Gesche, Tobias

**Working Paper**

# De-biasing strategic communication

Working Paper, No. 216

**Provided in Cooperation with:**
Department of Economics, University of Zurich

*Suggested Citation:* Gesche, Tobias (2021) : De-biasing strategic communication, Working Paper, No. 216, University of Zurich, Department of Economics, Zurich, https://doi.org/10.5167/uzh-121321

This Version is available at:
https://hdl.handle.net/10419/243113

University of Zurich

Department of Economics

Working Paper Series

ISSN 1664-7041 (print)
ISSN 1664-705X (online)

---

Working Paper No. 216

# De-Biasing Strategic Communication

Tobias Gesche

Revised version, September 2021

---

# De-Biasing Strategic Communication

**Tobias Gesche**
ETH Zurich*

September 2021

version accepted for publication in
*Games and Economic Behavior*

**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 and, 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 the 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, 8092 Zurich, Switzerland. Email: tgesche@ethz.ch

*I thank Nageeb Ali, Roman Inderst, Navin Kartik, Igor Letina, Ming Li, David Myatt, Nick Netzer, María Saéz-Martí, Deszö Szalay, Adrien Vigier, and Shihong Xiao for helpful discussions on previous versions. I also thank two anonymous referees and the advisory editor for helpful comments.*

# 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. This can hurt receivers of such information through two main channels. Firstly, some receivers may naively ignore the expert's COI and make poor choices by following biased messages. Secondly, other receivers might be aware of the COI but lack information about it, such as the COI's 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, so that 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 a 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. Disclosure is also an appealing option from a regulatory perspective as it is less paternalistic and less costly to regulators than direct supervision and de-biasing.[1]

The objective of this paper is to describe an economic mechanism which shows how disclosing COIs can often lead to consequences which are opposite to those intended. It does so by considering a communication game where the sender's private type is two-dimensional. This type consists of the superior information the sender has 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., a managed fund). The combination of these factors then unveils a simple economic mechanism through which disclosure can lead to more biased communication. This then hurts naive receivers who do not anticipate these 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, for example, via sales commissions. When commenting on the asset, he then faces a COI to overstate its value. The magnitude of this

[1] A prominent example of such a policy is contained in the Sarbanes-Oxley-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). Disclosure rules are also common to address scientific fraud and researchers' COIs (see Fanelli, 2009; Steen, 2011; Simonsohn, 2014). Fung et al. (2007) show further instances of disclosure rules and why they seem appealing (alongside with examples for their failure). This paper's framework can be applied to study their consequences.

bias is determined by equalizing the marginal costs of lying to the marginal return of doing so. The latter is given by the average marginal reaction of receivers (e.g., 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. Such de-biasing of the sender's message can then worsen things for the receiver when the expected bias differs from its actual value. Facing such strategic uncertainty, rational receivers will then act based on 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. 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 strategic uncertainty and, therefore, increases this weight. Thus, disclosure 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 then also increases with disclosure. Delegating or naive receivers who do not account for these strategic effects of biasing and de-biasing communication are then hurt by this increase in the bias following disclosure.

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 cannot be utilized 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 always a Pareto-improvement among receivers, except if all of them are fully rational. In fact, 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 adverse mechanism manifests and which allow to evaluate 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 when it is negative, there can be situations in which disclosure is a Pareto-improvement among *all* receivers and that full disclosure is only efficient in these cases.

**Contribution to the literature:** On the empirical side, the findings by Malmendier and Shanthikumar (2014) relate closely to this paper. They show that financial analysts *strategically* inflate their stock recommendations by tailoring it to the receivers' reactions. This feature is maintained in the following analysis. Malmendier and Shanthikumar use data covering a period before and after the Sarbanes-Oxley-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. Here, I 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 payment for the experts is based on the accuracy of the final estimates by clients, their advice and the resulting estimates are better than when the experts' payment is based on how high the 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 and end up making worse decisions than without disclosure. These findings on the adverse effects of disclosure have also been replicated (Koch and Schmidt, 2010; Inderst et al., 2010; Cain et al., 2011).[3] This paper's results are in line with them.

The present work 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, i.e., when lying costs are absent. In such equilibria, the sender's message identifies a partition of the state space while the magnitude of his commonly known COI steers the coarseness of communication as it limits the number of such partitions in the most informative equilibrium. This result applies independently of the specific meaning of language (i.e., how exactly states map to messages and back from messages into actions by receivers), as long as this mapping is common knowledge.[4] Often, however, the meaning of language is determined by the circumstances. For example, if a financial analyst announces "I expect share X will yield return Y

[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, when scandals about misleading advice by affiliated analysts became public and which contributed to the enactment of the Sarbanes-Oxley-Act shortly afterwards.

[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 inter-personal, psychological channels, which typically evolving within relatively close expert-client relationships. The results on the negative effects of disclosure presented in this paper do not require such close relationships – they also apply if expert and receiver interact very remotely and indirectly (such as in a market setting).

[4] See Sobel (2013) for an overview of the rich literature which has utilized and extended the partitioning result.See also the section on pragmatics therein for a further discussion on language and its meaning in the context of strategic communication. In Sobel (2020), he provides a classification of lying and deception in this context.

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 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 underlying 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, the absence of lying costs.[5] Capturing these insights, this paper assumes unbounded support, such as when the sender's type is normally distributed. It also assumes lying costs that grow (quadratically) in the size of the lie and naive receivers who follow the sender's message at face value. This connects the present work to Gordon and Nöldeke (2015), who also assume a quadratic-normal framework and naive receivers. However, they do not focus on disclosure but on how communication equilibria can yield different figures of speech (such as exaggeration, understatement, or irony) with differing levels of informativeness. They assume that receivers and the sender both have concave objective functions around, typically different, bliss points for receiver actions. Deviations from the respective bliss points are therefore increasingly costly – also for the sender. Under this assumption, naive receivers have similar effects as lying costs because senders are hurt if the over-react to too inflated messages. Similar points have also been observed in related settings by Kartik et al. (2007) and Ottaviani and Squintani (2006). This is different here, as given his COI, the sender's preferences are monotone in the receivers' actions.[6]

The above works also differ from the present one by assuming a one-dimensional sender type, where the his COI is common knowledge. In order to examine the consequences of disclosing it, I allow it to be the sender's private information. By assuming a two-dimensional sender type with an unknown COI, this work relates to Morgan and Stocken (2003). Within a cheap talk framework,

[5] Kartik et al. (2007) show for one-dimensional sender types 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 on the state space and lying costs. Kartik (2009) considers a compact state space with lying costs which enables equilibria of the "LSHP (low types separate and high types pool"-form: An upward-biased sender exaggerates via an invertible (and, therefore revealing) messaging strategy if the state is below a certain threshold. If the state surpasses the threshold, all such sender types pool and send a message that does not allow to extract further information.

[6] With a "bliss point"-utility for the upward-biased sender and the co-presence of naive and rational receivers, the sender wants to affect the latter actions by sending messages "*so inflated that the credulous receivers are deceived to take an action that is even higher than the one that is ideal for the sender*" (Kartik et al., 2007, p.96). Related to this, Ottaviani and Squintani (2006) show in a cheap talk setting that the presence of naive receivers can partially remove the partitional character of and result in equilibria which resemble those that emerge with lying costs (i.e., similiar to the LSHP-equilbira described in Kartik, 2009, see footnote 5 above). In the current paper, this is different as naive receivers do not limit the lie in the sender's message. Rather, the size of the lie typically increases with the share of naive receivers. The reason for this difference is that I do not assume a bliss point for the sender over receiver actions.

they find that the messaging strategy remains partitional when the sender's COI is described by a binary random variable. Using this setup, Li and Madarasz (2008) provide a first account of how disclosure can backfire. They show that communication can be more efficient when the *expected* COI is moderate, compared to when it is revealed to be either high or low. The intuition behind this result is that in the cheap talk framework they assume, communication becomes too coarse when the COI is known and large in magnitude. They also assume the sender to have a bliss point for receiver actions so that too coarse messages do then often not allow the sender to steer the receivers' actions close enough to his bliss point. Non-disclosure and with it, more precise communication, is then often better for *all* players (at least ex-ante and for players having sufficiently concave utility). The mechanism presented here is different. In particular, it does not rely on the sender having concave utility. Inderst and Ottaviani (2012) also feature two-dimensional sender types and model COIs as a bonus paid by producers to intermediaries who advise customers on which out of two products to choose; information and messages are therefore both binary. Disclosing COIs then reduces the provision of commissions but less so, in relative terms, for the inferior product. Consequently, the relative bias rises after disclosure and consumers make worse decisions. Here, I present a different channel for how disclosure can be harmful when COIs are not binary and communication is smooth. I demonstrate how in such a framework, disclosure fulfills the aim of enabling better information transmission to rational receivers. However, by doing so it amplifies a negative externality on naive ones. In particular, I show how lying costs – which are not considered in the above accounts of disclosure – play a crucial role in triggering this adverse effect.

