CESifo WORKING PAPERS

11181
2024

June 2024

Social Preferences, Trust, and Communication when the Truth Hurts

Jonathan Gehle, Ferdinand A. von Siemens

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 11181*

Social Preferences, Trust, and Communication when the Truth Hurts

Abstract

We investigate how heterogeneous social preferences affect the communication of painful information in social relationships. We characterize the existence conditions for a pooling equilibrium in which individuals conceal painful information because revealing the latter would signal that they are selfish, thereby leading to a loss of trust. We also find that compassionate individuals may then be more tempted to reveal bad news than selfish individuals because they benefit less from an intact social relationship. Moreover, there may be multiple equilibria with different degrees of information disclosure and standard equilibrium refinements have no bite. Coordination on an inefficient equilibrium could therefore lead to severe information frictions, even if the pain of receiving bad news is quite small.

JEL-Codes: D820, D830, D910.

Keywords: communication, painful information, social preferences, trust.

Jonathan Gehle
Goethe University Frankfurt
Theodor-W.-Adorno-Platz 3
Germany – 60323 Frankfurt
gehle@econ.uni-frankfurt.de

*Ferdinand A. von Siemens**
Goethe University Frankfurt
Theodor-W.-Adorno-Platz 3
Germany – 60323 Frankfurt
vonsiemens@econ.uni-frankfurt.de

*corresponding author

June 18, 2024
Declarations of interest: none.

1 Introduction

Although having the right information is critical to making the right decisions, some information causes an immediate negative emotional response, some information simply hurts the recipient. In consequence, individuals often avoid, ignore, or forget such unpleasant truths, see Golman, Hagman and Loewenstein (2017). However, not only the demand for, but also the supply of unpleasant information might be burdened by frictions. For instance, people might hesitate to share information with colleagues that challenges their deeply held beliefs, they might avoid to openly discuss certain sensitive topics like religion or politics during family gatherings, and they might refrain from giving honest but negative feedback to members of their sports team. Social preferences are likely to influence the decision to reveal unwelcome truths. For once, individuals with strong social preferences may choose not to disclose negative information simply to avoid causing harm to others, see Rosen and Tesser (1970) and Conlee and Tesser (1973). Furthermore, because social preferences are heterogeneous and private information, communicating a painful truth could be perceived as mean and selfish, damaging the social reputation of the bearer of bad news. Individuals may then choose to remain silent because they fear that revealing painful information may harm future cooperation in their social networks, see Milliken et al. (2003).

Despite the empirical evidence, our theoretical understanding of the importance of social preferences for communication in social relationships remains limited. In this paper, we thus analyze a strategic setting in which a speaker decides whether to pass on noisy, private information about a state of the world to a listener. Revealing bad news directly reduces the listener's utility. The listener observes the information passed on by the speaker and then chooses an action that produces payoffs that are higher if the action matches the state of the world. The speaker's and listener's interests are perfectly aligned with respect to these actions. To model their social relationship, the listener and the speaker then engage in a trust game in which the listener must decide whether or not to trust the speaker.

We then explore how heterogenous and privatly known social preferences might plausibly influence the disclosure of negative information. We assume that the listener always has egoistic preferences in the sense that she maximizes her material payoffs. However, the speaker is either egoistic or compassionate, and his preferences are private information. While a selfish speaker maximizes her payoffs, a compassionate speaker suffers from hurting the listener by revealing bad news and returns trust in the trust game. Because revealing bad news has different costs to a compassionate or egoistic speaker, it might reveal something about the speaker's trustworthiness. We derive all pure strategy perfect Bayesian equilibria of the above game. For certain parameter values, there indeed exists a pooling equilibrium in which both the egoistic and the compassionate speaker never reveal bad news. The listener then trusts if and only if she has not previously received bad news, because she interprets the latter out-of-equilibrium signal as an indication that the speaker is not trustworthy. There might also exist a pooling equilibrium in which both the egoistic and the compassionate speaker always reveal bad news, and the listener trusts the speaker both after receiving no news and bad news. Finally, there might also exist a separating equilibrium in which only a selfish speaker reveals bad news. There never exists a separating equilibrium in which only the compassionate speaker reveals bad news.

We use our theoretical model to derive comparative statics predictions with respect to the pain of bad news. Intuitively, we would expect more information to be revealed if the pain of receiving bad news decreases. If bad news causes very little pain, then revealing it is not very unkind or mean-spirited and even helps the listener to make a more informed decision. The unwanted truth is then actually wanted and should not endanger the listener's trust. If receiving bad news causes substantial pain to the listener, revealing bad news is clearly unkind and therefore more likely to result in a loss of trust. Our theoretical analysis confirms this intuition: higher pain of receiving bad news enlarges the set of parameter combinations for which we have equilibria with only partial or no revelation of bad news. Moreover, very high levels of pain destroy the existence of a pooling equilibrium involving the disclosure of bad news, while supporting equilibria characterized by partial or complete concealment of undesirable information.

The analysis also generates more subtle insights concerning the intricacies of social signaling. The pooling equliibrium with no information disclosure exists because people withhold painful information to appear trustworthy. However, in this equilibrium, more trustworthy individuals might actually have a stronger temptation to reveal bad news than more selfish individuals. The reason is that trustworthy individuals are less willing to exploit any social relationship and therefore benefit less from retaining a positive reputation. How reasonable is the latter result? One might argue that functioning social relationships require a certain balance between help provided and help received. More selfish individuals might be willing to push this relationship between asking and giving closer to the breaking point by requesting help more frequently. Our finding seems similar to the known result that more prosocial individuals might be less trusting because they fear being betrayed and exploited. However, the underlying mechanism is very different in the present context where prosocial individuals might exploit others while they can never be exploited themselves.

Further, the theory shows that there exist multiple equilibria for a wide range of parameters. This finding implies that people might coordinate on equilibria with strategic silence even if the pain from receiving bad news is relatively small. Refinements based on forward-induction arguments like Cho and Kreps (1987) have no bite. We only have out-of-equilibrium beliefs in the pooling equilibrium in which nobody every reveals bad news, but in this equilibrium an egoistic speaker might benefit the most from revealing bad news because this leads to a better decision while he does not suffer from inflicting pain on the listener. Revealing bad news therefore cannot be equilibrium-dominated for an egoistic speaker, which renders the standard equilibrium refinements ineffective. Finally, the analysis also reveals the conditions under which the pooling equilibrium with no information disclosure is Pareto-inefficient. This is the case if compassionate speakers have very strong social preferences and it is not that much better for the listener to make the optimal decision in case of the bad state of the world rather than choosing the outside option.

2 Related Literature

The current theory has four key elements: (1) the speaker communicates strategically to influence the listener's decision, (2) the listener suffers a utility loss from receiving bad news, (3) the speaker may have social preferences and thus care about the negative emotional effect they impose on the listener, and (4) the speaker cares about how communication affects her preference reputation in the trust game. The present paper is most closely related to strands of the literature that share one or more of these - or very similar - elements. In the following, we continue to use the terms speaker and listener to be consistent throughout the paper.

