\frac{\partial R(s,m,c)}{\partial m} = cD''(m) - 1 < 0$ at each optimally chosen m. Using the implicit function theorem, one then gets that

$$\frac{d\,m}{d\,s} = -\frac{\partial R(s, m, c)}{\partial s} \Big/ \frac{\partial R(s, c, m)}{\partial m} = \frac{1}{1 - cD''(m)} > 0$$

for an optimally chosen message m. $\square$

Proof of Lemma 2

When m is distributed according to F, it is jointly elliptically distributed with s. By E2, the resulting distribution of s conditional on m, denoted by its pdf $f(s|m)$, is then also elliptical. Furthermore, E3 implies that $f(s|m)$ is symmetric around $\mathrm{E}[s|m]$. By definition, it then has to hold that $d_r = \text{argmax}_{d \in S} \int_{\mathbb{R}} L(d-s) f(s|m) ds$. The necessary FOC for a candidate solution $d_r = \mathrm{E}[s|m]$ is given by

$$0 = \int_S L'(d_r - s) f(s|m]) ds = \int_{-\infty}^{+\infty} L'(\mathrm{E}[s|m] - s) f(s|m]) ds$$

and is also sufficient as L is strictly concave. To verify that this FOC applies for this candidate solution note that by being strictly concave, L is single peaked and symmetric around its bliss point s. Let $\Delta \leq 0$ be the absolute deviation of the candidate solution from the optimal choice, i.e. $\Delta = |d_r - s|$. By symmetry of L around zero it holds that $L'(\Delta) = -L'(-\Delta)$. Since $f(s|m)$ is symmetric around $\mathrm{E}[s|m] = d_r$ it then follows that

$$L'(\Delta) f(d_r - \Delta|m]) = -L'(-\Delta) f(d_r + \Delta|m]) \leqq 0$$

applies for any $\Delta \geqq 0$. Integrating over all $\Delta \in \mathbb{R}_+$ then validates that the above FOC actually holds. Since L is single-peaked, it is also the only solution. $\square$

Proof of Proposition 1

The following proof constructs a linear, pure strategy equilibrium when the implied demand $D(m)$. To do so, it proceeds in three steps. Step 1 solves the rational receiver's problem to choose his optimal action, given that the sender's message contains information about s. Step 2 determines how such signal extraction by rational receivers manifests in equilibrium when the sender anticipates this process. Step 3 combines these results to obtain equilibrium actions and beliefs.

Step 1: Consider a candidate equilibrium messaging strategy $\tilde{m}(s, c)$ such that $D(m) = \mu m + (1 - \mu)\tilde{d}_r(m)$ with $\tilde{d}_r(m) \equiv \text{argmax}_{d \in S} \mathrm{E}[L(d-s)|m]_{m=\tilde{m}(s,c)}$ exists and is twice differentiable w.r.t. m. By (4), each image $\tilde{m}$ has to be provided by the following messaging function.

$$\tilde{m}(s, c) = s + c(\mu + (1 - \mu)\tilde{d}'_r(\tilde{m})).$$

Note that for a given candidate equilibrium inference function $\tilde{d}_r$ and its derivative, $\tilde{d}'_r(\tilde{m}) = \tilde{d}'_r(m)|_{m=\tilde{m}}$ is a non-random image of the function $\tilde{d}'_r(m)$, i.e. a constant. Therefore, $\tilde{m}(s, c)$ is a linear combination of s and c and, by E1, distributed according to F. Lemma 2 then implies that $\tilde{d}_r(\tilde{m}) = \mathrm{E}[s|\tilde{m}]$ has to hold. Using E2 then yields that for any message m which is sent in an equilibrium

$$\tilde{d}_r(m) = \mathrm{E}[s] + \Big(m - \mathrm{E}[s] - \mathrm{E}[c](\mu + (1-\mu)\tilde{d}_r'(m)\Big) \frac{Cov[s,m]_{m=\tilde{m}(s,c)}}{Var[m]_{m=\tilde{m}(s,c)}} \tag{A.1}$$

has to hold (where $\mathrm{E}[m] = \mathrm{E}[s] + \mathrm{E}[c](\mu + (1-\mu)\tilde{d}_r'(m)$ has been used).

Step 2: Given the assumption that in any equilibrium, the associated inference coefficient $\tilde{\rho} \equiv \frac{Cov[s,m]_{m=\tilde{m}(s,c)}}{Var[m]_{m=\tilde{m}(s,c)}}$ is a constant, the function $\tilde{d}_r^{*\prime}(m)$ has to solve the following first-order linear differential equation :

$$\tilde{d}_r'(m) = \Big(1 - \mathrm{E}[c](1-\mu)\tilde{d}_r''(m)\Big)\,\tilde{\rho}.$$

When $\tilde{d}_r'(m) = 0$, it follows that $\tilde{d}_r'(m) = \tilde{\rho} = 0$. Similarly, if $\mathrm{E}[c] = 0$, then $\tilde{d}_r'(m) = \tilde{\rho}$. Now suppose that $\tilde{\rho}\mathrm{E}[c] \neq 0$. One then gets $\tilde{d}_r'(m)$ as the solution to the above differential equation, given by

$$\tilde{d}_r'(m) = \rho + \xi \cdot exp\left(-\frac{m}{(1-\mu)\mathrm{E}[c]\tilde{\rho}}\right),$$

where ξ is an integration factor. To determine its value, integrate the obtained $\tilde{d}_r'(m)$ over $M = \mathbb{R}$:

$$\mathrm{E}[s|\tilde{m}] = \int_{-\infty}^{+\infty} \tilde{d}_r'(m)dm = m\tilde{\rho} - \xi(1-\mu)\mathrm{E}[c]\tilde{\rho} \cdot exp\left(-\frac{m}{(1-\mu)\mathrm{E}[c]\tilde{\rho}}\right) + \tilde{K} \tag{A.2}$$

In the above, $\tilde{K}$ is a constant of integration. This can expression be plugged into the sender's expected utility to obtain $\mathrm{E}[U^S(s,c,m)|_{m=\tilde{m}(s,c)}] =$

$$c\mu m + c(1-\mu)\left[m\tilde{\rho} - \xi(1-\mu)\mathrm{E}[c]\tilde{\rho} \cdot exp\left(-\frac{m}{(1-\mu)\mathrm{E}[c]\tilde{\rho}}\right) + \tilde{K}\right] - \frac{1}{2}(m-s)^2 \tag{A.3}$$

To determine ξ, I start with the case that $c > 0$. In this case, $U^S(s,c,m)$ is increasing in $\mathrm{E}[s|\tilde{m}*]$, the term above in square brackets. If $\tilde{\rho}\mathrm{E}[c] > 0$, 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 would then maximize his 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 any $\xi > 0$, 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. It follows that with $c > 0$ and $\rho\mathrm{E}[c] > 0$, only $\xi = 0$ can be the equilibrium integration factor.

