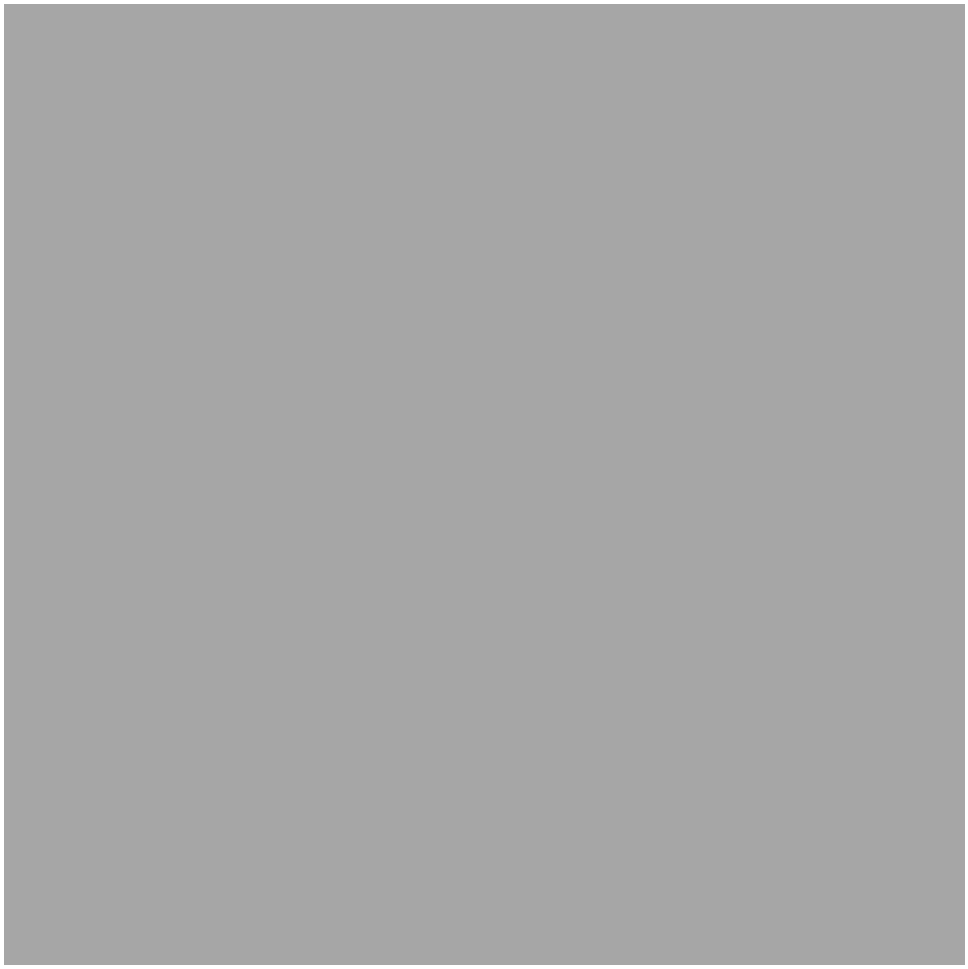

DISLIKING TO DISAGREE: IMPLICATIONS OF DISAGREEMENT AVERSION FOR INFORMATION DISCLOSURE

Kiryl Khalmetskiy
Mark T. Le Quement
Florian Hoffmann

Working Paper
No 293

NES Working
Paper series

January
2026

Disliking to disagree: Implications of disagreement aversion for information disclosure*

Kiryl Khalmetski† Mark T. Le Quement‡ Florian Hoffmann§

January 22, 2026

Abstract

We formalize the notion of belief homophily under asymmetric information by introducing a preference for perceived disagreement aversion. We study its implications for information sharing in a disclosure model where a privately informed sender and an uninformed receiver have heterogeneous priors, while the sender is averse to the receiver's perceived disagreement. Equilibrium disclosure is partial and tends to confirm the prior mean of the more confident player. The receiver learns more from senders whose prior means differ more from his own but whose prior variances are more similar. Perceived disagreement aversion implies qualitatively reverse predictions than aversion to actual disagreement.

Keywords: strategic disclosure, psychological games, disagreement aversion

JEL classification: D81, D83, D91

1 Introduction

Decentralized information exchange, for example in the form of informal political talk within social networks, is an important source of information for many. Yet, it is not bias free. An important role in generating these biases can be attributed to the tendency to avoid conflict in opinions. For example, a robust pattern is that people avoid conflictual topics and prefer to interact with individuals who share their worldview, for fear of creating interpersonal tensions (see, for example, Mutz, 2006; Sunstein, 2007). In particular,

*An earlier version of the paper was circulated under the title "Disliking to disagree."

†*Corresponding author.* New Economic School, ul. Nobel, 3, 121205 Moscow, Russia. E-mail: kkhalmetski@nes.ru.

‡School of Economics, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK. E-mail: m.le-quement@uea.ac.uk.

§Faculty of Economics and Business, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium. E-mail: florian.hoffmann@kuleuven.be.

62 percent of Americans agree that "The political climate these days prevents me from saying things I believe because others might find them offensive" (Ekins, 2020).[1] Generally, belief homophily or a preference for belief consonance with communication partners has been widely documented in social psychology, being often linked to fundamental individual motivations like maintaining harmonic social relations, protecting one's own or group identity, or aversion to cognitive dissonance (Golman et al., 2016; Bonomi et al., 2021).

Our first contribution is to introduce a simple and robust formalization of such disagreement aversion in the presence of asymmetric information. Second, we investigate the implications of these preferences for information sharing and acquisition in settings with heterogeneous prior belief distributions. Which information does a strategic sender with a preference for disagreement aversion disclose to receivers with a different prior belief distribution? Does a given receiver learn more from a sender whose prior belief is closer to his own? How does the structure of belief differences – e.g., differences in opinions (mean of prior distribution) vs. differences in confidences in one's opinion (dispersion of prior distribution) – affect information sharing and acquisition?

Guided by these questions, our model provides the following main positive predictions: First, aversion to ex post disagreement perceived by the communication partner leads to equilibrium disclosure that is partial while the disclosed information set is more aligned with the more confident player's prior opinion. Hence, the receiver is more likely to learn a new piece of information (not confirming his prior belief) from a more confident sender. Second, diversity in prior beliefs between communicating parties is beneficial for information disclosure. Third, such diversity does not arise naturally, as senders prefer to talk to receivers who share similar prior beliefs (assortative matching), which brings scope for regulatory interventions. Finally, the conflict between assortative matching and information disclosure disappears if the sender cares about actual (rather than perceived) disagreement with the receiver.

Concretely, we consider two players with different belief distributions over an unknown state of the world. Under symmetric information, i.e., when these belief distributions are common knowledge, disagreement corresponds to the distance between their means. Un-

[1] Similarly, Dorison et al. (2019) find that *"people who hold strong opinions on an issue rated policy discussion with holders of opposing views as more aversive than any other activity listed, including household chores, yard work, and a visit to the dentist."*

der asymmetric information, however, a given player might not know the exact actual belief of the other player, but only that it is drawn from a non-degenerate set according to a well-defined probability distribution. In this case, we then analogously define disagreement *perceived* by a particular player as the expected distance between this player's belief and the other player's belief if all private information were made public. Under asymmetric information, the degree of disagreement might, hence, be perceived differently across players and need not correspond to actual disagreement. We show that this measure of perceived disagreement is coherent in that it leaves out (spurious) disagreement driven only by asymmetric information. The sender's utility decreases in the receiver's perceived disagreement, capturing the idea that people avoid being *seen* as having conflicting opinions with the communication partner.

In order to analyze the implications of disagreement aversion for information sharing, our baseline setting considers a standard model of voluntary disclosure with uncertainty about the sender's information endowment (e.g., Dye, 1985; Jung and Kwon, 1988). The receiver is conventionally assumed to be willing to match the correct state of the world with his action. The two main novel ingredients of the model are as follows: (i) The sender and the receiver have heterogeneous but publicly known prior belief distributions about the unknown state of the world, i.e., there is ex ante disagreement, and (ii) the sender is averse to disagreement as perceived by the receiver. The sender, hence, chooses to disclose a privately observed signal informative about the state of the world if and only if it leads to lower perceived disagreement relative to non-disclosure. For tractability we choose a standard normal-normal environment in which both players' prior belief distributions about the unknown state of the world as well as the conditional distributions of signals are normal; accordingly, also the posterior beliefs about the state conditional on a signal disclosure are normally distributed. We refer to the mean of a belief distribution about the state as a player's *opinion*, and to its variance as the player's *confidence* in own opinion.

Our first set of findings characterizes which signals the sender discloses in equilibrium. The key driver of these results is rooted in the baseline properties of Bayesian updating: Given different levels of confidence in their prior opinions (different prior variances) the receiver and the sender update their opinions (expected values) with different intensities when new information is presented. As a consequence, there is a unique signal

yielding zero posterior disagreement, while signals beyond it gradually increase posterior disagreement. This implies that the equilibrium disclosure is partial and features what can be termed an opinion-corridor: A signal is disclosed if and only if it yields disagreement smaller than a unique threshold value. This translates into an interval of disclosed signals, which is symmetric around the zero disagreement signal. Further, the disclosure interval is biased towards the more confident player's prior opinion, its center being located closer to the prior mean of the latter.

Our second set of findings concerns the different parties' preferences over communication partners. Suppose that the receiver wishes to learn the state as precisely as possible and can freely pick a sender among a rich pool of perceived disagreement averse agents who are equally well informed but differ in prior belief distributions. Whom should the receiver pick? It turns out that the informativeness of a sender's disclosure *increases* in the difference in prior opinions (means), contrary to common intuition. It also decreases in the difference in prior confidence (variances). We also study a sender's preference over different receivers if she has to choose her communication partner before observing the private signal. This for example captures the choice of long term communication partners on social media. A sender creates a link with someone with an eye to sharing information in the future, being motivated by an image concern (to be perceived as agreeing and therefore liked) rather than a persuasion motive. We find that the sender prefers receivers with closer prior means. There is thus a conflict between senders' preference for minimizing the difference in prior means and receivers' preference for maximizing it. In particular, this may negatively affect receivers' learning if communication partners are matched based on the senders' preferences.

In our third set of findings, we contrast our previous results to the case when the sender is instead driven by actual (rather than perceived) disagreement aversion as in Che and Kartik (2009). Here, we consider a more general version of their disclosure model by allowing for different prior variances. We continue to assume that the receiver is only willing to learn the state of the world. Strikingly, most of our key results are reversed in the latter model: the equilibrium disclosure set features extreme signals beyond a certain interval (rather than a disclosure interval as before), in case when the equilibrium is unique.[2] Further, the receiver learns more from a sender with a closer prior mean

[2]In the generalized version of the model of Che and Kartik (2009), the nature of the equilibrium predictions depends on the relation of the prior variances of the players. The equilibrium is unique if

and a more distant prior variance, again reversing our results in the baseline setting. However, the sender still prefers to talk to receivers with a closer prior mean, as before. Hence, in contrast to perceived disagreement aversion, actual disagreement aversion does not lead to a conflict in preferences over communication partners between senders and receivers. Moreover, endogenous selection of communication partners does not impede the informativeness of subsequent communication.

In our fourth set of findings, we allow the sender to ex ante commit to predetermined disclosure strategies. We find that the optimal commitment strategy depends on the relation between the prior confidence levels. In particular, if the available commitment strategies include only full disclosure or no disclosure, the sender prefers full disclosure if she is relatively more confident than the receiver, and no disclosure otherwise. The benefit of commitment is that it mitigates perceived disagreement conditional on non-disclosure, which in a normal partial disclosure equilibrium is interpreted as evidence that the sender is possibly concealing relatively extreme signals (beyond the equilibrium disclosure interval). The possibility of commitment also reverses the role of prior disagreement, which is then conducive to rather less disclosure.

Finally, we show that perceived disagreement aversion on the receiver's side leads to similar inefficiencies in terms of social learning as perceived disagreement aversion on the sender's side, since it also pushes the receiver to select less informative senders.

Literature review The empirical motivation for our preference formulation lies in the social psychology literature on conformism and homophily (see Newcomb, 1961; Asch, 1955; Lazarsfeld and Merton, 1954; Goffman, 1959). In the experiments conducted by Asch (1955), subjects wrongly evaluated the length of a line in public after hearing others' (artificially induced) wrong assessment. Deutsch and Gerard (1955) show a weaker effect if judgments are reported privately, suggesting that *perceived* disagreement is agents' true concern. Bursztyn et al. (2020) and Prentice and Miller (1993) find that subjects align their expressed opinions with what they perceive as the mainstream opinion. Field studies conducted by Lazarsfeld and Merton's in the 1950s revealed a large degree of homophily in social networks in the USA and Mutz (2006) documents similar patterns today. A burgeoning contemporary literature in neuroscience and psychology studies disagreement aversion, as illustrated for example by Domínguez et al. (2016). Golman et al. (2016)

$\gamma_S \geq \gamma_R$. Otherwise, there can be multiple equilibria of different types.

provide an extensive overview of the behavioral and economic literature on the preference for belief consonance.

Besides providing a novel and coherent measure of perceived disagreement under asymmetric information, the main contribution of our paper resides in studying the implications of perceived disagreement aversion for information sharing among agents with different priors. Our model of information sharing builds on an extensive body of research on strategic disclosure of verifiable signals dating back to Grossman (1981) and Milgrom (1981) and extended to settings with heterogeneous priors in Banerjee and Somanathan (2001) and Kartik et al. (2021). In these papers incentives for strategic disclosure arise from misaligned preferences over the receiver's action conditional on the state, while it is driven by an aversion to disagreement in our setting.

As noted above, a model directly related to our analysis is the one of Che and Kartik (2009) (CK), with whom we also share the normal-normal setup. In their model, the sender wants the receiver to take an action matching the state. As we formally show, this is outcome-equivalent to a setting where the sender only cares about the actual disagreement with the receiver, i.e., the absolute distance between the posteriors following disclosure (which is a different preference relative to the perceived disagreement aversion considered in our model). CK examine the effect of prior belief misalignment (in means only) on the sender's incentives to both privately acquire and share hard information under equivalent prior variances. Prior misalignment unambiguously hurts disclosure but increases information acquisition, so that the receiver may benefit from some misalignment.[3] While endogenous information acquisition is not part of our main model, we draw important links to CK in terms of optimal disclosure, for which our equilibrium predictions differ starkly as discussed above. Overall, our analysis reveals that the nature of disagreement aversion on the side of the sender (actual vs. perceived) has drastic implications for the resulting patterns of optimal disclosure and selection of communication partners.

The reputational cheap talk literature features an endogenous preference for belief conformity. In Gentzkow and Shapiro (2006) and Ottaviani and Sørensen (2006a,b), the sender wishes to signal high expertise to the receiver (who observes the state with positive

[3] Ilinov et al. (2024) also examine a delegation setting in which there is a positive effect of misalignment in prior beliefs between the agent and the principal for the agent's incentives for costly information acquisition, though via a different mechanism than in CK.

probability). This incentivizes her to bias her cheap talk message towards the receiver's prior belief (which is identical to the sender's prior belief in Ottaviani and Sørensen, 2006a,b). In Visser and Swank (2007) and Levy (2007), deliberative committees whose members want to signal high expertise have an incentive to pretend to have similar signals (i.e., to agree) and to decide against the prior. As in Gentzkow and Shapiro (2006), in our setup if the sender is less confident than the receiver, she tends to omit signals contradicting the receiver's prior opinion, but here her only motivation is to mitigate perceived disagreement (her information quality being known). This latter motive will in contrast drive her to omit signals that confirm the receiver's prior if the sender is more confident than the receiver.

The rest of the paper is organized as follows. Section 2 introduces the setting. Section 3 provides the equilibrium characterization of the benchmark model of perceived disagreement aversion. Section 4 considers the optimal choice of communication partners for both the receiver and the sender if the sender is averse to perceived disagreement. Section 5 replicates the analysis for the case when the sender is driven by aversion to actual (rather than perceived) disagreement. Section 6 considers the problem of information design, i.e., optimal ex ante commitment to a disclosure rule for the sender under perceived disagreement aversion. Section 7 considers the implications for social learning if the receiver is also averse to perceived disagreement with the sender. Section 8 concludes. Appendix A provides an additional microfoundation for perceived disagreement aversion, Appendix B includes a supplementary numerical example, and online Appendix C contains all proofs.

2 Setup

We study a simple model of voluntary disclosure with a *perceived disagreement* averse sender. There are two players, a sender S (she) and a receiver R (he). The two players initially disagree, in the sense of having different priors over an underlying state of nature ω drawn from state space Ω. For concreteness, we assume $\Omega = \mathbb{R}$ and that each player $i \in \{S, R\}$ has a commonly known normal prior $N(\mu_i, \gamma_i^2)$, $\gamma_i > 0$. We call μ_i i's prior *opinion*, and $1/\gamma_i^2$ i's *confidence* in her prior opinion.

S can disclose information about ω, but as in Dye (1985) or Jung and Kwon (1988)

there is uncertainty about her information endowment. In particular, with probability φ, S privately observes an informative signal $\sigma = \omega + \varepsilon$, where ε is commonly known to be distributed according to $N(0, \gamma_\varepsilon^2)$. Thus, $1/\gamma_\varepsilon^2$ measures the precision of S's signal. If S observes no information about ω, we denote this by $\sigma = \varnothing$ (empty signal).

R does not observe signal σ nor whether S is informed or not. If $\sigma \neq \varnothing$, S decides whether to disclose the signal to R or not. We denote by d the disclosed information, where $d \in \{\sigma, \varnothing\}$ if S is informed and $d = \varnothing$ otherwise. R only observes the disclosed information d, based on which he updates his prior belief using Bayes rule. Subsequently, R takes an action $a \in \mathbb{R}$ so as to maximize the utility function:[4]

$$u_R(\omega, a) = -(\omega - a)^2 . \tag{1}$$

While R thus prefers a more informative disclosure strategy that allows for a lower expected loss, S's payoff is not affected by R's action choice. Instead, S values lower disagreement in posterior opinions as perceived by R. Concretely, for any disclosed information d, we define R's *ex post perceived disagreement* (PD) as

$$\widetilde{\Delta}_R(d) = E_R[\Delta(\sigma)\,|d] \tag{2}$$

where

$$\Delta(\sigma) = |E_S[\omega|\sigma] - E_R[\omega|\sigma]| \tag{3}$$

is the *actual ex post disagreement* between S and R given $\sigma \in \mathbb{R} \cup \varnothing$. Put differently, $\Delta(\sigma)$ reflects the magnitude by which S's posterior belief about ω given signal σ deviates from R's posterior belief given the same signal (i.e., the most accurate posterior mean conditional on σ from R's perspective).[5] The key novelty in our analysis is that S dislikes perceived disagreement on the part of R (i.e., dislikes being perceived as having incorrect beliefs), which we model by stipulating that S chooses disclosure d to minimize $\widetilde{\Delta}_R(d)$ defined in (2):

$$u_S(d) = -\widetilde{\Delta}_R(d). \tag{4}$$

Since R's information is controlled by S via her disclosure choices, R might not have

[4] As shall be made clear next, the utility function of R is irrelevant for the characterization of S's optimal disclosure strategy. It will become relevant when assessing the value of equilibrium disclosure for R and associated comparative statics, such as R's preferences over senders with heterogeneous priors.

[5] While our baseline model considers disagreement in means, the key features of our equilibrium characterization extend to more general measures of differences in posteriors such as the Kullback-Leibler divergence (the proof is available upon request).

sufficient information to compute actual disagreement. In particular, if S does not disclose any information (i.e., $d = \varnothing$), then R's perceived disagreement $\widetilde{\Delta}_R(\varnothing)$ is his expectation of actual disagreement $\Delta(\sigma)$ across all signals σ that induce non-disclosure $d = \varnothing$ in equilibrium, i.e., $\widetilde{\Delta}_R(\varnothing) = E_R[\Delta(\sigma)\,|d = \varnothing]$. Instead, if $d = \sigma \neq \varnothing$, R's perceived disagreement simply equals the actual disagreement, i.e., $\widetilde{\Delta}_R(d) = \Delta(\sigma)$. Thus, perceived disagreement $\widetilde{\Delta}_R(d)$ as defined in (2) is R's posterior estimate of the magnitude by which S's actual posterior belief about the state, i.e., $E_S[\omega|\sigma]$, deviates from what would be the most accurate belief from R's perspective given S's information, i.e., $E_R[\omega|\sigma]$. Put differently, $\widetilde{\Delta}_R(d)$ measures by how much S's posterior belief is on average incorrect as estimated by R. Another way to interpret our measure of perceived disagreement $\widetilde{\Delta}_R(d)$ is that it reflects R's estimate, given the disclosed information, of what would be the actual disagreement $\Delta(\sigma)$ if R observed the same information as S.[6]

To build further intuition about the concept of perceived disagreement, note how it differs from an alternative disagreement measure $\widehat{\Delta}_R(d) = E_R[|E_S[\omega|\sigma] - E_R[\omega|d]|]$, which is simply R's expectation of the absolute distance between players' posteriors after the disclosure stage. Suppose for example that S would be known to receive with probability one a perfectly informative signal. Assume also for simplicity that S cannot credibly disclose σ (e.g., since only cheap-talk communication is possible). Then, R's perceived disagreement according to our measure (2) would always be 0. Indeed, R would simply recognize that he would have exactly the same posterior as S had S's signal σ been disclosed to him. In contrast, the alternative disagreement measure $\widehat{\Delta}_R(d)$ would yield a positive value since in this case R would know that S's posterior belief conditional on her precise signal σ would (almost) always deviate from R's posterior belief conditional on non-disclosure. We find it more natural that R's perceived disagreement should be rather 0 in such scenarios (as our measure (2) predicts), since then the difference in posteriors is caused only by information asymmetries that can be potentially fixed by information sharing. In contrast, our measure (2) captures perceived disagreement that is caused purely by different interpretations of the *same* information due to different initial worldviews, i.e., prior belief distributions (see Barron et al., 2025, for empirical evidence

[6] Note that S's utility $u_S(d)$ does not depend on S's actual signal σ for given d. Inter alia, this suggests that the verifiable disclosure is a more natural setting to study the implications of our preferences than a cheap-talk setting with unverifiable messages. Indeed, the latter setting would lead to completely uninformative communication since all sender types would have an incentive to pool on the message yielding the lowest perceived disagreement.

that ex ante beliefs, i.e., priors in our setting, may reflect core individual values).[7]

We model perceived disagreement aversion purely in terms of preferences over beliefs, i.e., the sender's utility does not depend on the receiver's action conditional on disclosure. Indeed, as discussed in the introduction, disagreement aversion may have a purely psychological nature, e.g., being driven by a desire to maintain a good relationship with the conversation partner by avoiding confrontation on important matters, while not directly caring about her or his subsequent behavior. Another key example is media channels (including bloggers and influencers on social media). Consistent with evidence from media economics and social psychology, communicators often face incentives to cater to their audience's prior beliefs, as perceived disagreement risks audience disengagement. As a result, information disclosure is shaped not only by informational objectives but also by a desire to maintain belief alignment with the audience (Gentzkow and Shapiro, 2010; Stroud, 2011).

At the same time, aversion to perceived disagreement can be also seen as a reduced form of a richer preferences structure, where perceived disagreement aversion arises endogenously as a result of underlying economic incentives (where the players do care about each other's choices). For instance, a board of directors may want to grant authority over the choice of investment projects only to a CEO who prioritizes projects that the board of directors itself finds promising. As soon as the CEO strives for decision autonomy, this creates an incentive for her to convince the board, prior to the latter's delegation decision, that they share the same ranking of projects - even if the CEO's actual views differ. In this case, it is not the actual disagreement that matters for the CEO, but rather the disagreement perceived by the board as a result of their communication. Thus, the CEO endogenously becomes averse to perceived disagreement with the board. In Appendix A we formalize this setting, deriving aversion to perceived disagreement (under a quadratic measure of ex post disagreement) endogenously from the initial preferences and information structure.

[7] As Golman et al. (2016, p. 170) put it, "*If other people believe something different because they do not have access to the same information, it is easier to assume that they would believe the same as oneself if they had access to one's own information. If they have come to different beliefs from the same information, this poses a much greater challenge to one's own interpretation of reality.*"

3 Equilibrium analysis

In this section we characterize equilibria of the disclosure game. Our equilibrium concept throughout is sequential equilibrium.[8] An equilibrium of the disclosure game is characterized by a disclosure strategy of S (a probability of disclosing at each of S's information sets) and a decision rule of R (a distribution over $a \in \mathbb{R}$ at each of R's information sets) that are sequentially rational given players' beliefs. Beliefs are determined via Bayes rule whenever possible.[9] For simplicity of exposition and without loss of generality for the qualitative results, we assume that when indifferent, S prefers disclosure over non-disclosure, i.e., $d = \sigma$ if $\widetilde{\Delta}_R(\sigma) = \widetilde{\Delta}_R(\varnothing)$.[10]

3.1 Updating and ex post disagreement

As is well known, upon observing an informative signal $\sigma \in \mathbb{R}$, player i's posterior over the state of the world ω is also normal (conjugate prior property) with mean

$$E_i\left[\omega \,|\sigma\right] = \left(\frac{1}{1+\frac{\gamma_i^2}{\gamma_\varepsilon^2}}\right)\mu_i + \left(\frac{1}{1+\frac{\gamma_\varepsilon^2}{\gamma_i^2}}\right)\sigma \tag{5}$$

and variance

$$Var_i\left[\omega \,|\sigma\right] = \frac{1}{\frac{1}{\gamma_i^2}+\frac{1}{\gamma_\varepsilon^2}}. \tag{6}$$

Intuitively, the posterior mean in (5) is a convex combination of i's prior mean μ_i – i.e., his prior opinion – and the signal σ, where the weight on μ_i is decreasing in the signal to noise ratio $\frac{\gamma_i^2}{\gamma_\varepsilon^2}$. The posterior variance in (6) is decreasing in both i's confidence level $1/\gamma_i^2$ and signal precision $1/\gamma_\varepsilon^2$.

Following any informative signal $\sigma \in \mathbb{R}$, actual ex post disagreement as defined in (3) is given by the difference between S' and R's posterior opinion, which from (5) depends on their respective prior opinion and confidence. We first focus on the generic case of non-identical prior confidences. Figure 1 shows a numerical example of this case, depicting

[8] The concept of sequential equilibrium was extended for psychological games, as in our model, by Battigalli and Dufwenberg (2009). In particular, it implies that both first- and second-order beliefs of the players are consistent with equilibrium strategies.

[9] Off-equilibrium beliefs in our setting are trivial. In the sequential equilibrium, R's belief about σ is always equal to σ if $\sigma \neq \varnothing$ is disclosed, independently of whether such disclosure is on the equilibrium path or not. Non-disclosure $\varnothing$ is always on the equilibrium path since S is uninformed with a positive probability.

[10] The set of such signals has zero measure in equilibrium unless S and R have identical priors.

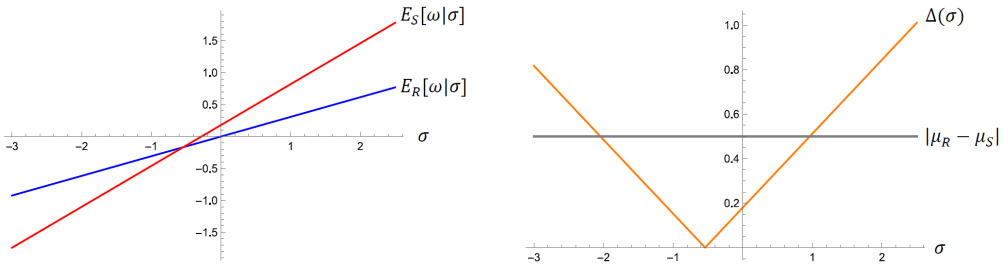

Figure 1: (*posterior disagreement*) The left panel plots the sender's (red line) and receiver's (blue line) posterior opinions as a function of the signal σ. The right panel depicts the implied ex post disagreement. The parameter values are $\mu_R = 0$, $\mu_S = 0.5$, $\gamma_R = 1$, $\gamma_S = 2$, $\gamma_\varepsilon = 1.5$.

the posterior opinions $E_S[\omega|\sigma]$ and $E_R[\omega|\sigma]$ (left panel) and their difference, ex post disagreement $\Delta(\sigma)$ (right panel), as functions of $\sigma \in \mathbb{R}$.

As is apparent from the left panel, $E_S[\omega|\sigma]$ (red line) and $E_R[\omega|\sigma]$ (blue line) are both linear functions of σ. Their intercepts differ, since S has a higher prior mean than R. Crucially, also their slopes differ as S and R update at different speeds. This is due to the fact that R is more confident in his prior opinion than S and, hence, puts relatively less weight on the signal (see (5)). Accordingly the red line is steeper in the signal σ than the blue line. As a consequence, the lines cross exactly once and the distance between posterior opinions – actual ex post disagreement $\Delta(\sigma)$ – has the V-shaped form shown in the right panel with minimum at $\widetilde{\sigma}$. That is, given S and R's priors, there is a unique signal

$$\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2 + \gamma_\varepsilon^2) - \mu_R(\gamma_S^2 + \gamma_\varepsilon^2)}{\gamma_R^2 - \gamma_S^2} \tag{7}$$

for which $\Delta(\widetilde{\sigma}) = 0$. We refer to this signal as the *zero disagreement signal*. Note that $\widetilde{\sigma}$ is never located between μ_S and μ_R, while it is closer to the prior opinion of the more confident party.

Now, since R and S update their prior at different (fixed) speeds for any given signal $\sigma \in \mathbb{R}$, actual ex post disagreement $\Delta(\sigma)$ increases linearly and tends to infinity as σ moves away from $\widetilde{\sigma}$. As a consequence, $\Delta(\sigma)$ is also symmetric around $\widetilde{\sigma}$. Further, there exist two signals $\underline{\sigma} < \overline{\sigma}$ such that actual ex post disagreement exactly equals prior disagreement, i.e., $\Delta(\underline{\sigma}) = \Delta(\overline{\sigma}) = |\mu_S - \mu_R|$. We refer to these signals as *status quo signals*. They satisfy $\underline{\sigma} < \min\{\mu_S, \mu_R\} < \max\{\mu_S, \mu_R\} < \overline{\sigma}$. For completeness, we

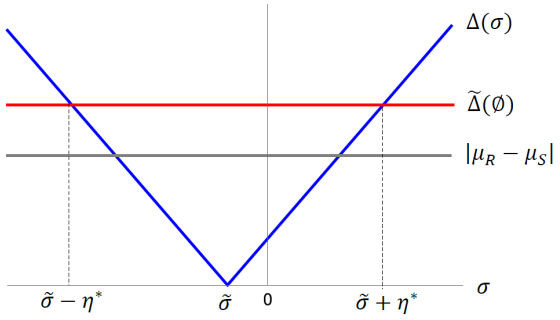

Figure 2: (*equilibrium disclosure*) The graph illustrates the equilibrium disclosure interval, with endpoints at the intersection of perceived disagreement upon disclosure (blue line) and upon non-disclosure (red line).

provide the algebraic expression of actual ex post disagreement:

$$\Delta(\sigma) = \begin{cases} k_1\sigma + k_2 \text{ for } \sigma \geq \widetilde{\sigma}, \\ -k_1\sigma - k_2 \text{ for } \sigma < \widetilde{\sigma}, \end{cases} \tag{8}$$

where

$$k_1 = \text{sgn}\,(\gamma_S - \gamma_R)\left(\frac{\gamma_S^2}{\gamma_S^2 + \gamma_\varepsilon^2} - \frac{\gamma_R^2}{\gamma_R^2 + \gamma_\varepsilon^2}\right), \tag{9}$$

$$k_2 = \text{sgn}\,(\gamma_S - \gamma_R)\left(\frac{\gamma_\varepsilon^2\mu_S}{\gamma_S^2 + \gamma_\varepsilon^2} - \frac{\gamma_\varepsilon^2\mu_R}{\gamma_R^2 + \gamma_\varepsilon^2}\right). \tag{10}$$

3.2 Equilibrium disclosure

We next characterize the equilibrium disclosure strategy. Full unraveling is prevented since a sender holding unfavorable information can claim to be uninformed. Hence, equilibrium features partial disclosure. Recalling that $\widetilde{\Delta}_R(\varnothing)$ is R's perceived disagreement conditional on non-disclosure, S will disclose signal $\sigma \neq \varnothing$ if and only if $\Delta(\sigma) \leq \widetilde{\Delta}_R(\varnothing)$. Given that $\Delta(\sigma)$ is V-shaped and symmetric around the zero disagreement signal $\widetilde{\sigma}$, the set of signals that are optimally disclosed is an interval symmetric around $\widetilde{\sigma}$, $I = [\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]$, as illustrated in Figure 2.

At the boundaries of the disclosure interval ($\widetilde{\sigma} \pm \eta$), S must be indifferent between disclosure and non-disclosure. The equilibrium value η^* hence satisfies

$$\underbrace{\Delta\left(\widetilde{\sigma} + \eta^*\right) = \Delta\left(\widetilde{\sigma} - \eta^*\right)}_{\text{PD if disclosing a boundary signal}} = \underbrace{\widetilde{\Delta}_R\left(\varnothing|\eta^*\right)}_{\text{PD given non-disclosure}}, \tag{11}$$

where $\widetilde{\Delta}_R(\varnothing|\eta)$ is a weighted average of prior disagreement and the disagreement condi-

tional on concealed signals, given disclosure interval $[\widetilde{\sigma}-\eta, \widetilde{\sigma}+\eta]$:

$$\begin{aligned}
\widetilde{\Delta}_R(\varnothing|\eta) &= E_R[\Delta(\sigma)\,|d=\varnothing,\eta] \\
&= \left[\frac{\varphi\left[F_R(\widetilde{\sigma}-\eta)\right]}{\varphi\left[F_R(\widetilde{\sigma}-\eta)+(1-F_R(\widetilde{\sigma}+\eta))\right]+1-\varphi}\right] E_R[\Delta(\sigma)\,|\sigma<\widetilde{\sigma}-\eta] \\
&\quad + \left[\frac{\varphi\left[1-F_R(\widetilde{\sigma}+\eta)\right]}{\varphi\left[F_R(\widetilde{\sigma}-\eta)+(1-F_R(\widetilde{\sigma}+\eta))\right]+1-\varphi}\right] E_R[\Delta(\sigma)\,|\sigma>\widetilde{\sigma}+\eta] \\
&\quad + \left[\frac{(1-\varphi)}{\varphi\left[F_R(\widetilde{\sigma}-\eta)+(1-F_R(\widetilde{\sigma}+\eta))\right]+1-\varphi}\right] |\mu_S-\mu_R|\,. \qquad (12)
\end{aligned}$$

Here F_R is the cumulative distribution function of $\sigma \in \mathbb{R}$ from R's ex ante perspective. The following proposition provides a characterization of equilibrium.

Proposition 1 *i. Let $\gamma_S \neq \gamma_R$. There exists a unique equilibrium and it features partial disclosure. In particular,*

(a) There exists a unique $\eta^* > 0$ pinned down by (11) such that S discloses signal σ if and only if it falls in the disclosure interval $I = [\widetilde{\sigma}-\eta^*, \widetilde{\sigma}+\eta^*]$.