Our analysis contributes to the literature on behavioral communication frictions. Caplin and Leahy (2004) study communication when the listener suffers a utility loss from receiving bad news and the speaker cares about the emotional impact his communication has on the listener. The listener takes no action after receiving information from the speaker. Caplin and Leahy (2004) argue that such an analysis requires the use of behavioral game theory, where the beliefs of the listener directly affect the utility of the speaker. They then show that communication could still work in these situations because no news would be interpreted as bad news, thus precluding strategic non-disclosure of painful information. Köszegi (2006) extends their analysis along several dimensions. First, he enriches the information structure so that the speaker can credibly argue that he may not have received information. Second, Köszegi (2006) considers an additional setup in which the listener takes an action after possibly receiving some information from the speaker. He shows that equilibria in these extended models might well involve strategic non-disclosure of bad news.

Further, Thaler (2023) studies the supply of unwanted information of a strategic speaker if the listener forms motivated beliefs. A speaker observes a state of the world that is either favorable or unfavorable for the listener and must then decide whether or not to reveal this information truthfully. The speaker cares for the probability with which the listener believes him to tell the truth. The listener finally receives additional utility from believing that the state of the world is favorable. He thus has an incentive to believe that the speaker tells the truth if his message is favorable, but that he lies if his message is unfavorable. Thaler (2023) shows that if the speaker and listener have different beliefs concerning the strength of motivated beliefs of the listener, there might exist an equilibrium in which the listener believes that the sender always speaks the truth while the sender always claims that the state of the world is favorable.

Unlike Caplin and Leahy (2004), Köszegi (2006), Thaler (2023), we study the effect of specific reputation concerns on the communication of unwanted information. The key distinguishing assumption of our paper is that the speaker has heterogeneous social preferences and these preferences are his private information. Therefore, the speaker's incentives to communicate depend on what the listener thereby infers about the speaker's preferences. Our new insight is that the existence of an equilibrium with information frictions may depend on the incentives of the speaker with stronger social preferences because the latter is less willing to exploit, and thus benefits less from, an intact social relationship.

Focusing on social prefernececs, our study also complements the literature on the effects of reputation concerns in cheap-talk games. In his seminal contribution, Sobel (1985) studies the communication behavior of a speaker who has private information about whether his preferences concerning the listener's action are aligned with the the listener. He shows that even speakers with non-aligned preferences may communicate truthfully to hide their conflict of interest with the listener. The reason is that the speaker may thus preserve her future influence over the listener. Bénabou and Laroque (1992) generalizes the model to a speaker with noisy private information, allowing for repeated manipulations and fluctuating reputations. Morris (2001) further extends their model by endogenizing the communication behavior of the speaker with aligned preferences. He finds that sometimes even a speaker with perfectly aligned preferences may have a reputational incentive to lie.

In contrast to Sobel (1985), Bénabou and Laroque (1992), and Morris (2001), information is not neutral in our setup, bad news directly reduce the listener's utility. Further, the speaker has private information about her social preferences regarding the pain caused by revealing an inconvenient truth. Thus, the communication of painful news may have effects on a rather distinct social interaction, modeled here as a trust game. Our results are consequently applicable to other, more private social situations. For example, people may hide painful information because they want to avoid a heated argument at the dinner table, not because they want to preserve their influence over their family's future decisions.

Finally, we contribute to a recent literature that examines the effect of reputation concerns on public discourse. Golman (2023) and Braghieri (2021) study communication when people want to express their true opinions but also care about their social reputation. Golman (2023) assumes that opinions are a function of beliefs and values, so expressing certain opinions may signal having certain values. He shows that the resulting signaling equilibria lead to endogenous communication norms. Further, because there can be multiple equilibria, changes in norms can abruptly affect communication. Finally, if some people can form motivated beliefs, different parts of the population can hold different beliefs, even if everyone observes the aggregate communication behavior of everyone involved in public discourse. Braghieri (2021) assumes that people suffer lying costs when they do not express their true opinions, but also care about their social reputation regarding the latter. He shows that miscommunication can occur in equilibrium when people are sufficiently concerned about their reputation. He also characterizes the conditions under which information is lost in public discourse. Unlike Golman (2023) and Braghieri (2021), we focus less on public discourse and more on private communication within close social networks. For example, our results apply to information frictions within work groups, families, and sports teams, where the relevant information may well be specific to the situation under consideration. For example, parents might be reluctant to provide their children with honest advise, where their private information is specific to the children. Furthermore, we consider information that has an emotional impact on the listener and explicitly model how the speaker might benefit from a favorable social reputation. The latter allows us to study which preference types have the strongest incentives to hide or reveal painful information.

3 Model

In this section, we describe the strategic situation that forms the foundation of our theory. We consider a dynamic game with two players and two stages. We call one player the speaker and the other player the listener. In the first stage, the speaker receives noisy information – no news or bad news – about a stochastic state of the world. The speaker decides whether to communicate bad news to the listener. Revealing bad news reduces the utility of the listener. After receiving the signal from the speaker, the listener chooses between actions whose payoffs depend on the state of the world. The preferences of both speaker and listener concerning these actions are perfectly aligned. The second stage is a binary trust game with the listener as the first-mover and the speaker as the second-mover. The speaker has private information on her social preferences: she is either egoistic or compassionate. If the speaker is compassionate, she suffers from hurting the listener by revealing bad news and is at the same time trustworthy in the trust game. If the speaker is egoistic, she does not care for the listener and is thus not trustworthy. Whether the speaker reveals bad news depends on her preferences and the expected behavioral response of the listener in the trust game. We analyze all pure-strategy perfect Bayesian equilibria of this game. We are particularly interested in equilibria in which the speaker might not reveal bad news in the first stage to appear trustworthy in the second stage. We next fill in all the details.

At the start of the strategic interaction, nature determines the state of the world. The state of the world $\theta \in \{g, b\}$ is either good or bad; it is good with prior probability ν_0. The speaker does not directly observe the state of the world but receives noisy information $i \in \{nn, bn\}$ that is either no news or bad news. If the state of the world is good, the speaker always receives no news. If the state is bad, the speaker receives bad news with probability $\lambda > 0$ and otherwise no news. Let $\tilde{\nu}_1(i)$ be the speaker's updated belief that the state of the world is good after receiving noisy information i. Bayes rule implies

$$\tilde{\nu}_1(bn) = 0 \quad \text{and} \quad \tilde{\nu}_1(nn) = \frac{\nu_0}{\nu_0 + (1 - \nu_0)(1 - \lambda)} \tag{1}$$

for the updated beliefs. Bayes' rule directly implies that bad news perfectly reveal the state to be bad. No news is, in some sense, good news for the speaker because it increases her belief that the state of the world is good. The precise posterior after no news depends on the probability λ with which there are no news even if the state of the world is bad.