I establish these findings in a framework of strategic communication where both, the state of the world and the sender's COI are represented by continuous variables. This connects closely to Fischer and Verrecchia (2000) who use this approach to study a manager who gets linear utility from influencing his company's share price through earnings announcements while facing quadratic costs of misreporting (an application that is, among others, also admissible here). Assuming that the manager's strategic incentives are uncorrelated with the state of the world, they show that decreasing uncertainty over the variable describing the sender's motives leads to an increase in the sender's lie. Frankel and Kartik (2019), as a part of their general treatment of projection-based signaling (i.e., one-dimensional signals sent by two-dimensional sender types), also consider the linear-quadratic framework and have similar results when allowing this correlation to be positive.[7] This work focuses

[7]Technically related, Bénabou and Tirole (2006) study projection of pro-social motivation in a linear-quadratic-normal framework. Frankel and Kartik (2018) use a similar setup to study how the information quality of a central bank about its type (the inflation target and real shocks) affects the signal it sends to the public via its (one-dimensional) monetary policy. Blume and Board (2013) and Giovannoni and Xiong (2019) look at multi-dimensional uncertainty with respect to players' language competence, the degree to which they are able to send and/or understand distinct messages. They show that removing such uncertainty often leads to inefficient communication. While Blume and Board (2013) focus on common-interest games, Giovannoni and Xiong (2019) show that this can also hold when there is a COI. However, they do not regard uncertainty with regard to a sender's COI (see also Footnote 4 in Blume and

on the effects of disclosing the sender's COI. To study these effects rigorously, it extends the analysis of communication games in this framework along three main dimensions.

First and foremost, this work incorporates the co-presence of rational and naive receivers, including their strategic effect on the sender's messaging strategy. This allows to measure the opposing effects of disclosure: On the one hand, there is the increase in the message's informativeness from which rational receivers benefit. On the other hand, there are the costs of the associated increase in the sender's bias: the deviation from a honest, literally meant message on which naive receivers rely. Second, this work studies the connection between disclosure and the correlation of the sender's COI with the state of the world. In doing so, it does explicitly consider the effect of disclosure when this correlation is negative. This does not only allow to capture several realistic settings (e.g., financial markets, see Appendix B) – it also shows that the effect of disclosure is not monotone: with a negative correlation, disclosure can be a Pareto-improvement among *all* receivers, including naive ones. This implication would be overlooked if one focused only on the non-negative case. Third and closely related to the preceding point, this paper also adds to the literature by explicitly modeling 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 essential in fully evaluating its consequences.

# 2 The model

## 2.1 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 an action $x \in S$. The payoff resulting from that action depends on how well it matches the realization of $s$. For example, $s$ might represent an asset's return and $x$ the receiver's optimal position into this asset. A receiver then suffers a loss which is the greater, the more $x$ and $s$ are misaligned. This is captured by her ex-post utility

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

where $L$ is a $\mathcal{C}^2$, strictly concave loss function that is symmetric around its of maximum $L(0)$. Different curvatures of $L$ can then capture different levels of risk aversion. The canonical example is the quadratic loss function with $u^R(x, s) = -\frac{1}{2}(x - s)^2$ put forward by Crawford and Sobel (1982)

Board, 2013, for a discussion of the relation between uncertainty regarding language competence and a sender's COI).

and used in much of the literature on strategic communication.[8]

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 functions $x_r(m)$ and $x_n(m)$ for rational and naive receivers' (re)actions, respectively. These functions 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 $1 - \mu$. The receivers' aggregate action can then be represented as follows:

$$X(m) = \mu x_n(m) + (1 - \mu) x_r(m) \tag{2}$$

Since the sender's message $m$ is supposed to reflect $s$, there is meaning in his message and stating $s$ falsely creates costs which are measured by $\frac{1}{2}(m - s)^2$. This functional form can capture lying costs based on social preferences, moral concerns against lying, or reputational concerns.[9] 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.

However, such strong influence of the sender on the receivers' actions can be exploited. The sender can face an incentive to induce either high or low actions. Such a COI of the sender manifests through an additional payoff $cX(m)$ with $c \in C \subseteq \mathbb{R}$. If $X(m)$ denotes aggregate demand, a value of $c > 0$ represents an incentive to generate high receiver actions (e.g., high demand through sales commissions). Conversely, $c < 0$ means that low actions are rewarded (e.g., when the sender wants to temporarily decrease the price of an asset because he would like to take a position in it). The magnitude of $c$ then denotes the strength of such incentives, relative to given lying costs. The sender's expected utility is therefore given by

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

Thus, the sender wants to influence $X$ through $m$ as much as possible up or down (depending

[8] Ottaviani (2000) shows that this specific function covers the case of a receiver with exponential utility who invests $x$ into a risky asset of which she knows its variance but not its expected value $s$.

[9] This cost function can also capture concerns for the utility of a receiver who follow the sender's message at face value, such as naive ones. Kartik (2009) uses it as a prominent example of lying costs (see, e.g., Erat and Gneezy, 2012; López-Pérez and Spiegelman, 2012; Abeler et al., 2014); Abeler et al. (2019) provide a recent meta-study on the determinants of lying costs. It can also proxy reputational costs (such as in Sobel, 1985; Morris, 2001): if ex-post, receivers learned $s$ and they used it to regress $s$ on the sender message, $(m-s)^2$ would negatively enter the associated coefficient of determination ($R^2$); the sender's credulity is thus decreasing in this squared distance. Frankel and Kartik (2019) use costs which are quadratic in how much the sender overstates 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 their footnote 19). I do not allow free downward deviation as in the settings considered here, both, an upward and a downward-bias can be considered as not telling the truth.

on $c$'s sign), subject to his lying costs. This setup yields an intuitive form for the sender's optimal message. To see this, suppose that $X(m)$ is a linear function. His optimal message is then

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

where $X'(m)$ is a constant. Thus, 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 $c$ times $X'(m)$, the receivers' reaction to the sender's message. For example, if no-one listens to the sender so that $X'(m) = 0$ holds, there is no point of lying and the bias equals zero. More generally, this reflects that, in the 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.

Also note that the message is strictly increasing in $s$. If $c$ was commonly known (I will later treat this as a special case), the receiver could invert the messaging strategy and recover $s$. Accordingly, the sender's message would be biased but Bayesian, rational receivers would not be hurt by it.[10] In the following, I will relax two assumptions, which allowing biased messages to be harmless: First, I will allow that $c$ is the sender's private information. The message $m$ is then a projection $S \times C \to M$ and rational receivers face strategic uncertainty when they try to recover $s$ from it. Second, I will also assume some receivers to be naive in the sense that they follow the sender's 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 even backfire.

## 2.2 Information structure

The state of the world $s$ and the COI $c$ are the sender's private information, from a multivariate normal distribution $\mathcal{N}(\boldsymbol{\eta}, \boldsymbol{\Sigma})$ with support $S \times C = \mathbb{R}^2$. The vector $\boldsymbol{\eta}$ represents the expected values while $\boldsymbol{\Sigma}$ denotes the variance-covariance matrix with finite, real-valued elements:[11]

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

[10] See Kartik et al. (2007) who establish this "biased but revealing"-insight in a related framework. Adapting their notation to the current one, that is $s$ corresponds to their $x$ and $X(m)$ to their $\hat{x}$ (the receiver action which the sender wants to influence), their assumption A.4 corresponds to $\partial^2 u^S(s, X(m), m)/(\partial s \partial X(m)) < 0$. This is easily verified to be violated here. Nevertheless, messages are biased but yet revealing (w.r.t. $s$ if $c$ is known).

[11] Note that while assuming normality implies unbounded support, the probability that realizations of $(s, c)$ are within some compact set can be made arbitrarily high. Also note that while the results here are, for the sake of an easier exposure, stated for the normal distribution, all essential steps can also be derived if the sender's type has a elliptical distribution. Examples for this include the heavier-tailed Laplace or Student-t-distribution (which are often used in financial and risk modeling, see Embrechts et al., 2002), the logistic distribution (which is often used to model latent processes underlying discrete outcomes). For recent related applications, see Deimen and Szalay (2019) and Frankel and Kartik (2019); Gómez et al. (2003) provide a technical overview.

When appropriate, I will refer to the correlation $Corr[s, c] = \frac{\sigma_{sc}}{\sigma_s \sigma_c}$. To make things interesting, I also assume $\sigma_s \sigma_c > 0$ and $|Corr[s, c]| < 1$ as otherwise, the receiver's inference problem would become effectively one-dimensional or vanish entirely (i.e., $\mathbf{\Sigma}$ is positive definite). I will refer to $\sigma_c^2$ as "strategic uncertainty". This name reflects that this parameter describes receivers' uncertainty with regard to the variable that shapes the sender's strategic motives when communicating. Uncertainty over this dimension thereby confounds inference over the value of $s$, in which receivers are ultimately interested. I will therefore refer to $\sigma_s^2$ as "fundamental uncertainty" as without such uncertainty, strategic uncertainty would not matter. Importantly, this framework also allows to handle the case when $s$ and $c$ are positively or negatively correlated, which is relevant, for example in financial markets (see Appendix B for an example of how $\sigma_{sc} \neq 0$ can arise in this context).

### 2.3 Rational and naive receivers

As shown in (4), 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, her action should be given by $x_r(m) = \arg\max_{x \in S} \mathrm{E}[L(x - s)|m]$. If $m$ and $s$ are jointly normally distributed and a rational receiver has a quadratic loss function, $x_r(m)$ would be her expectation over $s$, given the information $m$ she received in this regard. The following result shows that this also holds for the more general loss function $L$:

**Lemma 1.** *If $m$ and $s$ are jointly normally distributed, rational receivers choose $x_r(m) = \mathrm{E}[s|m]$.*

(The proof uses elements of Lemma 1's proof in Deimen and Szalay (2015); see Appendix A)

The optimal action $x_r(m)$ is for a fully rational, Bayesian receiver who is capable of updating her prior while adjusting for the effect of the sender's COI. In particular, Malmendier and Shanthikumar (2007) show that small investors such as private households follow analysts' optimistic recommendations more closely than bigger, institutional investors who behave more cautiously. To capture these observations, I allow for the possibility that share $\mu \in [0, 1)$ of the receivers are naive and take the sender's signal at face value. Their action is thus given by $x_n(m) = m$.[12] In consequence, the receivers' aggregate action (2) becomes

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

[12] By appropriate scaling of $\mu$, one can always account for situations where naive or delegating receivers do not react one-to-one, e.g. when $x_n(m)$ is a positive affine transformation with $x_n'(m) = r > 0$. As an example, suppose that there is a mass $0.5$ of naive receivers for whom, on average, $x_n(m) = 0.6m$ 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, a 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.