(b) The disclosure interval I contains the status quo signals, i.e., $[\underline{\sigma}, \overline{\sigma}] \subset I$.

(c) If $\mu_S \neq \mu_R$, then the center of the disclosure interval I is located closer to the prior opinion of the more confident player.

ii. Let $\gamma_S = \gamma_R$. There exists a unique equilibrium and it features full disclosure.

Part i.(a) of the Proposition states that equilibrium disclosure satisfies an opinion corridor property: Only signals that fall within a bounded interval are disclosed. So a limited range of "facts" are disclosed, while any more extreme facts remain "off limits".

By point i.(b), the disclosure interval contains all signals that improve on prior disagreement, which sets a lower bound on disclosure. Moreover, the critical disagreement level $\Delta(\widetilde{\sigma}+\eta^*)$ defining the boundaries of the disclosure interval is strictly larger than prior disagreement $|\mu_S-\mu_R|$, i.e. the disclosure interval always contains a limited range of signals that increase disagreement relative to the prior. The reason for this is that in equilibrium, R's perceived disagreement conditional on non-disclosure is a weighted average of prior disagreement $|\mu_S-\mu_R|$ – reflecting the possibility of an uninformed sender – and the disagreement conditional on concealed signals (see (12)). Since any signal S conceals must satisfy $\Delta(\sigma) > \widetilde{\Delta}_R(\varnothing|\eta^*)$ by optimality, for such signals it should hold $\Delta(\sigma) > \widetilde{\Delta}_R(\varnothing|\eta^*) > |\mu_S-\mu_R|$.

Point i.(c) implies that the set of signals that are disclosed is more aligned with the prior opinion of the more confident player. Intuitively, this follows from the basic property that the more confident player adjusts his opinion less to new facts (signals) than the less confident player. It follows that signals that contradict the more confident party's prior opinion generate more disagreement than signals that contradict the less confident party's prior opinion. Hence, the zero disagreement signal $\widetilde{\sigma}$ as defined in (7), which is the center of the disclosure interval, must be closer to the prior of the more confident player.

Note that if prior opinions μ_S and μ_R are very close, then the zero disagreement signal $\widetilde{\sigma}$ is also close to these prior opinions (in particular, if $\mu_S = \mu_R = \mu$, then $\widetilde{\sigma} = \mu$). Thus, the equilibrium disclosure interval $[\widetilde{\sigma} - \eta^*, \widetilde{\sigma} + \eta^*]$ will be centred around a point close to the prior opinions of both players. Put differently, if S and R are close in their prior opinion, then the disclosed information will tend to confirm this opinion on average. This mirrors the real-world phenomenon of "echo chambers", i.e., social groups where any information significantly deviating from the normative belief in this group is suppressed from social discourse.

Part ii of the proposition characterizes the knife-edge case in which both players are equally confident. Identical prior variances imply that S and R update their opinions at the same speed. Formally, posterior means $E_S[\omega|\sigma]$ and $E_R[\omega|\sigma]$ have identical slopes with respect to $\sigma \in \mathbb{R}$. Hence, ex post disagreement $\Delta(\sigma)$ is identical for all $\sigma \in \mathbb{R}$, with $\Delta(\sigma) = \frac{\gamma_\varepsilon^2}{\gamma_\varepsilon^2+\gamma_S^2}|\mu_S - \mu_R|$, which is smaller than the prior disagreement $|\mu_S - \mu_R|$. As a result, the unique equilibrium features full disclosure.

4 Preferences over communication partners

We have characterized the unique equilibrium of the disclosure game for any combination of priors. This allows us to discuss preferences over the prior characteristics of the communication partner from both S's and R's perspective. A receiver's preferences over different senders are determined by the informativeness of the respective equilibrium disclosure strategy in the sense of minimizing expected quadratic loss (1). In contrast, the sender prefers receivers for which ex post perceived disagreement in equilibrium is lowest in expectation as given by utility function (4).

4.1 Receiver preferences over senders

Below, we consider a given receiver – as characterized by a fixed prior (μ_R, γ_R) – and analyze how changes in the prior (μ_S, γ_S) of the faced sender affect the receiver's expected utility via its effect on the equilibrium disclosure strategy. This captures the case of a decision maker who picks an advisor from a large pool, where candidates differ in their prior belief distributions, have the same ex ante level of expertise (i.e., the same γ_ε^2 and φ) and are all perceived disagreement averse. The decision maker's objective is to maximize learning from the hired advisor.

4.1.1 Senders with different prior opinions

We first consider receiver preferences over senders with different prior opinions, μ_S, holding constant prior confidence $1/\gamma_S^2$.

Proposition 2 *Fix the receiver's prior (μ_R, γ_R) with $\gamma_R \neq \gamma_S$. The receiver's expected utility is strictly increasing in prior disagreement $|\mu_R - \mu_S|$. I.e., the receiver strictly prefers a sender whose prior opinion is more distant from his own.*

We next provide some intuition for why moving S's prior opinion away from μ_R can lead to more preferable equilibrium disclosure for R. Intuitively, a main driving force is that an increase in prior disagreement robustly increases *R's perceived disagreement upon non-disclosure*, $\widetilde{\Delta}_R(\varnothing|\eta^*)$. This in turn makes non-disclosure less attractive. Indeed, recall that conditional on non-disclosure, R assigns a strictly positive probability (bounded below by $1-\varphi$) to the event that S is uninformed, in which case actual disagreement equals prior disagreement (see the last line in (12)). Via this channel, higher prior disagreement pushes R's perceived disagreement conditional on non-disclosure upwards.

Higher prior disagreement however not only increases R's perceived disagreement upon non-disclosure, $\widetilde{\Delta}_R(\varnothing|\eta^*)$, but it also affects actual disagreement upon disclosure, $\Delta(\sigma)$, for $\sigma \neq \varnothing$. Whether $\Delta(\sigma)$ is increasing or decreasing in $|\mu_R - \mu_S|$ depends on the signal σ considered. Perhaps surprisingly, for some signals higher prior disagreement leads to lower posterior disagreement. In the example shown on Figure 3, the increase in $\mu_S > \mu_R$ yields a reduced posterior disagreement for high signal values and an increased posterior disagreement for low signal values. Indeed, the divergence of posterior beliefs after getting a sufficiently high signal is mitigated if S's prior mean (and hence her posterior mean)

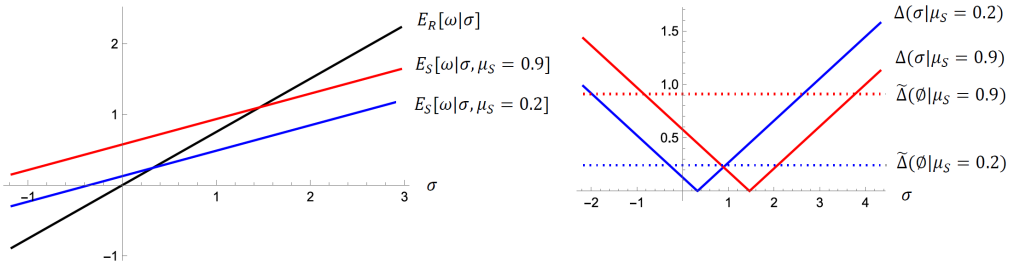

Figure 3: (*increasing prior disagreement*) The graph plots the effects of an increase in prior disagreement. The left panel illustrates the shift in the sender's posterior opinion due to an increase in μ_S from 0.2 (blue line) to 0.9 (red line) holding constant $\gamma_S = 0.3$, together with the receiver's posterior opinion (black line) for $\mu_R = 0$ and $\gamma_R = 0.7$. The right panel plots the effect of the increase in prior disagreement on posterior disagreement upon disclosure (shift from solid blue to solid red line) as well as on perceived disagreement upon non-disclosure (shift from dashed blue to dashed red line). The other parameter values are $\gamma_\varepsilon = 0.4$ and $\varphi = 0.3$.

is shifted in the direction of the updated posterior mean of R (who updates upwards at a higher speed), i.e., increases. Yet, simultaneously such shift in μ_S increases posterior disagreement for sufficiently low signals, where R's posterior is always lower.

Summarizing, an increase in prior disagreement has two effects on the sender's incentives. First, there is a robust effect that makes disclosure of any signal relatively more attractive by increasing R's perceived disagreement upon non-disclosure, $\widetilde{\Delta}_R(\varnothing|\eta^*)$ (see the upward shift of the dashed horizontal line in the right panel of Figure 3). Second, there is a more composite effect operating via $\Delta(\sigma)$, disagreement upon disclosure. $\Delta(\sigma)$ decreases at one end of the disclosure interval, where it incentivizes more disclosure, but $\Delta(\sigma)$ simultaneously increases at the other end of the disclosure interval, where it incentivizes less disclosure (see the shift of $\Delta(\sigma)$ at the upper and lower ends of the initial disclosure interval in the right panel of Figure 3). Eventually, the resulting change in the disclosure interval benefits R under the quadratic loss function. Thus, the total expected utility of R conditional on disclosure increases with prior disagreement.[11]

4.1.2 Senders with different prior confidence levels

We next consider receiver preferences over senders with different prior confidence levels $1/\gamma_S^2$, holding constant μ_S. By the fixed order of γ_S and γ_R we mean that the relation

[11] Note that the disclosure interval itself does not necessarily expand on both sides as a result of larger prior disagreement. At the same time, as shown in the proof of Proposition 2 in Appendix C, the probability of disclosure always increases. This (together with the particular way the disclosure interval shifts) eventually ensures that larger prior disagreement in means is beneficial for R.

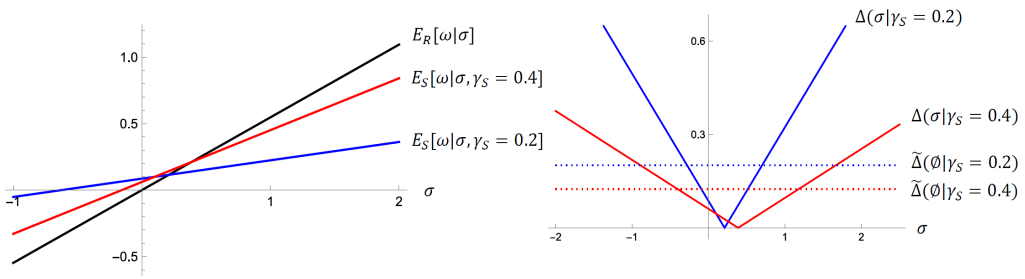

Figure 4: (*decreasing difference in prior confidence*) The graph plots the effects of a decrease in the difference of prior confidence. The left panel illustrates the shift in the sender's posterior opinion due to an increase in γ_S from 0.2 (blue line) to 0.4 (red line) holding constant $\mu_S = 0.1$, together with the receiver's posterior opinion (black line) for $\mu_R = 0$ and $\gamma_R = 0.55$. The right panel plots the effect of the decrease in the difference of prior confidence on posterior disagreement upon disclosure (shift from solid blue to solid red line) as well as on perceived disagreement upon non-disclosure (shift from dashed blue to dashed red line). The other parameter values are $\gamma_\varepsilon = 0.5$ and $\varphi = 0.5$.

$\gamma_S \geq \gamma_R$ or $\gamma_S \leq \gamma_R$ remains the same as γ_S changes.

Proposition 3 *Fix the receiver's prior* (μ_R, γ_R) *with* $\mu_R \neq \mu_S$. *Keeping the order of* γ_R *and* γ_S *fixed, the receiver's expected utility is strictly decreasing in the absolute difference* $|\gamma_R - \gamma_S|$. *I.e., the receiver strictly prefers a sender whose confidence level is less distant to his own.*

The proposition states that, preserving the ordinal ranking of γ_R and γ_S, an increase in the difference in prior variances (and hence, in confidences) has exactly the opposite effect than an increase in the difference in prior means (i.e., opinions) in that it reduces the expected value of equilibrium disclosure for R. As γ_S moves closer to γ_R the speeds of updating of the players converge to each other so that posterior disagreement conditional on disclosure remains comparatively small even for very high or low signals. Graphically, this is illustrated in the left panel of Figure 4. An increase in γ_S towards γ_R (from below) causes the slope of $E_S[\omega|\sigma]$ to increase (transition from blue to red line) and become closer to the slope of $E_R[\omega|\sigma]$ (black line). This maps into a flatter actual disagreement function $\Delta(\sigma)$ as shown in the right panel of Figure 4 (transition from solid blue to solid red line). Ceteris paribus, this makes disclosure of most signals more attractive for S and hence pushes towards a larger disclosure interval.

At the same time, as $|\gamma_R - \gamma_S|$ decreases, the above disclosure-enhancing effect is counteracted by a reduction in R's perceived disagreement conditional on non-disclosure $\widetilde{\Delta}_R(\varnothing|\eta^*)$ (see downward shift from the dashed blue to the dashed red line in Figure

4). This second effect makes non-disclosure relatively more attractive for S and follows from the same underlying intuition as the disclosure-enhancing effect. As implied by Proposition 3, the disclosure-enhancing effect brings a larger positive effect in terms of R's expected utility. This ultimately implies that R becomes better off under decreased difference in variances (preserving the ordinal ranking of γ_R and γ_S).[12]

Proposition 3 assumes unequal prior opinions, i.e., $\mu_R \neq \mu_S$. The case of $\mu_R = \mu_S$ is addressed next.

Proposition 4 *Fix the receiver's prior (μ_R, γ_R) with $\mu_R = \mu_S$.*

a) The receiver is indifferent between any γ'_S, γ''_S if both are unequal to γ_R.

b) The receiver strictly prefers $\gamma_S = \gamma_R$ over any $\gamma'_S \neq \gamma_R$.

We see that if $\mu_R = \mu_S$, S's prior confidence generically does not affect R's expected utility unless there is a shift from some $\gamma'_S \neq \gamma_R$ to $\gamma_S = \gamma_R$. In particular, $\gamma_S = \gamma_R$ implies full disclosure in equilibrium, yielding a strictly higher expected utility for R than partial disclosure under $\gamma'_S \neq \gamma_R$.

4.2 Sender preferences over receivers

Consider now sender preferences over receivers. From an ex ante perspective (i.e., before observing σ), senders care about their expectation of R's ex post perceived disagreement, i.e., $E_S[\widetilde{\Delta}_R(d)] = E_S[E_R[\Delta(\sigma)\,|d]]$, as pinned down by the disclosure equilibrium implied by the profile of priors under consideration. Senders prefer receivers who yield a lower value of this quantity.

Our main result refers to S's preferences over receivers with different prior opinions.

Proposition 5 *Fix the sender's prior (μ_S, γ_S). Then, $E_S[\widetilde{\Delta}_R(d)]$ is strictly decreasing in prior disagreement $|\mu_R - \mu_S|$. I.e., ex ante the sender strictly prefers communicating with a receiver with whom she has smaller prior disagreement.*

Comparing Proposition 5 with Proposition 2, we see that receiver and sender preferences over prior disagreement differ. While receivers prefer communicating with senders with whom prior disagreement in opinions is larger, senders always prefer a lower value of

[12] Yet, as in the case of prior means, reducing $|\gamma_S - \gamma_R|$ does not always lead to an expansion of the disclosure interval on both sides.

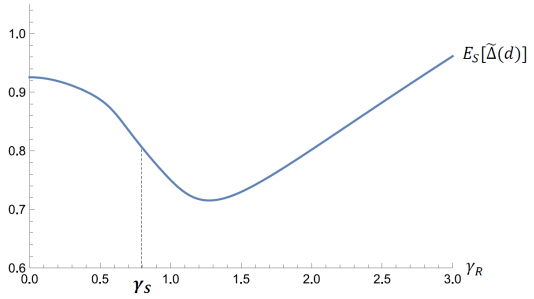

Figure 5: (*S's ex ante expected disagreement*) The graph plots S's ex ante expected perceived disagreement as a function of γ_R. The parameter values are $\mu_R = 0$, $\mu_S = 1$, $\gamma_S = 0.8$, $\gamma_\varepsilon = 1$, $\varphi = 0.5$.

$|\mu_R - \mu_S|$. This is intuitive since from the sender's ex ante perspective, expected posterior disagreement must decrease with prior disagreement. It is less trivial that this result holds independently of whether the reduction in prior disagreement makes equilibrium disclosure more or less informative for the receiver.

Note that Proposition 5 has important implications in settings where senders of information may select their communication partners, driven by minimizing ex post perceived disagreement. Then, senders choose the receiver who is closest in prior opinion to themselves. Yet, this allocation is the worst in terms of informativeness of communication from the perspective of the receiver by Proposition 2. Hence, receivers of information may generally prefer random matching mechanisms over selective matching induced by senders. Put differently, perceived disagreement aversion on the side of the sender creates frictions in social learning by precluding most efficient sender-receiver matches from the perspective of information transmission, and hence stands in direct conflict with the desire of the receiver to learn the correct state of the world. Moreover, as shown in Section 7, if the receiver himself cares about S's perception of their ex post disagreement (instead of about learning the correct state), the same tradeoff arises: the receiver then prefers to interact with senders who are closest to him in terms of prior beliefs, which again impedes information transmission.

Next, consider S's preferences over receivers who differ only in their prior confidence levels. Unlike in the above case of different prior opinions, it is impossible to say that S will always prefer an R with the prior confidence level either closer or further from γ_S. Figure 5 shows a numerical example where $E_S[\widetilde{\Delta}_R(d)]$ changes non-monotonically in response to γ_R and reaches its minimum at $\gamma_R \approx 1.27 > \gamma_S = 0.8$.

The reason for this non-monotonicity of $E_S[\widetilde{\Delta}_R(d)]$ in $|\gamma_S - \gamma_R|$ is that S's preference

over R's prior confidence levels is affected by two opposing effects. First, S tends to prefer similarly confident R because this reduces the part of expected disagreement driven by different speeds of updating (call this the similarity effect). At the same time, from the ex ante perspective, S expects to receive information that will lead R to update in the direction of S's prior, thus reducing disagreement. This updating towards μ_S will be stronger the less confident R is, i.e., the higher γ_R (call this the flexibility effect). So if $\gamma_R < \gamma_S$, an increase in γ_R implies that the two above effects (similarity and flexibility) push towards lower expected perceived disagreement. If instead $\gamma_R \geq \gamma_S$, an increase in γ_R implies both a more dissimilar and a less confident R so that the two above effects push in different directions in terms of expected perceived disagreement. In the example in Figure 5, the similarity effect overweighs the flexibility effect first when γ_R surpasses 1.27. One can show more generally that the flexibility effect dominates the similarity effect if $\mu_S \neq \mu_R$ and γ_R is sufficiently close to γ_S. Thus, R's perceived disagreement expected ex ante by S always decreases in γ_R for $\gamma_R \leq \gamma_S$, while the optimal confidence level of R is strictly larger than γ_S if $\mu_S \neq \mu_R$.

If $\mu_S = \mu_R$, then R's perceived disagreement expected by S is trivially 0 if also $\gamma_S = \gamma_R$, which is then the optimal confidence level in this case.

Proposition 6 *Fix S's prior (μ_S, γ_S).*

a) Let $\mu_S = \mu_R$. Then, $E_S[\widetilde{\Delta}_R(d)]$ strictly decreases in γ_R on $(0, \gamma_S]$ and strictly increases in γ_R otherwise.

b) Let $\mu_S \neq \mu_R$. Then, there exists $\gamma_R^* > \gamma_S$ such that $E_S[\widetilde{\Delta}_R(d)]$ strictly decreases in γ_R on $(0, \gamma_R^*]$, and strictly increases in γ_R for γ_R sufficiently large. Thus, $E_S[\widetilde{\Delta}_R(d)]$ is minimized at level(s) of γ_R strictly larger than γ_S.

5 Aversion to actual disagreement

5.1 Equilibrium disclosure

In this section, we turn to the situation where S cares about actual (rather than perceived) disagreement with R, i.e., about the absolute distance between the posteriors following the disclosure stage. S's utility function is then given by:

$$u_S(d) = -|E_S[\omega|\sigma] - E_R[\omega|d]|. \tag{13}$$

This can be interpreted as S caring about R learning the true state of the world. Intuitively, this concern may arise when S wants R to take the best action possible in terms of matching the correct state (e.g., if there is no conflict of interest between the players, but just a difference in prior belief distributions). As in the benchmark model, R also aims to match the correct state, i.e., his preferences are given by (1).

The distinction between actual and perceived disagreement aversion can be naturally microfounded in settings involving expert advice. When an expert provides information to a principal who will ultimately choose the action herself - such as a financial analyst advising an investor, a policy consultant advising a minister, or a doctor advising a patient - what matters for the expert's incentives is actual disagreement, because the principal's posterior belief directly determines the implemented choice. By contrast, an expert who seeks to persuade the principal to delegate authority to her - such as a manager lobbying for discretion, an entrepreneur seeking control from an investor, or a specialist aiming to secure decision rights within an organization - cares primarily about perceived disagreement, since delegation depends on the principal's belief about alignment rather than the actual difference in posteriors given potentially different information sets (see Appendix A). As we show below, this institutional distinction leads to markedly different disclosure incentives and can even reverse which experts are most desirable.

First, one can formally show that S's preferences (13) give rise to the same set of equilibria as when S cares about minimizing R's loss directly.

Proposition 7 *The set of equilibria when S's preferences are given by (13) is equivalent to the set of equilibria when S's preferences are given by $\widehat{u}_S(d) = -E_S[(a_R(d) - \omega)^2]$.*

Note that the latter preference structure is equivalent to the one considered in Che and Kartik (2009). They however focus on the special case of identical prior variances. In this section, we extend this analysis by allowing also different prior variances.

Next, proposition 8 provides equilibrium characterization under $\gamma_S \geq \gamma_R$.

Proposition 8 *Let $\gamma_S \geq \gamma_R$.*

a) If $\mu_S \neq \mu_R$, there exists a unique equilibrium and it features partial disclosure. In particular, S discloses signal σ if and only if it falls outside of the non-disclosure interval (σ_l, σ_h).

b) If $\mu_S = \mu_R$, there exists a unique equilibrium and it features full disclosure.

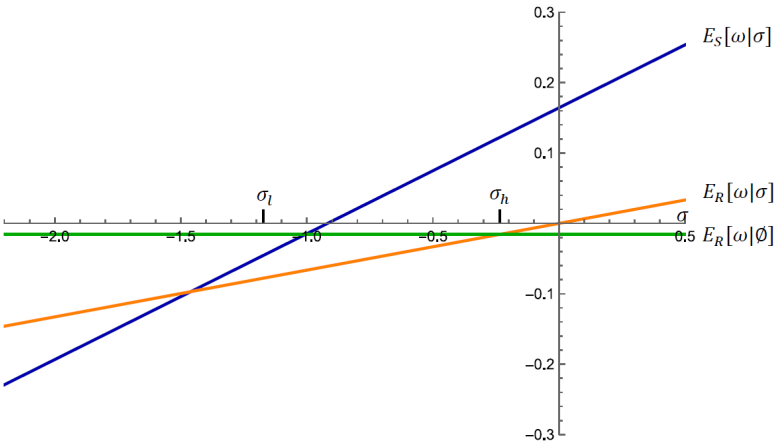

Figure 6: (*Equilibrium under aversion to actual disagreement*).The graph plots the sender's (blue line) and receiver's (orange line) posterior opinions as a function of the signal σ. The green line depicts the receiver's posterior opinion conditional on non-disclosure in equilibrium. The parameter values are $\mu_R = 0$, $\mu_S = 0.2$, $\gamma_R = 0.4$, $\gamma_S = 0.7$, $\gamma_\varepsilon = 1.5$, $\varphi = 0.7$.

The equilibrium structure is the reverse of the one under perceived disagreement aversion: Instead of only disclosing signals located *within* a certain interval as before, S only discloses signals *outside* of a certain interval. This is explained by the fact that under actual disagreement aversion, S does not want R's posterior belief to deviate from S's own *actual* posterior belief (rather than the one perceived by R). Hence, once the signal becomes sufficiently extreme, S would always prefer to make R's posterior aligned by letting him update in the direction of the signal instead of sticking to his belief conditional on non-disclosure (which is fixed at a constant level in equilibrium). This contrasts S's incentives under perceived disagreement aversion, in which case the extremity of S's private signal does not affect R's perceived disagreement (and hence S's utility) conditional on non-disclosure.

Figure 6 illustrates the equilibrium. It shows the posterior belief as a function of signal σ for both S and R (blue and orange lines, respectively), as well as R's posterior belief conditional on non-disclosure (green line). The actual disagreement conditional on disclosure (the distance between the blue and orange lines) is smaller than the disagreement conditional on non-disclosure (the distance between the blue and green lines) if and only if $\sigma \leq \sigma_l$ or $\sigma \geq \sigma_h$. Hence, S prefers to conceal all signals within (σ_l, σ_h).

To see the intuition behind full disclosure under common prior means $\mu_S = \mu_R$, note that S is less confident than R by assumption (i.e., $\gamma_S \geq \gamma_R$) and hence updates her belief (i.e. deviates from the common prior mean) at a higher speed after learning any

signal. Thus, S would always want R to "catch up" by disclosing the signal.

If $\gamma_S < \gamma_R$, then generally the equilibrium is not unique. Besides, equilibria may then be of different types depending on the parameters: i.e., there may exist equilibria characterized by non-disclosure intervals as given by Proposition 8, as well as equilibria defined by disclosure intervals as given by Proposition 1. Furthermore, simultaneously existing equilibria of the same type may be characterized by the opposite comparative statics with respect to a change in parameters, such as μ_S or γ_S. Appendix B provides corresponding numerical examples.

5.2 Preferences over communication partners

Let us now consider receiver preferences over senders with different priors as in Section 4.1, focusing on the case $\gamma_S \geq \gamma_R$ where the equilibrium is uniquely pinned down.[13] Which senders would the receiver prefer if the former are driven by aversion to actual disagreement? We obtain the following results.

Proposition 9 *Let $\gamma_S \geq \gamma_R$.*

a) Fix the receiver's prior (μ_R, γ_R) with $\gamma_R \neq \gamma_S$. The receiver's expected utility is strictly decreasing in prior disagreement $|\mu_R - \mu_S|$. I.e., the receiver strictly prefers a sender whose prior opinion is closer to his own.

b) Fix the receiver's prior (μ_R, γ_R) with $\mu_R \neq \mu_S$. The receiver's expected utility is strictly increasing in the absolute difference $|\gamma_S - \gamma_R|$. I.e., the receiver strictly prefers a sender whose confidence level is more distant to his own.

Thus, the results are exactly the reverse of those presented in Propositions 2 and 3 for the case of perceived disagreement aversion. In particular, prior disagreement is now detrimental to information disclosure. Intuitively, as prior disagreement in means diminishes, the posterior disagreement conditional on disclosure becomes driven by the difference in the speed of updating rather than the discrepancy in starting belief values. Since S always updates at a higher speed under $\gamma_S \geq \gamma_R$, her posterior belief is then more likely to move further than that of R in the direction of the signal. In such cases, she would prefer R to "catch up" with her belief by disclosing the signal (recall that the

[13] As noted above, for $\gamma_S < \gamma_R$, multiple existing equilibria may feature the opposite predictions in terms of comparative statics (see Appendix B).

extreme case $\mu_S = \mu_R$ leads to full disclosure by Proposition 8(b)). Similarly, increasing the difference in prior variances, i.e., in the updating speeds (keeping the prior means constant) again facilitates disclosure by the same argument. This implies that ceteris paribus R prefers an S with a higher prior variance (given $\gamma_S \geq \gamma_R$).

We also obtain a result analogous to Proposition 5 in that S prefers to be matched to a receiver with a closer prior mean.

Proposition 10 *Let $\gamma_S \geq \gamma_R$. Fix the sender's prior (μ_S, γ_S). Then, the actual disagreement expected ex ante by S is strictly decreasing in prior disagreement $|\mu_R - \mu_S|$. I.e., ex ante the sender strictly prefers communicating with a receiver with whom she has smaller prior disagreement.*

The result is intuitive as smaller difference in prior means generally implies smaller posterior disagreement on average.

Overall, receivers' preferences over senders are now in line with the latter' preferences over receivers: both type of players prefer to be matched to a communication partner with a closer prior mean. Hence, under actual disagreement aversion S-preferred selective matching *facilitates* the informativeness of communication for the receiver. This is in contrast to perceived disagreement aversion, where the effect was the opposite: while S still preferred assortative matching in prior means, this led to the worst outcome in terms of subsequent informativeness of disclosure (see Section 4.2). Thus, the subtle nature of sender's preferences over disagreement (actual vs. perceived) has drastic implications for the resulting conflict of interest between the players over the matching patterns.

6 Information design

Consider now a situation where S can commit to a certain disclosure rule ex ante (before observing the signal). The most intuitive examples are commitment to full disclosure, i.e., to disclosing any $\sigma \in \mathbb{R}$ (when, e.g., a company commits to full financial transparency by allowing credible audits by external parties), or to no disclosure, i.e., to disclosing no $\sigma \in \mathbb{R}$ (when, e.g., communication is simply avoided or blocked, no matter the obtained information). If S decides not to commit, then the prediction is the standard equilibrium characterized by Proposition 1. In what follows, we denote full (no) disclosure as FD

(ND). As in the benchmark case, S has utility function (4) (minimizing expected R's perceived disagreement). We obtain the following result.

Proposition 11 *Assume S can commit to either FD or ND, or not commit. Then, she prefers to commit to FD if $\gamma_S < \gamma_R$, and to ND if $\gamma_S > \gamma_R$. If $\gamma_S = \gamma_R$, S is indifferent between ND, FD and no commitment.*

To see why commitment to FD is preferred by S if $\gamma_S < \gamma_R$, recall that in equilibrium under no commitment (as well as under ND) S conceals relatively extreme signals that lead to large posterior disagreement. As a result, non-disclosure is costly in terms of resulting disagreement perceived by R. By committing to FD, S thus mitigates this cost by removing concealment of extreme signals from the possible scenarios considered by R upon non-disclosure (which now credibly reveals that S is uninformed, i.e., keeps R's perceived disagreement at the prior level). Moreover, because S treats extreme signal realizations as unlikely relative to R due to lower prior variance, S is better off taking a risk by ex ante committing to disclose these signals rather than letting R incorporate these signals into his assessment of disagreement conditional on non-disclosure.

The intuition behind S preferring commitment to ND if $\gamma_S > \gamma_R$ is analogous. By committing to ND, S mitigates the costs associated with non-disclosure by letting also moderate signals (leading to low posterior disagreement) be considered by R as possible scenarios upon non-disclosure. Since S also treats these moderate signals as unlikely relative to R due to larger prior variance, S is expectedly better off letting R bet on their realizations conditional on non-disclosure rather than allowing them to be actually disclosed under no commitment or FD.

Next, assume that S can commit to any disclosure rule of the form $[\widetilde{\sigma} - x, \widetilde{\sigma} + x]$ with $x > 0$, i.e., to disclosing σ if and only if $\sigma \in [\widetilde{\sigma} - x, \widetilde{\sigma} + x]$. Since $\Delta(\sigma)$ is V-shaped and symmetric around $\widetilde{\sigma}$, this is equivalent to assuming that S can commit to any maximum disagreement level conditional on disclosure, as disagreement conditional on any disclosed signal is then lower or equal to $\Delta(\widetilde{\sigma} - x) = \Delta(\widetilde{\sigma} + x)$. Denote the probability of non-disclosure assessed by player $i \in \{S, R\}$ given disclosure interval $[\widetilde{\sigma} - x, \widetilde{\sigma} + x]$ as

$$P_{\varnothing,i}(x) = \varphi F_i(\widetilde{\sigma} - x) + \varphi(1 - F_i(\widetilde{\sigma} + x)) + 1 - \varphi,$$

and $\partial P_{\varnothing,i}(x)/\partial x$ as $P'_{\varnothing,S}(x)$.

Proposition 12 *Assume S can commit to any disclosure rule of the form $[\widetilde{\sigma}-x, \widetilde{\sigma}+x]$ with $x>0$.*

a) Let $\gamma_S < \gamma_R$. Then, S prefers to commit to either FD or to disclosure interval $[\widetilde{\sigma}-\widehat{\eta}, \widetilde{\sigma}+\widehat{\eta}]$ where $\widehat{\eta}$ is implicitly given by

$$\frac{P'_{\varnothing,S}(\widehat{\eta})}{P_{\varnothing,S}(\widehat{\eta})} = \frac{P'_{\varnothing,R}(\widehat{\eta})}{P_{\varnothing,R}(\widehat{\eta})}. \tag{14}$$

b) Let $\gamma_S > \gamma_R$. Then, S prefers to commit to either ND or to disclosure interval $[\widetilde{\sigma}-\widehat{\eta}, \widetilde{\sigma}+\widehat{\eta}]$ where $\widehat{\eta}$ is implicitly given by (14).

Thus, the optimal commitment still allows for FD if $\gamma_S < \gamma_R$, and for ND if $\gamma_S > \gamma_R$.[14] At the same time, there is a possibility of interior solution defined by (14). The intuition behind this solution is as follows. By putting more moderate signals to the non-disclosure domain, from the ex ante perspective S replaces her own odds that these signals are realized and disclosed with the odds of R, who then assesses the likelihood of these signals while computing perceived disagreement upon non-disclosure. Once the odds of R are higher, reducing the disclosure interval becomes beneficial for S. Condition (14) defines the length of the disclosure interval where the marginal odds are equalized so that S becomes indifferent between expansion and contraction of the disclosure interval.

Whether S prefers a corner solution (FD or ND), or the interior solution $\widehat{\eta}$ eventually depends on the parameter values. In particular, if the prior disagreement in means is sufficiently small (large) under $\gamma_S < \gamma_R$ ($\gamma_S > \gamma_R$), then the corner solution is the unique prediction.

Proposition 13 *Assume S can commit to any disclosure rule of the form $[\widetilde{\sigma}-x, \widetilde{\sigma}+x]$ with $x>0$.*

a) Let $\gamma_S < \gamma_R$. If $|\mu_S - \mu_R|$ is sufficiently small, then S prefers to commit to FD.

b) Let $\gamma_S > \gamma_R$. If $|\mu_S - \mu_R|$ is sufficiently large, then S prefers to commit to ND.

To see the intuition behind Proposition 13(a), note that small prior disagreement in means ensures that only the difference in prior variances plays a role by determining the relative odds of signal realizations between the players. Then, for $\gamma_S < \gamma_R$, R assigns a higher likelihood than S to the realization of extreme signals that are concealed under

[14] Note that the problem of committing to a symmetric interval around the 0-disagreement signal is not well defined for $\gamma_S = \gamma_R$ as then $\Delta(\sigma)$ is constant (i.e., the 0-disagreement signal does not exist).

(any) partial disclosure rule or ND. Hence, the same intuition fostering disclosure of these signals by S (i.e., ex ante commitment to FD) as in Proposition 11 applies.