Continue to suppose that $c > 0$ but now $\rho\mathrm{E}[c] < 0$ holds. Inverse to the the preceding reasoning, $\mathrm{E}^*[s|\tilde{m}]$ 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. Also inverse to the above, any $\xi < 0$ would decrease the sender's expected utility so that $\xi = 0$ is chosen in any equilibrium with $c > 0$ and $\rho\mathrm{E}[c] < 0$.

For the case that $c < 0$, $U^S(s,c,m)$ is decreasing in $\mathrm{E}[s|\tilde{m}]$. 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 an equilibrium with $c < 0$ and $\tilde{\rho}\mathrm{E}[c] \neq 0$ holds.

Eventually, when $c = 0$ the inference $\mathrm{E}[s|m]$ does not enter $U^S(s,c,m)$; it is equivalent to $\xi = 0$. Thus, $\xi = 0$ is the unique integration factor and $\tilde{d}_r^*(m) = \tilde{\rho}$ holds.

Step 3: Given the above, one can determine the integration constant

$$\tilde{K} = \mathrm{E}[s] - (\mathrm{E}[s] + \mathrm{E}[c]\,(\mu + (1-\mu)\tilde{\rho}))\tilde{\rho}$$

by combining (A.1) and (A.2). Using $\xi = 0$ then allows to write (A.3) as

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

It is easily verified that the unique message which maximizes the above expression is given by $m = s + c\,(\mu + (1-\mu))\,\tilde{\rho}$. In equilibrium, it then has to hold that $m^*(s,c) = s + c\,(\mu + (1-\mu)\rho^*$ with $\rho^* = \tilde{\rho} = d_r^{*\prime}(m) = \frac{Cov[s,m]_{m=m^*(s,c)}}{Var[m]_{m=m^*(s,c)}}$, as stated in (9). Using $\tilde{\rho} = \rho^*$, $\xi = 0$, and the above expression for $\tilde{K}$ on (A.2) then yields the rational receivers belief and strategy as stated in (10). □

Proof of Lemma 3

The assumptions on ϵ as stated in the main text can be re-stated as follows:

$$\begin{pmatrix} s \\ c \\ \epsilon \end{pmatrix} \sim F\left(\begin{bmatrix} \bar{s} \\ \bar{c} \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_s^2 & \sigma_{sc} & 0 \\ \sigma_{sc} & \sigma_c^2 & 0 \\ 0 & 0 & \sigma_\epsilon^2 \end{bmatrix} \right)$$

Using E1 establishes that $\tilde{c} = c + \epsilon$ is distributed according to F and so is the random vector $(\tilde{c}, s, c)$. Note that because ϵ is not correlated with s or c and has an expected value of zero, $Cov[s, \tilde{c}] = \mathrm{E}[(s - \mathrm{E}[s])(c + \epsilon - \mathrm{E}[c])] = \mathrm{E}[(s - \bar{s})(c - \bar{c})] = \sigma_{sc}$, $Var[\tilde{c}] = \mathrm{E}[(c + \epsilon - \mathrm{E}[c])^2] = \mathrm{E}[(c + \epsilon - \bar{c})^2] = \sigma_c^2 + \sigma_\epsilon^2$, and $Cov[c, \tilde{c}] = \mathrm{E}[(c - \mathrm{E}[c])(c + \epsilon - \bar{c})] = \mathrm{E}[(c - \bar{c})(c - \bar{c})] = \sigma_c^2$. Therefore, it holds that

$$\begin{pmatrix} \tilde{c} \\ s \\ c \end{pmatrix} \sim F\left(\begin{bmatrix} \bar{c} \\ \bar{s} \\ \bar{c} \end{bmatrix}, \begin{bmatrix} \sigma_c^2 + \sigma_\epsilon^2 & \sigma_{sc} & \sigma_c^2 \\ \sigma_{sc} & \sigma_s^2 & \sigma_{sc} \\ \sigma_c^2 & \sigma_{sc} & \sigma_c^2 \end{bmatrix} \right).$$

Using E2 with the parameters from the above distribution then yields, after some rearranging, the stated conditional moments for $(s, c \mid \tilde{c})$. □

Proof of Proposition 2

By using $m^*(s,c) = s + c(\mu + (1-\mu)\rho)$ from Proposition 1 and the definition of $\rho^* = \frac{Cov[s,m^*]}{Var[m^*]}$, the latter must be a solution to

$$\rho = \frac{Cov[s, m^*]}{Var[m^*]} = \frac{Cov[s, m|\tilde{c}]_{m=m^*(s,c)}}{Var[m|\tilde{c}]_{m=m^*(s,c)}} = \frac{\mathrm{E}[(s - \mathrm{E}[s])[(s - \mathrm{E}[s]) + (\mu + (1-\mu)\rho)(c - \mathrm{E}[c])]|\tilde{c}]}{\mathrm{E}[((s - \mathrm{E}[s]) + (\mu + (1-\mu)\rho)(c - \mathrm{E}[c]))^2|\tilde{c}]}$$

Plugging in the elements of $F(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}})$ as described in Lemma 3 means that ρ^* has to solve

$$\rho = \frac{\left(1 - \psi(Corr[s,c])^2\right)\sigma_s^2 + (\mu + (1-\mu)\rho)(1-\psi)\sigma_{sc}}{(1 - \psi(Corr[s,c])^2)\sigma_s^2 + 2(\mu + (1-\mu)\rho)(1-\psi)\sigma_{sc} + (\mu + (1-\mu)\rho)^2(1-\psi)\sigma_c^2}$$

With full disclosure ($\psi = 1$), one gets $\rho^* = 1$. Otherwise, if $1 - \psi > 0$, one can factor this term which then yields, after substituting using $\phi = (1 - \psi(Corr[s,c])^2/(1-\psi)$, expression (11).