### 2.4 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 $\epsilon$ is an error term, which is jointly but independently distributed with $(s, c)$, has an expected value of zero, and variance $\sigma_\epsilon^2$. 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 random vector $(s, c, \epsilon)$ is drawn,

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 $s$, and then choose their action $x$.

## 3 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 $x_r^* : M \to S$ for rational receivers such that each player's expected utility is maximized, given the other players' strategies when 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 1, they choose $x_r^*(m) = \mathrm{E}[s|m^*] \equiv \mathrm{E}[s|m]|_{m=m^*(s,c)}$.[13] 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 $\tilde{c}$.

### 3.1 Updating after disclosure

It will be useful to express disclosure via the parameter $\psi$ defined as follows:

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

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 the degree of disclosure. 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 $\psi \in (0, 1)$ reflects imperfect disclosure ($\sigma_\epsilon^2 \in \mathbb{R}^+$). Thus, $\psi$ is key for the resulting posterior:

[13] 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.

**Lemma 2.** *The posterior distribution of* $(s, c \mid \tilde{c}\,)$ *is given by* $\mathcal{N}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}})$ *with*

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

(Proof in Appendix A)

First note that in the case of no disclosure ($\psi = 0$), the posterior equals the prior. If there is disclosure ($\psi > 0$), the signal $\tilde{c}$ directly affects the expected value of $c$, which is used to debias the received message. In consequence, it also reduces strategic uncertainty $\sigma_c^2$ 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 $\psi$). These effects would be overlooked if disclosure were modeled as an uni-variate comparative static over $\sigma_c^2$. Also note that full disclosure ($\psi = 1$) is a special case of the above: all but the first element in $\hat{\boldsymbol{\Sigma}}$ become zero while $c$ is known so that the sender's private type becomes effectively one-dimensional.

### 3.2 Equilibrium behavior

The sender has to take the above-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

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

where $x_r^{*\prime}(m) = \frac{\partial \mathrm{E}[s|m]}{\partial m}|_{m=m^*(s,c)}$. Thus, the sender's messaging strategy is not only determined by its effect on naive receivers and lying costs. It is also based on how strongly rational receivers react to the sender's (biased) equilibrium message, as captured by $x_r^{*\prime}(m)$.

To derive how rational receivers – and in response also senders – behave optimally, I define the "equilibrium inference coefficient" $\rho^*$. This parameter captures how well, given a sender's equilibrium messaging strategy $m^*(s, c)$ and rational receivers' posterior $\mathcal{N}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}})$, variations in equilibrium message $m^*$ capture variations in the underlying state of the world $s$:

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

Throughout this paper, I focus on the case that in equilibrium, $\rho^*$ is a real, strictly positive number. This precludes situations where the expert's message is completely uninformative ($\rho^* = 0$). It also does not cover 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 information market with experts, especially when this is an equilibrium feature.[14]

Now assume that the rational receivers' reaction is a linear function of the message. One can then show that, in equilibrium, the inference coefficient has to be equal to slope of their marginal reaction function (i.e., that $x_r^{*\prime}(m) = \rho^*$ holds). Proving this relationship constitutes the main building block for characterizing the players' equilibrium actions and relevant equilibrium beliefs:

**Proposition 1a.** *Every pure strategy Perfect Bayesian Equilibrium of the communication game with strategies $m^*(s,c)$ for the sender and linear strategies $x_r^*(m)$ for rational receivers takes the form of*

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

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

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

(Proof in Appendix A)

Note from (5) that assuming $x_r^*(m)$ to be linear means that the aggregate action $X(m)$ is linear too. It also means that equilibria are smooth – small changes in the sender's message and his information do not disproportionately affect receiver actions. Most importantly, however, it allows to demonstrate the main point of this paper, the adverse effects of disclosure. For this, consider the interplay of the sender' messaging and rational receivers' de-biasing strategies:

In equilibrium, the sender's message is the state of the world plus a bias, given by $m^*(s,c) - s = c\,(\mu + (1-\mu)\rho^*)$. This bias equals $cX'(m)$, the change in the aggregate receiver action to a message, scaled by $c$. Part of $X'(m)$ is the change in rational receivers' best response, given by $x_r^{*\prime}(m) = \rho^*$, weighted by their share $1-\mu$. This reflects that this best response $x_r^*(m)$ has two parts. The first is invariant to $m$. It equals the prior $\mathrm{E}[s]$ with weight $1-\rho^*$. The other part of rational receivers' best response is given by $m - \mathrm{E}[c]\,(\mu + (1-\mu)\rho^*)$, the received message corrected by the expected bias, weighted with $\rho^*$. Therefore, this coefficient measures how strongly rational receivers react to $m$. Figure 1 displays the interplay between the sender's message and rational receivers' associated inference and reaction through $\rho^*$; this will be crucial in the following.[15]

As a first step in the analysis of $\rho^*$, note that the correction by rational receivers is based on the *expected* commission. It can therefore be wrong in both, direction and magnitude. This is why they do often not react one-to-one to the corrected message: whenever $\rho^* \in (0,1)$, rational receivers

[14] In a similar (but not identical) setup, Gordon and Nöldeke (2015) analyze "ironic" communication equilibria where $m^*$ and $s$ are negatively correlated.

[15] This reasoning can also be understood by interpreting $\rho^*$ as the coefficient from a linear regression of $s$ on $m$: both, a regression coefficient and $\rho^*$, 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.

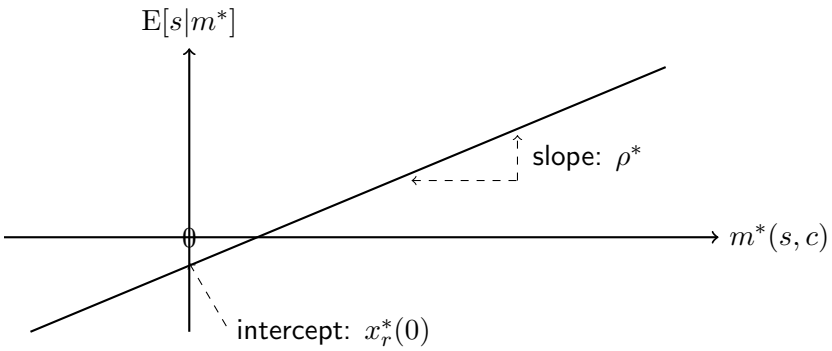


Figure 1: Graphical representation of rational receivers' inference $\mathrm{E}[s|m^*] = x_r^*(m)$ (see Proposition 1a).

strategically ignore a part of the sender's de-biased message and put weight on their prior $\bar{s}$ so that information is left unused. For illustration, consider the case that $\sigma_s^2 \to 0$ while $\sigma_c^2$ is sufficiently large. Almost all uncertainty is then not of fundamental, but of strategic nature. It follows that $\mathrm{Cov}[s, m^*]$ and, through it, also $\rho^*$ are almost zero. Rational receivers do then almost completely ignore $m^*$ and act based on their prior alone because almost all variation in the sender's message can only be due his COI. In equilibrium, the sender takes this non-reaction of rational receivers into account and scales down his bias to save lying costs. Also note that if $\rho^* \in (0, 1)$, a case that will be show to be often relevant, (9) implies that the sender's bias is increasing in the share of naive receivers $\mu$. This is different to the comparative statics presented in Ottaviani and Squintani (2006) and Kartik (2009) (see also footnote 6 and its preceding discussion).

Using the functional forms of equilibrium behavior as stated in Proposition 1a, one can then determine $\rho^*$ and how its value is affected by the game's informational parameters and disclosure:

**Proposition 1b.** *An equilibrium is the collection of strategies and beliefs as specified in Proposition 1a and an associated value of $\rho^*$ that is a fixed point to*

$$g(\rho) = \frac{\left(1 - \psi(\mathrm{Corr}[s,c])^2\right)\sigma_s^2 + (\mu + (1-\mu)\rho)(1-\psi)\sigma_{sc}}{(1-\psi(\mathrm{Corr}[s,c])^2)\sigma_s^2 + 2(\mu + (1-\mu)\rho)(1-\psi)\sigma_{sc} + (\mu + (1-\mu)\rho)^2(1-\psi)\sigma_c^2}. \quad (11)$$

*With no or imperfect disclosure ($\psi < 1$), there is a value $\tau^* < 0$ such that $\rho^* > 0$ if and only if $\sigma_{sc} > \tau^*$. If this holds, the following cases can emerge:*

*i) If $\sigma_{sc} > -\sigma_c^2$, there is a unique equilibrium with $\rho^* > 0$. It obeys $\rho^* < 1$.*

*ii) If $\sigma_{sc} < -\sigma_c^2$, there are either one or three equilibria with $\rho^* > 0$. They all obey $\rho^* > 1$.*

*iii) If $\sigma_{sc} = -\sigma_c^2$, there is a unique equilibrium with $\rho^* > 0$. It obeys $\rho^* = 1$.*

*With perfect disclosure ($\psi = 1$), there is a unique equilibrium with $\rho^* > 0$. It obeys $\rho^* = 1$.*

(Proof in Appendix A)

Note that fixed points to (11) correspond to the roots of a cubic polynomial in $\rho$. Thus, there can be up to three such positive roots. Case ii) shows this can only occur with *exactly* three, different values of $\rho^* > 1$. To analyze this and the other cases, I will restrict the analysis to "stable equilibria":