The intuition behind Proposition 13(b) is as follows. A sufficiently large prior disagreement ensures that under the partial disclosure rule $[\widetilde{\sigma} - \widehat{\eta}, \widetilde{\sigma} + \widehat{\eta}]$, non-disclosure leads to larger R's perceived disagreement than the set of disclosed signals on average. The reason is that any scenario implied by non-disclosure - concealment of extreme signals, or S being uninformed (in which case R's perceived disagreement is equal to the assumed large prior disagreement) is associated with a larger disagreement than the disclosed signals in the eyes of R. As a result, if $\gamma_S > \gamma_R$, S overall benefits from mitigating R's perceived disagreement conditional on non-disclosure by incorporating the disclosed signals into the non-disclosure domain, i.e., the same intuition fostering commitment to ND as in Proposition 11 applies.

Proposition 13 reveals that prior disagreement plays the opposite role for informativeness of disclosure relative to the benchmark case of no commitment considered in Section 4.1, as it is conducive to less disclosure (rather than more disclosure as implied by previous Proposition 2). In particular, low prior disagreement fosters commitment to full disclosure, while large prior disagreement fosters commitment to no disclosure (in the corresponding domains of prior variances). Hence, the effect of prior disagreement on information transmission under perceived disagreement aversion crucially depends on whether S can commit to a predetermined disclosure rule or not.

7 Receiver disagreement aversion

Just as it is natural to assume that S is averse to R's perceived disagreement, it is natural to assume that R is similarly averse to S's perceived disagreement (in particular, if communicating parties both belong to the same organization or social group). In what follows, we explore the corresponding implications for R's preference over different sender types.

In particular, assume that S is still averse to R's perceived disagreement as in Section 3. At the same time, R now does not take an action to match the state of the world (or puts sufficiently low weight on this concern), but instead cares only (or mostly) about S's perceived disagreement. Consistent with our definition (2), we define S's *perceived*

disagreement as

$$\widetilde{\Delta}_S(\sigma) = E_S[\Delta(\sigma)\,|\sigma\,] = \Delta(\sigma). \tag{15}$$

This is intuitive as S always knows her own information $\sigma \in \mathbb{R} \cup \varnothing$. Hence, she can always precisely assess what would be the actual disagreement had her signal been disclosed to R (the intuition behind perceived disagreement in Section 2), which is then $\Delta(\sigma)$. We now assume that R's utility is given by minus her expectation of S's perceived disagreement conditional on disclosure:

$$u_R(d) = -E_R[\widetilde{\Delta}_S(\sigma)|d] = -E_R[\Delta(\sigma)|d]. \tag{16}$$

Note that from the ex ante perspective, the expected utility $E_R[u_R] = E_R[-E_R[\Delta(\sigma)|d]] = -E_R[\Delta(\sigma)]$ does not depend on the equilibrium disclosure strategy: R knows that eventually S will assess disagreement with R based on her signal σ, independently of whether it is disclosed or not.

Assume now that R may potentially choose the most preferred sender from some available set (e.g., by choosing whom, among his friends, to follow on social media). Which sender would R prefer from the ex ante perspective? Our results essentially replicate the ones obtained in Section 4.2: like senders, receivers prefer communication partners with a closer prior mean, while the optimal confidence level of the sender is below the confidence level of the receiver (which follows from the same intuition as in Section 4.2).

Proposition 14 *Fix R's prior (μ_R, γ_R). Then, $E[u_R]$ is strictly decreasing in prior disagreement $|\mu_R - \mu_S|$. I.e., ex ante R strictly prefers facing a sender with whom he has smaller prior disagreement.*

Proposition 15 *Fix R's prior (μ_R, γ_R).*

a) Let $\mu_S = \mu_R$. Then, $E_R[u_R]$ is strictly decreasing in γ_S on $(0, \gamma_R]$ and strictly increasing otherwise.

b) Let $\mu_S \neq \mu_R$. Then, there exists $\gamma_S^* > \gamma_R$ such that $E_R[u_R]$ is strictly decreasing in γ_S on $(0, \gamma_S^*]$ and strictly increasing otherwise.

Thus, perceived disagreement aversion on the side of R potentially leads to the same inefficiencies in terms of information transmission as perceived disagreement aversion

on S's side once communication partners may be endogenously selected. In particular, assume that R would still have to take an action, while predominantly caring about avoiding S's perceived disagreement as assumed above. According to Proposition 14, R prefers to interact with senders who are closest to him in terms of prior beliefs (ceteris paribus). Yet, by Proposition 2, smaller prior disagreement would negatively affect the learning outcome for R in terms of matching the correct state of the world with his action. Thus, disagreement aversion does not only impose additional costs on R per se, but also precludes him from acting efficiently by pushing him to select the least informative senders.

8 Conclusion

This paper introduces a new type of belief-dependent preferences capturing an aversion to perceived disagreement, which is well documented in social psychology. Our analysis has identified a range of implications for strategic communication and patterns of social matching.

First, under perceived disagreement aversion, information disclosure features an "opinion corridor" property, where the disclosure interval is biased in the direction of the prior opinion of the more confident party. Thus, the receiver has a better chance to learn information contradicting his prior belief if the sender is relatively more confident. Otherwise, the sender will tend to comply to the prior belief of the receiver, thus diminishing the latter's opportunities to revise his prior belief in case it is biased in the wrong direction.

Second, information disclosure is relatively low when communication partners share similar prior opinions, while similar prior confidence levels instead have a positive effect on disclosure. Thus, diversity in prior opinions between communicating parties should be promoted and facilitated in settings where uninformed decision-makers have to rely on the opinion of informed experts who care about being perceived as holding a similar opinion or worldview, such as advising committees within the same political party, or managers relying on the opinion of their employees. Our model would also predict that in settings where a decision maker picks an advisor from a large pool in order to maximize learning, he may want to pick a more priorly disagreeing one. Here, one could think of voters choosing whom to follow on social media, or with whom to engage in political talk.

Further applications could be in the context of retail financial advise or management consulting services.[15]

Third, the diversity of prior opinions between endogenously matched parties does not arise naturally. Instead, experts caring about perceived disagreement prefer to talk to individuals whose prior opinion is closer to their own, thus giving rise to assortative matching in prior beliefs. The same matching patterns tend to arise also if receivers are affected by disagreement aversion in choosing experts. The resulting low prior disagreement is counterproductive for subsequent information disclosure given the previous point. Thus, promoting efficient social communication requires policy interventions aimed at avoiding assortative matching in beliefs (e.g., implementing purely random matching mechanisms). This may have implications for the design of matching and selection algorithms in policy committees, teams within organizations and online social networks.

Fourth, perceived disagreement aversion implies different patters of communication if the sender is able to ex ante commit to a predetermined disclosure rule (e.g., full disclosure or no disclosure). In particular, commitment to full disclosure yields a higher ex ante payoff for both the sender and the receiver than the unrestricted communication if the sender is more priorly confident than the receiver. In this situation, paradoxically, a way to address potential pitfalls of perceived disagreement aversion for both senders and receivers is to impose culture of complete transparency of social discourse, which eventually would lead to less expected perceived disagreement and higher willingness to communicate. This also provides additional justification for the existence and robustness of commitment mechanisms in real-world communication, such as obligatory disclosure of financial information by companies. In addition, the role of prior disagreement for informativeness is reversed if the sender can commit to a predetermined disclosure rule, as larger prior disagreement in means is then conducive to less disclosure. Thus, promoting diversity in prior opinions is less likely to foster informativeness in situations where the sender is able to stick to predetermined communication protocols.

Finally, if the sender cares not about reputation for having the same opinion but rather about actual disagreement with the receiver (or about him making the right action), then misalignment in prior opinions may hurt disclosure, while endogenous sender-preferred

[15] The conclusion of a positive effect of belief misalignment on communication draws some similarity to Che and Kartik (2009) who studied actual disagreement aversion. Yet, in their setting the positive effects stem from better incentives for information acquisition by the sender, while in our case diversity in priors facilitates information disclosure given a fixed quality of sender's information.

matching would facilitate it. Both effects are exactly reversed relative to the benchmark model of perceived disagreement aversion. Thus, the nature of the senders' motivation (persuasion vs. image concerns) has important implications for the design of optimal matching mechanisms aimed at promoting social learning.

Further empirical and theoretical research is called upon to test and refine the theory outlined in this paper. In a first step, experimental work should aim at deepening our understanding of the preference itself. How is the aversion affected by social context (e.g., whether S and R belong to the same or different social groups)? How do individuals quantify perceived disagreement and how exactly does it affect their utility? In a second step, experiments could test qualitative implications of the theory in terms of selective disclosure and matching patterns. From a theoretical perspective, perceived disagreement aversion could be studied within more general setups, for example featuring many senders and receivers, dynamic interaction, or richer matching processes.

Acknowledgments

We thank participants at various seminars and conferences for helpful comments as well as Peter Norman Sørensen (IAST Toulouse discussant) and Joel Sobel for detailed suggestions on an earlier draft.

Funding

Khalmetski gratefully acknowledges financial support of the German Research Foundation (DFG) through the Research Unit "Design and Behavior" (FOR 1371).

References

Aghion, P. and J. Tirole (1997). Formal and real authority in organizations. *Journal of political economy 105*(1), 1–29.

Asch, S. E. (1955). Opinions and social pressure. *Scientific American 193*(5), 31–35.

Banerjee, A. and R. Somanathan (2001). A simple model of voice. *The Quarterly Journal of Economics 116*(1), 189–227.

Barron, K., A. Becker, and S. Huck (2025). Motivated political reasoning: On the emergence of belief-value constellations. *European Economic Review 172*, 104929.

Battigalli, P. and M. Dufwenberg (2009). Dynamic psychological games. *Journal of Economic Theory 144*(1), 1–35.

Bonomi, G., N. Gennaioli, and G. Tabellini (2021). Identity, beliefs, and political conflict. *The Quarterly Journal of Economics 136*(4), 2371–2411.

Bursztyn, L., G. Egorov, and S. Fiorin (2020). From extreme to mainstream: The erosion of social norms. *The American Economic Review 110*(11), 3522–48.

Che, Y.-K. and N. Kartik (2009). Opinions as incentives. *Journal of Political Economy 117*(5), 815–860.

Dessein, W. (2002). Authority and communication in organizations. *The Review of Economic Studies 69*(4), 811–838.

Deutsch, M. and H. B. Gerard (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology 51*(3), 629.

Domínguez, D., F. Juan, S. A. Taing, and P. Molenberghs (2016). Why do some find it hard to disagree? an fmri study. *Frontiers in Human Neuroscience 9*, 718.

Dorison, C. A., J. A. Minson, and T. Rogers (2019). Selective exposure partly relies on faulty affective forecasts. *Cognition 188*, 98–107.

Dye, R. A. (1985). Disclosure of nonproprietary information. *Journal of Accounting Research 23*(1), 123–145.

Ekinci, E. and N. Theodoropoulos (2021). Disagreement and informal delegation in organizations. *International Journal of Industrial Organization 74*, 102696.

Ekins, E. E. (2020). Poll: 62% of americans say they have political views they are afraid to share. *Cato Institute, Survey Reports*.

Gentzkow, M. and J. M. Shapiro (2006). Media bias and reputation. *Journal of Political Economy 114*(2), 280–316.

Gentzkow, M. and J. M. Shapiro (2010). What drives media slant? evidence from us daily newspapers. *Econometrica 78*(1), 35–71.

Goffman, E. (1959). *The presentation of self in everyday life*. Garden City, NY: Doubleday Anchor Books.

Golman, R., G. Loewenstein, K. O. Moene, and L. Zarri (2016). The preference for belief consonance. *The Journal of Economic Perspectives 30*(3), 165–187.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality. *The Journal of Law and Economics 24*(3), 461–483.

Horrace, W. C. (2015). Moments of the truncated normal distribution. *Journal of Productivity Analysis 43*, 133–138.

Ilinov, P., A. Matveenko, M. Senkov, and E. Starkov (2024). Optimally biased expertise. *Working Paper*.

Jung, W.-O. and Y. K. Kwon (1988). Disclosure when the market is unsure of information endowment of managers. *Journal of Accounting Research 26*(1), 146–153.

Kartik, N., F. X. Lee, and W. Suen (2021). Information validates the prior: A theorem on bayesian updating and applications. *American Economic Review: Insights 3*(2), 165–82.

Lazarsfeld, P. F. and R. K. Merton (1954). Friendship as a social process: A substantive and methodological analysis. *Freedom and Control in Modern Society 18*(1), 18–66.

Levy, G. (2007). Decision making in committees: Transparency, reputation, and voting rules. *The American Economic Review 97*(1), 150–168.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics 12*(2), 380–391.

Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy.* Cambridge University Press.

Newcomb, T. M. (1961). *The acquaintance process.* Holt, Rinehart & Winston.

Ottaviani, M. and P. N. Sørensen (2006a). Reputational cheap talk. *The Rand Journal of Economics 37*(1), 155–175.

Ottaviani, M. and P. N. Sørensen (2006b). The strategy of professional forecasting. *Journal of Financial Economics 81*(2), 441–466.

Prentice, D. A. and D. T. Miller (1993). Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm. *Journal of Personality and Social Psychology 64*(2), 243.

Stroud, N. J. (2011). *Niche news: The politics of news choice.* Oxford University Press.

Sunstein, C. R. (2007). *Republic. com 2.0.* Princeton University Press.

Visser, B. and O. H. Swank (2007). On committees of experts. *The Quarterly Journal of Economics 122*(1), 337–372.

Appendix A: Microfoundation for perceived disagreement aversion

In this section we formalize an economic setting where the initial structure of preferences and information endogenously gives rise to a variant of our preferences for perceived disagreement aversion on the sender's side. That is, in this setting S cares only about her material payoff, but her optimal disclosure is the same as if she aimed to minimize R's perceived disagreement.

Assume that after the disclosure stage of our baseline model, R does not immediately take an action a, but instead has to decide whether to delegate decision rights to S or retain authority and take the decision himself. Formally, R chooses $\delta \in \{0, 1\}$, where $\delta = 0$ denotes no delegation and $\delta = 1$ denotes delegation. We next characterize S's and R's resulting expected utility which we denote by U_S^δ and U_R^δ respectively.

If R delegates ($\delta = 1$), authority transfers to S, who then chooses an action a_S trying to match the state of the world ω as closely as possible given her information set, i.e., to maximize

$$U_S^1(a_S) = -E_S\left[(\omega - a_S)^2 \,|\sigma\right].$$

R takes no further action and his expected payoff is then also determined by S's choice:

$$U_R^1(a_S) = -E_R\left[(\omega - a_S)^2 \,|d\right].$$

Instead, if R retains authority ($\delta = 0$) he has to take the action himself while S takes no further action. We assume that taking the action is costly for R, e.g., because he requires training or needs to acquire additional information to come to an informed decision. Concretely, we, hence, stipulate that R incurs a random cost of $c \geq 0$ supported on $\mathbb{R}^+$ and, thereby, gains access to S's information σ. R then takes the action a_R to maximize

$$U_R^0(a_R) = -E_R\left[(\omega - a_R)^2 \,|\sigma\right] - c,$$

while S's expected payoff is then

$$U_S^0(a_R) = -E_S\left[(\omega - a_R)^2 \,|\sigma\right],$$

i.e., S is just a passive beneficiary of R's choice.

Clearly, $U_S^1(a_S) \geq U_S^0(a_R)$ for all signal realizations σ, i.e., S always prefers delegation, where she is in full control of the optimal action rather than relying on R's choice affected by a different prior. R instead faces a trade-off between incurring the cost of acquiring information while choosing the optimal action himself, and relying on S's choice yet based on a different prior. Economic applications of this setting may include a CEO persuading the board of directors to grant her decision authority over the choice of an investment project; a financial advisor convincing a client to delegate authority over managing the client's investment funds; communication between a donor and a charity concerning the delegation of authority over the use of charitable funds, etc.[16]

We obtain that the equilibrium disclosure strategy of S is the same as under minimizing R's perceived disagreement as introduced in Section 2, yet with the ex post disagreement function $\Delta(\sigma)$ given by the quadratic rather than the absolute distance between beliefs:

Proposition 16 *S's chooses d to minimize $E_R[(E_S[\omega|\sigma] - E_R[\omega|\sigma])^2|d]$.*

Proof. Consider the strategy of R. Under no delegation, R's expected utility is

$$U_R^0(a_R) = -E_R[(a_R - \omega)^2|d] - c = -E_R[(E_R[\omega|\sigma] - \omega)^2|d] - c,$$

given that $a_i = E_i[\omega|\sigma]$ once player i minimizes $E_i[(a_i - \omega)^2|\sigma]$. Accordingly, under delegation, R expects to obtain

$$U_R^1(a_S) = -E_R[(a_S - \omega)^2|d] = -E_R[(E_S[\omega|\sigma] - \omega)^2|d].$$

Denote

$$\widetilde{\mu}_i(\sigma) := E_i[\omega|\sigma].$$

Hence, R chooses delegation if and only if

$$\begin{aligned} -E_R[(\widetilde{\mu}_R(\sigma) - \omega)^2|d] - c &\leq -E_R[(\widetilde{\mu}_S(\sigma) - \omega)^2|d] \\ \iff E_R[(\widetilde{\mu}_S(\sigma) - \omega)^2|d] - E_R[(\widetilde{\mu}_R(\sigma) - \omega)^2|d] &\leq c. \end{aligned} \tag{A.1}$$

Clearly, $U_S^1(a_S) \geq U_S^0(a_R)$ so that S always prefers delegation (i.e., the option to choose the optimal action herself). Then, since c is a random variable distributed on $\mathbb{R}^+$,

[16] For theoretical models involving strategic delegation under incomplete information see Aghion and Tirole (1997), Dessein (2002) and Ekinci and Theodoropoulos (2021).

S chooses disclosure to maximize the probability of delegation by R. By (A.1), this is equivalent to minimizing the term on the left-hand side:

$$\begin{aligned}
\widetilde{\Delta}_1(d) \quad : \quad & = E_R[(\widetilde{\mu}_S(\sigma)-\omega)^2|d] - E_R[(\widetilde{\mu}_R(\sigma)-\omega)^2|d] \\
= \quad & E_R[(\widetilde{\mu}_S(\sigma)-\omega)^2|d] - E_R[Var[\omega|\sigma]|d] \\
= \quad & (E_R[\widetilde{\mu}_S(\sigma)|d] - E_R[\omega|d])^2 \\
& + Var[\widetilde{\mu}_S(\sigma)|d] + Var[\omega|d] - 2Cov[\widetilde{\mu}_S(\sigma),\omega|d] \\
& - E_R[Var[\omega|\sigma]|d], \qquad \text{(A.2)}
\end{aligned}$$

where the second equality follows from $E_R[(\widetilde{\mu}_R(\sigma)-\omega)^2|d] = E_R[E_R[(\widetilde{\mu}_R(\sigma)-\omega)^2|\sigma]|d] = E_R[Var[\omega|\sigma]|d]$ by the law of iterated expectations.

At the same time, if S instead minimized R's perceived disagreement with the quadratic belief distance, then the corresponding expression to minimize would be

$$\begin{aligned}
\widetilde{\Delta}_2(d) \quad : \quad & = E_R[(\widetilde{\mu}_S(\sigma) - \widetilde{\mu}_R(\sigma))^2|d] \\
= \quad & (E_R[\widetilde{\mu}_S(\sigma)|d] - E_R[\widetilde{\mu}_R(\sigma)|d])^2 \\
& + Var[\widetilde{\mu}_S(\sigma)|d] + Var[\widetilde{\mu}_R(\sigma)|d] - 2Cov[\widetilde{\mu}_S(\sigma),\widetilde{\mu}_R(\sigma)|d] \\
= \quad & (E_R[\widetilde{\mu}_S(\sigma)|d] - E_R[\omega|d])^2 \\
& + Var[\widetilde{\mu}_S(\sigma)|d] + Var[\widetilde{\mu}_R(\sigma)|d] - 2Cov[\widetilde{\mu}_S(\sigma),\omega|d], \qquad \text{(A.3)}
\end{aligned}$$

where the last equality is by the law of iterated expectations. By (A.2) and (A.3),

$$\begin{aligned}
& \widetilde{\Delta}_1(d) - \widetilde{\Delta}_2(d) \\
= \quad & Var[\omega|d] - E_R[Var[\omega|\sigma]|d] - Var[\widetilde{\mu}_R(\sigma)|d] \\
= \quad & 0,
\end{aligned}$$

where the last equality is by the law of total variance. Thus, we finally obtain $\widetilde{\Delta}_1(d) = \widetilde{\Delta}_2(d)$, i.e., the equilibrium disclosure minimizes both $\widetilde{\Delta}_1(d)$ and $\widetilde{\Delta}_2(d)$. ■

Intuitively, R is more likely to delegate when his hypothetical belief about the state - formed under the assumption that he observes S's signal - is closer to S's posterior belief. In that case, R expects that S's action under delegation will be similar to the action R would take after acquiring S's signal under no delegation. Anticipating this, S has an incentive at the disclosure stage to minimize the expected gap between these beliefs -

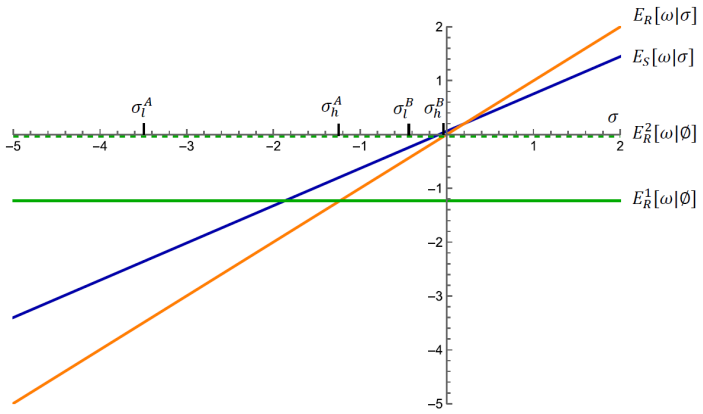

Figure B.1: (*Existence of multiple equilibria under aversion to actual disagreement if* $\gamma_S < \gamma_R$) The graph plots the sender's (blue line) and receiver's (orange line) posterior opinions as a function of the signal σ. The green lines depict the receiver's posterior opinions conditional on non-disclosure in two existing equilibria.

that is, to reduce R's perceived disagreement with S.

Thus, the delegation model provides an action-based microfoundation for perceived disagreement aversion with the quadratic distance between beliefs. While conceptually the quadratic measure of disagreement is similar to the one given by the absolute distance, in our analysis we retained the latter measure of disagreement for mathematical tractability.

Appendix B: Example of multiple equilibria under actual disagreement aversion

In this section we provide a numerical example demonstrating existence of multiple equilibria under aversion to actual disagreement considered in Section 5 if $\gamma_S < \gamma_R$.

Let $\mu_S = 0.2$, $\mu_R = 0$, $\gamma_S = 0.3$, $\gamma_R = 7$, $\gamma_\varepsilon = 0.2$, $\varphi = 0.9$. Then, there exist two equilibria: Equilibrium A with non-disclosure interval $(-3.51, -1.23)$ and equilibrium B with non-disclosure interval $(-0.41, -0.04)$. R's posterior belief conditional on non-disclosure is equal to -1.23 (-0.04) in equilibrium A(B) respectively. Figure B.1 illustrates both equilibria similar to Figure 6, with non-disclosure intervals for equilibria A and B denoted as (σ_l^A, σ_h^A) and (σ_l^B, σ_h^B), respectively. In particular, it shows that in equilibrium A, disclosing S's signal σ leads to a smaller actual disagreement than non-disclosure if and only if $\sigma \leq \sigma_l^A$ or $\sigma \geq \sigma_h^A$.

Notably both equilibria feature the opposite comparative statics with respect to both

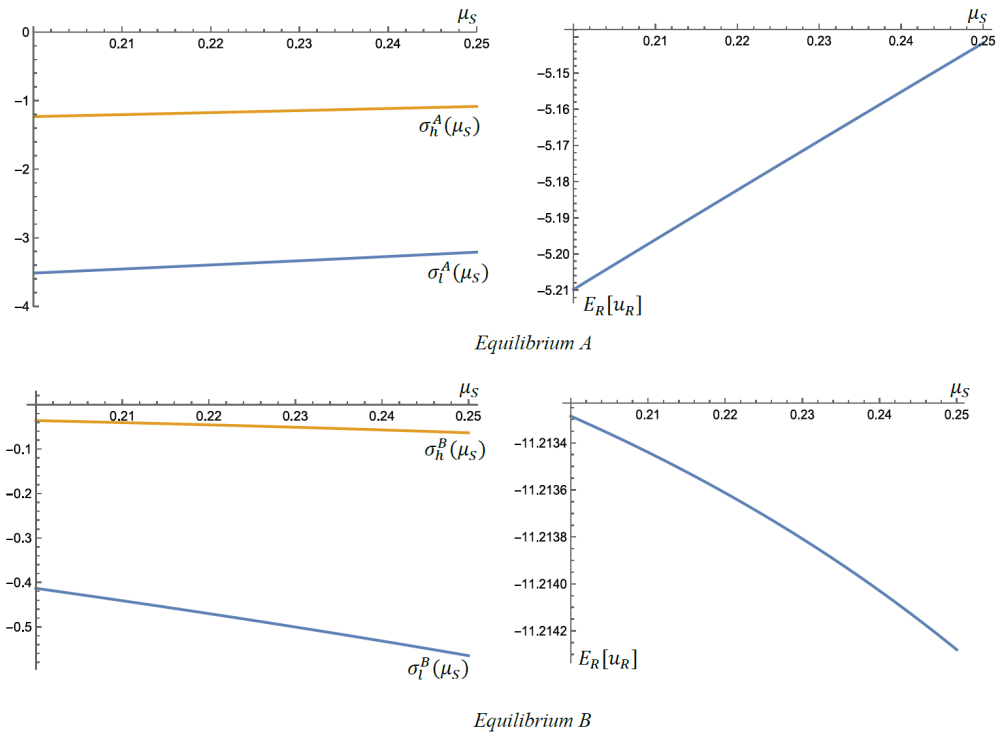

Figure B.2: (*Effect of μ_S on non-disclosure intervals and R's ex ante expected utility*) The graphs plot the effect of a change in μ_S on equilibrium non-disclosure intervals (left panel) and the receiver's expected utility from disclosure (right panel) in Equilibria A and B.

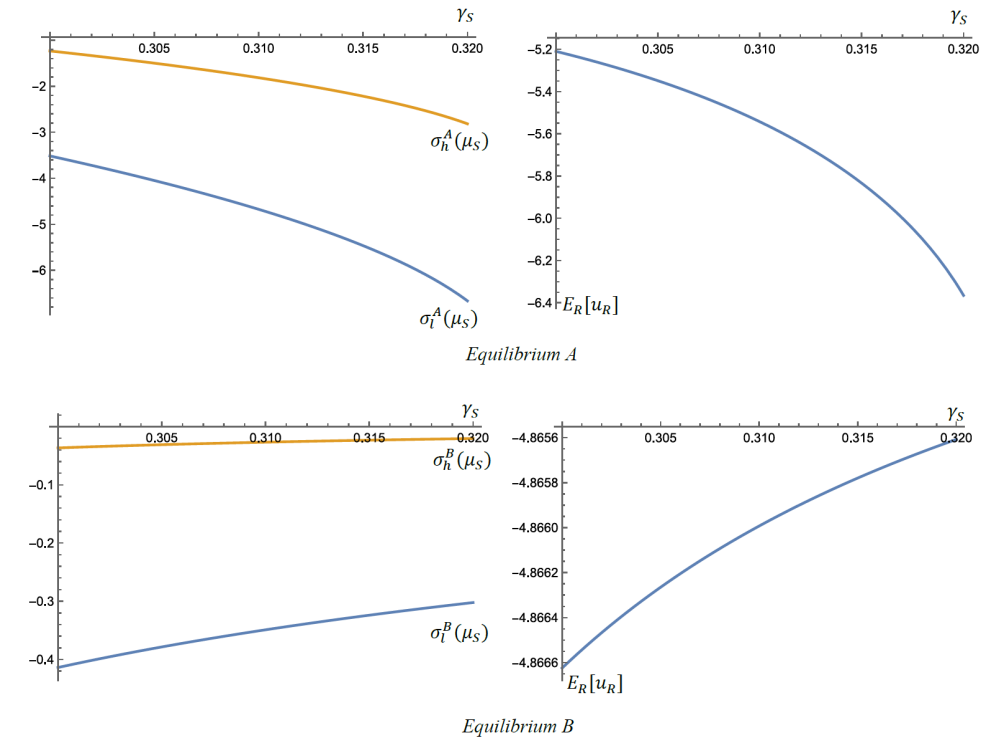

Figure B.3: (*Effect of γ_S on non-disclosure intervals and R's ex ante expected utility*) The graphs plot the effect of a change in γ_S on equilibrium non-disclosure intervals (left panel) and the receiver's expected utility from disclosure (right panel) in Equilibria A and B.

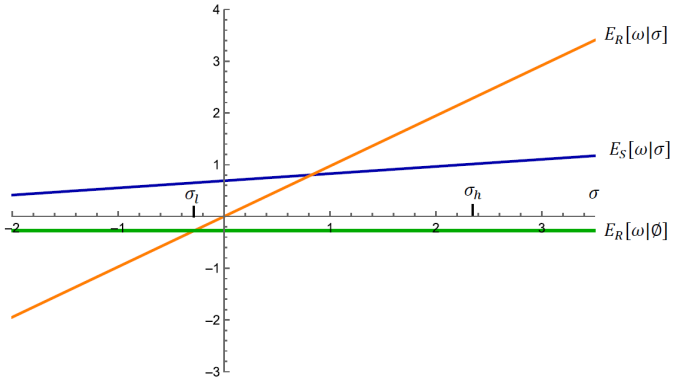

Figure B.4: (*Equilibrium with a compact disclosure interval under aversion to actual disagreement if* $\gamma_S < \gamma_R$) The graph plots the sender's (blue line) and receiver's (orange line) posterior opinions as a function of the signal σ. The green line depicts the receiver's posterior opinion conditional on non-disclosure in equilibrium.

μ_S and γ_S. In particular, increasing prior disagreement is beneficial for R in equilibrium A, and detrimental in equilibrium B. Figure B.2 shows the change in the non–disclosure interval as well as in R's expected utility in response to a change in μ_S from 0.2 to 0.25. One can see that the non-disclosure interval gets narrower (wider) in equilibrium A (B) as μ_S increases and, hence, gets more distant from $\mu_R = 0$. Accordingly, R would prefer a larger (smaller) prior disagreement with S in equilibrium A(B). R's preferences are similarly reversed between equilibria A and B with regards to a change in γ_S (see Figure B.3).

Actual disagreement aversion may also lead to equilibria of the same type (featuring compact disclosure intervals) as under perceived disagreement aversion. Figure B.4 shows posterior beliefs of S and R conditional on disclosure (blue and orange lines, respectively) and of R conditional on non-disclosure (green line) for one such equilibrium existing under parameter values $\mu_S = 0.8$, $\mu_R = 0$, $\gamma_S = 0.2$, $\gamma_R = 3$, $\gamma_\varepsilon = 0.5$ and $\varphi = 0.7$. One can see that actual disagreement is lower after disclosure than after non-disclosure if and only if $\sigma \in [\sigma_l, \sigma_h]$, so that S discloses only signals in this interval in equilibrium.

Appendix C (online): omitted proofs

C.1 Notational conventions

Throughout this appendix, we use the following notation.

Denote by $f_i(\sigma)$ the probability density functions of the distribution of S's signal σ in the eyes of player i, conditional on $\sigma \neq \varnothing$. Denote by $F_i(\sigma)$ the respective cumulative distribution function.

Besides, we define the following functions (where the agreement signal $\widetilde{\sigma}$ is given by (7)):

$$\sigma_l(\eta) \equiv \widetilde{\sigma} - \eta, \tag{C.1}$$

$$\sigma_h(\eta) \equiv \widetilde{\sigma} + \eta, \tag{C.2}$$

$$F_l(\eta) \equiv F_R(\sigma_l(\eta)), \tag{C.3}$$

$$F_h(\eta) \equiv F_R(\sigma_h(\eta)), \tag{C.4}$$

$$f_l(\eta) \equiv f_R(\sigma_l(\eta)), \tag{C.5}$$

$$f_h(\eta) \equiv f_R(\sigma_h(\eta)), \tag{C.6}$$

$$E_l(\eta) \equiv E_R[\sigma|\sigma < \sigma_l(\eta)] = \frac{1}{F_l(\eta)} \int_{-\infty}^{\sigma_l(\eta)} \sigma f_R(\sigma) d\sigma, \tag{C.7}$$

$$E_h(\eta) \equiv E_R[\sigma|\sigma > \sigma_h(\eta)] = \frac{1}{1 - F_h(\eta)} \int_{\sigma_h(\eta)}^{\infty} \sigma f_R(\sigma) d\sigma, \tag{C.8}$$

$$\Delta_0 \equiv |\mu_S - \mu_R|, \tag{C.9}$$

$$\widetilde{\Delta}_R(\varnothing|\eta) \equiv E_R[\Delta(\sigma)|\sigma \notin [\sigma_l(\eta), \sigma_h(\eta)]]. \tag{C.10}$$

Note, in particular, that $\widetilde{\Delta}_R(\varnothing|\eta^*)$ is R's perceived disagreement conditional on non-disclosure in the equilibrium defined by Proposition 1. The argument η in the above functions will be mostly suppressed below for notational convenience.

C.2 Equilibrium disclosure

C.2.1 Preliminaries

Lemma C.1 *For any $x \in \mathbb{R}$ and $i = S, R$ it holds*

$$E_i[\sigma|\sigma \ < \ x] = \mu_i - (\gamma_i^2 + \gamma_\varepsilon^2)\frac{f_i(x)}{F_i(x)}, \tag{C.11}$$

$$E_i[\sigma|\sigma \ > \ x] = \mu_i + (\gamma_i^2 + \gamma_\varepsilon^2)\frac{f_i(x)}{1 - F_i(x)}. \tag{C.12}$$

Proof. The result follows from the standard formula for the mean of the truncated normal distribution. ■

Lemma C.2 *Let $\gamma_S \neq \gamma_R$. The disclosure strategy in any equilibrium is defined by a finite interval $I(\eta^*) = [\widetilde{\sigma} - \eta^*, \widetilde{\sigma} + \eta^*]$ with $\eta^* > 0$ such that S discloses σ iff $\sigma \in I(\eta^*)$.*

Proof.

Step 1. In this step, we prove that for sufficiently high signals it holds $\Delta(\sigma) > \Delta_0$, i.e., disclosure increases disagreement relative to the prior level of disagreement. Recall that R's perceived disagreement is given by $\widetilde{\Delta}_R(d) = E_R[\Delta(\sigma)\,|d]$, where $\Delta(\sigma) = |E_S[\omega\,|\sigma\,] - E_R[\omega\,|\sigma\,]|$. Let us denote

$$D(\sigma) = E_S[\omega\,|\sigma\,] - E_R[\omega\,|\sigma\,], \tag{C.13}$$

so that $\Delta(\sigma) = |D(\sigma)|$. By (5) we have that for any $\sigma \in \mathbb{R}$,

$$D(\sigma) = \left(\frac{\gamma_S^2}{\gamma_S^2 + \gamma_\varepsilon^2} - \frac{\gamma_R^2}{\gamma_R^2 + \gamma_\varepsilon^2}\right)\sigma + \left(\frac{\gamma_\varepsilon^2\mu_S}{\gamma_S^2 + \gamma_\varepsilon^2} - \frac{\gamma_\varepsilon^2\mu_R}{\gamma_R^2 + \gamma_\varepsilon^2}\right). \tag{C.14}$$

Thus, $D(\sigma)$ is a linear function of $\sigma \in \mathbb{R}$, so that it has a unique root in $\mathbb{R}$ (once $\gamma_S \neq \gamma_R$). Consequently, $\Delta(\sigma) = |D(\sigma)|$ is V-shaped in $\sigma \in \mathbb{R}$, with the minimum value of 0 and tending to infinity as σ moves away from its unique root $\widetilde{\sigma}$. It follows immediately that for sufficiently high signals it holds $\Delta(\sigma) > \Delta_0$ (the same also holds for sufficiently low signals).