It will be useful to denote the above nominator and denominator in that expression by the following functions:

$$
\begin{aligned}
N(\rho) &\equiv \phi\sigma_s^2 + (\mu + (1-\mu)\rho)\sigma_{sc} \\
D(\rho) &\equiv \phi\sigma_s^2 + 2(\mu + (1-\mu)\rho)\sigma_{sc} + (\mu + (1-\mu)\rho)^2\sigma_c^2
\end{aligned}
$$

For the rational function (11), expressed as $g(\rho) \equiv N(\rho)/D(\rho)$, the following properties hold:

Property a) $N(\rho)/D(\rho)$ *is continuous with* $D(\rho) > 0$ *for all* $\rho \in \mathbb{R}$.

Proof: Since both, $D(\rho)$ and $N(\rho)$ are continuous in ρ, it is sufficient to show that $D(\rho) > 0$ always holds. Suppose to the contrary it would not. Rearranging $D(\rho)$, this would requiere that the quadratic function $\rho^2 + a\rho + b = 0$ with coefficients

$$
a = \frac{2(\sigma_{sc} + \mu\sigma_c^2)}{(1-\mu)\sigma_c^2} \qquad b = \frac{\phi\sigma_s^2 + 2\mu\sigma_{sc} + \mu^2\sigma_c^2}{(1-\mu)^2\sigma_c^2}
$$

has at least one real solution, thus that $(a/2)^2 - b \geq 0$ holds. Plugging in and rearranging, this yields $(\sigma_{sc}/(\sigma_c\sigma_s))^2 \geq \phi > 1$; a contraction to $|Corr[s,c]| < 1$. ☐

Property b) $\lim_{\rho\to+\infty}\left(\frac{N(\rho)}{D(\rho)}\right) = 0^-$ *if* $\sigma_{sc} < 0$ *and* $\lim_{\rho\to+\infty}\left(\frac{N(\rho)}{D(\rho)}\right) = 0^+$ *if* $\sigma_{sc} \geq 0$

Proof: $N(\rho)$ strictly decreases (weakly increases) linearly in ρ when $\sigma_{sc} < 0$ ($\sigma_{sc} \geq 0$) and attains negative (positive) values for ρ large enough. From a), $D(\rho)$ is strictly positive and it grows quadratically in ρ. Therefore, for large values of ρ, the ratio $N(\rho)/D(\rho)$ is negative (positive) and arbitrarily close to zero. ☐

Property c) $\frac{N(\rho)}{D(\rho)}$ *has at most two extreme points.*

Proof: Any extreme point has to set the first derivative

$$
\begin{aligned}
\left(\frac{N(\rho)}{D(\rho)}\right)' &= \frac{(1-\mu)\sigma_{sc}D(\rho) - 2(1-\mu)N(\rho)(\sigma_{sc} + (\mu + (1-\mu)\rho)\sigma_c^2)}{(D(\rho))^2} \\
&= \frac{(1-\mu)}{D(\rho)} \cdot \left(\sigma_{sc} - \frac{N(\rho)}{D(\rho)} \cdot 2(\sigma_{sc} + (\mu + (1-\mu)\rho)\sigma_c^2)\right)
\end{aligned}
$$

equal to zero. By a) and $\mu \in [0,1)$, the first factor is non-zero. Extreme points therefore have to solve

$$
\sigma_{sc}D(\rho) = N(\rho) \cdot 2(\sigma_{sc} + (\mu + (1-\mu)\rho)\sigma_c^2)
$$

Plugging in the functions for $N(\rho)$ and $D(\rho)$ yields an equation which is quadratic in ρ and, therefore, has at most two real solutions.

A solution ρ^* to (11) requires an intersection of the 45-degree line and $N(\rho)/D(\rho)$. Note that every such fixed point has to be a root of $k(\rho) = \rho D(\rho) - N(\rho)$, i.e. the following cubic equation:

$$
k(\rho) = \underbrace{(1-\mu)^2\sigma_c^2}_{\equiv A} \cdot\rho^3 + \underbrace{2(1-\mu)(\sigma_{sc} + \mu\sigma_c^2)}_{\equiv B} \cdot\rho^2 + \underbrace{\phi\sigma_s^2 + \mu^2\sigma_c^2 + (3\mu - 1)\sigma_{sc}}_{\equiv C} \cdot\rho \underbrace{-\phi\sigma_s^2 - \mu\sigma_{sc}}_{\equiv D} \tag{A.4}
$$

To examine multiplicity of roots to $k(\rho)$, I use the following result:

Theorem. (Descarte's rule of signs) *Consider a* n*-degree polynominal* $p(x) = \sum_{d=0}^{n} c_d \cdot x^d$ *with* *real coefficients. Order the non-zero coefficients c_k in an descending order of the exponent d. 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.*

It always holds that $A > 0$. Furthermore, in the proof of Lemma 5 it is shown that a solution $\rho > 0$ implies $D < 0$ because this is equivalent to $\sigma_{sc} > \tau(0)$ and $\tau(0) > \tau(\rho^*) = \tau^*$. By the sign rule, the only configuration for more than one sign change, given that $A > 0 > D$, is $C > 0 > B$. Thus, there are either one or three positive roots, which correspond to fixed points of $g(\rho)$.

Multiple fixed points require $B < 0$ and, therefore, $\sigma_{sc} < 0$. Suppose they exist. By property a) and b) derived in the first part of this proof, this means that $g(\rho) = N(\rho)/D(\rho)$ continuously approaches zero from below when ρ becomes large enough. Also, it has been shown that $N(0)/D(0) > 0$ (see proof of Lemma 5). Together, this implies that g has to have a negatively valued local minimum on $\mathbb{R}_{++}$ denoted by ρ_-, i.e. $g(\rho_-) < 0$. If ρ_- is the only extreme value over $\mathbb{R}_{++}$ this implies only one intersection with the 45-degree line, thus a unique fixed point. If it is not the unique extreme value then, by property c) derived in the first part of the proof, there is exactly one other extreme value of $g(\rho)$ over $\mathbb{R}_{++}$. Given that ρ_- is a local minimum, it has to be a local maximum and is denoted by ρ_+. It then follows from $\lim_{\rho\to+\infty} g(\rho) = 0^-$ that $0 < \rho_+ < \rho_+$ and $g(\rho_-) < 0 < g(\rho_+)$ have to hold. Accordingly, $g(\rho) = N(\rho)/D(\rho)$ is non-increasing on $[\rho^+, \rho^-]$. This, together with $N(0)/D(0) > 0$ implies that $\rho(\rho)$ cuts the 45-degree line exactly only once within this interval and never on $(\rho^-, +\infty)$. Thus, the unique fixed-point in (ρ^+, ρ^-) is the highest-valued one. As multiple, positively valued fixed points of $g(\rho)$ number to three, their coordinates can be denoted w.l.o.g. by $0 < \rho_1^* < \rho_2^* < \rho_3^*$. It then has to hold that