**Definition.** *An equilibrium is called a "stable equilibrium" if for the associated equilibrium inference coefficient $\rho^*$ which is a fixed point to $g(\rho)$, it holds that $\frac{d}{d\,\rho}\big(g(\rho)-\rho\big)\big|_{\rho=\rho^*} < 0$.*

Although originally a dynamic concept, this notion of stability has a long history of being used in the analysis of equilibria which originate from one-shot situations, e.g., for tâtonnement 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 a perpetuation, are locally unique, and can be found iteratively. An equilibrium which is not stable does not have these properties.[16] In the current setup, this applies only in one case:

**Lemma 3.** *Any equilibrium with $\rho^* > 0$ in the communication game that is unique is also 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 solution (i.e., with $\rho_2^*$) is not stable.* (Proof in Appendix A)

### 3.3 Linking equilibrium behavior to the information structure

The above shows that whether $\rho^*$ is larger or smaller than one has a special relevance. As the reference case, consider full disclosure ($\psi = 1$). In this situation, the signal $\tilde{c}$ precisely indicates the sender's COIs. His message can then be corrected for the bias it contains and be 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 $x_r^{*\prime}(m) = \rho^* = 1$. For no or imperfect disclosure, this does usually not hold. Different values of $\rho^*$ then reflect how the strategic interplay of senders and rational receivers is shaped by the game's informational parameters:

Case i) in Proposition 1b shows that whenever strategic uncertainty $\sigma_c^2$ is sufficiently high, specifically so high that $\sigma_{sc} > -\sigma_c^2$ holds, $\rho^* < 1$ applies. Rational receivers then partly ignore the de-biased message and put positive weight $1-\rho^*$ on their prior over $s$ (i.e., their action is a strictly convex combination of these two elements). Note that this above condition is equivalent to $Corr[s,c] > -\frac{\sigma_c}{\sigma_s}$. Strategic uncertainty exceeding fundamental uncertainty ($\frac{\sigma_c}{\sigma_s} > 1$) or a positive correlation $\sigma_{sc} \geq 0$ are therefore both sufficient conditions for such convex combinations.

Case ii) shows that when strategic uncertainty exceeds fundamental uncertainty, non-convex combinations with $\rho^* > 1$ are also possible. This happens when $\sigma_{sc} \in (\tau^*, -\sigma_c^2)$. Since then,

[16] This concept follows Hirsch and Smale (1974), pp. 185-188, who refer to such fixed points as "asymptotically stable". Blume and Board (2014) use this concept to examine endogenously chosen vagueness in a one-shot communication game. They also provide further references on how this notion of stability is relevant for one-shot situations, in particular, how it relates to Samuelson's correspondence principle. Gordon and Nöldeke (2015) employ a stability concept that is also based on the notion that after perpetuations, equilibria converge back to their original values.

$x_r^{*\prime}(m) = \rho^* > 1$ holds, a change in $m^*$ induces an over-proportional change in rational receivers' actions. To understand the economic intuition behind this, note that the condition requires $\sigma_{sc}$ to be sufficiently negative. This 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 $\rho^* = \frac{\mathsf{Cov}[s,m^*]}{\mathsf{Var}[m^*]}$ is positive, a higher message $m$ does, in expectation, still reflect a higher value of $s$. Rational receivers then utilize this 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* COI. Thus, when the COI $c$ is also too unpredictable relative to $s$ (e.g., if $\frac{\sigma_c}{\sigma_s} \geq 1$, see the discussion of Case i above), this condition cannot be fulfilled and over-reaction does not occur. The limits of expectation-based corrections are also reached if $\sigma_{sc} \leq \tau^*$. The expected bias is then so strong and opposed in direction to $s$ that the risk of a too strong mis-correction outweighs the benefits of over-reacting. A communication equilibrium with $\rho^* > 0$ can then not be established.

Finally, case iii) captures the case when the above effects just balance each other. Even without or prior to disclosure, $\rho^* = 1$ then emerges in equilibrium.

Figure 2 illustrates these findings. It shows values of solutions $\rho^* > 0$ for (11) over $\sigma_{sc}$ (or, equivalently, over $\mathit{Corr}[s,c]$), fixing other parameters. The three lines in the figure represent solutions for different values of $\sigma_s^2$, higher ones representing larger variance. This ordering reflects that higher variation in $s$ explains more variation in $m^*$ and thereby, stronger reaction of rational receivers to the de-biased message. Reflecting the above discussion, the graphs also show that whenever $\sigma_{sc} \geq 0$ or $\frac{\sigma_c}{\sigma_s} \geq 1$ (as for the lowest graph), any $\rho^* > 0$ is contained in the unit interval. The figure also portrays the normalized cutoff value $\tau^*/(\sigma_s\sigma_c)$ as vertical lines. For correlations left of these lines, an equilibrium with $\rho^* > 0$ does not exist. The upper two lines show that there are also values $\sigma_{sc} \in (\tau^*, -\sigma_c^2)$ where equilibria with $\rho^* > 1$ exist.

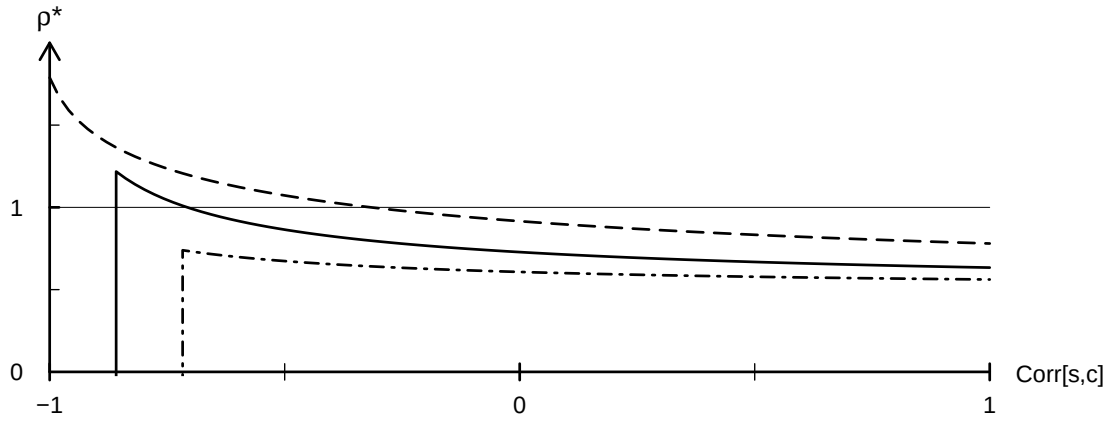


Figure 2: Stable, positive equilibrium inference coefficients $\rho^*$, positively-valued fixed points to (11) before disclosure ($\psi = 0$), over different values of $\mathit{Corr}[s,c] = \sigma_{sc}/(\sigma_s\sigma_c)$. Other parameters: $\mu = 0.5$, $\sigma_c^2 = 1$, and $\sigma_s^2 = 1/2/10$ for the bottom/middle/top line, respectively. If there are no values left of vertical lines, there are no positively-valued solutions (11) for the respective parameters and values of $\mathit{Corr}[s,c]$.

# 4 Consequences of disclosure

Given the preceding analysis of the communication game and its equilibrium, I will now look at the effects of disclosing COIs. For this, I consider an initial situation of no disclosure with $\psi = 0$ (i.e., $\sigma_\epsilon^2 \to \infty$). I then compare this to the situation "after disclosure". That is, I look at what happens after there has been the signal $\tilde{c} = c + \epsilon$ about the sender's COI through either imperfect disclosure (finite $\sigma_\epsilon^2 > 0$ resulting in $0 < \psi < 1$) or full disclosure ($\sigma_\epsilon^2 = 0$ resulting in $\psi = 1$; see Section 3.1) and players have adjusted equilibrium actions and beliefs to the signal.

Recall from Proposition 1b that will full disclosure, $\rho^* = 1$ holds. The following result generalizes this, showing that also with imperfect disclosure, $\rho^*$ moves closer towards this benchmark value:

**Lemma 4.** *After disclosure, the value $\rho^* > 0$ of a stable equilibrium changes as follows:*

*i) If $\sigma_{sc} > -\sigma_c^2$, the equilibrium inference coefficient increases towards $\rho^* = 1$.*

*ii) If $\sigma_{sc} < -\sigma_c^2$, the equilibrium inference coefficient decreases towards $\rho^* = 1$.*

*iii) If $\sigma_{sc} = -\sigma_c^2$, the equilibrium inference coefficient remains constant at $\rho^* = 1$.*

(Proof in Appendix A)

In the following, I will examine how this consequence of disclosure affects welfare, based on an ex-ante view, before $(s, c, \epsilon)$ is drawn. I start with looking at naive receivers. Recall that their utility decreases in the distance between their action and the state of the world. Since the sender's message equals $s$ plus a bias and naive receivers just follow this message, their (dis-)utility's argument equals this bias:

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

The second equality follows from $L$ being non-positive and symmetric around its maximum of zero. Thus, $\mathrm{E}[u_n^R]$ is monotonically decreasing in $\rho^*$. With Proposition 1b, one then gets the following:

**Corollary 1.** *The expected utility of naive receivers decreases in $\rho^*$. This means that in any stable equilibrium, the following happens after disclosure:*

*i) If $\sigma_{sc} > -\sigma_c^2$, the expected utility of naive receivers decreases.*

*ii) If $\sigma_{sc} < -\sigma_c^2$, the expected utility of naive receivers increases.*

*iii) If $\sigma_{sc} = -\sigma_c^2$, the expected utility of naive receivers remains constant.*

The above shows that 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 also needs to look at how disclosure affects rational receivers. These receivers de-bias the message based on what they expect to be the sender's bias. The expected damage caused by differences between actual realizations and expected values of the sender's bias depends on how much they rely on the corrected message and how volatile the bias is. To obtain a tractable measure for rational receivers' expected utility, one can exploit that the error $x_r^* - s$ in their action is a linear combination of normally distributed random variables and, therefore, in itself normally distributed. This allows to represent rational receivers' expected utility as mean-variance preferences. As their expected error is zero one can then get the following, single-argument representation:

**Lemma 5.** *The expected utility of rational receivers is given by*

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

*with $\mathcal{L}$ being strictly decreasing and concave. $\mathrm{E}[u_r^R] = 0$ holds if and only if there is full disclosure.*
(The proof adapts some techniques from Meyer (1987) and is in Appendix A.)