Step 2. Let us show that any equilibrium under $\gamma_S \neq \gamma_R$ features finite $\eta^* > 0$ such that S discloses σ if and only if $\sigma \in [\widetilde{\sigma} - \eta^*, \widetilde{\sigma} + \eta^*]$ (recall that $\widetilde{\sigma}$ is implicitly given by $D(\widetilde{\sigma}) = 0$).

First, note that full disclosure of all signals (FD) is never an equilibrium under $\gamma_S \neq \gamma_R$. Indeed, in this case $\widetilde{\Delta}_R(\varnothing) = \Delta_0$, as non-disclosure happens if and only if S is

uninformed (which is inferred by R in equilibrium). Then, S would have an incentive to deviate by concealing all signals σ such that $\Delta(\sigma) > \Delta_0$, which exist by Step 1.

Next, let us show that $\widetilde{\Delta}_R(\varnothing)$ must be finite and strictly positive. The fact that $\widetilde{\Delta}_R(\varnothing)$ should be finite can also be shown by contradiction. Suppose indeed that $\widetilde{\Delta}_R(\varnothing)$ is not finite. Then there would exist the FD-equilibrium, as S would always favour disclosing over not disclosing. But the FD-equilibrium cannot exist, as shown above.

The fact that $\widetilde{\Delta}_R(\varnothing)$ must be strictly positive is proved as follows. Recall that by definition it cannot be strictly negative. Suppose by contradiction that $\widetilde{\Delta}_R(\varnothing) = 0$. Then, since by (C.14) for any non-empty signal, $\Delta(\sigma) = |D(\sigma)| > 0$ if and only if $\sigma \neq \widetilde{\sigma}$ (under $\gamma_S \neq \gamma_R$), in equilibrium all signals must be concealed other than the signal $\widetilde{\sigma}$. This implies that after non-disclosure, R acknowledges that with a positive probability S holds a signal different from $\widetilde{\sigma}$. But for any such signal σ, given $\gamma_S \neq \gamma_R$, $\Delta(\sigma) > 0$. It follows immediately that $\widetilde{\Delta}_R(\varnothing) = E_R[\Delta(\sigma)|d = \varnothing] > 0$, which yields a contradiction.

Consider thus the only possible type of equilibrium with disclosure rule featuring a strictly positive and finite $\widetilde{\Delta}_R(\varnothing)$. In this case, given that, by (C.14), $\Delta(\sigma) = |D(\sigma)|$ as a function of $\sigma \in \mathbb{R}$ is symmetric around $\widetilde{\sigma}$ and linearly decreasing in σ for $\sigma \leq \widetilde{\sigma}$ (where $\Delta(\widetilde{\sigma}) = 0$), there exists a unique $\eta^* > 0$ (for given equilibrium level of $\widetilde{\Delta}_R(\varnothing)$) such that $\Delta(\widetilde{\sigma} - \eta^*) = \Delta(\widetilde{\sigma} + \eta^*) = \widetilde{\Delta}_R(\varnothing)$. The latter implies that for any $\sigma \in \mathbb{R}$, $\Delta(\sigma) \leq \widetilde{\Delta}_R(\varnothing)$ if and only if $\sigma \in [\widetilde{\sigma} - \eta^*, \widetilde{\sigma} + \eta^*]$. It follows immediately that the equilibrium disclosure set is given by $[\widetilde{\sigma} - \eta^*, \widetilde{\sigma} + \eta^*]$. ■

Lemma C.3 *Let $\gamma_S \neq \gamma_R$. A necessary and sufficient condition for $I(\eta) = [\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]$ to be an equilibrium disclosure interval at $\eta > 0$ is $\widetilde{\Delta}_R(\varnothing|\eta) = \Delta(\widetilde{\sigma} + \eta)$.*

Proof. Let us prove the necessity of this condition. Assume that the equilibrium disclosure interval is given by $I(\eta') = [\widetilde{\sigma} - \eta', \widetilde{\sigma} + \eta']$ for some $\eta' > 0$, so that R's perceived disagreement conditional on non-disclosure is given by $\widetilde{\Delta}_R(\varnothing|\eta')$ defined in (C.10). Then, S has incentive to disclose after obtaining a non-empty signal iff $\sigma \in I(\eta')$ or, equivalently,

$$\Delta(\sigma) \leq \widetilde{\Delta}_R(\varnothing|\eta') \text{ iff } \sigma \in [\widetilde{\sigma} - \eta', \widetilde{\sigma} + \eta'].$$

Since by (C.14), $\Delta(\sigma)$ as a function of $\sigma \in \mathbb{R}$ is symmetrically V-shaped around $\widetilde{\sigma}$ (with $\Delta(\widetilde{\sigma}) = 0$) if $\gamma_S \neq \gamma_R$, this implies that S is indifferent between disclosing and not

disclosing at both $\sigma_l = \widetilde{\sigma} - \eta'$ and $\sigma_h = \widetilde{\sigma} + \eta'$, i.e.,

$$\widetilde{\Delta}_R(\varnothing|\eta') = \Delta(\widetilde{\sigma} - \eta') = \Delta(\widetilde{\sigma} + \eta'). \tag{C.15}$$

Let us show the sufficiency of this condition, i.e., if $\widetilde{\Delta}_R(\varnothing|\eta) = \Delta(\widetilde{\sigma}+\eta)$ for some $\eta > 0$ while $\gamma_S \neq \gamma_R$, then there exists an equilibrium with disclosure interval $I = [\widetilde{\sigma}-\eta, \widetilde{\sigma}+\eta]$. Assume $\widetilde{\Delta}_R(\varnothing|\eta') = \Delta(\widetilde{\sigma} + \eta')$ for some $\eta' > 0$ while $\gamma_S \neq \gamma_R$. Since $\Delta(\sigma)$ as a function of $\sigma \in \mathbb{R}$ has a symmetric V-shape around $\widetilde{\sigma}$ if $\gamma_S \neq \gamma_R$, we have that for any non-empty signal σ,

$$\Delta(\sigma) \leq \Delta(\widetilde{\sigma} + \eta') \text{ iff } \sigma \in [\widetilde{\sigma} - \eta', \widetilde{\sigma} + \eta'].$$

This together with assumed $\Delta(\widetilde{\sigma} + \eta') = \widetilde{\Delta}_R(\varnothing|\eta')$ implies that for any non-empty signal σ,

$$\Delta(\sigma) \leq \widetilde{\Delta}_R(\varnothing|\eta') \text{ iff } \sigma \in [\widetilde{\sigma} - \eta', \widetilde{\sigma} + \eta'].$$

Consequently, there exists an equilibrium where S discloses σ if and only if $\sigma \in [\widetilde{\sigma} - \eta', \widetilde{\sigma} + \eta']$. ■

Lemma C.4 *Let $\gamma_S \neq \gamma_R$. A necessary and sufficient condition for $I(\eta) = [\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]$ to be an equilibrium disclosure interval at $\eta > 0$ is $\tau(\eta) = 0$, where*

$$\tau(\eta) = \varphi \begin{pmatrix} F_l(\sigma_l - \mu_R) \\ +(\gamma_R^2 + \gamma_\varepsilon^2)(f_l + f_h) \\ +(1 - F_h)(\mu_R - \sigma_h) \end{pmatrix} - (1 - \varphi)\left(\sigma_h + \frac{k_2 - \Delta_0}{k_1}\right) \tag{C.16}$$

with

$$\begin{aligned} k_1 &= sgn\,(\gamma_S - \gamma_R)\left(\frac{\gamma_S^2}{\gamma_S^2 + \gamma_\varepsilon^2} - \frac{\gamma_R^2}{\gamma_R^2 + \gamma_\varepsilon^2}\right), \\ k_2 &= sgn\,(\gamma_S - \gamma_R)\left(\frac{\gamma_\varepsilon^2 \mu_S}{\gamma_S^2 + \gamma_\varepsilon^2} - \frac{\gamma_\varepsilon^2 \mu_R}{\gamma_R^2 + \gamma_\varepsilon^2}\right). \end{aligned}$$

Proof. In what follows, we show that the necessary and sufficient equilibrium condition stated in Lemma C.3, $\Delta\left(\widetilde{\sigma} + \eta\right) = \widetilde{\Delta}_R\left(\varnothing|\eta\right)$, is equivalent to $\tau(\eta) = 0$.

By (C.10), we have

$$\begin{aligned}\widetilde{\Delta}_R(\varnothing|\eta) &= E_R[\Delta(\sigma)|\sigma \notin [\sigma_l, \sigma_h]] \\ &= \Pr[\sigma < \sigma_l] E_R[\Delta(\sigma)\,|\sigma < \sigma_l] + \Pr[\sigma > \sigma_h] E_R[\Delta(\sigma)\,|\sigma > \sigma_h] \\ &\quad + \Pr[\sigma = \varnothing]\Delta_0 \\ &= \frac{1}{\varphi(F_l + 1 - F_h) + 1 - \varphi} \begin{pmatrix} \varphi \int_{-\infty}^{\sigma_l} \Delta(\sigma) f_R(\sigma) d\sigma \\ +\varphi \int_{\sigma_h}^{\infty} \Delta(\sigma) f_R(\sigma) d\sigma \\ +(1-\varphi)\Delta_0 \end{pmatrix}. \end{aligned} \tag{C.17}$$

Note that by (C.13) and (C.14), for $\sigma \neq \varnothing$, $\Delta(\sigma)$ can be represented as

$$\Delta(\sigma) = |D(\sigma)| = \begin{cases} k_1\sigma + k_2 \text{ if } \sigma \geq \widetilde{\sigma}, \\ -k_1\sigma - k_2 \text{ if } \sigma < \widetilde{\sigma}, \end{cases} \tag{C.18}$$

where

$$k_1 = \text{sgn}\,(\gamma_S - \gamma_R)\left(\frac{\gamma_S^2}{\gamma_S^2 + \gamma_\varepsilon^2} - \frac{\gamma_R^2}{\gamma_R^2 + \gamma_\varepsilon^2}\right), \tag{C.19}$$

$$k_2 = \text{sgn}\,(\gamma_S - \gamma_R)\left(\frac{\gamma_\varepsilon^2\mu_S}{\gamma_S^2 + \gamma_\varepsilon^2} - \frac{\gamma_\varepsilon^2\mu_R}{\gamma_R^2 + \gamma_\varepsilon^2}\right). \tag{C.20}$$

Substituting this into (C.17) we obtain

$$\widetilde{\Delta}_R(\varnothing|\eta) = \frac{1}{\varphi(F_l + 1 - F_h) + 1 - \varphi} \begin{pmatrix} -\varphi k_1 F_l E_l - \varphi F_l k_2 \\ +\varphi k_1 (1 - F_h) E_h \\ +\varphi(1 - F_h)k_2 + (1-\varphi)\Delta_0 \end{pmatrix}. \tag{C.21}$$

This together with (C.18) implies that the equilibrium condition stated in Lemma C.3, $\Delta(\sigma_h) = \widetilde{\Delta}_R(\varnothing|\eta)$, is equivalent to:

$$\begin{aligned} 0 &= \frac{1}{\varphi(F_l + 1 - F_h) + 1 - \varphi} \begin{pmatrix} -\varphi k_1 F_l E_l - \varphi F_l k_2 + \varphi k_1 (1 - F_h) E_h \\ +\varphi(1 - F_h)k_2 + (1-\varphi)\Delta_0 \end{pmatrix} \\ &\quad -k_1\sigma_h - k_2. \end{aligned}$$

After rearrangements (using the result of Lemma C.1 and the fact that $\Delta(\sigma_l) = \Delta(\sigma_h) \Longleftrightarrow -k_1\sigma_l - k_2 = k_1\sigma_h + k_2$), we obtain

$$0 = \varphi \begin{pmatrix} F_l(\sigma_l - \mu_R) + (\gamma_R^2 + \gamma_\varepsilon^2)(f_l + f_h) \\ +(1 - F_h)(\mu_R - \sigma_h) \end{pmatrix} - (1 - \varphi)\left(\sigma_h + \frac{k_2 - \Delta_0}{k_1}\right). \tag{C.22}$$

Thus, we have shown that the necessary and sufficient equilibrium condition given by Lemma C.3, $\Delta(\sigma_h) = \widetilde{\Delta}_R(\varnothing|\eta)$, holds if and only if the last equality in (C.22) holds. ■

Lemma C.5 *Let $\gamma_S \neq \gamma_R$. Then, there exists at least one equilibrium.*

Proof. We show that if $\gamma_S \neq \gamma_R$, then there always exists $\eta > 0$ such that $\tau(\eta) = 0$, which is a necessary and sufficient equilibrium condition by Lemma C.4.

Step 1. This shows that $\tau(\eta) < 0$ for sufficiently large η. Indeed, by (C.16),

$$\begin{aligned}\tau(\eta) &= \varphi\left(\begin{array}{c} F_l(\widetilde{\sigma}-\eta-\mu_R)+(\gamma_R^2+\gamma_\varepsilon^2)(f_l+f_h) \\ +(1-F_h)(\mu_R-\widetilde{\sigma}-\eta)\end{array}\right)-(1-\varphi)\left(\widetilde{\sigma}+\eta+\frac{k_2-\Delta_0}{k_1}\right) \\ &= \lambda_1+\lambda_2, \end{aligned} \tag{C.23}$$

where

$$\begin{aligned}\lambda_1 &= \varphi\left(\begin{array}{c} F_l(\widetilde{\sigma}-\mu_R)+(\gamma_R^2+\gamma_\varepsilon^2)(f_l+f_h) \\ +(1-F_h)(\mu_R-\widetilde{\sigma})\end{array}\right), \\ \lambda_2 &= -\eta\left(\varphi(F_l+(1-F_h))+1-\varphi\right)-(1-\varphi)\left(\widetilde{\sigma}+\frac{k_2-\Delta_0}{k_1}\right).\end{aligned}$$

If $\eta \to \infty$, then $\sigma_l = \widetilde{\sigma}-\eta \to -\infty$ and $\sigma_h = \widetilde{\sigma}+\eta \to \infty$ so that $f_l \to 0$, $f_h \to 0$, $F_l \to 0$, $F_h \to 1$. Hence, $\lim_{\eta\to\infty}\lambda_1 = 0$ and $\lim_{\eta\to\infty}\lambda_2 = -\infty$. This implies by (C.23) that $\tau(\eta) < 0$ for sufficiently large η.

Step 2. This shows that there always exists $\eta > 0$ such that $\tau(\eta) = 0$.

Indeed, given that $\sigma_l = \widetilde{\sigma}-\eta$ and $\sigma_h = \widetilde{\sigma}+\eta$ by definition, we have $\lim_{\eta\to 0}\sigma_l = \lim_{\eta\to 0}\sigma_h = \widetilde{\sigma}$. Then, we obtain

$$\begin{aligned}\lim_{\eta\to 0}\tau(\eta) &= \varphi\lim_{\eta\to 0}\left(\begin{array}{c} F_l(\sigma_l-\mu_R)+(\gamma_R^2+\gamma_\varepsilon^2)(f_l+f_h) \\ +(1-F_h)(\mu_R-\sigma_h)\end{array}\right) \\ &\quad -(1-\varphi)\lim_{\eta\to 0}\left(\sigma_h+\frac{k_2-\Delta_0}{k_1}\right). \\ &= \varphi\left((\widetilde{\sigma}-\mu_R)(2F_R(\widetilde{\sigma})-1)+2(\gamma_R^2+\gamma_\varepsilon^2)f_R(\widetilde{\sigma})\right) \\ &\quad -(1-\varphi)\left(\widetilde{\sigma}+\frac{k_2-\Delta_0}{k_1}\right).\end{aligned}$$

The first term on the right-hand side is strictly positive since $F_R(\widetilde{\sigma}) \leq 0.5$ if and only if $\widetilde{\sigma} \leq \mu_R$. The second term $-(1-\varphi)\left(\widetilde{\sigma}+\frac{k_2-\Delta_0}{k_1}\right)$ is also positive since after substituting

for $\widetilde{\sigma}$ from (7) and for k_1 and k_2 from (C.19) and (C.20) we obtain

$$\widetilde{\sigma} + \frac{k_2 - \Delta_0}{k_1} = \text{sgn}[\gamma_S - \gamma_R] \frac{\Delta_0(\gamma_S^2 + \gamma_\varepsilon^2)(\gamma_R^2 + \gamma_\varepsilon^2)}{(\gamma_R^2 - \gamma_S^2)\gamma_\varepsilon^2} \leq 0.$$

Thus, $\lim_{\eta \to 0} \tau(\eta) > 0$. At the same time, by Step 1, $\tau(\eta) < 0$ for sufficiently large η. Consequently, by the continuity of $\tau(\eta)$ and the intermediate value theorem, there exists at least one $\eta > 0$ such that $\tau(\eta) = 0$. Then, by Lemma C.4 there should exist at least one equilibrium. ■

Lemma C.6 *Let $\gamma_S \neq \gamma_R$. Then, for any $\eta > 0$ it holds $\frac{d\tau(\eta)}{d\eta} = -(1 - \varphi(F_h - F_l)) < 0$, where $\tau(\eta)$ is defined in (C.16).*

Proof. By (C.16) we have

$$\tau(\eta) = \varphi \left(\begin{array}{c} F_l(\sigma_l - \mu_R) + (\gamma_R^2 + \gamma_\varepsilon^2)(f_l + f_h) \\ +(1 - F_h)(\mu_R - \sigma_h) \end{array} \right) - (1 - \varphi) \left(\sigma_h + \frac{k_2 - \Delta_0}{k_1} \right). \quad \text{(C.24)}$$

Taking the derivative with respect to η (noting that $\sigma_l = \widetilde{\sigma} - \eta$, $\sigma_h = \widetilde{\sigma} + \eta$, $f_i = f_R(\sigma_i)$ and $F_i = F_R(\sigma_i)$ are functions of η for $i = l, h$, while $\frac{df_R(\sigma)}{d\sigma} = f_R(\sigma)\frac{\mu_R - \sigma}{\gamma_R^2 + \gamma_\varepsilon^2}$ by the properties of the pdf of the normal distribution) we get

$$\begin{aligned} \frac{d\tau(\eta)}{d\eta} &= \varphi \left(\begin{array}{c} -f_l(\sigma_l - \mu_R) - F_l - f_l(\mu_R - \sigma_l) \\ +f_h(\mu_R - \sigma_h) - f_h(\mu_R - \sigma_h) - (1 - F_h) \end{array} \right) - (1 - \varphi) \\ &= -(1 - \varphi(F_h - F_l)) < 0. \end{aligned} \quad \text{(C.25)}$$

■

Proof of Proposition 1

Throughout the proof we let $\gamma_S \neq \gamma_R$.

Step 1. The existence of at least one equilibrium characterized by a partial disclosure interval $I = [\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]$ for some $\eta > 0$ (and non-existence of any other type of equilibrium) follows by Lemmas C.2 and C.5.

Next, Lemma C.6 implies that $\tau(\eta)$ defined in (C.16) is decreasing in η for $\eta > 0$. Hence, once a positive root of $\tau(\eta)$ exists, it must be unique. Consequently, since $\tau(\eta) = 0$ is a necessary and sufficient condition for $[\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]$ to be an equilibrium disclosure interval by Lemma C.4, the equilibrium is unique.

Step 2. This shows that the equilibrium disclosure interval strictly contains the status quo signals $\underline{\sigma}$ and $\overline{\sigma}$. I.e., the disclosure interval $[\widetilde{\sigma}-\eta^*, \widetilde{\sigma}+\eta^*]$ characterizing the unique equilibrium must be such that $\widetilde{\sigma}-\eta^* < \underline{\sigma}$ and $\overline{\sigma} < \widetilde{\sigma}+\eta^*$, where $\underline{\sigma}, \overline{\sigma}$ are such that $\Delta(\underline{\sigma}) = \Delta(\overline{\sigma}) = |\mu_S - \mu_R|$. This is equivalent to showing that $\widetilde{\Delta}_R(\varnothing|\eta^*) > \Delta(\underline{\sigma}) = \Delta(\overline{\sigma})$ so that S strictly prefers to disclose $\underline{\sigma}$ and $\overline{\sigma}$ over non-disclosure. The proof is by contradiction.

Suppose by contradiction that $\widetilde{\Delta}_R(\varnothing|\eta^*) \leq \Delta(\overline{\sigma}) = |\mu_S - \mu_R|$. No disclosure then implies two and only two possible scenarios that both have positive probability. Either σ is such that $\Delta(\sigma) > \widetilde{\Delta}_R(\varnothing|\eta^*)$ (i.e., S prefers non-disclosure over disclosing such signal). Or, S holds no signal, in which case disagreement is $|\mu_S - \mu_R| \geq \widetilde{\Delta}_R(\varnothing|\eta^*)$ by assumption. In other words, it must be true that $\widetilde{\Delta}_R(\varnothing|\eta^*)$ is a linear combination of

$$E\left[\Delta(\sigma) \,\middle|\, \Delta(\sigma) > \widetilde{\Delta}_R(\varnothing|\eta^*)\right],$$

and of $|\mu_S - \mu_R| \geq \widetilde{\Delta}_R(\varnothing|\eta^*)$. It follows immediately that $\widetilde{\Delta}_R(\varnothing|\eta^*) > \widetilde{\Delta}_R(\varnothing|\eta^*)$, which is a contradiction.

Step 3. This shows that if $\mu_S \neq \mu_R$, the center of the disclosure interval $\widetilde{\sigma}$ is closer to the prior mean of the more confident player (i.e., the player with the smaller prior variance). Assume $\mu_R > \mu_S$ (the proof for $\mu_R < \mu_S$ proceeds analogously and is omitted). Using (7), we obtain

$$\mu_S - \widetilde{\sigma} = \mu_S - \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2) - \mu_R(\gamma_S^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} = \frac{(\mu_R-\mu_S)(\gamma_S^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2}, \tag{C.26}$$

$$\mu_R - \widetilde{\sigma} = \mu_R - \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2) - \mu_R(\gamma_S^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} = \frac{(\mu_R-\mu_S)(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2}. \tag{C.27}$$

Hence, if $\gamma_R > \gamma_S$ we get $\mu_S - \widetilde{\sigma} > 0 \Rightarrow \widetilde{\sigma} < \mu_S < \mu_R$ (where the last inequality is by assumption), while if $\gamma_R < \gamma_S$ we get $\mu_R - \widetilde{\sigma} < 0 \Rightarrow \widetilde{\sigma} > \mu_R > \mu_S$. Thus, $\widetilde{\sigma} \notin [\mu_S, \mu_R]$ while $\widetilde{\sigma}$ is closer to the prior mean of the more confident player.

Step 4. This proves part ii of the proposition. Let $\gamma_S = \gamma_R$. Then, by (C.18) for any $\sigma \in \mathbb{R}$,

$$\Delta(\sigma) = \left|\frac{\gamma_\varepsilon^2}{\gamma_R^2+\gamma_\varepsilon^2}(\mu_S - \mu_R)\right| \leq |\mu_S - \mu_R|. \tag{C.28}$$

Consequently, there exists an equilibrium with full disclosure (as in such equilibrium non-disclosure reveals that S is indeed uninformed so that $\widetilde{\Delta}_R(\varnothing) = |\mu_S - \mu_R| \geq \Delta(\sigma)$ for

any σ).

Let us show that no other equilibrium exists. Assume by contradiction that there exists an equilibrium where some signal σ' is not disclosed. We have

$$\begin{aligned}\widetilde{\Delta}_R(\varnothing) &= \Pr(\sigma=\varnothing\,|d=\varnothing)\,|\mu_S-\mu_R| \\ &\quad + \Pr(\sigma\neq\varnothing\,|d=\varnothing)\left|\tfrac{\gamma_\varepsilon^2}{\gamma_R^2+\gamma_\varepsilon^2}\,(\mu_S-\mu_R)\right| \\ &\geq \left|\frac{\gamma_\varepsilon^2}{\gamma_R^2+\gamma_\varepsilon^2}\,(\mu_S-\mu_R)\right| = \Delta(\sigma'), \end{aligned} \tag{C.29}$$

where the second and the last equalities are by (C.28). Hence, S has incentive to deviate to disclosure after observing σ', which is a contradiction. ■

Corollary C.1 *Prior means always lie within the interior of the disclosure interval, i.e.,* $\mu_R \in (\widetilde{\sigma}-\eta^*, \widetilde{\sigma}+\eta^*)$ *and* $\mu_S \in (\widetilde{\sigma}-\eta^*, \widetilde{\sigma}+\eta^*)$.

Proof. If $d=\sigma=\mu_R$, then by (C.18) R's perceived disagreement is

$$\begin{aligned}\Delta(\mu_R) &= \left|\left(\frac{\gamma_S^2}{\gamma_S^2+\gamma_\varepsilon^2}-\frac{\gamma_R^2}{\gamma_R^2+\gamma_\varepsilon^2}\right)\mu_R+\left(\frac{\gamma_\varepsilon^2\mu_S}{\gamma_S^2+\gamma_\varepsilon^2}-\frac{\gamma_\varepsilon^2\mu_R}{\gamma_R^2+\gamma_\varepsilon^2}\right)\right| \\ &= \left|\frac{\gamma_\varepsilon^2}{\gamma_S^2+\gamma_\varepsilon^2}(\mu_S-\mu_R)\right| < |\mu_S-\mu_R|\,.\end{aligned}$$

Thus, $\Delta(\mu_R)<\Delta_0=\Delta(\underline{\sigma})=\Delta(\overline{\sigma})$, where the equalities are by definition of the status quo signals. Since $\Delta(\sigma)$ is V-shaped around $\widetilde{\sigma}$, this implies that $\mu_R\in(\underline{\sigma},\overline{\sigma})$. Consequently, by Proposition 1.i(b), $\mu_R\in(\widetilde{\sigma}-\eta^*,\widetilde{\sigma}+\eta^*)$.

The claim that $\mu_S\in(\widetilde{\sigma}-\eta^*,\widetilde{\sigma}+\eta^*)$ follows by the analogous argument as for μ_R. ■

C.3 Receiver preferences over senders

C.3.1 Preliminaries

Lemma C.7 *a) Let I' and I'' denote two equilibrium disclosure strategies such that S discloses iff $\sigma\in[\sigma'_l,\sigma'_h]$ and iff $\sigma\in[\sigma''_l,\sigma''_h]$ under I' and I'', respectively. Assume $[\sigma''_l,\sigma''_h]$ is a strict superset of $[\sigma'_l,\sigma'_h]$ such that $\sigma''_l<\sigma'_l<\sigma'_h<\sigma''_h$. Then, R's ex ante expected utility is strictly higher under I'' than under I'.*

b) R's ex ante expected utility is strictly higher under full disclosure than under partial disclosure.

Proof.

a) Consider two disclosure intervals $[\sigma_l', \sigma_h']$ and $[\sigma_l'', \sigma_h'']$ such that $\sigma_l'' < \sigma_l' < \sigma_h' < \sigma_h''$. Denote the optimal R's action given disclosure d as

$$a^*(d) = \arg\max_a E_R[u_R(\omega, a)|d] = \arg\max_a E_R[-(\omega - a)^2|d].$$

In particular, denote the optimal action conditional on non-disclosure given disclosure interval $[\sigma_l, \sigma_h]$ as $a^*(\varnothing, \sigma_l, \sigma_h)$. Recall also that S's signal space is given by $(-\infty, \infty) \cup \varnothing$, i.e., $\sigma \notin [\sigma_l, \sigma_h] \Leftrightarrow \sigma \in (-\infty, \sigma_l) \cup (\sigma_h, \infty) \cup \varnothing$. Then, R's ex ante expected utility conditional on disclosure interval $[\sigma_l'', \sigma_h'']$ is

$$\begin{aligned} E_R[u_R|\sigma_l'', \sigma_h''] &= \Pr[\sigma \notin [\sigma_l'', \sigma_h'']] E_R\left[u_R(\omega, a^*(\varnothing, \sigma_l'', \sigma_h''))|\sigma \notin [\sigma_l'', \sigma_h'']\right] \\ &\quad + \varphi \int_{\sigma_l''}^{\sigma_l'} E_R\left[u_R(\omega, a^*(\sigma)|\sigma\right] f_R(\sigma) d\sigma \\ &\quad + \varphi \int_{\sigma_h'}^{\sigma_h''} E_R\left[u_R(\omega, a^*(\sigma)|\sigma\right] f_R(\sigma) d\sigma \\ &\quad + \varphi \int_{\sigma_l'}^{\sigma_h'} E_R\left[u_R(\omega, a^*(\sigma))|\sigma\right] f_R(\sigma) d\sigma. \end{aligned}$$

Analogously, $E_R[u_R|\sigma_l', \sigma_h']$ can be decomposed as

$$\begin{aligned} E_R[u_R|\sigma_l', \sigma_h'] &= \Pr[\sigma \notin [\sigma_l'', \sigma_h'']] E_R\left[u_R(\omega, a^*(\varnothing, \sigma_l', \sigma_h'))|\sigma \notin [\sigma_l'', \sigma_h'']\right] \\ &\quad + \varphi \int_{\sigma_l''}^{\sigma_l'} E_R\left[u_R(\omega, a^*(\varnothing, \sigma_l', \sigma_h')|\sigma\right] f_R(\sigma) d\sigma \\ &\quad + \varphi \int_{\sigma_h'}^{\sigma_h''} E_R\left[u_R(\omega, a^*(\varnothing, \sigma_l', \sigma_h')|\sigma\right] f_R(\sigma) d\sigma \\ &\quad + \varphi \int_{\sigma_l'}^{\sigma_h'} E_R\left[u_R(\omega, a^*(\sigma))|\sigma\right] f_R(\sigma) d\sigma, \end{aligned}$$

where the second equality follows by the law of total expectations. Then,

$$E_R[u_R|\sigma_l'', \sigma_h''] - E_R[u_R|\sigma_l', \sigma_h'] = \lambda_1 + \lambda_2 + \lambda_3, \tag{C.30}$$

where

$$\lambda_1 = \Pr[\sigma \notin [\sigma_l'', \sigma_h'']] \begin{pmatrix} E_R\left[u_R(\omega, a^*(\varnothing, \sigma_l'', \sigma_h''))|\sigma \notin [\sigma_l'', \sigma_h'']\right] \\ -E_R\left[u_R(\omega, a^*(\varnothing, \sigma_l', \sigma_h'))|\sigma \notin [\sigma_l'', \sigma_h'']\right] \end{pmatrix},$$

$$\lambda_2 = \varphi \int_{\sigma_l''}^{\sigma_l'} \begin{pmatrix} E_R\left[u_R(\omega, a^*(\sigma)|\sigma\right] \\ -E_R\left[u_R(\omega, a^*(\varnothing, \sigma_l', \sigma_h')|\sigma\right] \end{pmatrix} f_R(\sigma)d\sigma,$$

$$\lambda_3 = \varphi \int_{\sigma_h'}^{\sigma_h''} \begin{pmatrix} E_R\left[u_R(\omega, a^*(\sigma)|\sigma\right] \\ -E_R\left[u_R(\omega, a^*(\varnothing, \sigma_l', \sigma_h')|\sigma\right] \end{pmatrix} f_R(\sigma)d\sigma.$$

We have that $\lambda_1 \geq 0$ since $a^*(\varnothing, \sigma_l'', \sigma_h'') = E[\omega|\sigma \notin [\sigma_l'', \sigma_h'']]$ is the unique maximum of $E_R[u_R(\omega, a)|\sigma \notin [\sigma_l'', \sigma_h'']]$ on $a \in \mathbb{R}$. Analogously, $\lambda_2 > 0$ and $\lambda_3 > 0$ since $a^*(\sigma)$ is the unique maximum of $E_R[u_R(\omega, a)|\sigma]$ on $a \in \mathbb{R}$. The inequalities are strict in the latter cases since, generally, $a^*(\sigma) \neq a^*(\varnothing, \sigma_l', \sigma_h')$ on a positive measure of signals. Given that $\lambda_1 \geq 0$, $\lambda_2 > 0$ and $\lambda_3 > 0$, (C.30) implies that $E_R[u_R|\sigma_l'', \sigma_h''] > E_R[u_R|\sigma_l', \sigma_h']$, i.e., R's ex ante expected utility is strictly higher under $[\sigma_l'', \sigma_h'']$.

b) The fact that R always prefers full disclosure over partial disclosure follows by the same arguments as in point a) while setting $\sigma_l'' = -\infty$ and $\sigma_h'' = \infty$. ■

Lemma C.8 *a) For any parameter $\kappa \in \{\mu_S, \gamma_S, \mu_R, \gamma_R\}$, if $\gamma_S \neq \gamma_R$ then*

$$\frac{d\sigma_l(\eta^*)}{d\kappa} = -\frac{1}{1-\varphi(F_h - F_l)} \frac{\partial \widetilde{\tau}(x,\kappa)}{\partial \kappa}_{|x=\sigma_l(\eta^*)},$$

where

$$\widetilde{\tau}(x,\kappa) = \varphi \begin{pmatrix} F_R[x](x-\mu_R) \\ +(\gamma_R^2+\gamma_\varepsilon^2)(f_R[x]+f_R[2\widetilde{\sigma}-x]) \\ +(1-F_R[2\widetilde{\sigma}-x])(\mu_R - 2\widetilde{\sigma}+x) \end{pmatrix} - (1-\varphi)\left(2\widetilde{\sigma} - x + \frac{k_2-\Delta_0}{k_1}\right).$$

b) For any parameter $\kappa \in \{\mu_S, \gamma_S, \mu_R, \gamma_R\}$, if $\gamma_S \neq \gamma_R$ then

$$\frac{d\sigma_h(\eta^*)}{d\kappa} = \frac{1}{1-\varphi(F_h - F_l)} \frac{\partial \widehat{\tau}(x,\kappa)}{\partial \kappa}_{|x=\sigma_h(\eta^*)},$$

where

$$\widehat{\tau}(x,\kappa) = \varphi \begin{pmatrix} F_R[2\widetilde{\sigma}-x](2\widetilde{\sigma}-x-\mu_R) \\ +(\gamma_R^2+\gamma_\varepsilon^2)(f_R[2\widetilde{\sigma}-x]+f_R[x]) \\ +(1-F_R[x])(\mu_R - x) \end{pmatrix} - (1-\varphi)\left(x + \frac{k_2-\Delta_0}{k_1}\right).$$

Proof. a) Consider $\widetilde{\tau}(x,\kappa)$ defined in the lemma for $\kappa \in \{\mu_S, \gamma_S, \mu_R, \gamma_R\}$ and assume

$\gamma_S \neq \gamma_R$. Note that at $x = \sigma_l(\eta^*)$ we obtain

$$\begin{aligned}\widetilde{\tau}(\sigma_l(\eta^*),\kappa) &= \varphi \begin{pmatrix} F_R[\sigma_l(\eta^*)](\sigma_l(\eta^*)-\mu_R) \\ +(\gamma_R^2+\gamma_\varepsilon^2)(f_R[\sigma_l(\eta^*)]+f_R[\sigma_h(\eta^*)]) \\ +(1-F_R[\sigma_h(\eta^*)])(\mu_R-\sigma_h(\eta^*)) \end{pmatrix} \\ &\quad -(1-\varphi)\left(\sigma_h(\eta^*)+\frac{k_2-\Delta_0}{k_1}\right) = \tau(\eta^*), \end{aligned} \tag{C.31}$$

where the first equality follow from $2\widetilde{\sigma} - \sigma_l(\eta) = 2\widetilde{\sigma} - (\widetilde{\sigma} - \eta) = \sigma_h(\eta)$. Since $\tau(\eta^*) = 0$ if $\gamma_S \neq \gamma_R$ by Lemma C.4, (C.31) implies

$$\begin{aligned}\widetilde{\tau}(\sigma_l(\eta^*),\kappa) &= \varphi \begin{pmatrix} F_R[\sigma_l(\eta^*)](\sigma_l(\eta^*)-\mu_R) \\ +(\gamma_R^2+\gamma_\varepsilon^2)(f_R[\sigma_l(\eta^*)]+f_R[2\widetilde{\sigma}-\sigma_l(\eta^*)]) \\ +(1-F_R[2\widetilde{\sigma}-\sigma_l(\eta^*)])(\mu_R-2\widetilde{\sigma}+\sigma_l(\eta^*)) \end{pmatrix} \\ &\quad -(1-\varphi)\left(2\widetilde{\sigma}-\sigma_l(\eta^*)+\frac{k_2-\Delta_0}{k_1}\right) = 0. \end{aligned}$$

Consequently, by the implicit function theorem, for any parameter $\kappa \in \{\mu_S, \gamma_S, \mu_R, \gamma_R\}$ at the equilibrium value of η^* it holds:

$$\frac{d\sigma_l(\eta^*)}{d\kappa} = -\frac{\partial\widetilde{\tau}(x,\kappa)/\partial\kappa}{\partial\widetilde{\tau}(x,\kappa)/\partial x}_{|x=\sigma_l(\eta^*)}, \tag{C.32}$$

where, in particular, $\frac{\partial\widetilde{\tau}(x,\kappa)}{d\kappa}_{|x=\sigma_l(\eta^*)}$ is the partial derivative of $\widetilde{\tau}(x,\kappa)$ with respect to κ keeping x constant at $\sigma_l(\eta^*)$ (which always exists for any $\kappa \in \{\mu_S, \gamma_S, \mu_R, \gamma_R\}$ if $\gamma_S \neq \gamma_R$).