$$0 < \rho_1^* < \rho_2^* < \rho_+ < \rho_3^* < \rho_-$$

As there is no further extreme point over $[\rho_1^*, \rho_2^*] \subset (0, \rho_+)$ while $0 < g(0) < g(\rho_+)$ holds, it follows that $g(\rho)$ is non-decreasing on $[\rho_1^*, \rho_2^*]$. Three fixed points of $g(\rho)$ at $\rho_1^* < \rho_2^*$ and $\rho_3 \in (\rho_+, \rho_-)$ then imply that $g(\rho) = N(\rho)/D(\rho)$ cuts the 45-degree line (which has slope 1) thrice: First from above at ρ_1^*, then from below at ρ_2^* (which requires a slope larger than 1), and then from above at ρ_3^*. It thus holds that

$$g(\rho)'|_{\rho=\rho_3^*} < 0 < g(\rho)'|_{\rho=\rho_1^*} < 1 < g(\rho)'|_{\rho=\rho_2^*}. \tag{A.5}$$

Using the fact that if this indeed an equilibrium, $\rho_1^* = g(\rho_1^*) = N(\rho_1^*)/D(\rho_1^*)$ has to hold, the requirement of a positive slope greater at $\rho_1^* > 0$ translates via (6) into

$$\left(\frac{N(\rho)}{D(\rho)}\right)'\Bigg|_{\rho=\rho_1^*} = \frac{(1-\mu)}{D(\rho_1^*)} \cdot \left(\sigma_{sc} - 2\rho_1^*(\sigma_{sc} + (\mu + (1-\mu)\rho_1^*)\sigma_c^2)\right) > 0.$$

For this to hold, $\sigma_{sc} + (\mu + (1-\mu)\rho_1^*)\sigma_c^2 < 0$ is a necessary condition as $\rho_1^* > 0 > \sigma_{sc}$. Multiplying by $(\mu + (1-\mu)\rho_1^*) > 0$ yields the equivalent necessary condition

$$(\mu + (1-\mu)\rho_1^*)\sigma_{sc} + (\mu + (1-\mu)\rho_1^*)^2\sigma_c^2 = D(\rho_1^*) - N(\rho_1^*) < 0.$$

Rearranging this inequality then yields $1 < N(\rho_1^*)/D(\rho_1^*) = \rho_1^* < \rho_2^* < \rho_3^*$ as a necessary condition for multiple fixed points ρ^*. □

Proof of Lemma 4

Fixed points to $g(\rho)$ as defined in (11) can be founds as roots to $f(\rho) = g(\rho) - \rho$. Such a fixed point ρ^* is then (asymptomatically) stable if $f'(\rho)|_{\rho=\rho_k^*} < 0$ (see Hirsch and Smale, 1974, pp. 185-188 and footnote 4.2). From the first part of the proof of Proposition 2 one gets that $f(0) = g(0) = N(0)/D(0) > 0$ holds. It follows that for f to have three roots, it has to cut the real line from above at ρ_1^*, from below at ρ_2^*, and again from above at ρ_3^*. This implies $f'(\rho_1^*) < 0$, $f'(\rho_2^*) > 0$, and $f'(\rho_3^*) < 0$ which proves stability of ρ_1^* and ρ_3^*, and that ρ_2^* is not stable. By the same reasoning, a unique root ρ^* has to obey $f'(\rho^*) < 0$ and is thus stable.

Proof of Lemma 5

Necessitiy: By property a) as derived above in the proof of Proposition 2, it follows that for a fixed point ρ^* which solves $g(\rho^*) = N(\rho^*)/D(\rho^*) > 0$, $N(\rho^*) > 0$ has to hold. This is equivalent to $\sigma_{sc} > \tau(\rho^*)$ where $\tau(\rho) = -\phi\sigma_s^2/(\mu + (1-\mu)\rho) < 0$ is defined for any $\rho > 0 \geq -\mu/(1-\mu)$ and for which $\tau'(\rho) > 0$ holds whenever $\rho \geq 0$. For $\rho^* > 0$ it therefore has to hold that $\sigma_{sc} > \tau^* \equiv \tau(\rho^*)$ with $\tau^* < 0$.

Sufficiency: To see that $\sigma_{sc} > \tau^*$ is also sufficient for (11) to have a solution $\rho^* > 0$, note that by the above reasoning $\sigma_{sc} > \tau(\rho^*) > \tau(0)$ holds and therefore, $Cov[s, m^*]|_{\rho=0} = N(0) > 0$ applies. Since $Var[m^*]_{\rho=0} = D(0) > 0$, it then follows that $g(0) = N(0)/D(0) > 0$. Together with continuity and a limit of zero of $g(\rho) = N(\rho)/D(\rho)$ as derived in properties a) and b) in the proof of Proposition 2, this means that there has to be at least one fixed point, i.e. at least one intersection of $g(\rho)$ with the 45-degree line over $\mathbb{R}_{++}$. □

Proof of Lemma 6

Necessitiy: From (11) one gets that$\rho^* \leq 1$, conditional on $\rho^* > 0$, holds if and only if

$$\left.\frac{\phi\sigma_s^2 + (\mu + (1-\mu)\rho)\sigma_{sc}}{\phi\sigma_s^2 + 2(\mu + (1-\mu)\rho)\sigma_{sc} + (\mu + (1-\mu)\rho)^2\sigma_c^2}\right|_{\rho=\rho^*} = \rho^* \leq 1$$

This condition simplifies to $\sigma_{sc} \geq -(\mu + (1-\mu)\rho^*))\sigma_c^2$ and becomes slacker for higher, positive values of ρ^*. Inserting $\rho^* = 1$, the upper bound on the desired value range, then yields $\sigma_{sc} \geq -\sigma_c^2$ as a necessary condition.