After receiving the noisy signal, the speaker sends a signal s to the listener. If the speaker receives no news, she can only forward no news to the listener so that s equals nn. If the speaker receives bad news, she can send signal $s \in \{nn, bn\}$ and must thus decide between revealing bad news or pretending to have received no news. Revealing bad news reduces the listener's utility by some positive amount p. This payoff reduction p captures how receiving undesirable information hurts the listener.

The listener observes the speaker's signal and subsequently chooses an action $a \in \{g, b, o\}$ where the letters stand for betting on the good state, betting on the bad state, or choosing an outside option. An action $a \in \{g, b, o\}$ generates for both the speaker and the listener an identical, positive payoff $\pi(a, \theta) \in \{\ell, m, h\}$ with $\ell < m < h$. Unless the listener chooses the outside option, the payoff depends on whether the action matches the state of the world. To be precise, speaker and listener receive the high payoff h if the action $a \in \{g, b\}$ matches the state of the world, and the low payoff ℓ if the action does not match the state. The outside option o generates the intermediate payoff m independently from the state of the world.

The speaker and the listener receive only limited information or feedback concerning the first stage before progressing to the second stage. In particular, the speaker only observes her private signal on the state of the world and then chooses her signal. The speaker does not observe what action the listener chooses in response to her signal. The listener only observes the signal of the speaker and then chooses her action. The speaker and the listener receive information on whether the listener's action matches the state of the world only after the second stage.

In the second stage, the speaker and the listener play a binary trust game, with the listener as the first-mover and the speaker as the second-mover. Both start with some strictly positive initial wealth of w. The listener then decides whether to transfer nothing or her entire wealth w to the speaker. The transfer $t \in \{0, w\}$ is multiplied by some factor k strictly larger than one. If the speaker receives no transfer, she has nothing to decide because there is nothing to return. If the speaker receives the transfer, she decides whether to return either nothing or the fair amount $(k+1)\,\frac{w}{2}$ that equalizes the payoffs of speaker and listener in the trust game.

We next describe the utility functions of the speaker and the listener. We are interested in whether the speaker might communicate his trustworthiness by not revealing bad news. To simplify the analysis, we assume that the listener is egoistic, so her utility equals her own payoff from her decision concerning the state of the world, the possible pain from receiving bad news, and her payoff from the trust game. Let x denote the payoff of the listener. The speaker might have social preferences and thus differ in her trustworthiness. Again for simplicity, there are only two types of preferences $\tau \in \{e, c\}$ so that the speaker is either egoistic or compassionate. If the speaker is egoistic, her utility equals her payoff from the decision of the listener concerning the state of the world and the payoff from the trust game. Let y denote this overall payoff. If the speaker is compassionate, she cares for her payoff, but in addition, suffers if he receives a higher payoff y than the payoff x of the listener. Following the functional form of Fehr and Schmidt (1999), the compassionate speaker receives overall utility

$$y - \beta \max\{y - x, 0\} \tag{2}$$

where the parameter $\beta > 0$ captures her strength of compassion. Suffering from unfavorable inequality is irrelevant in our setup as the speaker always receives a weakly higher payoff than the listener. Following the literature, we assume that β is strictly smaller than one, so the speaker is unwilling to destroy her own payoff without any transfer to the listener to reduce inequality. Further, we assume that β is strictly larger than one-half so that a compassionate speaker returns trust in the trust game. The speaker's preferences are private information where it is common knowledge that the speaker is compassionate with probability $\mu_0 \in]0, 1[$. The speaker and the listener maximize their respective expected utilities.

We finally complete the model by describing strategies and the used equilibrium concept. We only consider pure strategies. A pure strategy for the speaker has two components: (i) what signal she sends, and (ii) whether she returns trust in the trust game, both conditional on her type. Concerning the first component, let $s(\tau)$ be the signal that a speaker of type τ

sends after receiving bad news. It is unnecessary to describe what signal the speaker sends after receiving no news because the speaker can then only forward no news. Concerning the second component of her strategy, let $r(i, s, \tau)$ be the return behavior of a speaker of type τ after receiving the transfer of the listener in the trust game, conditional on her noisy information and signal from the first stage. Specifying the speaker's return after not being trusted is unnecessary because the speaker then has no choice because there is nothing to return. A pure strategy for the listener also has two components: (i) what action to take, conditional on the received signal, and (ii) whether to trust the speaker, conditional on the received signal and the action previously chosen by the listener herself. Let $a(s)$ be the action and $t(s, a)$ the amount sent in the trust game by the listener after receiving signal s and having chosen action a. Important will also be the beliefs the listener forms after seeing the signal of the speaker. Let $\mu_1(s)$ be the probability with which the listener believes the speaker to be compassionate, and let $\nu_1(s)$ be the probability with which the listener believes the state of the world to be good, both conditional on receiving signal s.

We use perfect Bayesian equilibrium as our equilibrium concept. In such a perfect Bayesian equilibrium, the equilibrium strategy $s^*(i, \tau)$ and $r^*(i, s, \tau)$ maximizes the expected utility of each type $\tau \in \{e, c\}$ of the speaker, for either information $i \in \{nn, bn\}$, for both signals $s \in \{nn, bn\}$, and given the equilibrium strategy of the listener. The equilibrium strategy $a^*(s)$ and $t^*(s, a)$ maximize the expected utility of the listener given the equilibrium strategy of the speaker and given the equilibrium beliefs $\mu_1^*(s)$ and $\nu_1^*(s)$ of the listener. Finally, the listener's equilibrium beliefs are consistent with the equilibrium strategy of the speaker and the prior probability for the speaker to be compassionate. The latter condition implies that the listener uses Bayes' rule to update her beliefs on the equilibrium path.

4 Results

We next derive all perfect Bayesian equilibria in pure strategies. We first analyze the trust game, then how the listener responds to the message received in the first stage of the strategic interaction, and finally the incentives of the speaker to reveal bad news.

4.1 Trust Game

We first analyze equilibrium behavior in the trust game. Due to the preferences and the structure of the strategic interaction, behavior in the first stage only affects optimal behavior in the trust game by changing the listener's beliefs. The listener is egoistic and considers the payoffs generated in the trust game in isolation. She optimally trusts if and only if she expects the speaker to be sufficiently trustworthy, no matter her expected payoffs from the first stage. The egoistic speaker never returns trust, again irrespectively of what she hopes to receive from the first stage.

The situation could be more complicated for the compassionate speaker because expected inequality from the first stage might affect her behavior in the second stage. However, the binary nature of the trust game precludes such complications. The speaker and the listener receive the same payoff generated by the listener's action. Any first-stage inequality arises only via the listener's suffering from receiving bad news. Further, the speaker always receives weakly more than the listener, no matter what happens in either stage. Any suffering from bad news induces favorable inequality from the speaker's perspective. Even if there is no additional inequality from the first stage, we assume a compassionate speaker to return trust. A compassionate speaker has even stronger incentives to return trust if there is additional inequality from the first stage. However, because the trust game is binary, she cannot return more than the fair amount $(k+1)\frac{w}{2}$ that equalizes payoffs in the trust game. Thus, the first-stage payoffs are essentially sunk and do not affect optimal behavior in the second stage in the theory, not even for a compassionate speaker. Intuitively, an egoistic speaker never, while a compassionate speaker always returns trust. And the listener optimally trusts the speaker if and only if she is sufficiently convinced that the speaker is compassionate. We summarize these findings in the following results. All formal proofs are in the appendix.