Next, consider the denominator in (C.32). Note that for any $x \in \mathbb{R}$,

$$\begin{aligned}\widetilde{\tau}(x,\kappa) &= \varphi \begin{pmatrix} F_R[\widetilde{\sigma}-(\widetilde{\sigma}-x)](\widetilde{\sigma}-(\widetilde{\sigma}-x)-\mu_R) \\ +(\gamma_R^2+\gamma_\varepsilon^2)(f_R[\widetilde{\sigma}-(\widetilde{\sigma}-x)]+f_R[\widetilde{\sigma}+(\widetilde{\sigma}-x)]) \\ +(1-F_R[\widetilde{\sigma}+(\widetilde{\sigma}-x)])(\mu_R-(\widetilde{\sigma}+(\widetilde{\sigma}-x))) \end{pmatrix} \\ &\quad -(1-\varphi)\left(\widetilde{\sigma}+(\widetilde{\sigma}-x)+\frac{k_2-\Delta_0}{k_1}\right) \\ &= \tau(\widetilde{\sigma}-x). \end{aligned} \tag{C.33}$$

It follows that $\frac{\partial\widetilde{\tau}(x,\kappa)}{\partial x} = \frac{d\tau(\widetilde{\sigma}-x)}{dx} = -\frac{d\tau(\eta)}{d\eta}_{|\eta=\widetilde{\sigma}-x}$. This implies that at $x = \sigma_l(\eta^*) = \widetilde{\sigma} - \eta^*$ we have

$$\frac{\partial\widetilde{\tau}(x,\kappa)}{\partial x}_{|x=\sigma_l(\eta^*)} = -\frac{d\tau(\eta^*)}{d\eta^*} = 1-\varphi(F_h - F_l),$$

where the last equality is by Lemma C.6. Substituting this into (C.32) we finally obtain

$$\frac{d\sigma_l(\eta^*)}{d\kappa} = -\frac{1}{1-\varphi(F_h - F_l)} \frac{\partial \widetilde{\tau}(x,\kappa)}{\partial \kappa}_{|x=\sigma_l(\eta^*)}.$$

b) The proof proceeds analogously to a) and is omitted. ■

C.3.2 Proof of Proposition 2

In what follows, we normalize μ_R to 0 without loss of generality.

Case 1: $\gamma_S < \gamma_R$.

Step 1. Let us show that R's expected utility strictly increases in μ_S for $\mu_S > 0$.

Step 1a. Denote R's expected quadratic loss as

$$L_R = -E_R[u_R] = E_R[(a-\omega)^2].$$

Let us show that at any equilibrium values of σ_l and σ_h, L_R strictly decreases in σ_h if σ_l adjusts in such a way that the probability of disclosure remains constant (formal definition is provided below).

Let $g_R(\omega)$ denote R's prior density function of ω, and $g_R(\omega|\sigma)$ denote R's posterior density function of ω conditional on σ. Then, given S's equilibrium disclosure strategy and R's optimal action conditional on disclosure $a_R(d) = E_R[\omega|d]$, we have

$$
\begin{aligned}
L_R &= E_R[(a_R-\omega)^2] \\
&= E_R[(E_R[\omega|d]-\omega)^2] \\
&= \varphi\int_{\sigma_l}^{\sigma_h}\int_{-\infty}^{\infty}\left(E_R[\omega|\sigma]-\omega)^2\right)g_R(\omega|\sigma)d\omega f_R(\sigma)d\sigma \\
&\quad +\varphi\int_{-\infty}^{\sigma_l}\int_{-\infty}^{\infty}\left(E_R[\omega|\varnothing]-\omega)^2\right)g_R(\omega|\sigma)d\omega f_R(\sigma)d\sigma \\
&\quad +\varphi\int_{\sigma_h}^{\infty}\int_{-\infty}^{\infty}\left(E_R[\omega|\varnothing]-\omega)^2\right)g_R(\omega|\sigma)d\omega f_R(\sigma)d\sigma \\
&\quad +(1-\varphi)\int_{-\infty}^{\infty}\left(E_R[\omega|\varnothing]-\omega)^2\right)g_R(\omega)d\omega \\
&= \varphi\Pr[\sigma\in[\sigma_l,\sigma_h]]\frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}} \\
&\quad +\varphi\int_{-\infty}^{\sigma_l}\int_{-\infty}^{\infty}\left(E_R[\omega|\varnothing]-\omega)^2\right)g_R(\omega|\sigma)d\omega f_R(\sigma)d\sigma \\
&\quad +\varphi\int_{\sigma_h}^{\infty}\int_{-\infty}^{\infty}\left(E_R[\omega|\varnothing]-\omega)^2\right)g_R(\omega|\sigma)d\omega f_R(\sigma)d\sigma \\
&\quad +(1-\varphi)\int_{-\infty}^{\infty}\left(E_R[\omega|\varnothing]-\omega)^2\right)g_R(\omega)d\omega,
\end{aligned}
$$

where the second equality is due to $\int_{-\infty}^{\infty}\left(E_R[\omega|\sigma]-\omega)^2\right)g_R(\omega|\sigma)d\omega$ being the posterior variance of ω conditional on disclosed signal σ, which is constant at $\frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}}$ for any σ by (6). This further simplifies to the following function of σ_l and σ_h (using Lemma C.1):

$$
\begin{aligned}
L_R(\sigma_l,\sigma_h) &= \varphi\Pr[\sigma\in[\sigma_l,\sigma_h]]\frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}} \\
&\quad +\varphi\begin{pmatrix} E_R[\omega|\varnothing,\sigma_l,\sigma_h]^2\left(F_R(\sigma_l)+1-F_R(\sigma_h)\right) \\ +2\gamma_R^2E_R[\omega|\varnothing,\sigma_l,\sigma_h](f_R(\sigma_l)-f_R(\sigma_h))\end{pmatrix} \\
&\quad +\varphi\begin{pmatrix} \int_{-\infty}^{\sigma_l}\int_{-\infty}^{\infty}\omega^2g_R(\omega|\sigma)d\omega f_R(\sigma)d\sigma \\ +\int_{\sigma_h}^{\infty}\int_{-\infty}^{\infty}\omega^2g_R(\omega|\sigma)d\omega f_R(\sigma)d\sigma\end{pmatrix} \\
&\quad +(1-\varphi)\left(E_R[\omega|\varnothing,\sigma_l,\sigma_h]^2+\int_{-\infty}^{\infty}\omega^2g_R(\omega)d\omega\right). \qquad \text{(C.34)}
\end{aligned}
$$

Denote by function $\widehat{\sigma}_l(P_D,\sigma_h)$ the level of σ_l such that the ex ante probability of disclosure for given σ_h from R's perspective is equal to P_D, i.e., such that

$$
\varphi\Pr[\sigma\in[\widehat{\sigma}_l(P_D,\sigma_h),\sigma_h]]=\varphi(F_R(\sigma_h)-F_R(\widehat{\sigma}_l(P_D,\sigma_h)))=P_D.
$$

Consider some equilibrium values of σ_l and σ_h, σ_l' and σ_h', and denote $P_D'=\Pr_R[\sigma\in$

$[\sigma'_l, \sigma'_h]]$. Let us take the derivative of $L_R(\sigma'_l, \sigma'_h)$ with respect to σ_h so that σ_l is adjusted to keep the probability of disclosure constant at P'_D. This is equivalent to the total derivative of $L_R(\widehat{\sigma}_l(P'_D, \sigma'_h), \sigma'_h)$ (given by (C.34)) with respect to σ_h. Taking this derivative (where $\frac{\partial \widehat{\sigma}_l(P'_D, \sigma'_h)}{\partial \mu_S} = \frac{f_h}{f_l}$ by the implicit function theorem) and simplifying, we obtain:

$$\frac{dL_R(\widehat{\sigma}_l(P'_D, \sigma'_h), \sigma'_h)}{d\sigma_h} = \varphi f_R(\sigma'_h) \int_{-\infty}^{\infty} (E_R[\omega|\varnothing] - \omega)^2 \left(g_R(\omega|\sigma'_l) - g_R(\omega|\sigma'_h)\right) d\omega. \quad \text{(C.35)}$$

Note that $g_R(\omega|\sigma'_l) - g_R(\omega|\sigma'_h)$ is the distance between two pdfs of the normal distributions with means $E_R[\omega|\sigma'_l]$ and $E_R[\omega|\sigma'_h]$ and the same variance. Hence, it is a symmetric function around $x = \frac{E[\omega|\sigma'_h] + E[\omega|\sigma'_l]}{2}$ in the sense that for any ω,

$$g_R(\omega|\sigma'_l) - g_R(\omega|\sigma'_h) = g_R(2x - \omega|\sigma'_h) - g_R(2x - \omega|\sigma'_l). \quad \text{(C.36)}$$

Consequently, the integral on the right-hand side of (C.35) further simplifies to

$$\begin{aligned}
& \int_{-\infty}^{\infty} (E_R[\omega|\varnothing] - \omega)^2 \left(g_R(\omega|\sigma'_l) - g_R(\omega|\sigma'_h)\right) d\omega \\
= & \int_{-\infty}^{x} (E_R[\omega|\varnothing] - \omega)^2 \left(g_R(\omega|\sigma'_l) - g_R(\omega|\sigma'_h)\right) d\omega \\
& + \int_{x}^{\infty} (E_R[\omega|\varnothing] - \omega)^2 \left(g_R(\omega|\sigma'_l) - g_R(\omega|\sigma'_h)\right) d\omega \\
= & \int_{-\infty}^{x} (E_R[\omega|\varnothing] - \omega)^2 \left(g_R(\omega|\sigma'_l) - g_R(\omega|\sigma'_h)\right) d\omega \\
& - \int_{-\infty}^{x} (E_R[\omega|\varnothing] - (2x - \omega))^2 \left(g_R(\omega|\sigma'_l) - g_R(\omega|\sigma'_h)\right) d\omega \\
= & \; 4 \left(E_R[\omega|\varnothing] - x\right) \int_{-\infty}^{x} (x - \omega) \left(g_R(\omega|\sigma'_l) - g_R(\omega|\sigma'_h)\right) d\omega, \quad \text{(C.37)}
\end{aligned}$$

where the second equality follows by (C.36) and integration by substitution. The sign of the last right-hand side coincides with the sign of $E_R[\omega|\varnothing] - x$ (as the integrand is clearly positive for any $\omega \leq x$).

Let us show that $E_R[\omega|\varnothing] - x < 0$. We have (given that μ_R is normalized to 0):

$$E_R[\omega|\varnothing] = \frac{\varphi}{\varphi(F_l + 1 - F_h) + 1 - \varphi} \left(\int_{-\infty}^{\sigma'_l} E_R[\omega|\sigma] f_R(\sigma) d\sigma + \int_{\sigma'_h}^{\infty} E_R[\omega|\sigma] f_R(\sigma) d\sigma \right).$$

Substituting for $E_R[\omega|\sigma]$ from (5) and simplifying using Lemma C.1 we obtain

$$E_R[\omega|\varnothing] = \frac{\varphi \gamma_R^2}{\varphi(F_l + 1 - F_h) + 1 - \varphi} \left(f_R(\sigma'_h) - f_R(\sigma'_l)\right) < 0, \quad \text{(C.38)}$$

where the inequality follows from the fact that $f_R(\sigma'_h) - f_R(\sigma'_l) = f_R(\widetilde{\sigma} + \eta) - f_R(\widetilde{\sigma} - \eta) < 0$

given that $\mu_R = 0$ and $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} > 0$ (since $\gamma_S < \gamma_R$ and $\mu_S > 0$ by assumption).

In turn, by (5)

$$x = \frac{E_R[\omega|\sigma'_h] + E_R[\omega|\sigma'_l]}{2} = \frac{\gamma_R^2}{\gamma_R^2+\gamma_\varepsilon^2}\frac{\sigma'_l+\sigma'_h}{2} = \frac{\gamma_R^2}{\gamma_R^2+\gamma_\varepsilon^2}\widetilde{\sigma} > 0, \tag{C.39}$$

again since $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} \geq 0$. (C.38) and (C.39) together imply that $E_R[\omega|\varnothing] - x < 0$, so that by (C.35) and (C.37) we finally obtain

$$\frac{dL_R(\widehat{\sigma}_l(P'_D, \sigma'_h), \sigma'_h)}{d\sigma_h} < 0. \tag{C.40}$$

Thus, at any equilibrium values of σ_l and σ_h, L_R strictly decreases in σ_h if σ_l is adjusted in such a way that the probability of disclosure remains constant.

Step 1b. Let us show that σ_h strictly increases in μ_S. By Lemma C.8,

$$\frac{d\sigma_h}{d\mu_S} = \frac{1}{1-\varphi(F_h-F_l)}\frac{\partial\widehat{\tau}(x,\mu_S)}{\partial\mu_S}_{|x=\sigma_h(\eta^*)}, \tag{C.41}$$

where

$$\widehat{\tau}(x,\kappa) = \varphi\begin{pmatrix} F_R[2\widetilde{\sigma}-x](2\widetilde{\sigma}-x-\mu_R) \\ +(\gamma_R^2+\gamma_\varepsilon^2)(f_R[2\widetilde{\sigma}-x]+f_R[x]) \\ +(1-F_R[x])(\mu_R-x) \end{pmatrix} - (1-\varphi)\left(x + \frac{k_2-\Delta_0}{k_1}\right). \tag{C.42}$$

Consider $\frac{\partial\widehat{\tau}(x,\mu_S)}{\partial\mu_S}_{|x=\sigma_h(\eta^*)}$. Taking the partial derivative of the right-hand side of (C.42) with respect to μ_S at $x = \sigma_h(\eta^*)$ (treating x as constant), and subsequently substituting $2\widetilde{\sigma} - \sigma_h(\eta^*) = \widetilde{\sigma} - \eta^* = \sigma_l(\eta^*)$ yields:

$$\begin{aligned}\frac{\partial\widehat{\tau}(x,\mu_S)}{\partial\mu_S}_{|x=\sigma_h(\eta^*)} &= \varphi\begin{pmatrix} 2f_l\frac{d\widetilde{\sigma}}{d\mu_S}\sigma_l(\eta^*) + 2F_l\frac{d\widetilde{\sigma}}{d\mu_S} \\ +2(\gamma_R^2+\gamma_\varepsilon^2)\frac{df_l}{d\sigma_l}\frac{d\widetilde{\sigma}}{d\mu_S} \end{pmatrix} \\ &\quad -(1-\varphi)\frac{\partial\frac{k_2-\Delta_0}{k_1}}{\partial\mu_S}.\end{aligned} \tag{C.43}$$

By the properties of the pdf of the normal distribution, $\frac{df_l}{d\sigma_l} = \frac{df_R(\sigma_l)}{d\sigma_l} = \frac{\mu_R-\sigma_l}{\gamma_R^2+\gamma_\varepsilon^2}f_l$. Substituting this into (C.43) yields

$$\frac{\partial\widehat{\tau}(x,\mu_S)}{\partial\mu_S}_{|x=\sigma_h(\eta^*)} = 2\varphi F_l\frac{d\widetilde{\sigma}}{d\mu_S} - (1-\varphi)\frac{\partial\frac{k_2-\Delta_0}{k_1}}{\partial\mu_S}.$$

Taking the derivatives on the right-hand side and simplifying, we eventually obtain:

$$\frac{\partial \widehat{\tau}(x,\mu_S)}{\partial \mu_S}_{|x=\sigma_h(\eta^*)} = \frac{(\gamma_R^2+\gamma_\varepsilon^2)\left(\gamma_S^2(1-\varphi)+2\gamma_\varepsilon^2(1-(1-F_l)\varphi)\right)}{(\gamma_R^2-\gamma_S^2)\gamma_\varepsilon^2} > 0. \tag{C.44}$$

It follows by (C.41) that the same is true for $\frac{d\sigma_h(\eta^*)}{d\mu_S} > 0$.

Step 1c. Let us show that the probability of disclosure increases in μ_S. We have

$$\begin{aligned} \frac{\partial P_D}{\partial \mu_S} &= \varphi \frac{\partial (F_h - F_l)}{\partial \mu_S} = \varphi\left(f_h \frac{\partial \sigma_h}{\partial \mu_S} - f_l \frac{\partial \sigma_l}{\partial \mu_S}\right) \\ &= \varphi\left(f_h \frac{\partial \sigma_h}{\partial \mu_S} - f_l\left(2\frac{d\widetilde{\sigma}}{d\mu_S} - \frac{d\sigma_h}{d\mu_S}\right)\right), \end{aligned} \tag{C.45}$$

where the last equality is by $\sigma_l = \widetilde{\sigma} - \eta = 2\widetilde{\sigma} - \sigma_h$. By (C.41) and (C.44),

$$\frac{d\sigma_h}{d\mu_S} = \frac{(\gamma_R^2+\gamma_\varepsilon^2)\left(\gamma_S^2(1-\varphi)+2\gamma_\varepsilon^2(1-(1-F_l)\varphi)\right)}{(1-\varphi(F_h-F_l))(\gamma_R^2-\gamma_S^2)\gamma_\varepsilon^2}.$$

Substituting this, together with $\frac{d\widetilde{\sigma}}{d\mu_S} = \frac{\gamma_R^2+\gamma_\varepsilon^2}{\gamma_R^2-\gamma_S^2}$, into (C.45) and simplifying we obtain

$$\frac{\partial P_D}{\partial \mu_S} = \frac{(\gamma_R^2+\gamma_\varepsilon^2)((1-\varphi)\gamma_S^2(f_l+f_h)+2\gamma_\varepsilon^2 X)}{(\gamma_R^2-\gamma_S^2)\gamma_\varepsilon^2(1-\varphi(F_h-F_l))}, \tag{C.46}$$

where

$$X = f_h - \varphi f_l(1-F_h) - \varphi f_h(1-F_l). \tag{C.47}$$

Let us show that $X > 0$ (so that $\frac{\partial P_D}{\partial \mu_S} > 0$ as well). We have:

$$\begin{aligned} X &= f_h - \varphi f_l(1-F_h) - \varphi f_h(1-F_l) \\ &> f_h - f_l(1-F_h) - f_h(1-F_l) \\ &= f_h F_l - f_l(1-F_h) \\ &= f_R(-\sigma_h)F_l - f_l F_R(-\sigma_h), \end{aligned} \tag{C.48}$$

where the last equality holds since μ_R is normalized to 0 (so that f_R is symmetric around 0). In turn, the expression on the right hand-side is positive if and only if

$$\frac{f_R(-\sigma_h)}{F_R(-\sigma_h)} > \frac{f_l}{F_l}, \tag{C.49}$$

which is true since the inverse Mills ratio $\frac{f_R(\sigma)}{F_R(\sigma)}$ is decreasing in σ and the fact that $\sigma_l > -\sigma_h$ since $\widetilde{\sigma}$ (i.e., the average of σ_l and σ_h) is positive in the considered parameter case. Consequently, $X > 0$ by (C.48). Then, by (C.46), the probability of disclosure

strictly increases in μ_S.

Step 1d. Let us finally show that R's expected utility strictly increases in μ_S for $\mu_S > 0$. Consider two means, $\mu_R < \mu'_S < \mu''_S$ and two corresponding equilibrium disclosure intervals $[\sigma'_l, \sigma'_h]$ and $[\sigma''_l, \sigma''_h]$ with two disclosure probabilities $P'_{D,R}$ and $P''_{D,R}$. By Step 1b, $\sigma''_h > \sigma'_h$. Consequently, by Step 1a,

$$u_R(\sigma'_l, \sigma'_h) < u_R(\widehat{\sigma}_l(P'_D, \sigma'_h), \sigma''_h). \tag{C.50}$$

Finally, by Step 1c, $P'_{D,R} < P''_{D,R}$ so that $[\sigma''_l, \sigma''_h]$ must yield a higher disclosure probability than $[\sigma'_l, \sigma'_h]$ and hence $[\widehat{\sigma}_l(P'_{D,R}, \sigma'_h), \sigma''_h]$ (as the two latter intervals yielding the same disclosure probability by construction). Consequently, $[\sigma''_l, \sigma''_h]$ must be a superset of $[\widehat{\sigma}_l(P'_{D,R}, \sigma'_h), \sigma''_h]$ and hence $\sigma''_l < \widehat{\sigma}_l(P'_D, \sigma'_h)$. Then, by Lemma C.7, $u_R(\widehat{\sigma}_l(P'_D, \sigma'_h), \sigma''_h) < u_R(\sigma''_l, \sigma''_h)$. This together with (C.50) implies $u_R(\sigma'_l, \sigma'_h) < u_R(\sigma''_l, \sigma''_h)$.

Step 2. By analogous arguments (and also symmetry considerations), R's expected utility strictly decreases in μ_S for $\mu_S < 0$.

Step 3. Let us show the claim of the proposition, i.e., that if $\gamma_S < \gamma_R$, then for any μ'_S, μ''_S it holds $E[u_R|\mu'_S] \lesseqgtr E[u_R|\mu''_S]$ if $|\mu_R - \mu'_S| \lesseqgtr |\mu_R - \mu''_S|$.

Without loss of generality, assume $\mu'_S < \mu''_S$. The claim for the fixed order of μ_S and μ_R, i.e., for the cases $\mu'_S < \mu''_S \leq \mu_R$ and $\mu_R \leq \mu'_S < \mu''_S$, holds by Steps 1 and 2 (given also the continuity of R's expected utility in μ_S at $\mu_S = \mu_R = 0$). Consider the last case $\mu'_S < \mu_R < \mu''_S$.

Denote $\widehat{\mu}'_S = -\mu'_S$ so that $\widehat{\mu}'_S$ and μ'_S are equidistant to $\mu_R = 0$. By symmetry considerations, the disclosure intervals under $\widehat{\mu}'_S$ and μ'_S are symmetric around μ_R. Then, by symmetry considerations, R obtains the same ex ante expected loss $E_R[(a-\omega)^2]$ under these symmetric disclosure intervals:

$$E[u_R|\mu'_S] = E[u_R|\widehat{\mu}'_S]. \tag{C.51}$$

Since, by assumption, $\mu'_S < \mu_R \Leftrightarrow \widehat{\mu}'_S > \mu_R$ and $\mu''_S > \mu_R$ (i.e., $\widehat{\mu}'_S$ and μ''_S are both on the same side of μ_R), by Steps 1 and 2 it holds

$$E[u_R|\widehat{\mu}'_S] \lesseqgtr E[u_R|\mu''_S] \text{ if } |\mu_R - \widehat{\mu}'_S| \lesseqgtr |\mu_R - \mu''_S| \Leftrightarrow |\mu_R - \mu'_S| \lesseqgtr |\mu_R - \mu''_S|.$$

Since $E[u_R|\widehat{\mu}'_S] = E[u_R|\mu'_S]$ by (C.51), this implies

$$E[u_R|\mu'_S] \lesseqgtr E[u_R|\mu''_S] \text{ if } |\mu_R - \mu'_S| \lesseqgtr |\mu_R - \mu''_S|.$$

Case 2: $\gamma_S > \gamma_R$.

The proof proceeds analogously and is omitted. ■

C.3.3 Proof of Proposition 3

In what follows, we normalize μ_R to 0 and let $\mu_S \geq 0$ without loss of generality.

Case 1: $\gamma_S < \gamma_R$.

Let us show that R's expected utility strictly increases in γ_S.

Step 1. Denote again R's expected quadratic loss as

$$L_R = -E_R[u_R] = E_R[(a-\omega)^2].$$

Denote also by function $\widehat{\sigma}_l(P_D, \sigma_h)$ the level of σ_l such that the ex ante probability of disclosure for given σ_h from R's perspective is equal to P_D, i.e., such that

$$\varphi \Pr[\sigma \in [\widehat{\sigma}_l(P_D, \sigma_h), \sigma_h]] = \varphi(F_R(\sigma_h) - F_R(\widehat{\sigma}_l(P_D, \sigma_h))) = P_D.$$

By Step 1a of the proof of Proposition 2 we have that at any equilibrium values of σ_l and σ_h, L_R strictly decreases in σ_h if σ_l adjusts in such a way that the probability of disclosure remains constant, i.e.,

$$\frac{dL_R(\widehat{\sigma}_l(P_D, \sigma_h), \sigma_h)}{d\sigma_h} < 0. \tag{C.52}$$

Step 2. Let us show that σ_h strictly increases in γ_S. By Lemma C.8,

$$\frac{d\sigma_h}{d\gamma_S} = \frac{1}{1-\varphi(F_h - F_l)} \frac{\partial \widehat{\tau}(x, \gamma_S)}{\partial \gamma_S}_{|x=\sigma_h(\eta^*)}, \tag{C.53}$$

where

$$\widehat{\tau}(x, \kappa) = \varphi \begin{pmatrix} F_R[2\widetilde{\sigma} - x](2\widetilde{\sigma} - x - \mu_R) \\ +(\gamma_R^2 + \gamma_\varepsilon^2)(f_R[2\widetilde{\sigma} - x] + f_R[x]) \\ +(1 - F_R[x])(\mu_R - x) \end{pmatrix} - (1-\varphi)\left(x + \frac{k_2 - \Delta_0}{k_1}\right). \tag{C.54}$$

Consider $\frac{\partial\widehat{\tau}(x,\gamma_S)}{\partial\gamma_S}_{|x=\sigma_h(\eta^*)}$. Taking the partial derivative of the right-hand side of (C.54) with respect to γ_S at $x=\sigma_h(\eta^*)$ (treating x as constant), and subsequently substituting $2\widetilde{\sigma}-\sigma_h(\eta^*)=\widetilde{\sigma}-\eta^*=\sigma_l(\eta^*)$ yields:

$$\begin{aligned}\frac{\partial\widehat{\tau}(x,\gamma_S)}{\partial\gamma_S}_{|x=\sigma_h(\eta^*)} &= \varphi\begin{pmatrix}2f_l\frac{d\widetilde{\sigma}}{d\gamma_S}(\sigma_l(\eta^*)-\mu_R)+2F_l\frac{d\widetilde{\sigma}}{d\gamma_S}\\ +2(\gamma_R^2+\gamma_\varepsilon^2)\frac{df_l}{d\sigma_l}\frac{d\widetilde{\sigma}}{d\gamma_S}\end{pmatrix}\\ &\quad -(1-\varphi)\frac{\partial\frac{k_2-\Delta_0}{k_1}}{\partial\gamma_S}. \qquad \text{(C.55)}\end{aligned}$$

By the properties of the pdf of the normal distribution, $\frac{df_l}{d\sigma_l}=\frac{df_R(\sigma_l)}{d\sigma_l}=\frac{\mu_R-\sigma_l}{\gamma_R^2+\gamma_\varepsilon^2}f_l$. Substituting this into (C.55) yields

$$\frac{\partial\widehat{\tau}(x,\gamma_S)}{\partial\gamma_S}_{|x=\sigma_h(\eta^*)}=2\varphi F_l\frac{d\widetilde{\sigma}}{d\gamma_S}-(1-\varphi)\frac{\partial\frac{k_2-\Delta_0}{k_1}}{\partial\gamma_S}.$$

Taking the derivatives on the right-hand side and simplifying, we eventually obtain:

$$\frac{\partial\widehat{\tau}(x,\gamma_S)}{\partial\gamma_S}_{|x=\sigma_h(\eta^*)}=\frac{2\gamma_S(\gamma_R^2+\gamma_\varepsilon^2)\mu_S(\gamma_R^2(1-\varphi)+2\gamma_\varepsilon^2(1-\varphi(1-F_l))}{(\gamma_R^2-\gamma_S^2)^2\gamma_\varepsilon^2}>0. \qquad \text{(C.56)}$$

It follows by (C.53) that the same is true for $\frac{d\sigma_h(\eta^*)}{d\gamma_S}>0$.

Step 3. Let us show that the probability of disclosure increases in γ_S. We have

$$\begin{aligned}\frac{\partial P_D}{\partial\gamma_S} &= \varphi\frac{\partial(F_h-F_l)}{\partial\gamma_S}=\varphi\left(f_h\frac{\partial\sigma_h}{\partial\gamma_S}-f_l\frac{\partial\sigma_l}{\partial\gamma_S}\right)\\ &= \varphi\left(f_h\frac{\partial\sigma_h}{\partial\gamma_S}-f_l\left(2\frac{d\widetilde{\sigma}}{d\gamma_S}-\frac{d\sigma_h}{d\gamma_S}\right)\right), \qquad \text{(C.57)}\end{aligned}$$

where the last equality is by $\sigma_l=\widetilde{\sigma}-\eta=2\widetilde{\sigma}-\sigma_h$. By (C.53) and (C.56),

$$\frac{d\sigma_h}{d\gamma_S}=\frac{2\gamma_S(\gamma_R^2+\gamma_\varepsilon^2)\mu_S(\gamma_R^2(1-\varphi)+2\gamma_\varepsilon^2(1-\varphi(1-F_l))}{(1-\varphi(F_h-F_l))(\gamma_R^2-\gamma_S^2)^2\gamma_\varepsilon^2}.$$

Substituting this, together with $\frac{d\widetilde{\sigma}}{d\gamma_S}=\frac{2\gamma_S(\gamma_R^2+\gamma_\varepsilon^2)\mu_S}{(\gamma_R^2-\gamma_S^2)^2}$, into (C.57) and simplifying we obtain

$$\frac{\partial P_D}{\partial\gamma_S}=\frac{2\gamma_S\mu_S(\gamma_R^2+\gamma_\varepsilon^2)((1-\varphi)\gamma_R^2(f_l+f_h)+2\gamma_\varepsilon^2X)}{(\gamma_R^2-\gamma_S^2)^2\gamma_\varepsilon^2(1-\varphi(F_h-F_l))}, \qquad \text{(C.58)}$$

where

$$X=f_h-\varphi f_l(1-F_h)-\varphi f_h(1-F_l). \qquad \text{(C.59)}$$

By the same argument as in Step 1c of the proof of Proposition 2, $X>0$. Then, by (C.58), $\frac{\partial P_D}{\partial\gamma_S}>0$ as well, i.e., the probability of disclosure strictly increases in γ_S.

Step 4. Let us finally show that R's expected utility strictly increases in γ_S. Consider two variances, $\gamma_S' < \gamma_S'' < \gamma_R$ and two corresponding equilibrium disclosure intervals $[\sigma_l', \sigma_h']$ and $[\sigma_l'', \sigma_h'']$ with two disclosure probabilities $P_{D,R}'$ and $P_{D,R}''$. By Step 2, $\sigma_h'' > \sigma_h'$. Consequently, by Step 1,

$$u_R(\sigma_l', \sigma_h') < u_R(\widehat{\sigma}_l(P_D', \sigma_h'), \sigma_h''). \tag{C.60}$$

Finally, by Step 3, $P_{D,R}' < P_{D,R}''$ so that $[\sigma_l'', \sigma_h'']$ must yield a higher disclosure probability than $[\sigma_l', \sigma_h']$ and hence $[\widehat{\sigma}_l(P_{D,R}', \sigma_h'), \sigma_h'']$ (as the two latter intervals yielding the same disclosure probability by construction). Consequently, $[\sigma_l'', \sigma_h'']$ must be a superset of $[\widehat{\sigma}_l(P_{D,R}', \sigma_h'), \sigma_h'']$ and hence $\sigma_l'' < \widehat{\sigma}_l(P_D', \sigma_h')$. Then, by Lemma C.7, $u_R(\widehat{\sigma}_l(P_D', \sigma_h'), \sigma_h'') < u_R(\sigma_l'', \sigma_h'')$. This together with (C.60) implies $u_R(\sigma_l', \sigma_h') < u_R(\sigma_l'', \sigma_h'')$.

Case 2: $\gamma_S > \gamma_R$.