Sufficiency: To see that that this condition is also sufficient first note from (11) that, for any $\sigma_{sc} \geq 0$, $\rho^* \in (0, 1)$ follows immediately. Second, from the above reasoning it also follows immediately that $\sigma_{sc} = -\sigma_c^2$ implies $\rho^* = 1$. Now suppose $\sigma_{sc} \in (-\sigma_c^2, 0)$, i.e., $\sigma_c^2 = -\sigma_{sc}/\lambda$ for some $\lambda \in (0, 1)$. To show that then, $\rho^* < 1$ follows, suppose the opposite and substitute into (11) to get

$$\left.\frac{\phi\sigma_s^2 + (\mu + (1-\mu)\rho)\sigma_{sc}}{\phi\sigma_s^2 + (\mu + (1-\mu)\rho^*)\sigma_{sc} \cdot \left(2 - \frac{\mu + (1-\mu)\rho^*}{\lambda}\right)}\right|_{\rho=\rho^*} \geq 1.$$

Since the above denominator represents, in equilibrium, $Var[m^*] = D(\rho^*) > 0$ (see property a) and the b) in the proof of Proposition 2), this simplifies to

$$0 \geq (\mu + (1-\mu)\rho)\sigma_{sc} \cdot \left.\left(1 - \frac{\mu + (1-\mu)\rho}{\lambda}\right)\right|_{\rho=\rho^*}$$

Clearly, this is a contradiction as with $\sigma_{sc} < 0$, $\rho = \rho^* \geq 1$ and $\lambda \in (0,1)$, both of the above RHS's factors will be strictly negative. $\square$

Proof of Lemma 7

Upon disclosure, ψ increases. This enters (11) via an increase in $\phi = (1 - \psi(Corr[s,c])^2)/(1-\psi)$:

$$\frac{\partial \phi(\cdot)}{\partial \psi} = \frac{-(Corr[s,c])^2(1-\psi) + (1 - \psi(Corr[s,c])^2)}{(1-\psi)^2} = \frac{1-(Corr[s,c])^2}{(1-\psi)^2} > 0$$

Denote this increased value with $\tilde{\phi} > \phi$, which means that the function $k(\rho)$ as used in part 2 of the proof of of Proposition 2 also changes.* Denote this new function with $\tilde{k}(\rho)$ as follows:

$$\tilde{k}(\rho) = \underbrace{(1-\mu)^2\sigma_c^2}_{\equiv \tilde{A}} \cdot \rho^3 + \underbrace{2(1-\mu)(\sigma_{sc} + \mu\sigma_c^2)}_{\equiv \tilde{B}} \cdot \rho^2 + \underbrace{\tilde{\phi}\sigma_s^2 + \mu^2\sigma_c^2 + (3\mu - 1)\sigma_{sc}}_{\equiv \tilde{C}} \cdot \rho \underbrace{-\tilde{\phi}\sigma_s^2 - \mu\sigma_{sc}}_{\equiv \tilde{D}} \tag{A.6}$$

Comparing these coefficients to those of $k(\rho)$ as stated in (A.4), one then gets $\tilde{A} = A > 0$, $\tilde{B} = B$, $\tilde{C} > C$, and $\tilde{D} < D < 0$. Applying Descartes' sign rule again implies that there are either one or three roots to $\tilde{k}(\rho)$, and therefore to $g(\rho)$ after disclosure. Denote these roots to $g(\rho)$ before and after disclosure by ρ^* and $\tilde{\rho}^*$, respectively. Then, the following holds:

a) When there is a solution $\rho^* \in (0,1)$, it is unique and there is a unique solution $\tilde{\rho}^* \in (\rho^*, 1)$.

b) When there is a solution $\rho^* = 1$, it is unique and there is a unique solution $\tilde{\rho}^* = 1$.

c) When there is a unique solution $\rho_U^* > 1$, there is a unique solution $\tilde{\rho}^* \in (1, \rho^*)$.

d) When there are three solutions $\rho_k^* > 1$ with $k \in \{1,2,3\}$, there is either
   a unique solution $\tilde{\rho}^*$ such that $1 < \tilde{\rho}^* < \rho_1^* < \rho_2^* < \rho_3^*$ or there are
   three such solutions $\tilde{\rho}_k^*$ such that $1 < \tilde{\rho}_1^* < \rho_1^* < \rho_2^* < \tilde{\rho}_2^* < \tilde{\rho}_3^* < \rho_3^*$.

To proof the above, recall from Proposition 2 that $g(\rho)$, for which $k(\rho)$ and $\tilde{k}(\rho)$ denotes the roots under different levels of disclosure, has either one or three fixed points with any solution $\rho^* \in (0,1]$ being unique. Also note from (A.4) and (A.6) that these function relate to each other as follows: $\tilde{k}(0) < k(0) < 0$ and $\tilde{k}'(\rho) = 3\tilde{A}\rho^2 + \tilde{B}\rho + \tilde{C} > k'(\rho) = 3A\rho^2 + B\rho + C$ for all $\rho \in \mathbb{R}_+$. Furthermore, $\tilde{k}(\rho) = k(\rho)$ if and only if $\rho = 1$. It then holds that $k(\rho) > \tilde{k}(\rho)$ if $\rho \in (0,1)$ and $k(\rho) < \tilde{k}(\rho)$ if $\rho > 1$. This means that if there is a (unique) root $\rho^* \in (0,1)$ of k, there must be a unique root $\tilde{k}$ on $(\rho^*, 1)$ and if $\rho^* = 1$, $\tilde{\rho}^* = 1$ applies. To see that a root $\tilde{\rho}^* < 1$ is unique, one can repeat the same reasoning as in the second part of the proof of Proposition 2 to show that multiple solutions require all of them to have a value larger than one. This proves cases a) and b).

For case c), thus when there is a unique $\rho^* > 1$, the above reasoning implies that $\tilde{k}(1) = k(1) < 0$. A unique root of k at $\rho^* > 1$ implies that $g(\rho)$ never cuts the real line again on $(\rho^*, +\infty)$. Neither does then $\tilde{k}$ since $\tilde{k}(\rho) > k(\rho)$ for $\rho > 1$. This, in addition with $\tilde{k}(1) = k(1) < 0$, means that $\tilde{k}$ cuts the real line once over $(1, \rho^*)$ which proves the case.