Result 1 (Trustworthiness) *In any perfect Bayesian equilibrium, the speaker returns nothing in the trust game if she is egoistic and returns the fair amount if she is compassionate,* $r^*(i,s,e) = 0$ *and* $r^*(i,s,c) = (k+1)\frac{w}{2}$.

Result 2 (Trust) *In any perfect Bayesian equilibrium, the listener after observing signal* $s \in \{nn, bn\}$ *optimally trusts the speaker if and only if she is sufficiently convinced that the speaker is compassionate,* $\mu_1^*(s) \geq \frac{2}{k+1}$.

Both an egoistic and a compassionate speaker prefers to be trusted because trust increases her second-stage utility. Note that an egoistic speaker's second-stage utility increases from w to $(1+k)\,w$ while a compassionate speaker's second-stage utility increases by a smaller amount from w to $(1+k)\frac{w}{2}$. The reason is that a compassionate speaker returns trust and thus benefits less from being trusted. Finally, we make the following simplifying assumption.

Assumption 1 (Trust) *Trust in the trust game is sufficiently efficient so that the listener optimally trusts the speaker given her prior belief,* $\mu_0 > 2/(k+1)$.

We are interested in an equilibrium in which the speaker hides bad news not to lose trust. If there is no trust given the listener's prior, there is nothing to lose for an egoistic speaker, and such a pooling equilibrium cannot exist.

4.2 Listener's Action

We next analyze how the listener responds to the speaker's signal in the first stage. How the listener responds to the signal cannot influence the speaker's behavior in the trust game because the latter does not observe the action chosen by the listener until the end of the interaction. The listener thus responds to the speaker's signal by maximizing her expected payoff in the second stage. A rational listener understands that bad news reveal the bad state. Thus, the listener optimally responds to bad news by choosing action b. How the listener responds to no news depends on how this signal affects her belief concerning the state of the world. We make the following assumption to make the environment reasonably informative.

Assumption 2 (Informativeness) *The prior probability* ν_0 *for the state of the world to be good is sufficiently low and the precision* λ *of the noisy information on the state of the world is sufficiently high so that*

$$\nu_0 < \frac{m-\ell}{h-\ell} < \frac{\nu_0}{\nu_0 + (1-\nu_0)(1-\lambda)} \tag{3}$$

holds.

These assumptions imply that the equilibrium behavior of the listener after receiving no news can depend on the speaker's behavior after receiving bad news. If the first inequality is not satisfied, it is always optimal for the listener to choose the good action after observing no news in any equilibrium. If the second inequality is not satisfied, it is never optimal for the listener to choose the good action after observing no news in any equilibrium, not even in an equilibrium in which the speaker always reveals bad news. The last condition holds if the precision of the speaker's information is sufficiently high. The assumption requires that the high payoff h is in an intermediate range: it is not too high so that the outside option is sometimes optimal, and it is not too low so that the good action can sometimes be optimal. We then get the following result.

Result 3 (Listener Action) *In any perfect Bayesian equilibrium, the listener chooses the bad action after receiving bad news, $a^*(bn) = b$. After receiving no news, the listener*

1. *optimally chooses the bad action if and only if she is sufficiently convinced that the state is bad, $\nu_1^*(nn) \leq (h - m)/(h - \ell)$,*

2. *optimally chooses the good action if and only if she is sufficiently convinced that the state is good, $\nu_1^*(nn) \geq (m - l)/(h - \ell)$, and*

3. *optimally chooses the outside option for intermediate beliefs $\nu_1^*(nn)$.*

The listener optimally chooses the bad action if she is sufficiently convinced that the state of the world is bad; she chooses the good action if she is sufficiently convinced that the state of the world is good. By the assumption that the outside option is optimal given the prior belief, there exists an intermediate belief range where the listener optimally chooses the outside option.

4.3 Speaker's Communication Choice

We next analyze whether the speaker reveals or hides bad news. First, we consider the pooling equilibrium in which both an egoistic and a compassionate speaker never reveal bad news. This equilibrium requires out-of-equilibrium beliefs such that the listener loses trust if a speaker reveals bad news. Specifically, the listener learns nothing about the state of the world and chooses the outside option after receiving no news. The speaker knows that the state is bad after receiving bad news. Revealing bad news leads the listener to change to the action exactly matching the bad state of the world. Thus, the speaker can improve the first-stage decision by revealing the truth. However, revealing bad news also has a downside. For an egoistic speaker, this downside is the loss of trust in the second stage. A compassionate speaker also suffers from a loss of trust, but this loss is smaller because she returns the fair amount. However, a compassionate speaker directly dislikes the pain he causes by revealing bad news. There are consequently two different incentive constraints, one for the egoistic and one for the compassionate speaker. We have the following result.

Result 4 (Pooling Equilibrium: Nobody Reveals Bad News) *There exists a perfect Bayesian pooling equilibrium satisfying Assumptions 1 and 2 in which an egoistic and compassionate speaker both never reveal bad news if and only if*

1. *losing trust is more important for an egoistic speaker than improving the first-stage decision,* $k\,w \geq h - m$*, and*

2. *losing trust and hiding bad news is more important for a compassionate speaker than improving the first-stage decision,* $\beta\,p + \frac{1}{2}(k-1)\,w \geq h - m$*.*

Next, we consider the pooling equilibrium in which both an egoistic and a compassionate speaker always reveal bad news. The listener then never learns anything about the speaker, no matter what news she observes. The speaker cannot affect the listener's trust. Because the speaker truthfully transmits all information, the listener always takes the universally preferred first-stage action. An egoistic speaker clearly has no reason to hide bad news. A compassionate speaker, however, might be tempted to hide bad news to avoid the pain she causes by revealing an inconvenient truth. The equilibrium thus exists if and only if avoiding empathetic suffering is less important for a compassionate speaker than ensuring the right action. We have the following result.