The proof that R's expected utility strictly decreases in γ_S in this case proceeds analogously to Case 1 and is omitted. ■

C.3.4 Proof of Proposition 4

Step 1. This proves point a) of the Proposition. By Lemma C.4, given priors (μ_S, γ_S) and (μ_R, γ_R) it holds in equilibrium

$$\varphi \begin{pmatrix} F_l(\sigma_l - \mu_R) + (\gamma_R^2 + \gamma_\varepsilon^2)(f_l + f_h) \\ +(1 - F_h)(\mu_R - \sigma_h) \end{pmatrix} - (1 - \varphi)\left(\sigma_h + \frac{k_2 - \Delta_0}{k_1}\right) = 0. \tag{C.61}$$

Rewriting the left-hand explicitly in terms of η we have

$$\varphi \begin{pmatrix} F_R[\widetilde{\sigma} - \eta](\widetilde{\sigma} - \eta - \mu_R) \\ +(\widetilde{\gamma}_R^2 + \gamma_\varepsilon^2)(f_R[\widetilde{\sigma} - \eta] + f_R[\widetilde{\sigma} + \eta]) \\ +(1 - F_R[\widetilde{\sigma} + \eta])(\mu_R - \widetilde{\sigma} - \eta) \end{pmatrix} - (1 - \varphi)\left(\widetilde{\sigma} + \eta + \frac{k_2 - \Delta_0}{k_1}\right) = 0. \tag{C.62}$$

Further, note that for $\mu_S = \mu_R$

$$\begin{aligned} \frac{k_2 - \Delta_0}{k_1} &= \frac{\text{sgn }(\gamma_S - \gamma_R)\left(\frac{\gamma_\varepsilon^2 \mu_R}{\gamma_S^2 + \gamma_\varepsilon^2} - \frac{\gamma_\varepsilon^2 \mu_R}{\gamma_R^2 + \gamma_\varepsilon^2}\right)}{\text{sgn }(\gamma_S - \gamma_R)\left(\frac{\gamma_S^2}{\gamma_S^2 + \gamma_\varepsilon^2} - \frac{\gamma_R^2}{\gamma_R^2 + \gamma_\varepsilon^2}\right)} = -\mu_R, \\ \widetilde{\sigma} &= \frac{\mu_R(\gamma_R^2 + \gamma_\varepsilon^2) - \mu_R(\gamma_S^2 + \gamma_\varepsilon^2)}{\gamma_R^2 - \gamma_S^2} = \mu_R. \end{aligned}$$

Substituting this into the equilibrium condition (C.62) yields

$$\varphi \begin{pmatrix} -F_R[\mu_R-\eta]\eta \\ +(\gamma_R^2+\gamma_\varepsilon^2)(f_R[\mu_R-\eta]+f_R[\mu_R+\eta]) \\ -(1-F_R[\mu_R+\eta])\eta \end{pmatrix} - (1-\varphi)\eta = 0.$$

Note that the left-hand side does not depend on γ_S. Consequently, the equilibrium solution for η (unique by Proposition 1.i(a)) does not depend on γ_S. It follows that the equilibrium disclosure interval (and hence R's ex ante expected utility), $[\widetilde{\sigma}-\eta^*, \widetilde{\sigma}+\eta^*] = [\mu_R-\eta^*, \mu_R+\eta^*]$, does not depend on γ_S.

Step 2. Let us show point b) of the proposition. By Propositions 1.ii and 1.i(a), $\gamma_S = \gamma_R$ ($\gamma_S \neq \gamma_R$) implies full (partial) disclosure in equilibrium. Consequently, R's expected utility is strictly higher under $\gamma_S = \gamma_R$ by Lemma C.7(b). ■

C.4 Sender preferences over receivers

C.4.1 Preliminaries

Lemma C.9 *Let $\mu_S \geq \mu_R$. Then,* $\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\mu_R} = \begin{cases} -\frac{\gamma_S^2(1-\varphi)+\gamma_\varepsilon^2(1-\varphi(F_h+F_l))}{(\gamma_S^2+\gamma_\varepsilon^2)(1-F_h\varphi+F_l\varphi)} & \text{if } \gamma_S > \gamma_R \\ -\frac{\gamma_S^2(1-\varphi)+\gamma_\varepsilon^2(1-\varphi(2-F_h-F_l))}{(\gamma_S^2+\gamma_\varepsilon^2)(1-F_h\varphi+F_l\varphi)} & \text{if } \gamma_S < \gamma_R \end{cases}$.

Proof. Since in equilibrium $\widetilde{\Delta}_R(\varnothing|\eta^*) = \Delta(\sigma_h)$ by Lemma C.3, we have (given that $\frac{dk_1}{d\mu_R} = 0$)

$$\begin{aligned} \frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\mu_R} &= \frac{d\Delta(\sigma_h)}{d\mu_R} = \frac{d(k_1\sigma_h+k_2)}{d\mu_R} \\ &= k_1\left(\frac{1}{1-\varphi(F_h-F_l)}\frac{\partial\widehat{\tau}(x,\mu_R)}{\partial\mu_R}_{|x=\sigma_h(\eta^*)}\right) + \frac{dk_2}{d\mu_R}, \end{aligned} \tag{C.63}$$

where the second equality is by (C.18), and the last equality is by Lemma C.8.

Let us compute $\frac{\partial\widehat{\tau}(x,\mu_R)}{\partial\mu_R}_{|x=\sigma_h(\eta^*)}$. By definition in Lemma C.8 we have

$$\widehat{\tau}(x,\mu_R) = \varphi \begin{pmatrix} F_R[2\widetilde{\sigma}-x](2\widetilde{\sigma}-x-\mu_R) \\ +(\gamma_R^2+\gamma_\varepsilon^2)(f_R[2\widetilde{\sigma}-x]+f_R[x]) \\ +(1-F_R[x])(\mu_R-x) \end{pmatrix} - (1-\varphi)\left(x+\frac{k_2-\Delta_0}{k_1}\right).$$

Taking the partial derivative of the right-hand side with respect to μ_R at $x = \sigma_h(\eta^*)$ (treating x as constant), and subsequently substituting $2\widetilde{\sigma}-\sigma_h(\eta^*) = \widetilde{\sigma}-\eta^* = \sigma_l(\eta^*)$

yields:

$$\frac{\partial\widehat{\tau}(x,\mu_R)}{\partial\mu_R}_{|x=\sigma_h(\eta^*)} = \varphi\left(\begin{array}{c}\left(\frac{\partial F_l}{\partial\mu_R}+2f_l\frac{d\widetilde{\sigma}}{d\mu_R}\right)(\sigma_l(\eta^*)-\mu_R)+F_l\left(2\frac{d\widetilde{\sigma}}{d\mu_R}-1\right)\\ +(\gamma_R^2+\gamma_\varepsilon^2)\left(\frac{\partial f_l}{\partial\mu_R}+2\frac{df_l}{d\sigma_l}\frac{d\widetilde{\sigma}}{d\mu_R}+\frac{\partial f_h}{\partial\mu_R}\right)\\ -\frac{\partial F_h}{\partial\mu_R}(\mu_R-\sigma_h(\eta^*))+(1-F_h)\end{array}\right)$$
$$-(1-\varphi)\frac{\partial\frac{k_2-\Delta_0}{k_1}}{\partial\mu_R}, \tag{C.64}$$

where $\frac{\partial f_k}{\partial\mu_R}=\frac{\partial f_R(\sigma_k,\mu_R)}{\partial\mu_R}$ and $\frac{\partial F_k}{\partial\mu_R}=\frac{\partial F_R(\sigma_k,\mu_R)}{\partial\mu_R}$ are the partial derivatives of, respectively, $f_R(\sigma_k,\mu_R)$ and $F_R(\sigma_k,\mu_R)$ with respect to μ_R keeping σ_k constant, for $k=l,h$. By the properties of the cdf and pdf of the normal distribution, $\frac{\partial F_l}{\partial\mu_R}=-f_l$, $\frac{\partial F_h}{\partial\mu_R}=-f_h$, $\frac{\partial f_l}{\partial\mu_R}=-f_l\frac{\mu_R-\sigma_l}{\gamma_R^2+\gamma_\varepsilon^2}$, $\frac{\partial f_h}{\partial\mu_R}=-f_h\frac{\mu_R-\sigma_h}{\gamma_R^2+\gamma_\varepsilon^2}$, $\frac{df_l}{d\sigma_l}=f_l\frac{\mu_R-\sigma_l}{\gamma_R^2+\gamma_\varepsilon^2}$. Substituting the above expressions into (C.64) yields

$$\frac{\partial\widehat{\tau}(x,\mu_R)}{\partial\mu_R}_{|x=\sigma_h(\eta^*)} = \varphi\left(F_l\left(2\frac{d\widetilde{\sigma}}{d\mu_R}-1\right)+(1-F_h)\right)-(1-\varphi)\frac{\partial\frac{k_2-\Delta_0}{k_1}}{\partial\mu_R}. \tag{C.65}$$

Substituting this into (C.63), taking the derivatives and simplifying, we finally get the expression for $\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\mu_R}$ stated in the lemma. ■

Lemma C.10 *Let $\mu_S\geq\mu_R=0$ and $\gamma_S\neq\gamma_R$. Then,*

$$\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\gamma_R} = sgn(\gamma_R-\gamma_S)\frac{(f_h+f_l)\gamma_R\gamma_\varepsilon^2(\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2)\varphi}{(\gamma_S^2+\gamma_\varepsilon^2)(\gamma_R^2+\gamma_\varepsilon^2)(1-F_h\varphi+F_l\varphi)}$$

Proof: Since in equilibrium $\widetilde{\Delta}_R(\varnothing|\eta^*)=\Delta(\sigma_h)$ by Lemma C.3, we have (noting that $\frac{dk_2}{d\gamma_R}=0$ as far as μ_R is normalized to 0)

$$\begin{aligned}\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\gamma_R} &= \frac{d\Delta(\sigma_h)}{d\gamma_R}=\frac{d(k_1\sigma_h+k_2)}{d\gamma_R}\\ &= k_1\frac{d\sigma_h}{d\gamma_R}+\frac{dk_1}{d\gamma_R}\sigma_h\\ &= k_1\left(\frac{1}{1-\varphi(F_h-F_l)}\frac{\partial\widehat{\tau}(x,\gamma_R)}{\partial\gamma_R}_{|x=\sigma_h(\eta^*)}\right)+\frac{dk_1}{d\gamma_R}\sigma_h,\end{aligned} \tag{C.66}$$

where the second equality is by (C.18), and the last equality is by Lemma C.8.

Let us compute $\frac{\partial \widehat{\tau}(x,\gamma_R)}{\partial \gamma_R}_{|x=\sigma_h(\eta^*)}$. By definition in Lemma C.8 we have

$$\begin{aligned}\widehat{\tau}(x,\gamma_R) &= \varphi \begin{pmatrix} F_R[2\widetilde{\sigma}-x](2\widetilde{\sigma}-x-\mu_R) \\ +(\gamma_R^2+\gamma_\varepsilon^2)(f_R[2\widetilde{\sigma}-x]+f_R[x]) \\ +(1-F_R[x])(\mu_R-x) \end{pmatrix} \\ &\quad -(1-\varphi)\left(x+\frac{k_2-\Delta_0}{k_1}\right). \end{aligned} \tag{C.67}$$

Taking the partial derivative of the right-hand side with respect to γ_R at $x=\sigma_h(\eta^*)$ (treating x as constant), and subsequently substituting $2\widetilde{\sigma}-\sigma_h(\eta^*)=\widetilde{\sigma}-\eta^*=\sigma_l(\eta^*)$ yields:

$$\begin{aligned}&\frac{\partial \widehat{\tau}(x,\gamma_R)}{\partial \gamma_R}_{|x=\sigma_h(\eta^*)} \\ &= \varphi \begin{pmatrix} \left(\frac{\partial F_l}{\partial \gamma_R}+2f_l\frac{d\widetilde{\sigma}}{d\gamma_R}\right)(\sigma_l(\eta^*)-\mu_R)+2F_l\frac{d\widetilde{\sigma}}{d\gamma_R} \\ +(\gamma_R^2+\gamma_\varepsilon^2)\left(\frac{\partial f_l}{\partial \gamma_R}+2\frac{df_l}{d\sigma_l}\frac{d\widetilde{\sigma}}{d\gamma_R}+\frac{\partial f_h}{\partial \gamma_R}\right) \\ +2\gamma_R(f_l+f_h)-\frac{\partial F_h}{\partial \gamma_R}(\mu_R-\sigma_h(\eta^*)) \end{pmatrix} \\ &\quad -(1-\varphi)\frac{\partial \frac{k_2-\Delta_0}{k_1}}{\partial \gamma_R}, \end{aligned} \tag{C.68}$$

where $\frac{\partial f_k}{\partial \gamma_R}=\frac{\partial f_R(\sigma_k,\gamma_R)}{\partial \gamma_R}$ and $\frac{\partial F_k}{\partial \gamma_R}=\frac{\partial F_R(\sigma_k,\gamma_R)}{\partial \gamma_R}$ are the partial derivatives of, respectively, $f_R(\sigma_k,\gamma_R)$ and $F_R(\sigma_k,\gamma_R)$ with respect to γ_R keeping σ_k constant, for $k=l,h$. By the properties of the cdf and pdf of the normal distribution (using $\sigma_l=2\widetilde{\sigma}-\sigma_h$),

$$\frac{\partial F_l}{\partial \gamma_R} = -f_l\frac{\gamma_R}{\gamma_R^2+\gamma_\varepsilon^2}\sigma_l, \tag{C.69}$$

$$\frac{\partial F_h}{\partial \gamma_R} = -f_h\frac{\gamma_R}{\gamma_R^2+\gamma_\varepsilon^2}\sigma_h, \tag{C.70}$$

$$\frac{\partial f_l}{\partial \gamma_R} = -f_l\frac{\gamma_R(\gamma_R^2+\gamma_\varepsilon^2-\sigma_l^2)}{(\gamma_R^2+\gamma_\varepsilon^2)^2}, \tag{C.71}$$

$$\frac{\partial f_h}{\partial \gamma_R} = -f_h\frac{\gamma_R(\gamma_R^2+\gamma_\varepsilon^2-\sigma_h^2)}{(\gamma_R^2+\gamma_\varepsilon^2)^2}, \tag{C.72}$$

$$\frac{df_l}{d\sigma_l} = -f_l\frac{2\widetilde{\sigma}-\sigma_h}{\gamma_R^2+\gamma_\varepsilon^2}. \tag{C.73}$$

Substituting these expressions into (C.68) yields

$$\frac{\partial \widehat{\tau}(x,\gamma_R)}{\partial \gamma_R}_{|x=\sigma_h(\eta^*)} = \varphi\left(2F_l\frac{d\widetilde{\sigma}}{d\gamma_R}+\gamma_R(f_l+f_h)\right)-(1-\varphi)\frac{\partial \frac{k_2-\Delta_0}{k_1}}{\partial \gamma_R}. \tag{C.74}$$

Next, to get an expression for σ_h, note that by Lemma C.4 in equilibrium it must

hold:

$$\begin{aligned}
& \tau(\eta^*) = 0 \\
\Leftrightarrow \quad & \varphi \left(\begin{array}{c} F_l(2\widetilde{\sigma} - \sigma_h - \mu_R) + (\gamma_R^2 + \gamma_\varepsilon^2)(f_l + f_h) \\ +(1 - F_h)(\mu_R - \sigma_h) \end{array} \right) - (1 - \varphi) \left(\sigma_h + \frac{k_2 - \Delta_0}{k_1} \right) = 0 \\
\Leftrightarrow \quad & \sigma_h = \frac{\varphi}{1 - F_h\varphi + F_l\varphi} \left(\begin{array}{c} -\frac{1-\varphi}{\varphi}\frac{k_2 - \Delta_0}{k_1} + (1 - F_h)\mu_R - F_l\mu_R \\ +(\gamma_R^2 + \gamma_\varepsilon^2)(f_l + f_h) + 2F_l\widetilde{\sigma} \end{array} \right). \qquad \text{(C.75)}
\end{aligned}$$

Substituting this together with (C.74) into (C.66), taking the derivatives and simplifying, we finally get

$$\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\gamma_R} = \operatorname{sgn}(\gamma_R - \gamma_S) \frac{(f_h + f_l)\gamma_R\gamma_\varepsilon^2(\gamma_S^2 + \gamma_R^2 + 2\gamma_\varepsilon^2)\varphi}{(\gamma_S^2 + \gamma_\varepsilon^2)(\gamma_R^2 + \gamma_\varepsilon^2)(1 - F_h\varphi + F_l\varphi)}. \qquad \text{(C.76)}$$

■

Lemma C.11 *The probability of non-disclosure expected ex ante by S is strictly larger (smaller) than the probability of non-disclosure expected ex ante by R if $\gamma_S > \gamma_R$ ($\gamma_S < \gamma_R$).*

Proof. In what follows, we assume $\mu_S \geq \mu_R$ without loss of generality. We consider separately two parameter cases: $\gamma_S > \gamma_R$ and $\gamma_S < \gamma_R$.

Case 1: $\gamma_S > \gamma_R$.

Denote the probabilities of non-disclosure in equilibrium from S's and R's ex ante perspectives, respectively, as $P_S^{\varnothing}$ and $P_R^{\varnothing}$. Given that disclosure happens in equilibrium if and only if $\sigma \in [\widetilde{\sigma} - \eta^*, \widetilde{\sigma} + \eta^*]$ by Proposition 1.i(b), we obtain

$$\begin{aligned}
P_S^{\varnothing} &= \Pr_S[d = \varnothing] = \varphi F_S(\widetilde{\sigma} - \eta^*) + \varphi(1 - F_S(\widetilde{\sigma} + \eta^*)) + 1 - \varphi, \\
P_R^{\varnothing} &= \Pr_R[d = \varnothing] = \varphi F_R(\widetilde{\sigma} - \eta^*) + \varphi(1 - F_R(\widetilde{\sigma} + \eta^*)) + 1 - \varphi.
\end{aligned}$$

Thus,

$$P_S^{\varnothing} - P_R^{\varnothing} = \varphi(F_S(\widetilde{\sigma} - \eta^*) - F_R(\widetilde{\sigma} - \eta^*) + F_R(\widetilde{\sigma} + \eta^*) - F_S(\widetilde{\sigma} + \eta^*))$$

Define function

$$\Delta P^{\varnothing}(\mu_S, \gamma_S, \mu_R, \gamma_R, s_1, s_2) = \varphi(F_S(s_1) - F_R(s_1) + F_R(s_2) - F_S(s_2)).$$

Hence, $\Delta P^{\varnothing}(\mu_S, \gamma_S, \mu_R, \gamma_R, \widetilde{\sigma} - \eta^*, \widetilde{\sigma} + \eta^*) = P_S^{\varnothing} - P_R^{\varnothing}$.

Let us show that $\Delta P^{\varnothing}(\mu_S, \gamma_S, \mu_R, \gamma_R, s_1, s_2) > 0$ for any s_1 and s_2 such that

$$s_1 \leq 2\mu_R - s_2 < \mu_R \leq \mu_S < s_2. \tag{C.77}$$

Indeed, for any given s_1 and s_2 satisfying (C.77) we have

$$\begin{aligned} \frac{\partial \Delta P^{\varnothing}(\mu_S, \gamma_S, \mu_R, \gamma_R, s_1, s_2)}{\partial \mu_S} &= \varphi\left(-f_S(s_1) + f_S(s_2)\right) \\ &\geq \varphi\left(-f_S(2\mu_S - s_2) + f_S(s_2)\right) \\ &= 0, \end{aligned} \tag{C.78}$$

where the inequality follows from $s_1 \leq 2\mu_R - s_2 \leq 2\mu_S - s_2 < \mu_S$ by (C.77). Besides, for any given s_1 and s_2 satisfying (C.77) we have

$$\begin{aligned} &\frac{\partial \Delta P^{\varnothing}(\mu_S, \gamma_S, \mu_R, \gamma_R, s_1, s_2)}{\partial \gamma_S} \\ = \; &\varphi\left(f_S(s_1)\frac{\gamma_S(\mu_S - s_1)}{\gamma_S^2 + \gamma_\varepsilon^2} - f_S(s_2)\frac{\gamma_S(\mu_S - s_2)}{\gamma_S^2 + \gamma_\varepsilon^2}\right) > 0, \end{aligned} \tag{C.79}$$

where the inequality is due to $s_1 < \mu_S < s_2$ by (C.77).

Let us now fix some arbitrary priors $\widetilde{\gamma}_S > \widetilde{\gamma}_R$ and $\widetilde{\mu}_S \geq \widetilde{\mu}_R$ and any s_1 and s_2 such that

$$s_1 \leq 2\widetilde{\mu}_R - s_2 < \widetilde{\mu}_R \leq \widetilde{\mu}_S < s_2. \tag{C.80}$$

We then have

$$\begin{aligned} \Delta P^{\varnothing}(\widetilde{\mu}_S, \widetilde{\gamma}_S, \widetilde{\mu}_R, \widetilde{\gamma}_R, s_1, s_2) &> \Delta P^{\varnothing}(\widetilde{\mu}_S, \widetilde{\gamma}_R, \widetilde{\mu}_R, \widetilde{\gamma}_R, s_1, s_2) \\ &\geq \Delta P^{\varnothing}(\widetilde{\mu}_R, \widetilde{\gamma}_R, \widetilde{\mu}_R, \widetilde{\gamma}_R, s_1, s_2) = 0, \end{aligned} \tag{C.81}$$

where the first inequality is by (C.79) and (C.80), and the second inequality is by (C.78) and the fact that condition (C.77) holds for any $\mu_S \in [\widetilde{\mu}_R, \widetilde{\mu}_S]$ by assumption (C.80). Thus, (C.81) implies that for any given $\widetilde{\gamma}_S > \widetilde{\gamma}_R$ and $\widetilde{\mu}_S \geq \widetilde{\mu}_R$ it holds

$$\begin{aligned} \Delta P^{\varnothing}(\widetilde{\mu}_S, \widetilde{\gamma}_S, \widetilde{\mu}_R, \widetilde{\gamma}_R, s_1, s_2) &> 0 \text{ for any} \\ (s_1, s_2) &: \; s_1 \leq 2\widetilde{\mu}_R - s_2 < \widetilde{\mu}_R \leq \widetilde{\mu}_S < s_2. \end{aligned} \tag{C.82}$$

Finally, for any given $\widetilde{\gamma}_S > \widetilde{\gamma}_R$ and $\widetilde{\mu}_S \geq \widetilde{\mu}_R$, it holds that $s_1 = \widetilde{\sigma} - \eta^*$ and $s_2 = \widetilde{\sigma} + \eta^*$ satisfy (C.77) since

$$\widetilde{\sigma} - \eta^* \leq 2\widetilde{\mu}_R - (\widetilde{\sigma} + \eta^*) < \widetilde{\mu}_R \leq \widetilde{\mu}_S < \widetilde{\sigma} + \eta^*,$$

where the first equality is due to $\widetilde{\sigma} - \widetilde{\mu}_R = -\frac{(\widetilde{\gamma}_R^2+\gamma_\varepsilon^2)(\widetilde{\mu}_S-\widetilde{\mu}_R)}{\widetilde{\gamma}_S^2-\widetilde{\gamma}_R^2} \leq 0$ for $\widetilde{\mu}_S \geq \widetilde{\mu}_R$ and $\widetilde{\gamma}_S > \widetilde{\gamma}_R$, and the second and fourth equalities hold by Corollary C.1. Consequently, $P_S^{\varnothing} - P_R^{\varnothing} = \Delta P^{\varnothing}(\widetilde{\mu}_S, \widetilde{\gamma}_S, \widetilde{\mu}_R, \widetilde{\gamma}_R, \widetilde{\sigma} - \eta^*, \widetilde{\sigma} + \eta^*) > 0$ by (C.82).

Case 2: $\gamma_S < \gamma_R$.

Let us show that $P_S^{\varnothing} - P_R^{\varnothing} = \Delta P^{\varnothing}(\mu_S, \gamma_S, \mu_R, \gamma_R, s_1, s_2) < 0$ for any s_1 and s_2 such that

$$s_2 \geq 2\mu_S - s_1 > \mu_S \geq \mu_R > s_1. \tag{C.83}$$

Indeed, for any given s_1 and s_2 satisfying (C.83) we have

$$\begin{aligned} \frac{\partial \Delta P^{\varnothing}(\mu_S, \gamma_S, \mu_R, \gamma_R, s_1, s_2)}{\partial \mu_S} &= \varphi\left(-f_S(s_1) + f_S(s_2)\right) \\ &\leq \varphi\left(-f_S(s_1) + f_S(2\mu_S - s_1)\right) \\ &= 0, \end{aligned} \tag{C.84}$$

where the inequality follows from $s_2 \geq 2\mu_S - s_1 > \mu_S$ by (C.83). Besides, for any given s_1 and s_2 satisfying (C.83) we have

$$\begin{aligned} &\frac{\partial \Delta P^{\varnothing}(\mu_S, \gamma_S, \mu_R, \gamma_R, s_1, s_2)}{\partial \gamma_S} \\ = \ &\varphi\left(f_S(s_1)\frac{\gamma_S(\mu_S - s_1)}{\gamma_S^2 + \gamma_\varepsilon^2} - f_S(s_2)\frac{\gamma_S(\mu_S - s_2)}{\gamma_S^2 + \gamma_\varepsilon^2}\right) > 0, \end{aligned} \tag{C.85}$$

where the inequality is due to $s_1 < \mu_S < s_2$ by (C.83).

Let us now fix some arbitrary priors $\widetilde{\gamma}_S < \widetilde{\gamma}_R$ and $\widetilde{\mu}_S \geq \widetilde{\mu}_R$ and any s_1 and s_2 such that

$$s_2 \geq 2\widetilde{\mu}_S - s_1 > \widetilde{\mu}_S \geq \widetilde{\mu}_R > s_2. \tag{C.86}$$

We then have

$$\begin{aligned} \Delta P^{\varnothing}(\widetilde{\mu}_S, \widetilde{\gamma}_S, \widetilde{\mu}_R, \widetilde{\gamma}_R, s_1, s_2) &< \Delta P^{\varnothing}(\widetilde{\mu}_S, \widetilde{\gamma}_R, \widetilde{\mu}_R, \widetilde{\gamma}_R, s_1, s_2) \\ &\leq \Delta P^{\varnothing}(\widetilde{\mu}_R, \widetilde{\gamma}_R, \widetilde{\mu}_R, \widetilde{\gamma}_R, s_1, s_2) = 0, \end{aligned} \tag{C.87}$$

where the first inequality is by (C.85) and (C.83), and the second inequality is by (C.84) and the fact that condition (C.83) holds for any $\mu_S \in [\widetilde{\mu}_R, \widetilde{\mu}_S]$ by assumption (C.86).

Thus, (C.87) implies that for any given $\widetilde{\gamma}_S < \widetilde{\gamma}_R$ and $\widetilde{\mu}_S \geq \widetilde{\mu}_R$ it holds

$$\begin{aligned} \Delta P^{\varnothing}(\widetilde{\mu}_S, \widetilde{\gamma}_S, \widetilde{\mu}_R, \widetilde{\gamma}_R, s_1, s_2) &< 0 \text{ for any} \\ (s_1, s_2) &: \quad s_2 \geq 2\widetilde{\mu}_S - s_1 > \widetilde{\mu}_S \geq \widetilde{\mu}_R > s_2. \end{aligned} \tag{C.88}$$

Finally, for any given $\widetilde{\gamma}_S < \widetilde{\gamma}_R$ and $\widetilde{\mu}_S \geq \widetilde{\mu}_R$, it holds that $s_1 = \widetilde{\sigma} - \eta^*$ and $s_2 = \widetilde{\sigma} + \eta^*$ satisfy (C.83) since

$$\widetilde{\sigma} + \eta^* \geq 2\widetilde{\mu}_S - (\widetilde{\sigma} - \eta^*) > \widetilde{\mu}_S \geq \widetilde{\mu}_R > \widetilde{\sigma} - \eta^*,$$

where the first equality is due to $\widetilde{\sigma} - \widetilde{\mu}_S = -\frac{(\widetilde{\gamma}_S^2 + \gamma_\varepsilon^2)(\widetilde{\mu}_S - \widetilde{\mu}_R)}{\widetilde{\gamma}_S^2 - \widetilde{\gamma}_R^2} \geq 0$ for $\widetilde{\mu}_S \geq \widetilde{\mu}_R$ and $\widetilde{\gamma}_S < \widetilde{\gamma}_R$, and the second and fourth equalities hold by Corollary C.1. Consequently, $P_S^{\varnothing} - P_R^{\varnothing} = \Delta P^{\varnothing}(\widetilde{\mu}_S, \widetilde{\gamma}_S, \widetilde{\mu}_R, \widetilde{\gamma}_R, \widetilde{\sigma} - \eta^*, \widetilde{\sigma} + \eta^*) < 0$ by (C.88). ■

C.4.2 Proof of Proposition 5

In what follows, we show that for any given R's prior means denoted as μ'_R and μ''_R and fixed R's prior variance, S prefers R with the prior mean closer to μ_S.

Throughout the proof, we assume $\mu_S \geq \max\{\mu'_R, \mu''_R\}$ (the proof for other possible constellations of prior means then follows by symmetry considerations). In this case, the claim is equivalent to showing that R's perceived disagreement expected ex ante by S $E_S[\widetilde{\Delta}_R(d)]$ strictly decreases with μ_R for any $\mu_R \leq \mu_S$ (so that S prefers an R with a prior mean closer to μ_S). In turn, here we separately consider three possible parameter cases: $\gamma_S > \gamma_R$, $\gamma_S < \gamma_R$ and $\gamma_S = \gamma_R$. In what follows, the argument d in $E_S[\widetilde{\Delta}_R(d)]$ is suppressed for notational simplicity.

Case 1: $\gamma_S > \gamma_R$.
As before, denote by η^* the value of η corresponding to the equilibrium disclosure interval $[\widetilde{\sigma} - \eta^*, \widetilde{\sigma} + \eta^*]$. Denote also

$$P_S^{\varnothing} = \Pr_S[d = \varnothing] = \varphi F_S(\widetilde{\sigma} - \eta^*) + \varphi(1 - F_S(\widetilde{\sigma} + \eta^*)) + 1 - \varphi$$

(the probability of non-disclosure from S's ex ante perspective). Then, we have:

$$\begin{aligned} & E_S[\widetilde{\Delta}_R] = E_S[E_R[\Delta(\sigma)|d]] \\ = \; & P_S^{\varnothing}\widetilde{\Delta}_R(\varnothing|\eta^*) + (1-P_S^{\varnothing})\frac{1}{F_S(\widetilde{\sigma}+\eta^*)-F_S(\widetilde{\sigma}-\eta^*)}\int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}+\eta^*}\Delta(\sigma)f_S(\sigma)d\sigma \\ = \; & P_S^{\varnothing}\widetilde{\Delta}_R(\varnothing|\eta^*) + \varphi\begin{pmatrix} \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}} -(k_1\sigma+k_2)f_S(\sigma)d\sigma \\ +\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}(k_1\sigma+k_2)f_S(\sigma)d\sigma \end{pmatrix}, \end{aligned} \tag{C.89}$$

where the last equality is by (C.18).

We need to show that $\frac{dE_S[\widetilde{\Delta}_R]}{d\mu_R} < 0$ for any $\mu_R \leq \mu_S$. Taking the derivative, by standard calculations we obtain

$$\frac{dE_S[\widetilde{\Delta}_R]}{d\mu_R} = P_S^{\varnothing}\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\mu_R} + \varphi\frac{dk_2}{d\mu_R}\left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*} f_S(\sigma)d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}} f_S(\sigma)d\sigma\right). \tag{C.90}$$

Next, by Lemma C.9 we have

$$\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\mu_R} = -\frac{\gamma_S^2(1-\varphi)+\gamma_\varepsilon^2(1-\varphi(F_h+F_l))}{(\gamma_S^2+\gamma_\varepsilon^2)(1-F_h\varphi+F_l\varphi)}. \tag{C.91}$$

Note that $F_h + F_l \leq 1$ since

$$F_h = F_R(\widetilde{\sigma}+\eta^*) \leq F_R(\mu_R+\eta^*) = 1 - F_R(\mu_R-\eta^*) \leq 1 - F_R(\widetilde{\sigma}-\eta^*) = 1 - F_l,$$

where the inequalities are due to $\widetilde{\sigma} \leq \mu_R$ in the considered parameter case by (C.27), and the second equality follows by the symmetry of f_R around μ_R. Consequently, by (C.91)

$$\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\mu_R} < 0. \tag{C.92}$$

Next, consider the following term on the right-hand side of (C.90):

$$\frac{dk_2}{d\mu_R}\left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*} f_S(\sigma)d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}} f_S(\sigma)d\sigma\right) = -\frac{\gamma_\varepsilon^2}{\gamma_R^2+\gamma_\varepsilon^2}\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\left(f_S(\sigma) - f_S(2\widetilde{\sigma}-\sigma)\right)d\sigma \leq 0, \tag{C.93}$$

where the inequality follows from $f_S(\sigma) \geq f_S(2\widetilde{\sigma}-\sigma)$ for any $\sigma \geq \widetilde{\sigma}$ due to $\widetilde{\sigma} \leq \mu_S$ in the current parameter case by (C.26). Combining (C.93) and (C.92) with (C.90) we obtain $\frac{dE_S[\widetilde{\Delta}_R]}{d\mu_R} < 0$.

Case 2: $\gamma_S < \gamma_R$. The proof proceeds analogously to Case 1 and is omitted.

Case 3: $\gamma_S = \gamma_R$.

In this case, (C.18) implies that at any signal σ the disagreement function is equal to the

constant $\frac{\gamma_\varepsilon^2(\mu_S-\mu_R)}{\gamma_R^2+\gamma_\varepsilon^2}$. Moreover, by Proposition 1.ii full disclosure is the only equilibrium (so that R's perceived disagreement conditional on non-disclosure is equal to the prior disagreement $\mu_S - \mu_R$). Hence,

$$E_S[\widetilde{\Delta}_R|\gamma_S = \gamma_R] = (1-\varphi)(\mu_S - \mu_R) + \varphi\frac{\gamma_\varepsilon^2(\mu_S - \mu_R)}{\gamma_R^2 + \gamma_\varepsilon^2} \tag{C.94}$$

and

$$\frac{dE_S[\widetilde{\Delta}_R|\gamma_S = \gamma_R]}{d\mu_R} = -(1-\dot{\phi}) - \varphi\frac{\gamma_\varepsilon^2}{\gamma_R^2 + \gamma_\varepsilon^2} < 0.$$

■

C.4.3 Proof of Proposition 6

Without loss of generality, throughout the proof we assume $\mu_S \geq \mu_R$ and normalize μ_R to 0. In what follows, the argument d in $E_S[\widetilde{\Delta}_R(d)]$ is suppressed for notational simplicity.