Now consider case d), i.e that there are three positively-valued fixed points to $g(\rho)$. By Proposition 2, their coordinates have to obey $1 < \rho_1^* < \rho_2^* < \rho_3^*$. The continuous, cubic function k obeys $k(0) < 0$ (see the second part of the proof of Proposition 2). This implies that k cuts the real line

*One can, w.l.o.g. assume that $\tilde{\phi} > \phi = 1$ which then reflects the situation before disclosure with $\sigma_\epsilon^2 \to \infty$.

from below at ρ_1^*, from above at ρ_2^*, and again from below at ρ_3^*. Since it is a continuous polynomial, it has to have a local maximum and minimum in between these points. They are denoted by ρ_-^k and ρ_+^k, respectively so that

$$1 < \rho_1^* < \rho_+^k < \rho_2^* < \rho_-^k < \rho_3^*$$

holds. If $\tilde{k}$ also has three roots, denoted by $\tilde{\rho}_1^* < \tilde{\rho}_2^* < \tilde{\rho}_3^*$, it is a similarly-shaped polynomial by analogous reasoning. Therefore, $\tilde{k}$ cuts the real line from below at $\tilde{\rho}_1^*$, from above at $\tilde{\rho}_2^*$, and from below at $\tilde{\rho}_3^*$. From $\tilde{k}(1) = k(1) < 0$ and $\tilde{k}(\rho) > k(\rho)$ for $\rho > 1$, it follows that when $\tilde{k}$ cuts the real line from below (above), it has to do so at lower (higher) values than k. For three roots of $\tilde{k}$, this implies that

$$1 < \tilde{\rho}_1^* < \rho_1^* < \rho_2^* < \tilde{\rho}_2^* < \tilde{\rho}_3^* < \rho_3^*$$

which proves the second part of case d). If $\tilde{k}$ has only one root (two have been ruled out by the sign rule), $\tilde{k}(1) = k(1) < 0$ and $\tilde{k}(\rho) > k(\rho)$ again imply that it cuts the real line from below, i.e. at a lower value of ρ than for k. It follows that $1 < \tilde{\rho}^* < \rho_1^* < \rho_2^* < \rho_3^*$ which proves the first part of case d).

Now let the subscript $l \in \{1, 2, 3\}$ capture the above-described inference coefficients for potentially up to three equilbria. W.l.o.g., let $l = 1$ be used when such a coefficient is unique. Going over the above cases for all stable equilbria, that is for $l \neq 2$, disclosure leads to i) $1 > \tilde{\rho}_1^* > \rho_1^*$ if and only if $\rho^* < 1$, ii) $1 = \tilde{\rho}_1^* = \rho_1^*$ if and only if $\rho^* = 1$, iii) $1 < \tilde{\rho}_1^* < \rho_1^*$ and $1 < \tilde{\rho}_3^* < \rho_3^*$ if and only if $\rho^* > 1$. Using Lemma 6 to express these conditions on ρ^* via the game's fundamentals then yields the stated proposition. □

Proof of Lemma 8

The argument of rational receivers' (expected) utility is given by

$$\begin{aligned} z \equiv d_r^*(m) - s &= (1-\rho^*)\mathrm{E}[s] + \rho^*\left[m^*(s,c) - \mathrm{E}[c]\left(\mu + (1-\mu)\rho^*\right)\right] - s \\ &= -(s - \mathrm{E}[s]) + \rho^*\left[m^*(s,c) - \mathrm{E}[s] - \mathrm{E}[c]\left(\mu + (1-\mu)\rho^*\right)\right] \\ &= -(s - \mathrm{E}[s]) + \left(m^*(s,c) - \mathrm{E}[m^*(s,c)]\right)\rho^* \end{aligned} \tag{A.7}$$

Note that by (9), $m^*(s,c)$ is a linear transformation of the vector (s,c). By E1, it is therefore distributed according to F. Similarly, s is also distributed according to F. By the same argument, z is then distributed according to $F(\mathrm{E}[z], Var[z])$. Using $\sigma_z = \sqrt{Var[z]}$, one can normalize z via the linear transformation $\hat{z}(z) = z/\sigma_z - \mathrm{E}[z]$ such that $\hat{z}$ follows $F(0,1)$. The associated probability density function will be denoted $f(\hat{z})$. The expected utility of rational receivers can then be expressed as

$$\mathrm{E}[L(z)] = \int_{-\infty}^{+\infty} L\left(\mathrm{E}[z]\sigma_z + \hat{z}\sigma_z\right) f(\hat{z}) d\hat{z} \equiv V\left(\mathrm{E}[z], \sigma_z\right) \leq 0.$$

From (A.7) it follows that $\mathrm{E}[z] = 0$. Using $\sigma_z = \sqrt{Var[z]}$, one can define the univariate function $\mathcal{L}(\sigma_z) \equiv V(0, \sigma_z) \leq 0$ which denotes a rational receiver's expected utility with the first derivative

$$\mathcal{L}'(\sigma_z) = \frac{\partial V(\mathrm{E}[z], \sigma_z)}{\partial \sigma_z}\Big|_{\mathrm{E}[z]=0} = \int_{-\infty}^{+\infty} \left[\hat{z} \cdot L'\left(\hat{z}\sigma_z\right)\right] f(\hat{z}) d\hat{z}.$$

Because L is strictly concave and symmetric around zero, $\mathrm{sgn}[\hat{z}] = -\,\mathrm{sgn}\left[L'\left(\hat{z}\sigma_z\right)\right]$ holds. Therefore, the above derivative is negative (zero) for $\sigma_z > (=)\ 0$. From this then also follows that

$$\mathcal{L}''(\sigma_z) = \int_{-\infty}^{+\infty} \left[\hat{z}^2 \cdot L''(\hat{z}\sigma_z)\right) f(\hat{z}) d\hat{z} < 0.$$

To see that full disclosure is also necessary for $\mathcal{L}(0) = 0$ to hold, note from the above that this requires $\sigma_z = 0$ and therefore $d_r^*(m) = s$. Suppose that this held under imperfect disclosure. For $d_r^*(m) = s$ to apply in this case, (10) requires both, $\rho^* = 1$ and $c = \mathrm{E}[c]$ to hold simultaneously for *any* realization (s, c). This is a contradiction to the fact that under imperfect disclosure with $\psi \in (0,1)$, $Var[c|\tilde{c}] > 0$ and $Var[s|\tilde{c}] > 0$ applies (see Lemma 3).

For the alternative representation of the argument $Var[z]$ for the function $\mathcal{L}$, note that by using the definition of ρ^*, one gets the following:

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

From the law of total variance and using again the definition of ρ^*, it also holds that

$$\begin{aligned}
\mathrm{E}\left[Var[s|m^*]\right] &= Var[s] - Var[\mathrm{E}[s|m^*]] \\
&= \sigma_s^2 - \mathrm{E}[(d_r^*(m) - \mathrm{E}[s]])^2] \\
&= \sigma_s^2 - \mathrm{E}[((m^* - \mathrm{E}[m^*])\,\rho^*)^2] \\
&= \sigma_s^2 - (\rho^*)^2 Var[m^*] \\
= Var[z] &= \sigma_s^2 - \rho^* Cov[s, m^*] \\
&= \sigma_s^2 - \frac{Cov[s, m^*]^2}{Var[m^*]} \\
&= \sigma_s^2 \left(1 - Corr[s, m^*]^2\right) \geq 0
\end{aligned}$$

where $Corr[s, m^*] = Corr[s, m]_{m=m^*(s,c)} = Cov[s, m^*]/\left(\sigma_s \sqrt{Var[m^*]}\right)$. □