Result 5 (Pooling Equilibrium: Everybody Reveals Bad News) *There exists a perfect Bayesian pooling equilibrium satisfying Assumptions 1 and 2 in which an egoistic and a compassionate speaker always reveal bad news if and only if hiding bad news is less important for a compassionate speaker than ensuring the optimal first-stage decision,* $\beta\,p \leq h - \ell$*.*

Next, we consider the separating equilibrium in which only an egoistic speaker reveals bad news. Observing no news then makes the listener more – but not completely – convinced that the speaker is compassionate. The reason is that an egoistic speaker sometimes sends no news because she sometimes receives no news. Assumption 1 implies that the listener already trusts the speaker given her prior. In the separating equilibrium, the listener, therefore, trusts the speaker after observing no news. Bad news reveals that the speaker is egoistic, and the listener responds by not trusting. The speaker must thus take into account that revealing bad news destroys trust. However, revealing bad news also has a positive consequence because it improves the listener's choice in the first stage. How much depends on what the listener chooses after observing no news. In contrast to the pooling equilibrium with the revelation of bad news, receiving no news in the separating equilibrium is a diffused signal because a compassionate speaker never reveals bad news. The listener optimally chooses the outside option after observing no news in this separating equilibrium if the speaker is ex-ante sufficiently likely to be compassionate. Otherwise, the listener chooses the good action. The separating equilibrium exists if only a selfish speaker has incentives to reveal bad news. Such an equilibrium might exist because a compassionate speaker has weaker incentives to reveal bad news. The reason is that although revealing bad news improves the listener's fist-stage decision, a compassionate speaker wants to avoid causing the listener pain. We derive the following result.

Result 6 (Separating Equilibrium: Only Egoist Reveals Bad News) *There exists a perfect Bayesian separating equilibrium satisfying Assumptions 1 and 2 in which only an egoistic speaker reveals bad news if and only if*

1. *improving the first-stage decision is more important for a egoistic speaker than gaining trust,* $h - \pi(a^*(nn), b) \geq kw$*, and*

2. *gaining trust and hiding bad news is more important for a compassionate speaker than improving the first-stage decision,* $\beta\, p + \frac{1}{2}(k-1)\, w \geq h - \pi(a^*(nn), b)$.

Finally, it is easy to show that there never exists a separating equilibrium in which only a compassionate speaker reveals bad news. The listener then chooses the outside option after observing no news, and she trusts the speaker after observing bad news. An egoistic speaker then has incentives to deviate and reveal bad news. The reason is that revealing bad news strictly improves the listener's first-stage decision with no downside as it either leaves trust unaffected or even increases trust. We summarize this result as follows.

Result 7 (Separating Equilibrium: Only Compassionate Reveals Bad News) *There never exists a perfect Bayesian separating equilibrium satisfying Assumptions 1 and 2 in which only a compassionate speaker reveals bad news.*

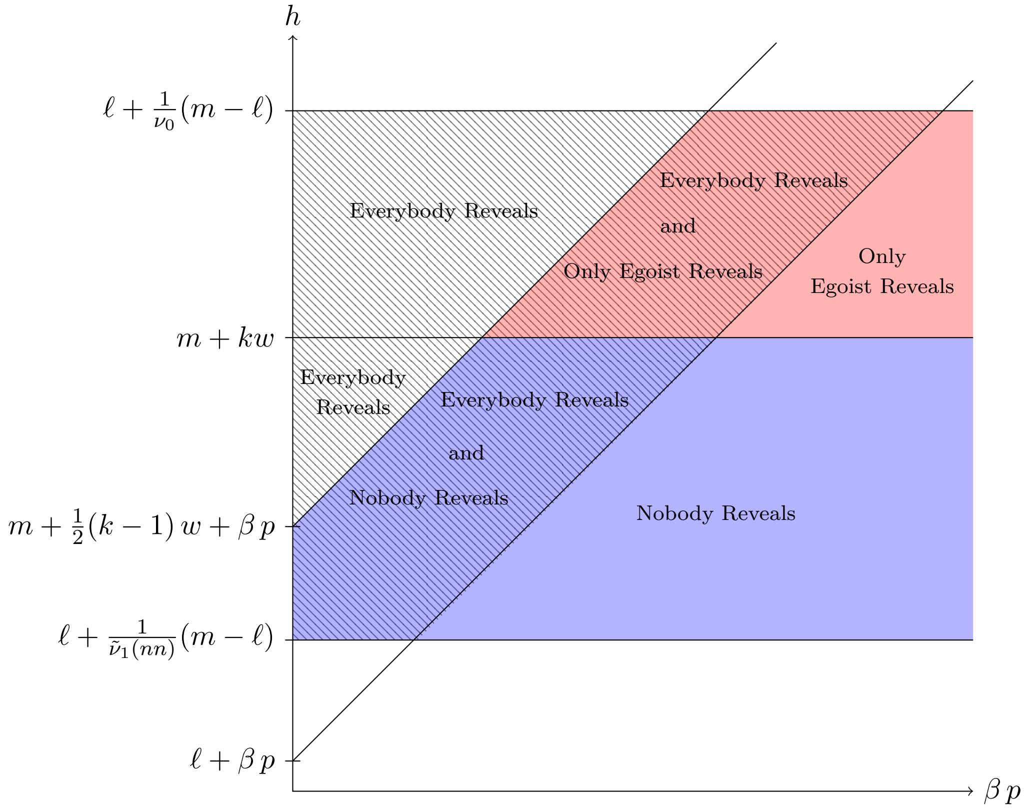

Figure 1: Equilibrium Existence.

Figure 1 illustrates our theoretical results for the case in which the speaker is sufficiently likely to be compassionate so that the listener chooses the outside option after receiving no news in the separating equilibrium. Whether a particular equilibrium exists then depends on (i) the gains from implementing the optimal first-stage action by revealing bad news, (ii) the pain a compassionate speaker feels when revealing bad news, and (iii) the second-stage gains from appearing trustworthy. We put the high payoff h on the vertical axis because it directly translates into the gains from implementing the right first-stage action by revealing bad news. We put the compassionate pain from revealing bad news $\beta\,p$ on the horizontal axis. Note that the pain p that the listener feels when receiving bad news can be arbitrarily small or large. The benefits of appearing trustworthy depend on the wealth w, which in the above illustration only determines specific cutoff values but is not varied to keep the illustration simple. Assumption 2 requires that the high payoff h is in an intermediate range, as indicated by the upper and lower horizontal lines. Remember that $\tilde{\nu}_1(i)$ is the speaker's updated belief that the state of the world is good after receiving no news.

Consider the incentives of an egoistic speaker to reveal bad news. These incentives depend on whether the listener believes the bearer of bad news to be egoistic, an endogenous property of the equilibrium. In the pooling equilibrium where everybody reveals bad news, there is no cost for an egoistic speaker to do so because it improves the listener's first-stage action without eroding trust. In the separating and the pooling equilibrium where nobody reveals bad news, an egoistic speaker faces a trade-off: revealing bad news improves the listener's first-stage action, but also destroys trust. It is then optimal for an egoistic speaker to hide bad news if the benefit $h - m$ from ensuring the optimal first-stage action is smaller than the benefit $k\,w$ from preserving trust. In Figure 1, this is the case if and only if h is weakly above the horizontal line $m + k\,w$. Note that the incentives for an egoistic speaker do not depend on the pain p caused by the disclosure of bad news.

The incentives of a compassionate speaker to reveal bad news are quite different from those of an egoistic speaker. In the pooling equilibrium where everyone reveals bad news, doing so has negative side effects to a compassionate speaker because he suffers from inflicting pain on the listener. It is optimal for a compassionate speaker to reveal bad news if and only if the positive effect of revealing bad news on the listener's first-stage action $h - \ell$ exceeds the empathic suffering from bringing bad news $\beta\,p$. In Figure 1, this is the case if and only if h is weakly above the diagonal $\ell + \beta\,p$.