Thus, the expected utility of rational receivers can be expressed as a decreasing function of $\sigma_s^2$, scaled down by the squared correlation between $s$ and the equilibrium message $m^*$.[17]

This formulation of the expected utility for rational receivers helps in analyzing the opposing effects of disclosure: if $\sigma_{sc} > -\sigma_c^2$ holds and $\rho^*$ increases after disclosure, rational receivers react more to the message. However, the sender then does also increase the bias' magnitude so that the net effect on rational receiver's expected payoff is not clear. Conversely, when $\rho^*$ decreases, so does the bias. But does such a decrease in the inference coefficient then not also imply that the message's informativeness and, with it, rational receivers' utility decreases? Using Lemma 5, the following result shows that the net effect of disclosure for them is always positive:

**Proposition 2.** *After disclosure, the expected utility of rational receivers increases in every stable equilibrium.* (Proof in Appendix A)

Note that the above result implies that $\rho^*$ does *not* monotonically vary with the message's informativeness: If $\rho^* > 1$, the equilibrium inference coefficient decreases after disclosure, even though disclosure helps rational receivers to extract more information from $m^*$. More generally, these results mean that while disclosure is always good news from the perspective of rational receivers, naive ones are often hurt by it. The following is then a direct consequence:

**Corollary 2.** *In any stable equilibrium with $\rho^* > 0$ with features naive and rational receivers ($\mu > 0$), disclosure is a Pareto-improvement among all receivers if and only if $\sigma_{sc} \leq -\sigma_c^2$.*

[17] 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 footnotes 9 and 15).

Only when the conditions for the inference coefficient to be at least one are fulfilled (see Proposition 1b), then all receivers benefit from disclosure. If this is not the case, naive receivers will suffer so that based on a Pareto-criterion, disclosure is not optimal.

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 $w_n \cdot w_k > 0$:

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

The first two terms in the above expression denote 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 not considered as an informational but rather 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 $cX(m)$ is a welfare-irrelevant transfer (see also the example in Appendix B).

The above welfare function can reflect different sources of dis-utility. If $w_k = 0$, this is with regard to receiver welfare only (e.g., with weights $w_n = \mu$ and $w_r = 1-\mu$). Setting $w_k > 0$ includes the sender's lying costs into efficiency considerations, for example because these costs matter per se or because they capture reputational losses (see footnote 9). Note that when $w_k > 0$, all efficiency results also apply even if naive receivers are absent or do not matter for welfare considerations (i.e., $\mu = 0$ and/or $w_n = 0$). The reason for this is that the sender's expected lying costs are a special, parameterized case of the naive receivers' expected utility as presented in (12).

Once weights for $W$ are chosen, one can use this paper's framework to evaluate the welfare gains of disclosure. Of course, this requires more assumptions on the utility functions and distributional parameters. However, the following results show that some policy-relevant statements concerning the effect of disclosure on $W$ can be made even when exact parameters are unknown:

**Proposition 3.** *If $\sigma_{sc} \leq (>) - \sigma_c^2$, full disclosure always (never) maximizes welfare $W$ in any stable equilibrium.* (Proof in Appendix A)

The above shows that full disclosure is often inefficient. This happens whenever naive receivers are hurt by disclosure (see Corollary 2). The reason for this is that receivers have smooth, strictly concave utility: when there is full disclosure and rational receivers are at their optimum (see Lemma 5) one can add some sufficiently low noise to the perfect signal $\tilde{c} = c$ so the resulting loss for them can be made arbitrarily small. In contrast, the relative gain for naive receivers that comes with the associated decrease in the reaction by rational receivers – and therefore the sender's bias – is bigger.

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. However, determining whether this holds or what, if applicable, the optimal interior level is requires concrete assumptions on the receivers' loss functions and the game's parameters (Appendix C contains an example where no disclosure is optimal).[18]

Also note in order to assess the impact of disclosure it is not necessary to know whether $\sigma_{sc} > \sigma_c^2$ holds, the condition which has been shown to crucially determine its consequences. For an observer who wants to get testable predictions and make informed decisions it can also be sufficient to just observe how receivers react to new information. To see this, note that since naive receivers follow the sender one-to-one, one can combine (5) and (10) to get that $X'(m) < 1$ is equivalent to $\rho^* < 1$. Since the inference coefficient decreases after disclosure (Lemma 4), this means that naive receivers are hurt by disclosure (see Proposition 1b and Corollary 1). Conversely, if $X'(m) > 1$ holds, receivers react, on average, stronger or equal than one-to-one to a change in the sender's message. It then follows that full disclosure is optimal and benefits all receivers. Thus, knowing the elasticity of the receivers' average reactions to new messages can be sufficient to assess the impact of disclosure.

# 5 Discussion and 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 that allows to accommodate various situations where strategic communication affects market behavior (e.g., financial markets) and where different receivers can have different levels of risk aversion, I study the effects of disclosing conflicts 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 receivers. It does so because, in equilibrium, the average reaction to the biased signal affects the sender's bias. After disclosure, when rational receivers got 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

[18] Even under assumptions 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. 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 w.r.t. the optimal level of disclosure are feasible. However, changes in other parameters (i.e., $\mu, \sigma_s^2, \sigma_c^2, \sigma_{sc}$), do also affect $\rho^*$, as specified in (11). These indirect effects typically oppose the direct effect on receivers' marginal utilities and thereby prevent generally-valid, clear-cut comparative statics.

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 strategic bias in communication.

In light of these results, it is important to note that naive receivers who follow the sender one-to-one are strategically equivalent to receivers who delegate their action or decision $x$ to the sender, for example a managed fund. This can be either because of blind trust (Gennaioli et al., 2015) or because the cost of handling the assets oneself and acquiring the information is too costly relative to the informational gain from acting rationally (Sims, 2003). Similarly, the current setting does not only capture situations where there is a mass $\mu$ of naive receivers. It also captures scenarios where a risk-neutral sender faces a single receiver but does not know whether this receiver is naive (with probability $\mu$) or rational.[19] Finally, note that the exact form of receivers' loss function $L$ does not matter and can also be different for different receiver types. This could, for example, capture a lower risk aversion by institutional, rational investors compared to naively acting, private investors.

This paper also determines precisely when and how the adverse effects of disclosure on naive or delegating receivers manifest. In terms of economic fundamentals, this is always the case when the state of the world and the sender's COI are weakly positively correlated. Another sufficient condition for disclosure to backfire is when strategic uncertainty regarding the sender's COI exceeds fundamental uncertainty regarding the state of the world. In terms of observed behavior, the elasticity of the receivers' (re)action can be used as a criterion: disclosure backfires on naive receivers if the expert's message does not induce at least a one-to-one average reaction among receivers. Only when they react at least one-to-one, then disclosure is an improvement among all, rational *and* naive (or delegating), 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 of this paper show that when some people do not have the ability or 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 better outcomes than merely announcing them.

[19] This is different to other works which also assume naive receivers but with non-monotonic "bliss-point"-preferences for the sender (see, e.g., Ottaviani and Squintani, 2006; Kartik et al., 2007; Chen, 2011; Gordon and Nöldeke, 2015, see also footnote 6). With such preferences, it makes a difference whether the sender talks to a pool of naive and rational receivers (so that only their average reaction matters) or whether he talks to one receiver who is either rational or naive. Kartik et al. (2007) discuss these differences in their examples 1 and 3.

## 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 (2019). Preferences for Truth-Telling. *Econometrica 87*(4), 1115–1153.

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

Blume, A. and O. Board (2013). Language Barriers. *Econometrica 81*(2), 781–812.

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

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 (2015). Information, Authority, and Smooth Communication in Organizations. *CEPR Discussion Papers* (10969).

Deimen, I. and D. Szalay (2019). Delegated expertise, authority, and communication. *American Economic Review 109*(4), 1349–1374.

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.

Frankel, A. and N. Kartik (2018). What kind of central bank competence? *Theoretical Economics 13*(2), 697–727.

Frankel, A. and N. Kartik (2019). Muddled Information. *Journal of Political Economy 127*(4), 1739–1776.

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.

Giovannoni, F. and S. Xiong (2019). Communication under language barriers. *Journal of Economic Theory 180*, 274–303.

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 (2015). 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., 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.

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.

Sobel, J. (2020). Lying and Deception in Games. *Journal of Political Economy 128*(3), 907–947.

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.