Proof of Proposition 3

Lemma 8 shows that the expected utility of rational receivers strictly increases in $Corr[s, m^*]^2$. For equilibria with $\rho^* > 0$, and therefore $Cov[s, m^*] > 0$, it is then sufficient to show that $Corr[s, m^*] > 0$ increases upon disclosure. For this note that

$$Corr[s, m^*] = \frac{Cov[s, m^*]}{Var[m^*]} \cdot \frac{\sqrt{Var[m^*]}}{\sigma_s} = \rho^* \cdot \frac{\sqrt{Var[m^*]}}{\sigma_s}. \tag{A.8}$$

First consider the case that $\sigma_{sc} > -\sigma_c^2$: From Lemma 7 and Lemma 6, this means that the equilibrium inference after disclosure (denoted by $\tilde{\rho}^*$) is larger than before, i.e., $1 > \tilde{\rho}^* > \rho^* > 0$. Also, the value of ϕ increases, i.e., $\tilde{\phi} > \phi \geq 1$ (see proof of Lemma 7).

Since the first factor on the RHS of (A.8) increases upon disclosure, it is then sufficient to show that also the second increases, i.e., that $D(\tilde{\rho}^*.\tilde{\phi}) > D(\rho^*, \phi)$ holds, with

$$D(\phi, \rho^*) = Var[m^*] = \phi\sigma_s^2 + 2(\mu + (1-\mu)\rho^*)\sigma_{sc} + (\mu + (1-\mu)\rho^*)^2\sigma_c^2.$$

While an increase in ϕ clearly increases the above term, the indirect effect via ρ^* is not that clear. However, from the fact that $\sigma_{sc} + \sigma_c^2 > 0$ is a necessary and sufficient condition for $\rho^* \in (0, 1)$ (see Lemma 6) it follows that in this case also $\sigma_{sc} + (\mu + (1-\mu)\rho^*)\sigma_c^2 > 0$ holds and therefore

$$\partial D(\rho^*, \phi)/\partial \rho^*|_{\rho^* \in (0,1)} = 2(1-\mu) \cdot \left(\sigma_{sc} + (\mu + (1-\mu)\rho^*)\sigma_c^2\right) > 0.$$

Now consider the case of $\sigma_{sc} = -\sigma_c^2$: By Lemma 7 and Lemma 6, means that $\tilde{\rho}^* = \rho^* = 1$ holds before and after disclosure. While this term increases, $\tilde{\phi} > \phi$ holds nevertheless. Therefore, $Corr[s, m^*]$ increases.

Finally, consider the case of $\sigma_{sc} < -\sigma_c^2$: From Lemma 7 and Lemma 6 it then follows that for such equilibria, $1 < \tilde{\rho}^* < \rho^*$ holds, thus disclosure *decreases* ρ^* while ϕ still increases. While an increase in ϕ increases $Corr[s, m^*]$ (see above), the effect of a decrease in ρ^* is not that clear. To examine this effect, first note that in this case, the following holds:

$$\partial D(\rho^*, \phi)/\partial \rho^*|_{\rho^* > 1} = 2(1-\mu) \cdot \left(\sigma_{sc} + (\mu + (1-\mu)\rho^*)\sigma_c^2\right) < 0$$

Again, the inequality follows from the fact that $\sigma_{sc} + \sigma_c^2 \leq 0$ is a necessary and sufficient condition for $\rho^* > 1$ (see Lemma 6) which implies $\sigma_{sc} + (\mu + (1-\mu)\rho^*)\sigma_c^2 < 0$. Using that $\rho^* = N(\rho^*, \phi)/D(\rho^*, \phi)$ with $N(\rho^*, \phi) = \sigma_s^2 + (\mu + (1-\mu)\rho^*)\sigma_{sc}$ means

$$\frac{\partial Corr[s, m^*]}{\partial \rho^*}\Big|_{\rho^* > 1} = \partial \left(\frac{N(\rho^*, \phi)}{D(\rho^*, \phi)}\right) \Big/ \partial \rho^* \cdot \frac{\sqrt{D(\rho^*, \phi)}}{\sigma_s} + \frac{N(\rho^*, \phi)}{D(\rho^*, \phi)} \cdot \frac{\partial D(\rho^*, \phi)/\partial \rho^*}{2\sigma_s\sqrt{D(\rho^*, \phi)}}.$$

Multiplying the above with $\sigma_s\sqrt{D(\rho^*, \phi)} > 0$, re-substituting ρ^*, and simplifying then yields

$$\begin{aligned} \operatorname{sgn}\left[\frac{\partial Corr[s, m^*]}{\partial \rho^*}\Big|_{\rho^* > 1}\right] &= \operatorname{sgn}\left[\frac{\partial N(\rho^*, \phi)}{\partial \rho^*} - \frac{\rho^*}{2} \cdot \frac{\partial D(\rho^*, \phi)}{\partial \rho^*}\right] \\ &= \operatorname{sgn}\left[\sigma_{sc} - \rho^*(\sigma_{sc} + (\mu + (1-\mu)\rho^*)\sigma_c^2)\right] \end{aligned}$$

Multiplying the above with $D(\rho^*, \phi) > 0$ and substituting ρ^* with the RHS of (11) at $\rho = \rho^*$ then yields, after some transformations that the sign of the above equals

$$\operatorname{sgn}\left[\sigma_{sc}^2 - \sigma_c^2\sigma_s^2\right] = \operatorname{sgn}\left[(Corr[s, c])^2 - 1\right] < 0.$$

Given the decrease to $\tilde{\rho}^* < \rho^*$, $Corr[s, m^*]$ therefore increases upon disclosure. □

Proof of Proposition 4

I start with $w_k = 0$ and denote, with slight abuse of notation, $W(\psi) \equiv W(\rho^*(\phi(\psi, \cdot), \cdot))$ via the analogously defined $\mathrm{E}[u_r^R(\psi)] \equiv \mathrm{E}[u_r^R(\rho^*(\phi(\psi, \cdot), \cdot))]$ and $\mathrm{E}[u_n^R(\psi)] \equiv \mathrm{E}[u_n^R(\rho^*(\phi(\psi, \cdot), \cdot))]$. This reflects that the effect of disclosure, as measured via an increase in ψ affects ρ^* via ϕ (i.e., via $\partial\phi(\psi, \cdot)/\partial\psi > 0$, see the proof of Lemma 7). To determine this effect, consider (11) and let $N(\rho^*)$

and $D(\rho^*)$ denote its RHS to get the following:

$$\frac{\partial \rho^*(\phi,\cdot)}{\partial \phi} = \frac{\partial \left(N(\rho^*)/D(\rho^*)\right)}{\partial \phi} = \frac{\sigma_s^2 D(\rho^*) - N(\rho^*)\sigma_s^2}{D(\rho^*)^2} = \frac{(1-\rho^*)\sigma_s^2}{D(\rho^*)}$$