Claim 1. *R's perceived disagreement expected ex ante by S $E_S[\widetilde{\Delta}_R]$ strictly decreases in γ_R if $\gamma_R < \gamma_S$.*

Proof: By (C.89),

$$E_S[\widetilde{\Delta}_R] = P_S^{\varnothing}\widetilde{\Delta}_R(\varnothing|\eta^*) + \varphi\begin{pmatrix} \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}} -(k_1\sigma + k_2)f_S(\sigma)d\sigma \\ + \int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}(k_1\sigma + k_2)f_S(\sigma)d\sigma \end{pmatrix}.$$

We have (noting that $\frac{dk_2}{d\gamma_R} = 0$ since μ_R is normalized to 0):

$$\begin{aligned}
&\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} \\
= \;& \frac{dP_S^{\varnothing}}{d\gamma_R}\widetilde{\Delta}_R(\varnothing|\eta^*) + P_S^{\varnothing}\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\gamma_R} \\
& + \varphi\frac{dk_1}{d\gamma_R}\left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma\right) \\
& + \varphi\frac{d\widetilde{\sigma}}{d\gamma_R}\Delta(\widetilde{\sigma})f_S(\widetilde{\sigma}) - \varphi\left(\frac{d\widetilde{\sigma}}{d\gamma_R} - \frac{d\eta^*}{d\gamma_R}\right)\Delta(\widetilde{\sigma}-\eta^*)f_S(\widetilde{\sigma}-\eta^*) \\
& + \varphi\left(\frac{d\widetilde{\sigma}}{d\gamma_R} + \frac{d\eta^*}{d\gamma_R}\right)\Delta(\widetilde{\sigma}+\eta^*)f_S(\widetilde{\sigma}+\eta^*) - \varphi\frac{d\widetilde{\sigma}}{d\gamma_R}\Delta(\widetilde{\sigma})f_S(\widetilde{\sigma}) \\
= \;& \varphi\begin{pmatrix} f_S(\widetilde{\sigma}-\eta^*)\left(\frac{d\widetilde{\sigma}}{d\gamma_R} - \frac{d\eta^*}{d\gamma_R}\right) \\ -f_S(\widetilde{\sigma}+\eta^*)\left(\frac{d\widetilde{\sigma}}{d\gamma_R} + \frac{d\eta^*}{d\gamma_R}\right)\end{pmatrix}\widetilde{\Delta}_R(\varnothing|\eta^*) \\
& + P_S^{\varnothing}\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\gamma_R} \\
& + \varphi\frac{dk_1}{d\gamma_R}\left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma\right) \\
& - \varphi\Delta(\widetilde{\sigma}+\eta^*)\begin{pmatrix} \left(\frac{d\widetilde{\sigma}}{d\gamma_R} - \frac{d\eta^*}{d\gamma_R}\right)f_S(\widetilde{\sigma}-\eta^*) \\ -\left(\frac{d\widetilde{\sigma}}{d\gamma_R} + \frac{d\eta^*}{d\gamma_R}\right)f_S(\widetilde{\sigma}+\eta^*)\end{pmatrix} \\
= \;& \varphi\left(\widetilde{\Delta}_R(\varnothing|\eta^*) - \Delta(\widetilde{\sigma}+\eta^*)\right) \\
& \times \begin{pmatrix} \left(\frac{d\widetilde{\sigma}}{d\gamma_R} - \frac{d\eta^*}{d\gamma_R}\right)f_S(\widetilde{\sigma}-\eta^*) \\ -\left(\frac{d\widetilde{\sigma}}{d\gamma_R} + \frac{d\eta^*}{d\gamma_R}\right)f_S(\widetilde{\sigma}+\eta^*)\end{pmatrix} \\
& + P_S^{\varnothing}\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\gamma_R} + \varphi\frac{dk_1}{d\gamma_R}\left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma\right) \\
= \;& P_S^{\varnothing}\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\gamma_R} + \varphi\frac{dk_1}{d\gamma_R}\left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma\right), \qquad \text{(C.95)}
\end{aligned}$$

where the second equality follows by $\Delta(\widetilde{\sigma}+\eta^*) = \Delta(\widetilde{\sigma}-\eta^*)$, and the last equality follows due to $\widetilde{\Delta}_R(\varnothing|\eta^*) - \Delta(\widetilde{\sigma}+\eta^*) = 0$ by Lemma C.3.

Further, by Lemma C.10 for $\gamma_R < \gamma_S$ it holds

$$\begin{aligned}
\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\gamma_R} &= -\frac{(f_h+f_l)\gamma_R\gamma_\varepsilon^2(\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2)\varphi}{(\gamma_S^2+\gamma_\varepsilon^2)(\gamma_R^2+\gamma_\varepsilon^2)(1-F_h\varphi+F_l\varphi)} \\
&= -\frac{(f_h+f_l)\gamma_R\gamma_\varepsilon^2(\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2)\varphi}{(\gamma_S^2+\gamma_\varepsilon^2)(\gamma_R^2+\gamma_\varepsilon^2)P_R^{\varnothing}}, \qquad \text{(C.96)}
\end{aligned}$$

where $P_R^{\varnothing} = (1 - F_h\varphi + F_l\varphi)$ is the ex ante probability of non-disclosure from R's per-

spective. Substituting it back to (C.95) and taking the derivative $\frac{dk_1}{d\gamma_R}$ we obtain

$$\begin{aligned}\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} &= -\frac{P_S^{\varnothing}}{P_R^{\varnothing}}\frac{(f_h+f_l)\gamma_R\gamma_\varepsilon^2(\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2)\varphi}{(\gamma_S^2+\gamma_\varepsilon^2)(\gamma_R^2+\gamma_\varepsilon^2)} \\ &\quad -\varphi\frac{2\gamma_R\gamma_\varepsilon^2}{(\gamma_R^2+\gamma_\varepsilon^2)^2}\left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma\right). \qquad \text{(C.97)}\end{aligned}$$

The subsequent analysis distinguishes two cases: $\widetilde{\sigma}+\eta^* \geq -\widetilde{\sigma}$ and $\widetilde{\sigma}+\eta^* < -\widetilde{\sigma}$.

Case 1: $\widetilde{\sigma}+\eta^* \geq -\widetilde{\sigma}$.

Consider the second term in (C.97). We have:

$$\begin{aligned}&\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma \\ &= \int_{\widetilde{\sigma}}^{-\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma + \left(\int_{-\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma\right). \qquad \text{(C.98)}\end{aligned}$$

Since, $\widetilde{\sigma}+\eta^* \geq -\widetilde{\sigma}$ by assumption and $\widetilde{\sigma} \leq 0 \leq -\widetilde{\sigma}$ in the considered parameter case, the term in brackets in (C.98) is strictly positive:

$$\int_{-\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma > 0. \qquad \text{(C.99)}$$

Next, using integration by substitution we obtain:

$$\begin{aligned}\int_{\widetilde{\sigma}}^{-\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma &= \int_{\widetilde{\sigma}}^{0}\sigma f_S(\sigma)d\sigma + \int_0^{-\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma \\ &= \int_0^{-\widetilde{\sigma}}(-x)f_S(-x)dx + \int_0^{-\widetilde{\sigma}}x f_S(x)dx \\ &= \int_0^{-\widetilde{\sigma}}x(f_S(x)-f_S(-x))dx \geq 0, \qquad \text{(C.100)}\end{aligned}$$

where the inequality follows from the fact that $\mu_S \geq \mu_R = 0$ so that $f_S(x) - f_S(-x) \geq 0$ for any $x \geq 0$ (recall that $0 \leq -\widetilde{\sigma}$). (C.100) and (C.99) finally imply that the whole right-hand side of (C.98) is strictly positive, which together with (C.97) yields $\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} < 0$ for $\gamma_R < \gamma_S$. Thus, Claim 1 is proven for the case $\widetilde{\sigma}+\eta^* \geq -\widetilde{\sigma}$.

Case 2: $\widetilde{\sigma}+\eta^* < -\widetilde{\sigma}$.

The proof of Claim 1 for this case proceeds in three steps.

Step 1. First, let us decompose $\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R}$ in two components, denoted further as A and B, as follows. From (C.97) we have

$$\begin{aligned}\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} &= -\frac{P_S^{\varnothing}}{P_R^{\varnothing}}\frac{(f_h+f_l)\gamma_R\gamma_\varepsilon^2(\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2)\varphi}{(\gamma_S^2+\gamma_\varepsilon^2)(\gamma_R^2+\gamma_\varepsilon^2)} \\ &\quad -\varphi\frac{2\gamma_R\gamma_\varepsilon^2}{(\gamma_R^2+\gamma_\varepsilon^2)^2}\left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma\right). \quad \text{(C.101)}\end{aligned}$$

Note that

$$\begin{aligned}f_h &= \frac{1-F_h}{\gamma_R^2+\gamma_\varepsilon^2}(\gamma_R^2+\gamma_\varepsilon^2)\frac{f_h}{1-F_h} \\ &= \frac{1-F_h}{\gamma_R^2+\gamma_\varepsilon^2}E_R[\sigma|\sigma\geq\widetilde{\sigma}+\eta^*] \\ &= \frac{1}{\gamma_R^2+\gamma_\varepsilon^2}\int_{\widetilde{\sigma}+\eta^*}^{\infty}\sigma f_R(\sigma)d\sigma \\ &= \frac{1}{\gamma_R^2+\gamma_\varepsilon^2}\left(\int_{-\infty}^{\infty}\sigma f_R(\sigma)d\sigma - \int_{-\infty}^{\widetilde{\sigma}+\eta^*}\sigma f_R(\sigma)d\sigma\right) \\ &= -\frac{1}{\gamma_R^2+\gamma_\varepsilon^2}\int_{-\infty}^{\widetilde{\sigma}+\eta^*}\sigma f_R(\sigma)d\sigma,\end{aligned}$$

where the second equality is by (C.12), and the last equality follows from $\int_{-\infty}^{\infty}\sigma f_R(\sigma)d\sigma = \mu_R = 0$. After substituting this into (C.101) and rearranging we obtain:

$$\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} = \varphi\frac{\gamma_R\gamma_\varepsilon^2}{(\gamma_R^2+\gamma_\varepsilon^2)^2}\begin{pmatrix} -\frac{P_S^{\varnothing}}{P_R^{\varnothing}}\frac{f_l(\gamma_R^2+\gamma_\varepsilon^2)(\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2)}{\gamma_S^2+\gamma_\varepsilon^2} \\ +\frac{P_S^{\varnothing}}{P_R^{\varnothing}}\frac{\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2}{\gamma_S^2+\gamma_\varepsilon^2}\int_{-\infty}^{\widetilde{\sigma}+\eta^*}\sigma f_R(\sigma)d\sigma \\ -2\left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma\right)\end{pmatrix}.$$

In turn, the right-hand side can now be decomposed as follows:

$$\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} = \varphi\frac{\gamma_R\gamma_\varepsilon^2}{(\gamma_R^2+\gamma_\varepsilon^2)^2}(A+B) \quad \text{(C.102)}$$

where

$$
\begin{aligned}
A &= -\frac{P_S^{\varnothing}}{P_R^{\varnothing}}\frac{f_l(\gamma_R^2+\gamma_\varepsilon^2)(\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2)}{\gamma_S^2+\gamma_\varepsilon^2} \\
&\quad +\frac{P_S^{\varnothing}}{P_R^{\varnothing}}\frac{\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2}{\gamma_S^2+\gamma_\varepsilon^2}\int_{-\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}+\eta^*}\sigma f_R(\sigma)d\sigma \\
&\quad -2\int_{-\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma+2\int_{\widetilde{\sigma}-\eta^*}^{3\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma, \qquad \text{(C.103)} \\
B &= \frac{P_S^{\varnothing}}{P_R^{\varnothing}}\frac{\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2}{\gamma_S^2+\gamma_\varepsilon^2}\int_{-\infty}^{-\widetilde{\sigma}-\eta^*}\sigma f_R(\sigma)d\sigma \\
&\quad -2\int_{\widetilde{\sigma}}^{-\widetilde{\sigma}-\eta^*}\sigma f_S(\sigma)d\sigma+2\int_{3\widetilde{\sigma}+\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma \qquad \text{(C.104)}
\end{aligned}
$$

In Steps 2 and 3, we show that both A and B are strictly negative, respectively.

Step 2. Let us show that $A<0$. The first term is clearly strictly negative. The second term is 0, since $\int_{-\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}+\eta^*}\sigma f_R(\sigma)d\sigma=0$ (as $f_R(\sigma)$ is symmetric around $\mu_R=0$).

Consider the third term of A. Using integration by substitution we obtain

$$
\begin{aligned}
\int_{-\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma &= \int_{-\widetilde{\sigma}-\eta^*}^{0}\sigma f_S(\sigma)d\sigma+\int_0^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma \\
&= \int_0^{\widetilde{\sigma}+\eta^*}(-x)f_S(-x)dx+\int_0^{\widetilde{\sigma}+\eta^*}xf_S(x)dx \\
&= \int_0^{\widetilde{\sigma}+\eta^*}x(f_S(x)-f_S(-x))dx\geq 0, \qquad \text{(C.105)}
\end{aligned}
$$

where the inequality follows from the fact that $\mu_S\geq\mu_R=0$ so that $f_S(x)-f_S(-x)\geq 0$ for any $x\geq 0$. This implies that the third term of A, $-2\int_{-\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma$, is negative.

Finally, consider the last term of A, $2\int_{\widetilde{\sigma}-\eta^*}^{3\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma$. We have

$$
\widetilde{\sigma}-\eta^*<3\widetilde{\sigma}+\eta^*<0, \qquad \text{(C.106)}
$$

where the first inequality is by $3\widetilde{\sigma}+\eta^*-(\widetilde{\sigma}-\eta^*)=2(\widetilde{\sigma}+\eta^*)>0$, which in turn holds since $\widetilde{\sigma}+\eta^*>\mu_R=0$ by Corollary C.1 The second inequality is by $3\widetilde{\sigma}+\eta^*=\widetilde{\sigma}+(2\widetilde{\sigma}+\eta^*)<0$, which in turn holds since $\widetilde{\sigma}=\frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2}\leq 0$ (as $\gamma_R<\gamma_S$ by assumption) while $2\widetilde{\sigma}+\eta^*<0\Leftrightarrow\widetilde{\sigma}+\eta^*<-\widetilde{\sigma}$ by assumption of Case 2. (C.106) immediately implies that the last term of A, $2\int_{\widetilde{\sigma}-\eta^*}^{3\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma$, is strictly negative.

Thus, all terms of A are negative or 0 (with at least the first and the last terms strictly negative), which implies $A<0$.

Step 3. Let us show that B defined in (C.104) is strictly negative.

Recall

$$\begin{aligned} B &= \frac{P_S^{\varnothing}}{P_R^{\varnothing}} \frac{\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2}{\gamma_S^2+\gamma_\varepsilon^2} \int_{-\infty}^{-\widetilde{\sigma}-\eta^*} \sigma f_R(\sigma) d\sigma \\ &\quad -2\int_{\widetilde{\sigma}}^{-\widetilde{\sigma}-\eta^*} \sigma f_S(\sigma) d\sigma + 2\int_{3\widetilde{\sigma}+\eta^*}^{\widetilde{\sigma}} \sigma f_S(\sigma) d\sigma. \end{aligned}$$

Consider the first term. Note that $\frac{P_S^{\varnothing}}{P_R^{\varnothing}} > 1$ by Lemma C.11, while $\int_{-\infty}^{-\widetilde{\sigma}-\eta^*} \sigma f_R(\sigma) d\sigma < 0$ due to $-\widetilde{\sigma}-\eta^* = -(\widetilde{\sigma}+\eta^*) < \mu_R = 0$ by Lemma C.1. This implies

$$\begin{aligned} B &< \frac{\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2}{\gamma_S^2+\gamma_\varepsilon^2} \int_{-\infty}^{-\widetilde{\sigma}-\eta^*} \sigma f_R(\sigma) d\sigma \\ &\quad -2\int_{\widetilde{\sigma}}^{-\widetilde{\sigma}-\eta^*} \sigma f_S(\sigma) d\sigma + 2\int_{3\widetilde{\sigma}+\eta^*}^{\widetilde{\sigma}} \sigma f_S(\sigma) d\sigma. \end{aligned} \tag{C.107}$$

Next, define function

$$\begin{aligned} \Upsilon(z) &= \frac{\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2}{\gamma_S^2+\gamma_\varepsilon^2} \int_{-\infty}^{z} \sigma f_R(\sigma) d\sigma \\ &\quad -2\int_{\widetilde{\sigma}}^{z} \sigma f_S(\sigma) d\sigma + 2\int_{2\widetilde{\sigma}-z}^{\widetilde{\sigma}} \sigma f_S(\sigma) d\sigma. \end{aligned} \tag{C.108}$$

Then, (C.107) can be rewritten as

$$B < \Upsilon(-\widetilde{\sigma}-\eta^*). \tag{C.109}$$

Below, we show that $\Upsilon(z) < 0$ for any $z \in (\widetilde{\sigma}, 0)$ that will be sufficient to show that $B < 0$ in the considered case ($\widetilde{\sigma}+\eta^* < -\widetilde{\sigma} \Leftrightarrow -\widetilde{\sigma}-\eta^* > \widetilde{\sigma}$).

Rearranging, we obtain

$$\begin{aligned}
\Upsilon(z) &= \frac{\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2}{\gamma_S^2+\gamma_\varepsilon^2}\int_{-\infty}^{z}\sigma f_R(\sigma)d\sigma - 2\int_{\widetilde{\sigma}}^{z}\sigma f_S(\sigma)d\sigma + 2\int_{2\widetilde{\sigma}-z}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma \\
&= \frac{\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2}{\gamma_S^2+\gamma_\varepsilon^2}\int_{-\infty}^{z}\sigma f_R(\sigma)d\sigma - 2\left(\int_{-\infty}^{z}\sigma f_S(\sigma)d\sigma - \int_{-\infty}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma\right) \\
&\quad +2\left(\int_{-\infty}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma - \int_{-\infty}^{2\widetilde{\sigma}-z}\sigma f_S(\sigma)d\sigma\right) \\
&= \frac{\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2}{\gamma_S^2+\gamma_\varepsilon^2}\left(F_R(z)E_R[\sigma|\sigma\le z]\right) \\
&\quad -2\left(F_S(z)E_S[\sigma|\sigma\le z] - F_S(\widetilde{\sigma})E_S[\sigma|\sigma\le\widetilde{\sigma}]\right) \\
&\quad +2\left(F_S(\widetilde{\sigma})E_S[\sigma|\sigma\le\widetilde{\sigma}] - F_S(2\widetilde{\sigma}-z)E_S[\sigma|\sigma\le 2\widetilde{\sigma}-z]\right) \\
&= \frac{\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2}{\gamma_S^2+\gamma_\varepsilon^2}\left(-(\gamma_R^2+\gamma_\varepsilon^2)f_R(z)\right) \\
&\quad -2\left(\mu_S F_S(z) - (\gamma_S^2+\gamma_\varepsilon^2)f_S(z) - \mu_S F_S(\widetilde{\sigma}) + (\gamma_S^2+\gamma_\varepsilon^2)f_S(\widetilde{\sigma})\right) \\
&\quad +2\left(\mu_S F_S(\widetilde{\sigma}) - (\gamma_S^2+\gamma_\varepsilon^2)f_S(\widetilde{\sigma}) - \mu_S F_S(2\widetilde{\sigma}-z) + (\gamma_S^2+\gamma_\varepsilon^2)f_S(2\widetilde{\sigma}-z)\right),
\end{aligned}$$

where the last equality follows from Lemma C.1. Substituting explicit expressions for F_R, F_S, f_R and f_S and simplifying we obtain

$$\begin{aligned}
\Upsilon(z) &= \frac{1}{\sqrt{2\pi}\widehat{\gamma}_S^2}\begin{pmatrix} 2\widehat{\gamma}_S^3 e^{-\frac{(z-\mu_S)^2}{2\widehat{\gamma}_S^2}} - 4\widehat{\gamma}_S^3 e^{-\frac{\widehat{\gamma}_S^2\mu_S^2}{2(\widehat{\gamma}_S^2-\widehat{\gamma}_R^2)^2}} \\ +2\widehat{\gamma}_S^3 e^{-\frac{\left(z+\frac{(\widehat{\gamma}_S^2+\widehat{\gamma}_R^2)\mu_S}{\widehat{\gamma}_S^2-\widehat{\gamma}_R^2}\right)^2}{2\widehat{\gamma}_S^2}} - e^{-\frac{z^2}{2\widehat{\gamma}_R^2}}\widehat{\gamma}_R\left(\widehat{\gamma}_S^2+\widehat{\gamma}_R^2\right) \end{pmatrix} \\
&\quad -\mu_S\begin{pmatrix} -2\,\mathrm{Erfc}\left[\frac{\widehat{\gamma}_S\mu_S}{\sqrt{2}(\widehat{\gamma}_S^2-\widehat{\gamma}_R^2)}\right] + \mathrm{Erfc}\left[\frac{-z+\mu_S}{\sqrt{2}\widehat{\gamma}_S}\right] \\ +\mathrm{Erfc}\left[\frac{z+\frac{(\widehat{\gamma}_S^2+\widehat{\gamma}_R^2)\mu_S}{\widehat{\gamma}_S^2-\widehat{\gamma}_R^2}}{\sqrt{2}\widehat{\gamma}_S}\right] \end{pmatrix},
\end{aligned} \tag{C.110}$$

where $\widehat{\gamma}_S = \sqrt{\gamma_S^2+\gamma_\varepsilon^2}$ and $\widehat{\gamma}_R = \sqrt{\gamma_R^2+\gamma_\varepsilon^2}$. Taking the derivative with respect to z and rearranging we obtain

$$\frac{\partial\Upsilon(z)}{\partial z} = \frac{1}{\sqrt{2\pi}\widehat{\gamma}_S^2}\left(\frac{2e^{-\frac{\left(z+\frac{(\widehat{\gamma}_S^2+\widehat{\gamma}_R^2)\mu_S}{\widehat{\gamma}_S^2-\widehat{\gamma}_R^2}\right)^2}{2\widehat{\gamma}_S^2}}\widehat{\gamma}_S}{\widehat{\gamma}_S^2-\widehat{\gamma}_R^2}\lambda_1(z) + ze^{-\frac{(z-\mu_S)^2}{2\widehat{\gamma}_S^2}}\lambda_2(z)\right), \tag{C.111}$$

where

$$\begin{aligned}\lambda_1(z) &= z(\widehat{\gamma}_R^2-\widehat{\gamma}_S^2)-2\widehat{\gamma}_R^2\mu_S,\\ \lambda_2(z) &= \frac{e^{\frac{1}{2}\left(-\frac{z^2}{\widehat{\gamma}_R^2}+\frac{(z-\mu_S)^2}{\widehat{\gamma}_S^2}\right)}(\widehat{\gamma}_S^2+\widehat{\gamma}_R^2)}{\widehat{\gamma}_R}-2\widehat{\gamma}_S.\end{aligned}$$

Let us show that $\lambda_1(z)<0$ and $\lambda_2(z)>0$ for $z\in(\widetilde{\sigma},0)$. Consider $\lambda_1(z)$. Since $\gamma_S>\gamma_R$ by assumption, it is decreasing in z. Hence, $\lambda_1(z)$ reaches its maximum on $(\widetilde{\sigma},0)$ at $z=\widetilde{\sigma}$, i.e.,

$$\begin{aligned}\forall z &\in (\widetilde{\sigma},0):\lambda_1(z)<\lambda_1(\widetilde{\sigma})=\frac{\mu_S\widehat{\gamma}_R^2}{\widehat{\gamma}_R^2-\widehat{\gamma}_S^2}(\widehat{\gamma}_R^2-\widehat{\gamma}_S^2)-2\widehat{\gamma}_R^2\mu_S\\ &= -\widehat{\gamma}_R^2\mu_S\le-\widehat{\gamma}_R^2\mu_R=0.\end{aligned}\tag{C.112}$$

It follows that $\lambda_1(z)<0$ for any $z\in(\widetilde{\sigma},0)$.

Consider $\lambda_2(z)$. We have

$$\frac{\partial\lambda_2(z)}{\partial z}=\frac{e^{\frac{1}{2}\left(-\frac{z^2}{\widehat{\gamma}_R^2}+\frac{(z-\mu_S)^2}{\widehat{\gamma}_S^2}\right)}(\widehat{\gamma}_S^2+\widehat{\gamma}_R^2)(z(\widehat{\gamma}_R^2-\widehat{\gamma}_S^2)-2\widehat{\gamma}_R^2\mu_S)}{\widehat{\gamma}_S^2\widehat{\gamma}_R^3}<0,$$

since $z(\widehat{\gamma}_R^2-\widehat{\gamma}_S^2)-2\widehat{\gamma}_R^2\mu_S=\lambda_1(z)<0$ for any $z\in(\widetilde{\sigma},0)$ as shown above. Since $\lambda_2(z)$ is decreasing in z, it reaches its minimum on $(\widetilde{\sigma},0)$ at $z=0$, i.e.,

$$\begin{aligned}\forall z &\in (\widetilde{\sigma},0):\lambda_2(z)>\lambda_2(0)=\frac{e^{\frac{1}{2}\frac{\mu_S^2}{\widehat{\gamma}_S^2}}(\widehat{\gamma}_S^2+\widehat{\gamma}_R^2)}{\widehat{\gamma}_R}-2\widehat{\gamma}_S\\ &\ge \frac{\widehat{\gamma}_S^2+\widehat{\gamma}_R^2}{\widehat{\gamma}_R}-2\widehat{\gamma}_S=\frac{(\widehat{\gamma}_S-\widehat{\gamma}_R)^2}{\widehat{\gamma}_R}>0.\end{aligned}$$

It follows that $\lambda_2(z)>0$ for $z\in(\widetilde{\sigma},0)$. Consequently, the term $ze^{-\frac{(z-\mu_S)^2}{2\widehat{\gamma}_S^2}}\lambda_2(z)<0$ for $z\in(\widetilde{\sigma},0)$. This together with $\lambda_1(z)<0$ by (C.112) implies by (C.111) that $\frac{\partial\Upsilon(z)}{\partial z}<0$. Hence, $\Upsilon(z)$ reaches its maximum on $z\in(\widetilde{\sigma},0)$ at $z=\widetilde{\sigma}$, i.e.,

$$\forall z\in(\widetilde{\sigma},0):\Upsilon(z)<\Upsilon(\widetilde{\sigma}).\tag{C.113}$$

Substituting $\widetilde{\sigma}=\frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2}=\frac{\mu_S\widehat{\gamma}_R^2}{\widehat{\gamma}_R^2-\widehat{\gamma}_S^2}$ into (C.110) and simplifying we obtain

$$\Upsilon(\widetilde{\sigma})=-\frac{e^{-\frac{\widehat{\gamma}_R^2\mu_S^2}{2\left(\widehat{\gamma}_S^2-\widehat{\gamma}_R^2\right)^2}}\widehat{\gamma}_R\left(\widehat{\gamma}_S^2+\widehat{\gamma}_R^2\right)}{\sqrt{2\pi}\widehat{\gamma}_S^2}<0.\tag{C.114}$$

This together with (C.113) implies

$$\forall z \in (\widetilde{\sigma}, 0) : \Upsilon(z) < 0. \tag{C.115}$$

Now, recall that we consider the case $\widetilde{\sigma}+\eta^* < -\widetilde{\sigma}$, which together with $\widetilde{\sigma}+\eta^* > \mu_R = 0$ (by Corollary C.1) implies $-\widetilde{\sigma}-\eta^* \in (\widetilde{\sigma}, 0)$. This together with (C.115) yields $\Upsilon(-\widetilde{\sigma}-\eta^*) < 0$. This together with (C.109) finally leads to $B < 0$.

Overall, we have shown that both A and B from the right-hand side of (C.102) are strictly negative. Consequently, in the considered case of $\widetilde{\sigma} + \eta^* < -\widetilde{\sigma}$, we obtain $\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} < 0$. As this has also been shown for the other possible case $\widetilde{\sigma} + \eta^* \geq -\widetilde{\sigma}$, the proof of Claim 1 is complete.

Claim 2. *R's perceived disagreement expected ex ante by S* $E_S[\widetilde{\Delta}_R]$ *is continuous in* γ_R *at any* $\gamma_R \in (0, \infty)$.

Proof. We show the continuity of $E_S[\widetilde{\Delta}_R]$ at a given value of γ_R in two separate cases: $\gamma_R \neq \gamma_S$ and $\gamma_R = \gamma_S$.

Case 1. $\gamma_R \neq \gamma_S$.

Recall that by (C.89) for $\gamma_R \neq \gamma_S$ we have

$$\begin{aligned} E_S[\widetilde{\Delta}_R] &= (\varphi F_S(\sigma_l) + \varphi(1 - F_S(\sigma_h)) + 1 - \varphi)\, \widetilde{\Delta}_R(\varnothing|\eta^*) \\ &\quad + \varphi \begin{pmatrix} \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}} -(k_1\sigma + k_2) f_S(\sigma) d\sigma \\ + \int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*} (k_1\sigma + k_2) f_S(\sigma) d\sigma \end{pmatrix}. \end{aligned} \tag{C.116}$$

Note that if $\gamma_R \neq \gamma_S$, then $\widetilde{\Delta}_R(\varnothing|\eta^*)$ is differentiable and hence continuous in γ_R by Lemma C.10. $\widetilde{\sigma}$, k_1 and k_2 are also continuous in γ_R as far as $\gamma_R \neq \gamma_S$ by (7), (C.19) and (C.20), respectively. Finally, since in equilibrium $\widetilde{\Delta}_R(\varnothing|\eta^*) = \Delta(\sigma_h) = k_1\sigma_h + k_2$ by Lemma C.3 and (C.18), we have

$$\sigma_h = \frac{\widetilde{\Delta}_R(\varnothing|\eta^*) - k_2}{k_1}. \tag{C.117}$$

Since $k_1 \neq 0$ and $\widetilde{\Delta}_R(\varnothing|\eta^*)$, k_1 and k_2 are all continuous in γ_R as far as $\gamma_R \neq \gamma_S$ as mentioned above, it follows from (C.117) that both σ_h and $\sigma_l = 2\widetilde{\sigma} - \sigma_h$ are also continuous in γ_R if $\gamma_R \neq \gamma_S$.

Thus, $\widetilde{\Delta}_R(\varnothing|\eta^*)$, $\widetilde{\sigma}$, k_1, k_2, σ_l and σ_h are all continuous in γ_R at any $\gamma_R \neq \gamma_S$. Consequently, the same is true for the right-hand side of (C.116) and thus $E_S[\widetilde{\Delta}_R]$.

Case 2. $\gamma_R = \gamma_S$.

We need to show that $E_S[\widetilde{\Delta}_R]$ is continuous in γ_R at $\gamma_R = \gamma_S$, i.e.,

$$\lim_{\gamma_R \to \gamma_S^-} E_S[\widetilde{\Delta}_R|\gamma_R] = \lim_{\gamma_R \to \gamma_S^+} E_S[\widetilde{\Delta}_R|\gamma_R] = E_S[\widetilde{\Delta}_R|\gamma_R = \gamma_S].$$

The proof proceeds by two steps corresponding to the cases $\mu_S > \mu_R$ and $\mu_S = \mu_R$ (recall that assumption $\mu_S \geq \mu_R$ is without loss of generality).

Step 1. In this step, we prove Claim 2 for the case $\mu_S > \mu_R$.

Claim 2.1: *If* $\mu_S > \mu_R$, *then* $\lim_{\gamma_R \to \gamma_S^\pm} \sigma_l = -\infty$, $\lim_{\gamma_R \to \gamma_S^\pm} \sigma_h = \infty$, $\lim_{\gamma_R \to \gamma_S^-} \widetilde{\sigma} = -\infty$, $\lim_{\gamma_R \to \gamma_S^+} \widetilde{\sigma} = \infty$, $\lim_{\gamma_R \to \gamma_S^\pm} \eta^* = \infty$.

Proof. By definition of the upper status quo signal $\overline{\sigma}$ and (C.18), we have (recall that μ_R is normalized to 0):

$$\begin{aligned} \Delta(\overline{\sigma}) &= \Delta_0 \Leftrightarrow \\ k_1\overline{\sigma} + k_2 &= \Delta_0 \Leftrightarrow \\ \overline{\sigma} &= \frac{\Delta_0 - k_2}{k_1} = \begin{cases} \frac{\gamma_S^2\left(\gamma_R^2+\gamma_\varepsilon^2\right)\mu_S}{\left(\gamma_S^2-\gamma_R^2\right)\gamma_\varepsilon^2} \text{ if } \gamma_S > \gamma_R \\ \frac{\left(\gamma_S^2+2\gamma_\varepsilon^2\right)\left(\gamma_R^2+\gamma_\varepsilon^2\right)\mu_S}{\left(\gamma_R^2-\gamma_S^2\right)\gamma_\varepsilon^2} \text{ if } \gamma_S < \gamma_R \end{cases}, \end{aligned} \tag{C.118}$$

so that

$$\lim_{\gamma_R \to \gamma_S^\pm} \overline{\sigma} = \infty \tag{C.119}$$

as $\mu_S > \mu_R = 0$ by assumption.

Analogously, for the lower status quo signal we have

$$\begin{aligned} \Delta(\underline{\sigma}) &= \Delta_0 \Leftrightarrow \\ -k_1\underline{\sigma} - k_2 &= \Delta_0 \Leftrightarrow \\ \underline{\sigma} &= \frac{\Delta_0 + k_2}{-k_1} = \begin{cases} \frac{\left(\gamma_S^2+2\gamma_\varepsilon^2\right)\left(\gamma_R^2+\gamma_\varepsilon^2\right)\mu_S}{\left(\gamma_R^2-\gamma_S^2\right)\gamma_\varepsilon^2} \text{ if } \gamma_S > \gamma_R \\ \frac{\gamma_S^2\left(\gamma_R^2+\gamma_\varepsilon^2\right)\mu_S}{\left(\gamma_S^2-\gamma_R^2\right)\gamma_\varepsilon^2} \text{ if } \gamma_S < \gamma_R \end{cases}, \end{aligned} \tag{C.120}$$

so that

$$\lim_{\gamma_R \to \gamma_S^\pm} \underline{\sigma} = -\infty \tag{C.121}$$

as $\mu_S > \mu_R = 0$ by assumption. Further, by (7)

$$\lim_{\gamma_R \to \gamma_S^-} \widetilde{\sigma} = \lim_{\gamma_R \to \gamma_S^-} \frac{\mu_S(\gamma_R^2 + \gamma_\varepsilon^2)}{\gamma_R^2 - \gamma_S^2} = -\infty, \tag{C.122}$$

$$\lim_{\gamma_R \to \gamma_S^+} \widetilde{\sigma} = \lim_{\gamma_R \to \gamma_S^+} \frac{\mu_S(\gamma_R^2 + \gamma_\varepsilon^2)}{\gamma_R^2 - \gamma_S^2} = \infty. \tag{C.123}$$

Since the status quo signals lie within the equilibrium disclosure interval $[\sigma_l, \sigma_h]$ by Proposition 1.i(b), (C.119) and (C.121) yield $\lim_{\gamma_R \to \gamma_S^\pm} \sigma_h = \infty$ and $\lim_{\gamma_R \to \gamma_S^\pm} \sigma_l = -\infty$. At the same time, since $\widetilde{\sigma}$ converges to either positive or negative infinity as shown above, it must hold that $\lim_{\gamma_R \to \gamma_S^\pm} \eta^* = \infty$ (as otherwise either $\sigma_l = \widetilde{\sigma} - \eta^* \to -\infty$ or $\sigma_h = \widetilde{\sigma} + \eta^* \to \infty$ would yield a contradiction).

Claim 2.2: If $\mu_S > \mu_R$, then for $i = S, R$: $\lim_{\gamma_R \to \gamma_S^\pm} f_i(\widetilde{\sigma} - \eta^*) = \lim_{\gamma_R \to \gamma_S^\pm} f_i(\widetilde{\sigma} + \eta^*) = \lim_{\gamma_R \to \gamma_S^\pm} F_i(\widetilde{\sigma} - \eta^*) = \lim_{\gamma_R \to \gamma_S^\pm}(1 - F_i(\widetilde{\sigma} + \eta^*)) = 0, \lim_{\gamma_R \to \gamma_S^-} F_i(\widetilde{\sigma}) = 0, \lim_{\gamma_R \to \gamma_S^+} F_i(\widetilde{\sigma}) = 1$.