Further, in the pooling equilibrium where nobody reveals bad news, a compassionate speaker might have stronger incentives to reveal bad news than an egoistic speaker. The reason is that a compassionate speaker benefits less from trust because he does not exploit the listener and instead returns the fair share of the transfer. The equilibrium thus exists if the benefit $h - m$ from ensuring the optimal first-stage action smaller than the benefit from preserving trust $\frac{1}{2}(k-1)\,w$ and the avoided empathetic suffering $\beta\,p$. In Figure 1, this is the case if and only if h is below the diagonal $m + \frac{1}{2}(k-1)\,w + \beta\,p$. This argument explains why there is this triangle of parameter combinations to the left in which the pooling equilibrium with hiding of bad news breaks down due to the compassionate speaker speaking up.

Combining the above arguments, the pooling equilibrium in which both an egoistic and a compassionate speaker never reveal bad news exists for parameter combinations in the blue area. The pooling equilibrium in which both an egoistic and a compassionate speaker reveal bad news exists for parameter combinations in the hatched area. The separating equilibrium in which only an egoistic speaker reveals bad news exists for parameter combinations in the red area.

We can now explore the comparative statics concerning the level of pain p induced by the disclosure of negative information. Intuitively, we would expect social preferences to more strongly influence communication only when there is more pain, when the bad news more strongly affects the recipient's utility. In fact, if the disclosure of bad news causes minimal or no distress to the listener, it ceases to be "bad" news and becomes instead useful and welcomed information. In such cases, a compassionate speaker should not experience a loss in their own utility from sharing the negative news, and there is no compelling reason to interpret the act of revealing bad news as an indication of the speaker's selfishness. Figure 1 confirms this intuition. First, we observe that a higher value of p expands the blue area, which in turn broadens the range of parameter combinations for which there exists a pooling equilibrium with no disclosure of bad news. Second, the pooling equilibrium with full disclosure only exists for relatively small, and the pooling equilibrium with no disclosure exists only for relatively large values of p.

Further, Figure 1 shows that multiple equilibria coexist for a wide range of parameter combinations. Standard equilibrium refinements based on forward induction like equilibrium dominance and the intuitive criterion by Cho and Kreps (1987) have no bite in the present setup. These refinements work by restricting out-of-equilibrium beliefs. There are no out-of-equilibrium beliefs in the separating and pooling equilibrium where all reveal bad news, so these refinements never rule out these equilibria. In the pooling equilibrium where everybody hides bad news, the revelation of bad news is indeed an out-of-equilibrium action, and the corresponding pessimistic out-of-equilibrium beliefs are indeed crucial for equilibrium existence. But the best possible response to revealing bad news for an egoistic speaker is that the listener bets on the bad state of the world and retains trust. The resulting maximum rationalizable deviation payoff is $h + (k+1)\,w$ which is strictly larger than the equilibrium payoff $m + (k+1)\,w$ of an egoistic speaker. Consequently, revealing bad news can never be equilibrium dominated for the egoistic speaker, so the pessimistic beliefs after observing bad news survive the standard belief refinements.

Interestingly, the pooling equilibrium with revelation of bad news might Pareto-dominate the pooling equilibrium with no information disclosure if compassionate speakers have very strong social preferences and it is not that much better for the listener to make the optimal decision in case of the bad state of the world as compared to choosing the outside option. Second-stage equilibrium behavior and thus the related expected equilibrium payoffs are identical in both equilibria: the listener trusts and only a compassionate speaker returns trust. Consider the first stage. If the speaker receives no news, the resulting signal leads to the listener betting on the good state in the pooling equilibrium with revelation of bad news, while she chooses the outside option in the pooling equilibrium with no information disclosure. Assumption 2 implies that, conditional on receiving no news, betting on the good states maximizes the expected payoffs for speaker and listener. The pooling equilibrium with information disclosure leads to better decisions after no news as compared to the pooling equilibrium with no information disclosure.

To complete the argument, we must take into account that receiving bad news reduces the utility of the listener. In the pooling equilibrium with information disclosure, the speaker reveals bad news and the listener correctly bets on the bad state of the world. The first-stage equilibrium payoffs are h for an egoistic speaker, $h - \beta\, p$ for a compassionate speaker, and $h - p$ for the listener. In the pooling equilibrium with no information disclosure, the speaker hides the bad news, the listener chooses the outside option, and the equilibrium payoffs are m for both types of speaker and the listener. An egoistic speaker clearly prefers the pooling equilibrium with information disclosure because h exceeds m. Also a compassionate speaker might prefer the pooling equilibrium with information disclosure if the intermediate payoff m from choosing the outside option is close to the low payoff ℓ from betting on the wrong state of the world. This actually follows from the incentive constraint: a compassionate speaker in this equilibrium must prefer revealing bad news over transmitting no news, which holds if and only if $h - \beta\, p$ exceeds ℓ. If this inequality is strict, and if m is sufficiently close to ℓ, then $h - \beta\, p$ also exceeds m. Finally, even the listener prefers to receive bad news rather than new news if a compassionate speaker has very strong social preferences and strictly prefers the pooling equilibrium with information disclosure. Formally, $h - \beta\, p > m$ implies $h - p > m$ for parameter values β sufficiently close to one. A very compassionate speaker almost fully internalizes the pain he imposes on the speaker, so his interests are almost perfectly aligned with the latter.

5 Conclusion

We develop a strategic setup to study the influence of social preferences on the disclosure of painful truths in social relationships. Our theory shows that social reputation concerns might generate communication frictions. Our analysis confirms the intuitive argument and derives the precise parameter combinations for which different perfect Bayesian equilibria with more or less information disclosure exist. The model suggests that we might observe severe information frictions in situations where people care for their perceived trustworthiness, extending arguments from the existing literature to different social contexts where it is more important to come across as friendly and prosocial rather than competent and knowledgable. The analysis also suggests that we should observe stronger communication frictions if unwelcome truths are more painful, that more trustworthy individuals might actually be more tempted to reveal bad news because they benefit less from a functioning social relationship, and that communication problems might be unnecessarily severe because there might be multiple equilibria and people might coordinate on an inefficient equilibrium.

As a final afterthought, the inefficient equilibria with strategic silence might be quite persistent because they could be supported by wrong out-of-equilibrium beliefs, see Fudenberg and Levine (1993) and Esponda and Pouzo (2016). If people wrongly believe that disclosing bad news leads to a breakdown of trust, they remain silent. But then they can actually never learn that speaking up might have no negative consequences. Such non-equilibrium arguments are, by construction, not explicitly considered in the present equilibrium analysis but could be an interesting avenue for fruitful future research.

References

Bénabou, R. and Laroque, G. (1992). Using privileged information to manipulate markets: Insiders, gurus, and credibility, *Quarterly Journal of Economics* **107**(3): 921–958.