From the fact that ϕ increases in ψ, it follows that

$$\operatorname{sgn}\left[\frac{\partial \rho^*(\phi(\psi,\cdot),\cdot)}{\partial \psi}\right] = \operatorname{sgn}\left[\frac{\partial \rho^*(\phi,\cdot)}{\partial \phi}\right] = \operatorname{sgn}\left[1-\rho^*\right]$$

Since $\frac{\mathrm{E}[u_n^R(\rho^*)]}{\partial \rho^*}$ is negative and ρ^* increases (decreases) upon disclosure if and only if $\rho^* < 1$ ($\rho^* > 1$) one then gets the following:

$$\operatorname{sgn}\left[\frac{\mathrm{E}[u_n^R(\rho^*(\phi,\cdot),\cdot))]}{\partial \psi}\right] = \operatorname{sgn}\left[\frac{\mathrm{E}[u_n^R(\rho^*)]}{\partial \rho^*} \cdot \frac{\partial \rho^*(\phi,\cdot)}{\partial \phi}\right] = \operatorname{sgn}\left[\rho^* - 1\right]. \tag{A.9}$$

When $\rho^* \in (0,1)$, every increase in ϕ therefore hurts naive receivers. In contrast, it has been shown that when there is full disclosure, i.e. $\phi = 1$, rational receivers achieve their maximum utility.

The first part of the proposition (that full disclosure is never optimal) can then be established by showing the following: When $\rho^* \in (0,1)$, there exists a $\Delta > 0$ such that starting from full disclosure with $\phi = 1$, a gradual decrease of disclosure to $\psi = 1 - \Delta$ increases $W(\psi) = w_r \cdot \mathrm{E}[u_r^R(\psi)] + w_n \cdot \mathrm{E}[u_n^R(\psi)]$. This is equivalent to showing that $\lim_{\Delta \to 0^+}\left(W(1) - W(1-\Delta)\right)$ is negative, i.e. that

$$\begin{aligned}
\operatorname{sgn}\left[\lim_{\Delta \to 0^+}\left(\frac{W(1) - W(1-\Delta)}{\Delta}\right)\right] &= \operatorname{sgn}\left[\sum_{j=r,n} w_j \cdot \lim_{\Delta \to 0^+}\left(\frac{\mathrm{E}[u_j^R(1)] - \mathrm{E}[u_j^R(1-\Delta)]}{\Delta}\right)\right] \\
&= \operatorname{sgn}\left[w_r \cdot \left.\frac{\partial \mathrm{E}[u_r^R(\psi)]}{\partial \psi}\right|_{\psi=1} + w_n \cdot \left.\frac{\partial \mathrm{E}[u_n^R(\psi)]}{\partial \psi}\right|_{\psi=1}\right] \\
&= \operatorname{sgn}\left[w_n \cdot \left.\frac{\partial \mathrm{E}[u_n^R(\psi)]}{\partial \psi}\right|_{\psi=1}\right] = \operatorname{sgn}\left[\rho^* - 1\right] < 0
\end{aligned}$$

holds. The second-last equality in the above follows from the fact that by Lemma 8, rational receivers expected utility w.r.t to ψ is maximized under full disclosure, i.e. $\frac{\partial \mathrm{E}[u_r^R(\psi)]}{\partial \psi}\Big|_{\psi=1} = 0$, while the last one follows from (A.9).

For the case that $w_k > 0$, note that the above proof applies for any loss function $u_n^R(\cdot) = L(\cdot)$ which is strictly concave and symmetric around zero. It therefore also holds when in addition to $\mathrm{E}[u_n^R(\sigma_\psi^2)]$, positive weight is assigned to $-\mathrm{E}[c(\mu + \rho^*(phi(\psi,\cdot),\cdot)(1-\mu))^2]$. This then yield the first part of the proposition.

The second statement is an immediate consequence of the fact that when $\rho^* > 1$, according to (A.9), increasing the level of disclosure ψ helps naive receivers and that full disclosure maximizes the utility of rational receivers (see Lemma 8). □

Example for non-disclosure to be optimal

As a concrete example for a scenario where non-disclosure is optimal, consider the parameters $\sigma_s^2 = \sigma_c^2 = 1$, $\bar{s} = \bar{c} = \sigma_{sc} = 0$, together with $\mu = w_n = w_r = 0.5$, $w_k = 0$, and the loss function $L(d-s) = -(d-s)^2$. Plugging these parameters into (11) and solving yields $\rho^* \approx 0.6$. Following

Lemma 7, disclosure then increases this inference coefficient. Using Proposition 1 on (13) yields

$$
\begin{aligned}
W &= -0.5\left(\mathrm{E}[(\rho[m^*(s,c)-\bar{c}(\mu+(1-\mu)\rho)]+(1-\rho)\bar{s}-s)^2]+\mathrm{E}[(m^*(s,c)-s)^2]\right)\Big|_{\rho=\rho^*} \\
&= -0.5\left(\mathrm{E}[(\rho m^*(s,c)-s)^2]+\mathrm{E}[(m^*(s,c)-s)^2]\right)\Big|_{\rho=\rho^*} \\
&= -0.5\left(\mathrm{E}[(s(\rho-1)+c\rho(0.5+0.5\rho))^2]+\mathrm{E}[(c(0.5+0.5\rho))^2]\right)\Big|_{\rho=\rho^*} \\
&= -0.5\left((\rho-1)^2\mathrm{E}[s^2]+2(\rho-1)\rho(0.5+0.5\rho)\mathrm{E}[sc]+(\rho^2+1)(0.5+0.5\rho)^2\mathrm{E}[c^2]\right|_{\rho=\rho^*} \\
&= -0.5\left((\rho-1)^2+(\rho^2+1)(0.5+0.5\rho)^2\right)\Big|_{\rho=\rho^*}
\end{aligned}
$$

which is easily verified to be strictly decreasing in ρ if $\rho > 0.4$. Therefore, W is maximized by non-disclosure.