For Online Publication

# De-biasing Strategic Communication

## Appendix

Tobias Gesche*

Contents:

- Appendix A: Proofs for results in the main text
- Appendix B: Example for $\sigma_{sc} \neq 0$ in the context of financial markets
- Appendix C: Example for non-disclosure to be optimal

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

# Appendix A: Proofs for results in the main text

## Features of normally distributed random variables

For the following proofs, it will be useful to list three well-established properties of (jointly) normally distributed variables. For this, consider a random vector $\mathbf{x} \in \mathbb{R}^n$ with $n \geq 2$, which is normally distributed according to $\mathcal{N}(\boldsymbol{\eta}, \boldsymbol{\Sigma})$. Also consider two non-empty partitions $\mathbf{x}_1$ and $\mathbf{x}_2$ of $\mathbf{x}$ so that $\boldsymbol{\eta}$ consists of $\boldsymbol{\eta}_1$ and $\boldsymbol{\eta}_2$ while $\boldsymbol{\Sigma}$ is partitioned into four blocks $\boldsymbol{\Sigma}_{11}$, $\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Sigma}_{21}$, and $\boldsymbol{\Sigma}_{22}$. Then, the following holds:

P1: *linear combinations of elements of* $\mathbf{x}$ *are normally distributed*

P2: $(\mathbf{x}_2|\mathbf{x}_1)$ *is normally distributed according to* $\mathcal{N}(\boldsymbol{\eta}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\eta}_1), \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12})$

P3: $\mathbf{x}$ *and linear combination formed from its elements are symetrically distributed around their respective expected values.*

It will also be useful to note that if $\mathbf{x} \in \mathbb{R}^2$, P2 implies that $(x_2|x_1)$ is distributed according to

$$\mathcal{N}\left(\eta_{x_2|x_1} = \mathrm{E}[x_2] + (x_1 - \mathrm{E}[x_1])\,\frac{\mathsf{Cov}[x_1, x_2]}{\mathsf{Var}[x_1]}, \Sigma_{x_2|x_1} = \mathsf{Var}[x_2]\left(1 - (\mathit{Corr}[x_1, x_2])^2\right)\right).$$

## Proof of Lemma 1

If $m$ is normally distributed, P2 applies so that $s|m$ is also normally distributed and, by P3, symmetric around $\mathrm{E}[s|m]$. Using $\varphi(s|m)$ to denote its pdf it then holds that $x_r(m) = \mathrm{argmax}_{d \in S} \int_{\mathbb{R}} L(d - s)\varphi(s|m)ds$ . Therefore, the necessary FOC for a candidate solution $x_r = \mathrm{E}[s|m]$ is given by

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

It 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. $\Delta = |x_r - s|$. By symmetry of $L$ around zero it holds that $L'(\Delta) = -L'(-\Delta)$. Since $\varphi(s|m)$ is symmetric around $\mathrm{E}[s|m] = x_r$ it then follows that

$$L'(\Delta)\varphi(x_r - \Delta|m]) = -L'(-\Delta)\varphi(x_r + \Delta|m]) \leqq 0$$

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. □

## Proof of Lemma 2

The assumptions on $\epsilon$ means that the expected values and the variance-covariance matrix for $(\mathrm{s}, \mathrm{c}, \epsilon)$ are given by

$$\begin{bmatrix} \bar{s} \\ \bar{c} \\ 0 \end{bmatrix} \text{ and } \begin{bmatrix} \sigma_s^2 & \sigma_{sc} & 0 \\ \sigma_{sc} & \sigma_c^2 & 0 \\ 0 & 0 & \sigma_\epsilon^2 \end{bmatrix}.$$

Since $\tilde{c} = c + \epsilon$ is normally distributed (see P1), so is the random vector $(\tilde{c}, s, c)$. Note that because $\epsilon$ is not correlated with $s$ or $c$ and has an expected value of zero, $\text{Cov}[s, \tilde{c}] = \text{E}[(s - \text{E}[s])(c + \epsilon - \text{E}[c])] = \text{E}[(s - \bar{s})(c - \bar{c})] = \sigma_{sc}$, $\text{Var}[\tilde{c}] = \text{E}[(c + \epsilon - \text{E}[c])^2] = \text{E}[(c + \epsilon - \bar{c})^2] = \sigma_c^2 + \sigma_\epsilon^2$ , and $\text{Cov}[c, \tilde{c}] = \text{E}[(c - \text{E}[c])(c + \epsilon - \bar{c})] = \text{E}[(c - \bar{c})(c - \bar{c})] = \sigma_c^2$. Putting this together means that the vector of expected values and the associated variance-covariance matrix for $(\tilde{c}, s, c)$ are given by

$$\begin{bmatrix} \bar{c} \\ \bar{s} \\ \bar{c} \end{bmatrix} \text{ and } \begin{bmatrix} \sigma_c^2 + \sigma_\epsilon^2 & \sigma_{sc} & \sigma_c^2 \\ \sigma_{sc} & \sigma_s^2 & \sigma_{sc} \\ \sigma_c^2 & \sigma_{sc} & \sigma_c^2 \end{bmatrix}.$$

Using P2 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 1a

Consider a candiate equilibrium messaging strategy $\tilde{m}(s, c)$ such that $\tilde{X}(m) = \mu m + (1 - \mu)\tilde{x}_r(m)$ where $\tilde{x}_r(m)$ is a linear function of $m = \tilde{m}(s, c)$. In consequence, $\tilde{X}(m)$ is also a linear function of $m$ so that by (4), $\tilde{m}(s, c) = s + c\,(\mu + (1 - \mu)\tilde{x}_r'(m))$ has to hold. Linearity of $\tilde{x}_r(m)$ also means that $\tilde{x}_r'(m)$ is a constant (i.e, $\tilde{x}_r''(m) = 0$). Therefore, $\tilde{m}(s, c)$ is a linear combination of $s$ and $c$ and, by P1, normally distributed. Using P2 yields

$$\tilde{x}_r(m) = \text{E}[s|m]_{m=\tilde{m}(s,c)} = \text{E}[s] + \left(m - \text{E}[s] - \text{E}[c](\mu + (1 - \mu)\tilde{x}_r'(m))\right)\tilde{\rho} \qquad \text{(A.1)}$$

where P2 was also used to replace $\text{E}[m]$ with $\text{E}[s] + \text{E}[c](\mu + (1 - \mu)\tilde{c}$. In addition, (8) was used to replace $\text{Cov}[s, m]_{m=\tilde{m}(s,c)}/\text{Var}[m]_{m=\tilde{m}(s,c)}$ with $\tilde{\rho}$. It then follows that $\tilde{x}_r'(m) = \tilde{\rho}$ has to hold.

Given the above, the intercept for $\tilde{x}_r(m)$ can be obtained by integrating. That is, $\tilde{x}_r(m) = \int_M \tilde{x}_r'(m)dm = \tilde{\rho}m + \tilde{K}$ has to hold, where $\tilde{K}$ is an integration constant. To determine this constant, plug $\tilde{x}_r(m) = \tilde{\rho}m + \tilde{K}$ into (A.1) to get $\tilde{K} = \text{E}[s] - (\text{E}[s] + \text{E}[c]\,(\mu + (1 - \mu)\,\tilde{\rho}))\,\tilde{\rho}$.

Thus, the sender's expected utility is a quadratic function of $m$:

$$\text{E}[U^S(s, c, m)] = mc\,(\mu + (1 - \mu)) - \frac{1}{2}(m - s)^2 + c(1 - \mu)\tilde{K}].$$

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 $\rho^* = \tilde{\rho} = x_r^{*\prime}(m)$, as stated in (9). Using $\tilde{\rho} = \rho^*$ and the above expression for $\tilde{K}$ on $\tilde{x}_r(m) = \tilde{\rho}m + \tilde{K}$ then yields the rational receiver's belief and strategy $x_r^*(m)$ as stated in (10). □

## Proof of Proposition 1b

This proof describes how $\rho^*$ is determined and how this relates to the games' parameters. It does so in four steps, which build on each other: Step 1) derives the fixed point expression as presented in (11). It also derives equivalent functional forms and their properties that are used in the following steps and proofs. Step 2) derives conditions for the existence of solutions with $\rho^* > 0$ while Step 3) relates the parameters in $\boldsymbol{\Sigma}$ to the value range of $\rho^*$, in particular relative to $\rho^* = 1$. In Step 4), multiplicity and uniqueness of positive fixed points to (11) are examined.

### Step 1) Determining $g(\rho)$ and its properties:

Using $m^*(s,c)$ from Proposition 1a and the definition of $\rho^*$, the latter must solve

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

Plugging in the elements of $\mathcal{N}(\hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\Sigma}})$ means that $\rho^*$ has to solve the expression stated in (11). If $\psi = 1$, one gets $\rho^* = 1$. If $1 - \psi > 0$, this term can be factored out. Using

$$\phi \equiv (1 - \psi(\mathit{Corr}[s,c])^2/(1-\psi)$$

allows to write (11) as follows:

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

The above fixed point expression will be used in this and the following proofs as an alternative to (11) when $\psi < 1$. Note that this expression corresponds to a rational function $g(\rho) \equiv N(\rho)/D(\rho)$ with nominator $N(\rho) \equiv \phi\sigma_s^2 + (\mu+(1-\mu)\rho)\sigma_{sc}$ and denominator $D(\rho) \equiv \phi\sigma_s^2 + 2(\mu+(1-\mu)\rho)\sigma_{sc} + (\mu+(1-\mu)\rho)^2\sigma_c^2$. For this function, three properties hold:

- Property a): $N(\rho)/D(\rho)$ *is continuous with* $D(\rho) > 0$ *for all* $\rho \in \mathbb{R}$.
- Property b): $\lim_{\rho\to+\infty}\left(\frac{N(\rho)}{D(\rho)}\right) = 0^-$ *if* $\sigma_{sc} < 0$ *and* $\lim_{\rho\to+\infty}\left(\frac{N(\rho)}{D(\rho)}\right)$ *if* $\sigma_{sc} \geq 0$
- Property c): $\frac{N(\rho)}{D(\rho)}$ *has at most two extreme points*.