Braghieri, L. (2021). Political correctness, social image, and information transmission, mimeo, Bocconi University pp. 1–100.

Caplin, A. and Leahy, J. (2004). The supply of information by a concerned expert, *Economic Journal* **114**: 487–505.

Cho, I.-K. and Kreps, D. M. (1987). Signaling games and stable equilibria, *Quarterly Journal of Economics* **102**(2): 179–221.

Conlee, M. C. and Tesser, A. (1973). The effects of recipient desire to hear on news transmission, *Sociometry* **36**(4): 588–599.

Esponda, I. and Pouzo, D. (2016). Berk-nash equilibrium: A framework for medling agents with misspecified models, *Econometrica* **84**(3): 1093–1130.

Fudenberg, D. and Levine, D. K. (1993). Self-confirming equilibrium, *Econometrica* **61**(3): 523–545.

Golman, R. (2023). Acceptable discourse: Social norms of beliefs and opinions, *European Economic Review* **160**: 104588.

Golman, R., Hagman, D. and Loewenstein, G. (2017). Information avoidance, *Journal of Economic Literature* **55**(1): 96–135.

Köszegi, B. (2006). Emotional agency, *Quarterly Journal of Economics* **121**(1): 121–155.

Milliken, F. J., Morrison, E. W. and Hewlin, P. F. (2003). An exploratory study of employee silence: Issues that employees don't communicate upward and why, *Journal of Management Studies* **40**(6): 1453–1476.

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

Rosen, S. and Tesser, A. (1970). On reluctance to communicate undesirable information: The MUM effect, *Sociometry* **33**(3): 253–263.

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

Thaler, M. (2023). The supply of motivated beliefs, mimeo, University College London .

Appendix A: Formal Proofs

Proof of Result 1 (Trustworthiness)

Let $E\pi(a,\theta)$ be the speaker's expected first-stage payoff ignoring any pain from revealing bad news. An egoistic speaker optimally returns nothing in the trust game if and only if

$$E\pi(a,\theta) + (k+1)w \geq E\pi(a,\theta) + \frac{(k+1)\,w}{2} \tag{4}$$

irrespectively of the signal she sent. This condition holds with a strictly inequality because we assume the initial wealth w in the trust game to be strictly positive. Let $\mathbb{I}(s)$ be an indicator function that is equal to one if and only if the speaker has revealed bad news and zero otherwise. A compassionate speaker then optimally returns the fair amount if and only if

$$E\pi(a,\theta) + \frac{(k+1)\,w}{2} - \mathbb{I}(s)\beta\, p \geq E\pi(a,\theta) + (k+1)\,w - \beta(k+1)\,e - \mathbb{I}(s)\beta\, p \tag{5}$$

which holds with strict inequality because β is strictly larger than one-half and w is strictly larger than zero by assumption. The value of the indicator function does not matter because the empathetic suffering from revealing bad news cannot be reduced by returning more than the fair amount. These arguments imply the equilibrium behavior of the speaker. *Q.E.D.*

Proof of Result 2 (Trust)

Let $E\pi(a,\theta)$ be the listener's expected first-stage payoff ignoring any pain from revealing bad news, and let $\mathbb{I}(s)$ be again an indicator function that is equal to one if and only if the speaker has revealed bad news. Given the equilibrium strategy of the speaker from Result 1 and the listener's own belief $\mu_1^*(s)$ concerning the speaker's type, the listener optimally trust the speaker if and only if

$$E\pi(a,\theta) + \mu_1^*(n)\frac{k+1}{2}w - \mathbb{I}(s)\,p \geq E\pi(a,\theta) + w - \mathbb{I}(s)\,p \tag{6}$$

where the indicator function simply cancels. Solving the inequality gives the conditions for the optimal trusting behavior as stated in the result. *Q.E.D.*

Proof of Result 3 (Listener Action)

Because the listener is egoistic and cannot influence the speaker's behavior in the next stage, her optimal action maximizes her expected first-stage payoff. The listener optimally chooses the bad action after bad news because bad news perfectly reveal the bad state and

$$h - p > m - p. \tag{7}$$

After receiving no news, the best alternative to the outside option depends on the speaker's belief $\nu_1^*(nn)$. The listener prefers the good action over the bad action if and only if the good state is weakly more likely, $\nu_1^*(nn) \geq 1/2$. The overall optimal action follows from comparing the maximum payoff from the good or the bad action with the constant payoff m from the outside option. If the bad state is more likely, the best alternative to the outside option is the bad action, and this bad action is overall optimal if and only if

$$h - \nu_1^*(nn)(h - \ell) \geq m \tag{8}$$

which yields the lower cutoff in the result. If the good state is more likely, the best alternative to the outside option is the good action, and this good action is overall optimal if and only if

$$\ell + \nu_1^*(nn)(h - \ell) \geq m \tag{9}$$

which yields the upper cutoff in the result. Finally, Assumption 2 implies

$$m > \max\{\ell + \nu_0(h - \ell),\, h - \nu_0(h - \ell)\} \tag{10}$$

so that the listener prefers the outside option to the best alternative given her prior, which also implies that the upper cutoff in the result is strictly larger than the lower cutoff. Thus, there exists a range of intermediary beliefs in which the listener optimally chooses the outside option *Q.E.D.*

Proof of Result 4 (Nobody Reveals Bad News)

In the equilibrium in which the speaker never reveals bad news,

$$s^*(e) = nn \quad \text{and} \quad s^*(c) = nn \tag{11}$$

$$r^*(i, s, e) = 0 \quad \text{and} \quad r^*(i, s, c) = \frac{(k+1)\,w}{2} \tag{12}$$

for all $s \in \{nn, bn\}$ and $i \in \{nn, bn\}$ characterize the behavior of the speaker, and

$$a^*(nn) = o \quad \text{and} \quad a^*(bn) = b \tag{13}$$

$$t^*(nn, a) = w \quad \text{and} \quad t^*(bn, a) = 0 \tag{14}$$

for any $a \in \{g, b, o\}$ characterize the behavior and

$$\mu_1^*(nn) = \mu_0 \quad \text{and} \quad \mu_1^*(bn) \leq 2/(k+1) \tag{15}$$

$$\nu_1^*(bn) = 0 \quad \text{and} \quad \nu_1^*(nn) = \nu_0 \tag{16}$$

characterize the beliefs of the listener. The listener's beliefs are consistent with her prior belief and the equilibrium behavior of the speaker. Observing bad news is out-of-equilibrium so that the belief $\mu_1^*(bn)$ is arbitrary; it only needs to be pessimistic enough to make not trusting optimal after observing bad news. The listener's action given the observed signal is optimal given her conditional beliefs and Result 3, and her trust behavior is optimal given her beliefs and Result 2. The return behavior of the speaker in the trust game is optimal given Result 1. The equilibrium, thus, exists if and only if the signaling behavior is optimal given the equilibrium behavior of the listener. Not revealing bad news is optimal for an egoistic speaker if and only if

$$m + (k+1)\,w \geq h + w \tag{17}$$

and it is optimal for a compassionate speaker if and only if

$$m + \frac{(k+1)\,w}{2} \geq h + w - \beta p \tag{18}$$

which yields the two conditions in the result after rearranging. *Q.E.D.*

Proof of Result 5 (Everybody Reveals Bad News)

In the equilibrium in which the speaker always reveals bad news,

$$s^*(e) = bn \quad \text{and} \quad s^*(c) = bn \tag{19}$$

$$r^*(i,s,e) = 0 \quad \text{and} \quad r^*(i,s,c) = \frac{(k+1)\,w}{2} \tag{20}$$

for all $s \in \{nn, bn\}$ and $i \in \{nn, bn\}$ characterize the behavior of the speaker, and

$$a^*(nn) = g \quad \text{and} \quad a^*(bn) = b \tag{21}$$

$$t^*(nn,a) = w \quad \text{and} \quad t^*(bn,a) = w \tag{22}$$

for any $a \in \{g, b, o\}$ characterize the behavior and

$$\mu_1^*(nn) = \mu_0 \quad \text{and} \quad \mu_1^*(bn) = \mu_0 \tag{23}$$

$$\nu_1^*(bn) = 0 \quad \text{and} \quad \nu_1^*(nn) = \frac{\nu_0}{\nu_0 + (1-\nu_0)(1-\lambda)} \tag{24}$$

characterize the beliefs of the listener. Just as in the proof of the previous results, the listener's beliefs are consistent with her prior belief and the equilibrium behavior of the speaker, the behavior of the listener is optimal given her beliefs and the equilibrium behavior of the speaker, and the behavior of the speaker in the trust game is optimal given the equilibrium behavior of the listener. Equilibrium existence boils down to the question whether always revealing bad news is optimal. Revealing bad news is optimal for an egoistic speaker if and only if

$$h + (k+1)\,w \geq \ell + w \tag{25}$$

which is always satisfied. Revealing bad news is optimal for a compassionate speaker if and only if

$$h - \beta p + \frac{(k+1)\,w}{2} \geq \ell + \frac{(k+1)\,w}{2} \tag{26}$$

which yields the condition in the result after rearranging. *Q.E.D.*

Proof of Result 6 (Only Egoist Reveals Bad News)

In the equilibrium in which only an egoistic speaker reveals bad news,

$$s^*(e) = bn \quad \text{and} \quad s^*(c) = nn \tag{27}$$

$$r^*(i,s,e) = 0 \quad \text{and} \quad r^*(i,s,c) = \frac{(k+1)\,w}{2} \tag{28}$$

for all $s \in \{nn, bn\}$ and $i \in \{nn, bn\}$ characterize the behavior of the speaker, and $a^*(bn) = b$ and

$$a^*(nn) \begin{cases} = o & \text{if} \quad \nu_1^*(nn) < (h-m)/(h-\ell) \\ \in \{o,g\} & \text{if} \quad \nu_1^*(nn) = (h-m)/(h-\ell) \\ = g & \text{if} \quad \nu_1^*(nn) > (h-m)/(h-\ell) \end{cases} \tag{29}$$

and

$$t^*(nn,a) = w \quad \text{and} \quad t^*(bn,a) = 0 \tag{30}$$

for any $a \in \{g, b, o\}$ characterize the behavior and

$$\mu_1^*(bn) = 0 \quad \text{and} \quad \mu_1^*(nn) = \frac{\mu_0}{\mu_0 + (1-\mu_0)(1-\lambda)} \tag{31}$$

$$\nu_1^*(bn) = 0 \quad \text{and} \quad \nu_1^*(nn) = \frac{\nu_0}{\nu_0 + (1-\nu_0)\big(\mu_0 + (1-\mu_0)(1-\lambda)\big)} \tag{32}$$

characterize the beliefs of the listener. Observing no news is a noisy indication that the speaker is more likely to be compassionate. Because the listener optimally trusts given her prior belief by Assumption 1, she optimally trusts in this equilibrium after observing no news because the latter implies even more positive beliefs. The listener's optimal action after observing no news depends on how much this convinces her that the state of the world is more likely to be good, where Assumption 2 does not imply that the listener then optimally chooses the good action. Just as in the proof of the previous results, the listener's beliefs are consistent with her prior and the equilibrium behavior of the speaker, the behavior of the listener is optimal given her beliefs and the equilibrium behavior of the speaker, and the behavior of the speaker in the trust game is optimal given the equilibrium behavior of the listener.

Equilibrium existence boils down to the signaling incentives for the speaker. Revealing bad news is optimal for an egoistic speaker if and only if

$$h + w \geq \pi(a^*(nn), b) + (k+1)\, w \tag{33}$$

while not revealing bad news is optimal for a compassionate speaker if and only if

$$\pi(a^*(nn), b) + \frac{(k+1)\, w}{2} \geq h - \beta p + w \tag{34}$$

which yields the conditions in the result after rearranging. The precise conditions actually not reported in the result follow from plugging in the equilibrium action chosen by the listener after no news into the payoff function π given that the state of the world is bad. *Q.E.D.*

Proof of Result 7 (Only Compassionate Reveals Bad News)

In the equilibrium in which only a compassionate speaker reveals bad news,

$$s^*(e) = nn \quad \text{and} \quad s^*(c) = bn \tag{35}$$

$$r^*(i, s, e) = 0 \quad \text{and} \quad r^*(i, s, c) = \frac{(k+1)\, w}{2} \tag{36}$$

for all $s \in \{nn, bn\}$ and $i \in \{nn, bn\}$ characterize the behavior of the speaker, and

$$a^*(nn) \in \{g, o\} \quad \text{and} \quad a^*(bn) = b \tag{37}$$

$$t^*(nn, a) \in \{0, w\} \quad \text{and} \quad t^*(bn, a) = w \tag{38}$$

for any $a \in \{g, b, o\}$ characterize the behavior and

$$\mu_1^*(nn) = \mu_0 \quad \text{and} \quad \mu_1^*(bn) = 1 \tag{39}$$

$$\nu_1^*(bn) = 0 \quad \text{and} \quad \nu_1^*(nn) = \frac{\nu}{\nu + (1-\nu)\big(\mu_0\,(1-\lambda) + (1-\mu_0)\big)} \tag{40}$$

characterize the beliefs of the listener. Note that the listener's optimal action and trust decision after no news depend on her belief after receiving no news. However, we need not make a case distinction here because the precise behavior is irrelevant for the remaining proof. The separating equilibrium only exists if an egoistic speaker has incentives no to reveal bad news, and thus only if

$$\ell + w + t^*(nn, a^*(nn))\, k \geq h + w + k\, w \tag{41}$$

holds. This condition is always violated because the listener never chooses the optimal bad action after receiving no news, and revealing bad news at least weakly increases trust. Thus, such a separating equilibrium never exists. *Q.E.D.*