*Proof of Property a)*: Since both, $D(\rho)$ and $N(\rho)$ are continuous in $\rho$, 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 $|Corr[s,c]| < 1$. □

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

*Proof of Property c)*: 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 Property 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 quadratic equation with at most two real solutions. □

**Step 2) Conditions for existence of a solution $\rho^* > 0$**

A solution $\rho^* > 0$ requires $g(\rho^*) = N(\rho^*)/D(\rho^*) > 0$. By Property a), this requires $N(\rho^*) > 0$. This can be translated to $\sigma_{sc} > \tau(\rho^*)$ where $\tau(\rho) = -\phi\sigma_s^2/(\mu + (1-\mu)\rho) < 0$ is defined for any $\rho > 0 \geq -\mu/(1-\mu)$ with $\tau'(\rho) > 0$. Therefore, $\sigma_{sc} > \tau^* \equiv \tau(\rho^*)$ with $\tau^* < 0$ is a necessary condition for $\rho^* > 0$.

To see that this condition is also sufficient note that if $\sigma_{sc} > \tau^* = \tau(\rho^*)$, it also holds that $\sigma_{sc} > \tau(0)$ and, therefore, $N(0) > 0$. Since $D(0) > 0$, it follows that $g(0) = N(0)/D(0) > 0$. Together with continuity and $\lim_{\rho\to+\infty}(g(\rho) = N(\rho)/D(\rho)) = 0$ as derived in properties a) and b), this means that there has to be at least one intersection of $g(\rho)$ with the 45-degree line over $\mathbb{R}_{++}$.

**Step 3) Values $\rho^* > 0$ and the threshold value of 1**

First consider the following necessary condition for $\rho^* \leq 1$:[a] from (A.2), this holds 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 $\rho^*$. Inserting $\rho^* = 1$, the upper bound on the desired value range, then yields $\sigma_{sc} > (=) - \sigma_c^2$ as a necessary condition for $\rho^* < (=) 1$ and, therefore, $\sigma_{sc} < -\sigma_c^2$ for $\rho^* > 1$.

To see that that these conditions are also sufficient first note from (A.2) that for any $\sigma_{sc} \geq 0$, $\rho^* \in (0, 1)$ follows immediately. Similiarly, inserting $\sigma_{sc} = -\sigma_c^2$ into (A.2) shows that this is a sufficient condition for a fixed point $\rho^* = 1$. 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 (A.2) 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 $D(\rho^*) > 0$, this simplifies to

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

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. Finally, by the above results, the case of $\sigma_{sc} \in (\tau^*, -\sigma_c^2)$ means that in an equilbirum with a positive inference coefficient, $\rho^* > 1$ has to hold for this coefficient.

**Step 4) Multiplicity of equilibria**

The results of the preceding step yield Case i) through iii) of the proposition. To complete its proof, one has to look at multiplicity for different parameter constellations. For this, note that a fixed point to (A.2) requires an intersection of the 45-degree line and $N(\rho)/D(\rho)$. Thus, $\rho^*$ has to be a root of $k(\rho) = \rho D(\rho) - N(\rho)$, the following cubic equation:

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

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

[a] Even though this necessary condition is not stated in the proposition, it will be useful in deriving the sufficient condition stated there. Furthermore, the necessary condition is also used in the proof of Lemma 4.

**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.*

For the polynomial coefficients in (A.3), $A > 0$ always holds. Furthermore, it has been shown above that a solution $\rho > 0$ implies $D < 0$ because this is equivalent to $\sigma_{sc} > \tau(0)$ with the above-defined function $\tau(\rho)$ where $\tau(0) > \tau(\rho^*) = \tau^*$ and $\rho^* > 0$ implies $\sigma_{sc} > \tau^*$. By the rule of signs, 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, corresponding to fixed points of $g(\rho)$.

Suppose multiple positive fixed points exist. Denote their location w.l.o.g by $0 < \rho_1^* < \rho_2^* < \rho_3^*$. Note that this requires $B < 0$ and, therefore, $\sigma_{sc} < 0$. By property a) and b), this means that $g(\rho) = N(\rho)/D(\rho)$ continuously approaches zero from below when $\rho$ becomes large. Also, it has been shown that $N(0)/D(0) > 0$. Together, this implies that $g$ has a negatively valued local minimum over $\mathbb{R}_{++}$. Denote it by $\rho_-$ (i.e., $g(\rho_-) < 0$ with $\rho_- > 0$). If $\rho_-$ were the only extreme value over $\mathbb{R}_{++}$, there would be only one intersection with the 45-degree line, thus a unique fixed point. Thus, by Property c), there needs to be one other extreme value of $g(\rho)$ over $\mathbb{R}_{++}$. Given that $\rho_-$ is a local minimum, this extreme value has to be a local maximum and is denoted by $\rho_+$. It then follows from $\lim_{\rho \to +\infty} g(\rho) = 0^-$ that $0 < \rho_+ < \rho_-$ and $g(\rho_-) < 0 < g(\rho_+)$. Accordingly, $g(\rho) = N(\rho)/D(\rho)$ is non-increasing over $[\rho_+, \rho_-]$. This, together with $N(0)/D(0) > 0$, implies that $g(\rho)$ cuts the 45-degree line once within this interval and never on $(\rho_-, +\infty)$. Thus, the unique fixed-point over $(\rho_+, \rho_-)$ is also the highest-valued one. It then holds 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 over $[\rho_1^*, \rho_2^*]$. 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 thrice: First from above at $\rho_1^*$, then from below at $\rho_2^*$ (which requires a slope larger than 1), and then from above at $\rho_3^*$:

$$g(\rho)'|_{\rho=\rho_3^*} < 0 < g(\rho)'|_{\rho=\rho_1^*} < 1 < g(\rho)'|_{\rho=\rho_2^*} \tag{A.4}$$

Using the fact that in equilibrium, $\rho^* = g(\rho^*) = N(\rho^*)/D(\rho^*)$ has to hold, $g(\rho)'|_{\rho=\rho_1^*} > 0$ translates 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 $1 < N(\rho_1^*)/D(\rho_1^*) = \rho_1^* < \rho_2^* < \rho_3^*$ and, therefore, $\sigma_{sc} < -(\mu + (1-\mu)\rho_1^*)\sigma_c^2 < -\sigma_c^2$ (i.e., Case ii) as a necessary condition for multiple fixed points $\rho^*$. □

## Proof of Lemma 3

Fixed points to $g(\rho)$ as defined in (11) can be found as roots to $f(\rho) = g(\rho) - \rho$. Such a fixed point $\rho^*$ is then (asymptotically) stable if $f'(\rho)|_{\rho=\rho_k^*} < 0$. If $\psi = 1$, one gets $f(0) = g(0) > 0$. Similarly, if $\psi < 1$, one gets from the first part of the proof of Proposition 1b that $f(0) = g(0) = N(0)/D(0) > 0$ also holds. It follows that for $f$ to have three roots, it has to cut the real line from above at $\rho_1^*$, from below at $\rho_2^*$, and again from above at $\rho_3^*$. This implies $f'(\rho_1^*) < 0$, $f'(\rho_2^*) > 0$, and $f'(\rho_3^*) < 0$ which proves stability of $\rho_1^*$ and $\rho_3^*$, and that $\rho_2^*$ is not stable. By the same reasoning, a unique root $\rho^*$ has to obey $f'(\rho^*) < 0$ and is thus stable. □

## Proof of Lemma 4

After disclosure, $\psi < 1$ increases. This enters (11) or, equivalently, (A.2) via an increase in $\phi = (1 - \psi(\mathit{Corr}[s,c])^2)/(1-\psi)$:

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

Denote this increased value with $\tilde{\phi} > \phi$, which means that the function $k(\rho)$ as used in Part 2 of the proof of Proposition 1b also changes.[b] Then, one can define the new function with $\tilde{k}(\rho)$ as follows:

$$\tilde{k}(\rho) = \underbrace{(1-\mu)^2\sigma_c^2}_{\equiv \tilde{A}}\cdot\rho^3 + \underbrace{2(1-\mu)(\sigma_{sc} + \mu\sigma_c^2)}_{\equiv \tilde{B}}\cdot\rho^2 + \underbrace{\tilde{\phi}\sigma_s^2 + \mu^2\sigma_c^2 + (3\mu-1)\sigma_{sc}}_{\equiv \tilde{C}}\cdot\rho \underbrace{-\tilde{\phi}\sigma_s^2 - \mu\sigma_{sc}}_{\equiv \tilde{D}} \tag{A.5}$$

Comparing these coefficients to those of $k(\rho)$ as stated in (A.3), 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 fixed point to $g(\rho)$ after disclosure. Denote these fixed point before and after disclosure by $\rho^*$ and $\tilde{\rho}^*$, respectively. Then, the following (sub-)cases, based on the cases presented in Proposition 1b, exist:

i) If there is a unique solution $\rho^* \in (0,1)$, then $\tilde{\rho}^* \in (\rho^*, 1)$ and this solution $\tilde{\rho}^*$ is unique.

ii-1) If there is a unique solution $\rho^* > 1$, then $\tilde{\rho}^* \in (1, \rho^*)$ and this solution $\tilde{\rho}^*$ is unique.

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

ii-2) If there are three solutions $1 < \rho_1^* < \rho_2^* < \rho_3^*$ 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 $1 < \tilde{\rho}_1^* < \rho_1^* < \rho_2^* < \tilde{\rho}_2^* < \tilde{\rho}_3^* < \rho_3^*$.

iii) If there is a unique solution $\rho^* = 1$, then $\tilde{\rho}^* = 1$ and this solution $\tilde{\rho}^*$ is unique.

To proof the above, recall from Proposition 1b 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.3) and (A.5) 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 1b to show that multiple solutions require all of them to have a value larger than one. This proves cases i) and iii).

For Case ii-1), with a unique $\rho^* > 1$, the above implies $\tilde{k}(1) = k(1) < 0$. A unique root of $k$ at $\rho^* > 1$ means that $g(\rho)$ never cuts the real line again on $(\rho^*, +\infty)$. Neither does $\tilde{k}$ as $\tilde{k}(\rho) > k(\rho)$ for $\rho > 1$. This, in addition with $\tilde{k}(1) = k(1) < 0$, means that $\tilde{k}$ cuts the real line once over $(1, \rho^*)$.

Case ii-2) captures three positively-valued fixed points to $g(\rho)$. By Proposition 1b, their coordinates have to obey $1 < \rho_1^* < \rho_2^* < \rho_3^*$. The continuous, cubic function $k$ obeys $k(0) < 0$ (see Stage 4 of the proof of Proposition 1b). This implies that $k$ cuts the real line from below at $\rho_1^*$, from above at $\rho_2^*$, and again from below at $\rho_3^*$. Since it is a continuous polynomial, it has to have a local maximum and minimum in between these points. They are denoted by $\rho_-^k$ and $\rho_+^k$, 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 analogously-shaped polynomial. 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 ii-2). 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 $\rho$ than for $k$. It follows that $1 < \tilde{\rho}^* < \rho_1^* < \rho_2^* < \rho_3^*$ which proves the first part of case d).

Now, use subscript $l \in \{1, 2, 3\}$ to denote the above-described inference coefficients for up to three equilibria. W.l.o.g., use $l = 1$ if a coefficient is unique. Going over the above cases for all stable equilbria (i.e., for $l \neq 2$), disclosure leads to i) $1 > \tilde{\rho}_1^* > \rho_1^*$ if and only if $\rho^* < 1$, ii) $1 < \tilde{\rho}_1^* < \rho_1^*$ and $1 < \tilde{\rho}_3^* < \rho_3^*$ if and only if $\rho^* > 1$, and iii) $\tilde{\rho}_1^* = \rho_1^*$ if and only if $\rho^* = 1$. Using the proof

of Proposition 1b, which establishes necessary and sufficient conditions to link values of $\rho^*$ to the game's fundamentals (see Foonote a), then yields the stated lemma. □

## Proof of Lemma 5

The argument for rational receivers' (expected) utility $\mathrm{E}[L(z)]$ is given by

$$\begin{aligned} z \equiv x_r^*(m) - s &= (1-\rho^*)\mathrm{E}[s] + \rho^* \left[m^*(s,c) - \mathrm{E}[c]\left(\mu + (1-\mu)\rho^*\right)\right] - s \\ &= -(s - \mathrm{E}[s]) + \left(m^*(s,c) - \mathrm{E}[m^*(s,c)]\right)\rho^*. \end{aligned} \tag{A.6}$$

Note that $m^*(s,c)$ is a linear function of $s$ and $c$ and therefore normally distributed (see P1). Since $s$ is also normally distributed, so is $z$. Using $\sigma_z = \sqrt{\mathrm{Var}[z]}$, one can then normalize $z$ via the linear transformation $\hat{z}(z) = z/\sigma_z - \mathrm{E}[z]$ such that $\hat{z}$ follows $\mathcal{N}(0,1)$. The associated probability density function will be denoted $\varphi(\hat{z})$. The expected utility of rational receivers can then be expressed as

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

From (A.6) it follows that $\mathrm{E}[z] = 0$. Using $\sigma_z = \sqrt{\mathrm{Var}[z]}$, one can define the univariate function $\mathcal{L}\left(\sigma_z\right) \equiv V\left(0, \sigma_z\right) \leq 0$, which denotes a rational receiver's expected utility. It then holds that

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

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, it then also follows that

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

To see that full disclosure is necessary for $\mathcal{L}(0) = 0$ to hold, note from the above that this requires $\sigma_z = 0$ and therefore $x_r^*(m) = s$. Suppose that this held under imperfect disclosure. For $x_r^*(m) = s$ 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 $\psi \in (0,1)$, $\mathrm{Var}[c|\tilde{c}] > 0$ and $\mathrm{Var}[s|\tilde{c}] > 0$ applies (see Lemma 2). With full disclosure, the sender's message is revealing so that $x_r^*(m) = s$ holds. This establishes sufficiency.

To see how the argument $\mathsf{Var}[z] = \mathsf{Var}[x_r^*(m) - s]$ in $\mathcal{L}$ can be alternatively expressed, note that

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

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

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

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

## Proof of Proposition 2

Lemma 5 shows that the expected utility of rational receivers strictly increases in $\mathit{Corr}[s, m^*]^2$. For equilibria with $\rho^* > 0$, and therefore $\mathsf{Cov}[s, m^*] > 0$, it is then sufficient to show that $\mathit{Corr}[s, m^*] > 0$ increases after disclosure. For this note that

$$\mathit{Corr}[s, m^*] = \frac{\mathsf{Cov}[s, m^*]}{\mathsf{Var}[m^*]} \cdot \frac{\sqrt{\mathsf{Var}[m^*]}}{\sigma_s} = \rho^* \cdot \frac{\sqrt{\mathsf{Var}[m^*]}}{\sigma_s}. \tag{A.7}$$

Now, consider the three different parameter constellations as described in Proposition 1b:

Case i) means that $\sigma_{sc} > -\sigma_c^2$. From Lemma 4, it then follows 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 $\phi$ increases, i.e., $\tilde{\phi} > \phi \geq 1$ (see proof of Lemma 4). Since the first factor on the RHS of (A.7) increases after disclosure, it is then sufficient to show that also the second increases. Using $D(\phi, \rho < *)$ as defined in the proof of Proposition 1b, this mean that $D(\tilde{\rho}^*.\rho^*)$ holds, with

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

While an increase in $\phi$ clearly increases the above term, the indirect effect via $\rho^*$ is not that clear. However, from the fact that $\sigma_{sc} + \sigma_c^2 > 0$ is a necessary and sufficient condition for $\rho^* \in (0, 1)$ it

follows that in this case also $\sigma_{sc} + (\mu + (1-\mu)\rho^*)\sigma_c^2 > 0$ holds and therefore

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

Case ii) means $\sigma_{sc} < -\sigma_c^2$. By Proposition1b and Lemma 4, it therefore holds that $1 < \tilde{\rho}^* < \rho^*$. Thus, disclosure *decreases* $\rho^*$ while $\phi$ still increases. While an increase in $\phi$ also shifts $\mathit{Corr}[s, m^*]$ (see above), the effect of a decrease in $\rho^*$ is not that clear. To examine this effect, note that

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

Again, the inequality follows from the fact that $\sigma_{sc} + \sigma_c^2 \leq 0$ is a necessary and sufficient condition for $\rho^* > 1$. This 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 that

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

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

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

Multiplying the above again, this time with $D(\rho^*, \phi) > 0$ and substituting $\rho^*$ with the RHS of (A.2) at $\rho = \rho^*$ then yields, after some transformations, that the sign of the above equals

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

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

Case iii) captures the case of $\sigma_{sc} = -\sigma_c^2$: By Proposition 1b and Lemma 4, $\tilde{\rho}^* = \rho^* = 1$ holds before and after disclosure while $\tilde{\phi} > \phi$ holds nevertheless. Therefore, $\mathit{Corr}[s, m^*]$ increases. □

## Proof of Proposition 3

I start with $w_k = 0$ and denote, with slight abuse of notation, $W(\psi) \equiv W(\rho^*(\phi(\psi, \cdot), \cdot))$ through 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 from an initial value $\psi < 1$ affects $\rho^*$ via $\phi$ (i.e., via $\partial \phi(\psi, \cdot)/\partial \psi > 0$, see the proof of Lemma 4). To determine this effect,

consider (A.2) 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 $\phi$ increases in $\psi$, 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 $\rho^*$ increases (decreases) after disclosure if and only if $\rho^* < 1$ ($\rho^* > 1$) one then gets the following:

$$\operatorname{sgn}\left[\frac{\mathrm{E}[u_n^R(\rho^*(\phi,\cdot),\cdot))]}{\partial \psi}\right] = \operatorname{sgn}\left[\frac{\mathrm{E}[u_n^R(\rho^*)]}{\partial \rho^*} \cdot \frac{\partial \rho^*(\phi,\cdot)}{\partial \phi}\right] = \operatorname{sgn}\left[\rho^* - 1\right]. \tag{A.8}$$

When $\rho^* \in (0,1)$, every increase in $\phi$ 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

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

holds. The second-last equality in the above follows from the fact that by Lemma 5, rational receivers expected utility w.r.t to $\psi$ 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.8).

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.8), increasing the level of disclosure $\psi$ helps naive receivers and that full disclosure maximizes the utility of rational receivers (see Lemma 5). □

## Appendix B: Example for $\sigma_{sc} \neq 0$ in the context of financial markets

In the following, I will lay out an example how in a setting where the sender sends a report $m$ on a financial asset and $X(m)$ is the demand for this asset, situations with either $\sigma_{sc} > 0$ or $\sigma_{sc} < 0$ can occur:

As the reference case, consider a third party (i..e, neither the sender nor the receiver), 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: $p(m) = \beta X(m)$ with $\beta > 0$. Normalize the reservation price of the third party to zero and allow it to transfer share $\tau$ 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 $\tau$).

Now consider the same situation but assume that the third party has also information about the asset's fundamentals $s$, (e.g., 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.

To see the case for a positive correlation, consider the same situation, but in order to sell, put options are now 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.[c] Thus, the present framework allows to analyze different modes of how COIs can arise, in particular those which are relevant for strategic communication in financial markets.

[c] If the third party wants to buy quantity $q > 0$ (i.e., sell quantity $q < 0$), analogous arguments can be derived.

## Appendix C: Example for non-disclosure to be optimal

As a concrete example for a scenario with $\psi < 1$ 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 (A.2) and solving yields $\rho^* \approx 0.6$. Following Proposition 1b and Lemma 4, disclosure then increases this inference coefficient. Using $m^*(s, c)$ as defined in Proposition 1a on (13) then yields

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

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