Proof. Follows immediately from Claim 2.1.

Claim 2.3: If $\mu_S > \mu_R$, then $\lim_{\gamma_R \to \gamma_S^\pm} \widetilde{\Delta}_R(\varnothing|\eta^*) = \Delta_0$.

Proof. By (C.21) and Lemma C.1 (recall that μ_R is normalized to 0),

$$\begin{aligned} & \widetilde{\Delta}_R(\varnothing|\eta) \\ = & \frac{1}{\varphi(F_l + 1 - F_h) + 1 - \varphi} \\ & \times \begin{pmatrix} \varphi k_1(\gamma_R^2 + \gamma_\varepsilon^2) f_l - \varphi F_l k_2 + \varphi k_1(\gamma_R^2 + \gamma_\varepsilon^2) f_h \\ +\varphi(1 - F_h)k_2 + (1 - \varphi)\Delta_0 \end{pmatrix}. \end{aligned} \tag{C.124}$$

Then, by Claim 2.2 we obtain (given that $\lim_{\gamma_R \to \gamma_S^\pm} k_1 = 0$ and $\lim_{\gamma_R \to \gamma_S^\pm} k_2 = \text{sgn}\,(\gamma_S - \gamma_R)\frac{\gamma_\varepsilon^2}{\gamma_S^2 + \gamma_\varepsilon^2}\Delta_0$, i.e., these limits are bounded):

$$\lim_{\gamma_R \to \gamma_S^\pm} \widetilde{\Delta}_R(\varnothing|\eta^*) = \Delta_0.$$

Claim 2.4. If $\mu_S > \mu_R$, then $\lim_{\gamma_R \to \gamma_S^\pm} E_S[\widetilde{\Delta}_R|\gamma_R] = E_S[\widetilde{\Delta}_R|\gamma_R = \gamma_S]$.

Proof. By (C.89), for $\gamma_R \neq \gamma_S$ we have

$$\begin{aligned}
E_S[\widetilde{\Delta}_R] &= P_S^{\varnothing}\widetilde{\Delta}_R(\varnothing|\eta^*) + \varphi \begin{pmatrix} \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}} -(k_1\sigma + k_2) f_S(\sigma)d\sigma \\ + \int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*} (k_1\sigma + k_2) f_S(\sigma)d\sigma \end{pmatrix} \\
&= P_S^{\varnothing}\widetilde{\Delta}_R(\varnothing|\eta^*) \\
&\quad +\varphi k_1 \begin{pmatrix} \int_{-\infty}^{\widetilde{\sigma}+\eta^*} \sigma f_S(\sigma)d\sigma - \int_{-\infty}^{\widetilde{\sigma}} \sigma f_S(\sigma)d\sigma \\ -\left(\int_{-\infty}^{\widetilde{\sigma}} \sigma f_S(\sigma)d\sigma - \int_{-\infty}^{\widetilde{\sigma}-\eta^*} \sigma f_S(\sigma)d\sigma\right) \end{pmatrix} \\
&\quad +\varphi k_2 \begin{pmatrix} \int_{-\infty}^{\widetilde{\sigma}+\eta^*} f_S(\sigma)d\sigma - \int_{-\infty}^{\widetilde{\sigma}} f_S(\sigma)d\sigma \\ -\left(\int_{-\infty}^{\widetilde{\sigma}} f_S(\sigma)d\sigma - \int_{-\infty}^{\widetilde{\sigma}-\eta^*} f_S(\sigma)d\sigma\right) \end{pmatrix} \\
&= (\varphi F_S(\widetilde{\sigma}-\eta^*) + \varphi(1 - F_S(\widetilde{\sigma}+\eta^*) + 1 - \varphi)\widetilde{\Delta}_R(\varnothing|\eta^*) \\
&\quad +\varphi k_1 \begin{pmatrix} F_S(\widetilde{\sigma}+\eta^*)E_S[\sigma|\sigma \leq \widetilde{\sigma}+\eta^*] - 2F_S(\widetilde{\sigma})E_S[\sigma|\sigma \leq \widetilde{\sigma}] \\ +F_S(\widetilde{\sigma}-\eta^*)E_S[\sigma|\sigma \leq \widetilde{\sigma}-\eta^*] \end{pmatrix} \\
&\quad +\varphi k_2 \left(F_S(\widetilde{\sigma}+\eta^*) - 2F_S(\widetilde{\sigma}) + F_S(\widetilde{\sigma}-\eta^*)\right) \\
&= (\varphi F_S(\widetilde{\sigma}-\eta^*) + \varphi(1 - F_S(\widetilde{\sigma}+\eta^*) + 1 - \varphi)\widetilde{\Delta}_R(\varnothing|\eta^*) \\
&\quad +\varphi k_1 \begin{pmatrix} F_S(\widetilde{\sigma}+\eta^*)\mu_S - (\gamma_S^2+\gamma_\varepsilon^2)f_S(\widetilde{\sigma}+\eta^*) \\ -2F_S(\widetilde{\sigma})\mu_S + 2(\gamma_S^2+\gamma_\varepsilon^2)f_S(\widetilde{\sigma}) \\ +F_S(\widetilde{\sigma}-\eta^*)\mu_S - (\gamma_S^2+\gamma_\varepsilon^2)f_S(\widetilde{\sigma}-\eta^*) \end{pmatrix} \\
&\quad +\varphi k_2 \left(F_S(\widetilde{\sigma}+\eta^*) - 2F_S(\widetilde{\sigma}) + F_S(\widetilde{\sigma}-\eta^*)\right), \qquad \text{(C.125)}
\end{aligned}$$

where the last equality follows by Lemma C.1. Then, taking the limit and using Claims 2.2 and 2.3 we obtain (given that $\lim_{\gamma_R \to \gamma_S^{\pm}} k_1 = 0$ and $\lim_{\gamma_R \to \gamma_S^{\pm}} k_2 = \text{sgn}\,(\gamma_S - \gamma_R)\frac{\gamma_\varepsilon^2}{\gamma_S^2+\gamma_\varepsilon^2}\Delta_0$):

$$\begin{aligned}
\lim_{\gamma_R \to \gamma_S^{\pm}} E_S[\widetilde{\Delta}_R] &= (1-\varphi)\Delta_0 + \varphi\frac{\gamma_\varepsilon^2\mu_S}{\gamma_S^2+\gamma_\varepsilon^2} \\
&= E_S[\widetilde{\Delta}_R|\gamma_R = \gamma_S],
\end{aligned}$$

where the last equality is by (C.94) (recall that μ_R is normalized to 0).

Claim 2.4 implies that for $\mu_S > \mu_R$, $E_S[\widetilde{\Delta}_R|\gamma_R]$ is continuous in γ_R at $\gamma_R = \gamma_S$.

Step 2. In this step, we prove the continuity of $E_S[\widetilde{\Delta}_R|\gamma_R]$ in γ_R at $\gamma_R = \gamma_S$ if $\mu_S = \mu_R = 0$.

Claim 2.5. If $\mu_S = \mu_R = 0$, *then* $\lim_{\gamma_R \to \gamma_S^{\pm}} \widetilde{\Delta}_R(\varnothing|\eta^*) = 0$.

Proof. By (C.124), if $\mu_S = \mu_R = 0$ then

$$\begin{aligned}
& \widetilde{\Delta}_R(\varnothing|\eta) \\
= & \frac{1}{\varphi(F_l + 1 - F_h) + 1 - \varphi} \\
& \times \begin{pmatrix} \varphi k_1(\gamma_R^2 + \gamma_\varepsilon^2) f_l - \varphi F_l k_2 + \varphi k_1(\gamma_R^2 + \gamma_\varepsilon^2) f_h \\ +\varphi(1 - F_h)k_2 + (1 - \varphi)\Delta_0 \end{pmatrix} \\
= & \varphi k_1(\gamma_R^2 + \gamma_\varepsilon^2) \frac{f_l + f_h}{\varphi(F_l + 1 - F_h) + 1 - \varphi},
\end{aligned}$$

where the last equality follows from $k_2 = \text{sgn}\,(\gamma_S - \gamma_R)\left(\frac{\gamma_\varepsilon^2 \mu_S}{\gamma_S^2 + \gamma_\varepsilon^2} - \frac{\gamma_\varepsilon^2 \mu_R}{\gamma_R^2 + \gamma_\varepsilon^2}\right) = 0$ and $\Delta_0 = \mu_S - \mu_R = 0$ by assumption. Then, since $\lim_{\gamma_R \to \gamma_S^\pm} k_1 = 0$, we obtain

$$\mu_S = \mu_R = 0 : \lim_{\gamma_R \to \gamma_S^\pm} \widetilde{\Delta}_R(\varnothing|\eta^*) = 0.$$

Claim 2.6. If $\mu_S = \mu_R = 0$, *then* $\lim_{\gamma_R \to \gamma_S^\pm} E_S[\widetilde{\Delta}_R] = E_S[\widetilde{\Delta}_R | \mu_R = \mu_S, \gamma_R = \gamma_S]$.

Proof. By (C.125), for $\gamma_S \neq \gamma_R$ we have

$$\begin{aligned}
E_S[\widetilde{\Delta}_R] = & (\varphi F_S(\widetilde{\sigma} - \eta^*) + \varphi(1 - F_S(\widetilde{\sigma} + \eta^*)) + 1 - \varphi)\widetilde{\Delta}_R(\varnothing|\eta^*) \\
& + \varphi k_1 \begin{pmatrix} F_S(\widetilde{\sigma} + \eta^*)\mu_S - (\gamma_S^2 + \gamma_\varepsilon^2) f_S(\widetilde{\sigma} + \eta^*) \\ -2F_S(\widetilde{\sigma})\mu_S + 2(\gamma_S^2 + \gamma_\varepsilon^2) f_S(\widetilde{\sigma}) \\ +F_S(\widetilde{\sigma} - \eta^*)\mu_S - (\gamma_S^2 + \gamma_\varepsilon^2) f_S(\widetilde{\sigma} - \eta^*) \end{pmatrix} \\
& + \varphi k_2 \left(F_S(\widetilde{\sigma} + \eta^*) - 2F_S(\widetilde{\sigma}) + F_S(\widetilde{\sigma} - \eta^*)\right). \qquad \text{(C.126)}
\end{aligned}$$

Since for $\mu_S = \mu_R = 0$ we have $\lim_{\gamma_R \to \gamma_S^\pm} k_1 = 0$, $k_2 = 0$ and $\lim_{\gamma_R \to \gamma_S^\pm} \widetilde{\Delta}_R(\varnothing|\eta^*) = 0$ (by Claim 2.5), the above expression implies

$$\lim_{\gamma_R \to \gamma_S^\pm} E_S[\widetilde{\Delta}_R] = 0 = E_S[\widetilde{\Delta}_R | \mu_R = \mu_S, \gamma_R = \gamma_S],$$

where the last equality follows from the fact that under identical priors $\Delta(\sigma) = 0$ for any σ.

Thus, Claim 2.6 implies that for $\mu_S = \mu_R$, $E_S[\widetilde{\Delta}_R|\gamma_R]$ is again continuous in γ_R at $\gamma_R = \gamma_S$. This completes the proof of Claim 2 for the case $\gamma_R = \gamma_S$.

Claim 3. *R's perceived disagreement expected ex ante by S* $E_S[\widetilde{\Delta}_R]$ *strictly decreases in*

γ_R *if* $\gamma_R > \gamma_S$, $\mu_S > \mu_R$ *and* γ_R *is sufficiently close to* γ_S.

Proof. Assume $\gamma_R > \gamma_S$ and $\mu_S > \mu_R$. Note that the derivations in (C.95) hold independently from whether $\gamma_S > \gamma_R$ or not, thus we still have

$$\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} = P_S^{\varnothing} \frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\gamma_R} + \varphi \frac{dk_1}{d\gamma_R} \left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*} \sigma f_S(\sigma) d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}} \sigma f_S(\sigma) d\sigma \right). \quad \text{(C.127)}$$

By Lemma C.10 for $\gamma_R > \gamma_S$ we have

$$\frac{d\widetilde{\Delta}_R(\varnothing|\eta^*)}{d\gamma_R} = \frac{(f_h + f_l)\gamma_R \gamma_\varepsilon^2 (\gamma_S^2 + \gamma_R^2 + 2\gamma_\varepsilon^2)\varphi}{(\gamma_S^2 + \gamma_\varepsilon^2)(\gamma_R^2 + \gamma_\varepsilon^2)(1 - F_h\varphi + F_l\varphi)}. \quad \text{(C.128)}$$

Substituting it back to (C.127) and taking the derivative $\frac{dk_1}{d\gamma_R}$ in the considered parameter case we obtain

$$\begin{aligned} \frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} &= P_S^{\varnothing} \frac{(f_h + f_l)\gamma_R \gamma_\varepsilon^2 (\gamma_S^2 + \gamma_R^2 + 2\gamma_\varepsilon^2)\varphi}{(\gamma_S^2 + \gamma_\varepsilon^2)(\gamma_R^2 + \gamma_\varepsilon^2)(1 - F_h\varphi + F_l\varphi)} \\ &\quad + \varphi \frac{2\gamma_R \gamma_\varepsilon^2}{(\gamma_R^2 + \gamma_\varepsilon^2)^2} \left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*} \sigma f_S(\sigma) d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}} \sigma f_S(\sigma) d\sigma \right). \quad \text{(C.129)} \end{aligned}$$

Consider the limits of each term as γ_R goes to γ_S from the right. For the first term, we have

$$\lim_{\gamma_R \to \gamma_S^+} \frac{(f_h + f_l)\gamma_R \gamma_\varepsilon^2 (\gamma_S^2 + \gamma_R^2 + 2\gamma_\varepsilon^2)\varphi}{(\gamma_S^2 + \gamma_\varepsilon^2)(\gamma_R^2 + \gamma_\varepsilon^2)(1 - F_h\varphi + F_l\varphi)} = 0, \quad \text{(C.130)}$$

since $\lim_{\gamma_R \to \gamma_S^+} f_l = \lim_{\gamma_R \to \gamma_S^+} f_h = 0$ by Claim 2.2 from the proof of Claim 2 (given that $\mu_S > \mu_R$ by assumption).

Consider the second term of (C.129). First, we have

$$\lim_{\gamma_R \to \gamma_S^+} \left(- \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}} \sigma f_S(\sigma) d\sigma \right) = - \int_{-\infty}^{\infty} \sigma f_S(\sigma) d\sigma = -\mu_S < 0, \quad \text{(C.131)}$$

as $\lim_{\gamma_R \to \gamma_S^+} \widetilde{\sigma} - \eta^* = -\infty$ and $\lim_{\gamma_R \to \gamma_S^+} \widetilde{\sigma} = \infty$ by Claim 2.1 from the proof of Claim 2

(given that $\mu_S > \mu_R$ by assumption). Next,

$$\begin{aligned}
&\lim_{\gamma_R \to \gamma_S^+} \int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*} \sigma f_S(\sigma) d\sigma \\
= &\lim_{\gamma_R \to \gamma_S^+} \left(\int_{-\infty}^{\widetilde{\sigma}+\eta^*} \sigma f_S(\sigma) d\sigma - \int_{-\infty}^{\widetilde{\sigma}} \sigma f_S(\sigma) d\sigma \right) \\
= &\lim_{\gamma_R \to \gamma_S^+} (F_S(\widetilde{\sigma}+\eta^*) E_S[\sigma | \sigma \leq \widetilde{\sigma}+\eta^*] - F_S(\widetilde{\sigma}) E_S[\sigma | \sigma \leq \widetilde{\sigma}]) \\
= &\lim_{\gamma_R \to \gamma_S^+} \begin{pmatrix} F_S(\widetilde{\sigma}+\eta^*)\mu_S - (\gamma_S^2+\gamma_\varepsilon^2) f_S(\widetilde{\sigma}+\eta^*) \\ -F_S(\widetilde{\sigma})\mu_S + (\gamma_S^2+\gamma_\varepsilon^2) f_S(\widetilde{\sigma}) \end{pmatrix} \\
= &\ 0 \qquad \text{(C.132)}
\end{aligned}$$

where the third equality follows by Lemma C.1, and the last equality holds since $\lim_{\gamma_R \to \gamma_S^+} \widetilde{\sigma} + \eta^* = \infty$ and $\lim_{\gamma_R \to \gamma_S^+} \widetilde{\sigma} = \infty$ by Claim 2.1 from the proof of Claim 2 (given that $\mu_S > \mu_R$ by assumption).

(C.129)-(C.132) together imply that for $\mu_S > \mu_R = 0$ we have

$$\lim_{\gamma_R \to \gamma_S^+} \frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} = -\varphi \frac{2\gamma_R \gamma_\varepsilon^2}{(\gamma_R^2+\gamma_\varepsilon^2)^2} \mu_S < 0,$$

which completes the proof of Claim 3.

Claim 4. *R's perceived disagreement expected ex ante by S* $E_S[\widetilde{\Delta}_R]$ *strictly increases with* γ_R *if* $\mu_S = \mu_R$ *and* $\gamma_R > \gamma_S$.

Proof. Recall from (C.129) that for $\gamma_R > \gamma_S$ we have

$$\begin{aligned}
\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} = &\ P_S^{\varnothing} \frac{(f_h + f_l)\gamma_R \gamma_\varepsilon^2 (\gamma_S^2 + \gamma_R^2 + 2\gamma_\varepsilon^2)\varphi}{(\gamma_S^2+\gamma_\varepsilon^2)(\gamma_R^2+\gamma_\varepsilon^2)(1 - F_h\varphi + F_l\varphi)} \\
&+ \varphi \frac{2\gamma_R\gamma_\varepsilon^2}{(\gamma_R^2+\gamma_\varepsilon^2)^2} \left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*} \sigma f_S(\sigma) d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}} \sigma f_S(\sigma) d\sigma \right). \qquad \text{(C.133)}
\end{aligned}$$

Note that $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} = 0$ as far as $\mu_S = \mu_R$ (given that μ_R is normalized to 0). Besides, $\widetilde{\sigma} - \eta^* < \mu_R = 0 < \widetilde{\sigma} + \eta^*$ by Corollary C.1. It follows

$$\begin{aligned}
&\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*} \sigma f_S(\sigma) d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}} \sigma f_S(\sigma) d\sigma \\
= &\int_0^{\widetilde{\sigma}+\eta^*} \sigma f_S(\sigma) d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{0} \sigma f_S(\sigma) d\sigma > 0.
\end{aligned}$$

This together with (C.133) implies $\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} > 0$ for any $\gamma_R > \gamma_S$ if $\mu_S = \mu_R = 0$.

Claim 5. *Let $\mu_S > \mu_R$. Then, there exists γ'_R, such that R's perceived disagreement expected ex ante by S $E_S[\widetilde{\Delta}_R]$ strictly increases in γ_R if $\gamma_R > \gamma'_R$.*

Proof. Recall from (C.129) that for $\gamma_R > \gamma_S$ we have

$$\begin{aligned}\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} &= P_S^{\varnothing}\frac{(f_h+f_l)\gamma_R\gamma_\varepsilon^2(\gamma_S^2+\gamma_R^2+2\gamma_\varepsilon^2)\varphi}{(\gamma_S^2+\gamma_\varepsilon^2)(\gamma_R^2+\gamma_\varepsilon^2)(1-F_h\varphi+F_l\varphi)} \\ &\quad +\varphi\frac{2\gamma_R\gamma_\varepsilon^2}{(\gamma_R^2+\gamma_\varepsilon^2)^2}\left(\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma-\int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma\right). \quad \text{(C.134)}\end{aligned}$$

The first term is strictly positive. Let us show that the second term is also positive for sufficiently large γ_R. Recall that by Proposition 1.i(b), it holds $\widetilde{\sigma}-\eta^* < \underline{\sigma} \le \overline{\sigma} < \widetilde{\sigma}+\eta^*$, where $\underline{\sigma}$ and $\overline{\sigma}$ are the lower and upper status quo signals, respectively. For the case $\gamma_R > \gamma_S$, by (C.118) and (C.120) these signals are equal to:

$$\begin{aligned}\overline{\sigma} &= \frac{(\gamma_S^2+2\gamma_\varepsilon^2)(\gamma_R^2+\gamma_\varepsilon^2)\mu_S}{(\gamma_R^2-\gamma_S^2)\gamma_\varepsilon^2} > 0, \\ \underline{\sigma} &= \frac{\gamma_S^2(\gamma_R^2+\gamma_\varepsilon^2)\mu_S}{(\gamma_S^2-\gamma_R^2)\gamma_\varepsilon^2} < 0.\end{aligned}$$

Then, we have

$$\begin{aligned}&\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma-\int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma \\ =\ &\int_{\widetilde{\sigma}}^{\overline{\sigma}}\sigma f_S(\sigma)d\sigma-\int_{\underline{\sigma}}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma+\int_{\overline{\sigma}}^{\widetilde{\sigma}+\eta^*}\sigma f_S(\sigma)d\sigma-\int_{\widetilde{\sigma}-\eta^*}^{\underline{\sigma}}\sigma f_S(\sigma)d\sigma \\ >\ &\int_{\widetilde{\sigma}}^{\overline{\sigma}}\sigma f_S(\sigma)d\sigma-\int_{\underline{\sigma}}^{\widetilde{\sigma}}\sigma f_S(\sigma)d\sigma, \quad \text{(C.135)}\end{aligned}$$

where the inequality follows from $\overline{\sigma} > 0$ and $\underline{\sigma} < 0$. Note that

$$\begin{aligned}\lim_{\gamma_R\to\infty}\widetilde{\sigma} &= \lim_{\gamma_R\to\infty}\frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2}=\mu_S, \quad \text{(C.136)} \\ \lim_{\gamma_R\to\infty}\overline{\sigma} &= \frac{(\gamma_S^2+2\gamma_\varepsilon^2)\mu_S}{\gamma_\varepsilon^2}, \quad \text{(C.137)} \\ \lim_{\gamma_R\to\infty}\underline{\sigma} &= -\frac{\gamma_S^2\mu_S}{\gamma_\varepsilon^2}=2\mu_S-\lim_{\gamma_R\to\infty}\overline{\sigma}. \quad \text{(C.138)}\end{aligned}$$

Denote $l = \lim_{\gamma_R \to \infty} \overline{\sigma}$. Then, taking the limit of the right-hand side of (C.135) we obtain

$$\begin{aligned}
&\lim_{\gamma_R \to \infty} \left(\int_{\widetilde{\sigma}}^{\overline{\sigma}} \sigma f_S(\sigma) d\sigma - \int_{\underline{\sigma}}^{\widetilde{\sigma}} \sigma f_S(\sigma) d\sigma \right) \\
= &\int_{\mu_S}^{l} \sigma f_S(\sigma) d\sigma - \int_{2\mu_S - l}^{\mu_S} \sigma f_S(\sigma) d\sigma \\
= &\int_{\mu_S}^{l} \sigma f_S(\sigma) d\sigma - \int_{\mu_S}^{l} (2\mu_S - \sigma) f_S(2\mu_S - \sigma) d\sigma \\
= &\int_{\mu_S}^{l} \sigma f_S(\sigma) d\sigma - \int_{\mu_S}^{l} (2\mu_S - \sigma) f_S(\sigma) d\sigma \\
= &\; 2 \int_{\mu_S}^{l} (\sigma - \mu_S) f_S(\sigma) d\sigma > 0,
\end{aligned}$$

where the first equality is by (C.136)-(C.138), the second equality is obtained by integration by substitution, and the inequality is due to $l = \frac{(\gamma_S^2 + 2\gamma_\varepsilon^2)\mu_S}{\gamma_\varepsilon^2} > \mu_S$. Thus,

$$\lim_{\gamma_R \to \infty} \left(\int_{\widetilde{\sigma}}^{\overline{\sigma}} \sigma f_S(\sigma) d\sigma - \int_{\underline{\sigma}}^{\widetilde{\sigma}} \sigma f_S(\sigma) d\sigma \right) > 0,$$

which together with (C.135) implies that there exists sufficiently large γ'_R such that $\int_{\widetilde{\sigma}}^{\widetilde{\sigma}+\eta^*} \sigma f_S(\sigma) d\sigma - \int_{\widetilde{\sigma}-\eta^*}^{\widetilde{\sigma}} \sigma f_S(\sigma) d\sigma > 0$ for any $\gamma_R > \gamma'_R$. Finally, by (C.134) this yields that there exists sufficiently large γ'_R such that $\frac{dE_S[\widetilde{\Delta}_R]}{d\gamma_R} > 0$ for any $\gamma_R > \gamma'_R$.

To sum up, by the preceding claims we have shown the following properties of R's perceived disagreement expected ex ante by S $E_S[E_R[\Delta(\sigma)|d]] = E_S[\widetilde{\Delta}_R]$ (given that assumption $\mu_S \geq \mu_R$ is without loss of generality):

- $E_S[\widetilde{\Delta}_R]$ strictly decreases in γ_R if $\gamma_R < \gamma_S$ (Claim 1);
- $E_S[\widetilde{\Delta}_R]$ is continuous in γ_R at any $\gamma_R \in (0, \infty)$ (Claim 2);
- $E_S[\widetilde{\Delta}_R]$ strictly decreases in γ_R if $\mu_S \neq \mu_R$, $\gamma_R > \gamma_S$ and γ_R is sufficiently close to γ_S (Claim 3);
- $E_S[\widetilde{\Delta}_R]$ strictly increases in γ_R if $\mu_S = \mu_R$ and $\gamma_R > \gamma_S$ (Claim 4);
- $E_S[\widetilde{\Delta}_R]$ strictly increases in γ_R if $\mu_S \neq \mu_R$ and γ_R is sufficiently large (Claim 5).

Claims 1, 2 and 4 lead to the claim of Proposition 6(a). Claims 1, 2, 3 and 5 lead to the claim of Proposition 6(b). ■

C.5 Aversion to actual disagreement

C.5.1 Proof of Proposition 7

We need to show that the set of equilibria under S's loss function $L_1(d,\sigma) = |E_R[\omega|d] - E_S[\omega|\sigma]|$ is the same as under the alternative loss function $L_2(d,\sigma) = E_S[(a_R(d)-\omega)^2|d,\sigma]$.

Assume there exists an equilibrium under $L_1(d,\sigma)$ with the set of disclosed signals D, such that σ is disclosed if and only if $\sigma \in D$. Note that this uniquely pins down R's equilibrium actions to $a_R(\varnothing) = E_R[\omega|\varnothing]$ and $a_R(\sigma) = E_R[\omega|\sigma]$ (as follows from the first-order condition for minimizing R's loss function $(a_R - \omega)^2$). Let us show that the same equilibrium exists under $L_2(d,\sigma)$, i.e., for the same R's strategy $a_R(d)$ and any signal $\sigma \in R$, $L_1(\sigma,\sigma) \geq L_1(\varnothing,\sigma)$ holds if and only if $L_2(\sigma,\sigma) \geq L_2(\varnothing,\sigma)$ (i.e., S prefers to disclose the same range of signals given $a_R(d)$ under both loss functions).

We have:

$$\begin{aligned} & L_2(\sigma,\sigma) - L_2(\varnothing,\sigma) \\ = \; & a_R^2(\sigma) - 2E_S[\omega|\sigma]a_R(\sigma) + E_S[\omega^2] \\ & -a_R^2(\varnothing) + 2E_S[\omega|\sigma]a_R(\varnothing) - E_S[\omega^2] \\ = \; & (a_R(\sigma) - a_R(\varnothing))(a_R(\sigma) + a_R(\varnothing) - 2E_S[\omega|\sigma]). \end{aligned} \tag{C.139}$$

Let us show that $L_1(\sigma,\sigma) \geq L_1(\varnothing,\sigma)$ holds if and only if $L_2(\sigma,\sigma) \geq L_2(\varnothing,\sigma)$ in all parameter cases.

Case 1: $E_R[\omega|\sigma] - E_S[\omega|\sigma] \geq 0$ and $E_R[\omega|\varnothing] - E_S[\omega|\sigma] \geq 0$.

In this case, we have

$$L_1(\sigma,\sigma) - L_1(\varnothing,\sigma) = E_R[\omega|\sigma] - E_R[\omega|\varnothing] = a_R(\sigma) - a_R(\varnothing).$$

Note that by assumption, $a_R(\sigma) + a_R(\varnothing) - 2E_S[\omega|\sigma] \geq 0$. Consequently, by (C.139), $L_1(\sigma,\sigma) \geq L_1(\varnothing,\sigma)$ holds if and only if $L_2(\sigma,\sigma) \geq L_2(\varnothing,\sigma)$.

Case 2: $E_R[\omega|\sigma] - E_S[\omega|\sigma] \geq 0$ and $E_R[\omega|\varnothing] - E_S[\omega|\sigma] < 0$.

In this case, we have

$$\begin{aligned} & L_1(\sigma,\sigma) - L_1(\varnothing,\sigma) \\ = \; & E_R[\omega|\sigma] + E_R[\omega|\varnothing] - 2E_S[\omega|\sigma] \\ = \; & a_R(\sigma) + a_R(\varnothing) - 2E_S[\omega|\sigma]. \end{aligned}$$

Note that by assumption, $E_R[\omega|\sigma] > E_R[\omega|\varnothing] \Leftrightarrow a_R(\sigma) > a_R(\varnothing)$. Consequently, by (C.139), $L_1(\sigma,\sigma) \geq L_1(\varnothing,\sigma)$ holds if and only if $L_2(\sigma,\sigma) \geq L_2(\varnothing,\sigma)$.

Case 3: $E_R[\omega|\sigma] - E_S[\omega|\sigma] < 0$ and $E_R[\omega|\varnothing] - E_S[\omega|\sigma] \geq 0$.

In this case, we have

$$\begin{aligned} & L_1(\sigma,\sigma) - L_1(\varnothing,\sigma) \\ = \; & -E_R[\omega|\sigma] - E_R[\omega|\varnothing] + 2E_S[\omega|\sigma] \\ = \; & -(a_R(\sigma) + a_R(\varnothing) - 2E_S[\omega|\sigma]). \end{aligned}$$

Note that by assumption, $E_R[\omega|\sigma] < E_R[\omega|\varnothing] \Leftrightarrow a_R(\sigma) < a_R(\varnothing)$. Consequently, by (C.139), $L_1(\sigma,\sigma) \geq L_1(\varnothing,\sigma)$ holds if and only if $L_2(\sigma,\sigma) \geq L_2(\varnothing,\sigma)$.

Case 4: $E_R[\omega|\sigma] - E_S[\omega|\sigma] < 0$ and $E_R[\omega|\varnothing] - E_S[\omega|\sigma] < 0$.

In this case, we have

$$\begin{aligned} & L_1(\sigma,\sigma) - L_1(\varnothing,\sigma) \\ = \; & -E_R[\omega|\sigma] + E_R[\omega|\varnothing] \\ = \; & -(a_R(\sigma) - a_R(\varnothing)) \end{aligned}$$

Note that by assumption, $a_R(\sigma) + a_R(\varnothing) - 2E_S[\omega|\sigma] < 0$. Consequently, by (C.139), $L_1(\sigma,\sigma) \geq L_1(\varnothing,\sigma)$ holds if and only if $L_2(\sigma,\sigma) \geq L_2(\varnothing,\sigma)$.

Thus, we have shown that whenever an equilibrium exists under $L_1(d,\sigma)$, an equilibrium with the same S's disclosure strategy (and hence the same R's strategy and beliefs) exists under $L_2(d,\sigma)$. The reverse claim can be shown by analogous argument. ■

C.5.2 Proof of Proposition 8(a)

Proposition: Let $\gamma_S \geq \gamma_R$.

a) If $\mu_S \neq 0$, *then there exists a unique equilibrium characterized by a non-disclosure interval* (σ_l, σ_h) *such that* S *discloses obtained signal* σ *if and only if* $\sigma \notin (\sigma_l, \sigma_h)$.

b) *If* $\mu_S = 0$, *then the only equilibrium features full disclosure.*

In what follows, we normalize $\mu_R = 0$ and let $\mu_S > 0$ without loss of generality.

Claim 1. *Any equilibrium features a non-disclosure interval* (σ_l, σ_h) *such that* S *discloses obtained signal* σ *if and only if* $\sigma \notin (\sigma_l, \sigma_h)$, *with* σ_l *and* σ_h *satisfying*

$$E_R[\omega|\varnothing] - E_S[\omega|\sigma_l] = E_S[\omega|\sigma_l] - E_R[\omega|\sigma_l], \tag{C.140}$$

$$E_R[\omega|\varnothing] = E_R[\omega|\sigma_h]. \tag{C.141}$$

Proof. The result for the case of equal variances has been established by Proposition 1 in Che and Kartik (2009), given that the sender's preferences in their model give rise to the same set of equilibria by Proposition 7.[17] Let us extend the result for the case of $\gamma_S > \gamma_R$.

Let $a_\varnothing$ be the equilibrium R's action conditional on non-disclosure (that happens with a positive probability as S is uninformed with a positive probability). As before, denote $\widetilde{\sigma}$ the unique 0-disagreement signal such that $E_R[\omega|\widetilde{\sigma}] = E_S[\omega|\widetilde{\sigma}]$. Consider now separately three cases depending on the relation between $E_R[\omega|\widetilde{\sigma}]$ and $E_R[\omega|\varnothing]$.

Case 1: $E_R[\omega|\widetilde{\sigma}] < E_R[\omega|\varnothing]$.

By (5), both $E_S[\omega|\sigma]$ and $E_R[\omega|\sigma]$ are both increasing linear function of σ, with $E_S[\omega|\sigma]$ being steeper since $\gamma_S > \gamma_R$ by assumption. Then, the signal space can be split into four adjacent non-empty intervals of signals s_1, s_2, s_3 and s_4 such that:

1) $E_S[\omega|\sigma] \leq E_R[\omega|\sigma] < E_R[\omega|\varnothing]$ for all $\sigma \in s_1$.

2) $E_R[\omega|\sigma] < E_S[\omega|\sigma] \leq E_R[\omega|\varnothing]$ for all $\sigma \in s_2$.

3) $E_R[\omega|\sigma] < E_R[\omega|\varnothing] < E_S[\omega|\sigma]$ for all $\sigma \in s_3$.

4) $E_R[\omega|\varnothing] \leq E_R[\omega|\sigma] < E_S[\omega|\sigma]$ for all $\sigma \in s_4$.

Let $L(d, \sigma) = |E_R[\omega|d] - E_S[\omega|\sigma]|$ be S's loss function conditional on disclosure $d \in \{\sigma, \varnothing\}$ and observed signal σ. Clearly, $L(\sigma, \sigma) \leq L(\varnothing, \sigma)$ in both s_1 and s_4, i.e., the sender prefers to disclose σ in these intervals. Next, in s_2, S prefers non-disclosure (i.e.,

[17] They assumed that S prefers non-disclose in case of indifference so that the non-disclosure interval $[\sigma_l, \sigma_h]$ includes its boundaries, but this is inconsequential.

$L(\sigma,\sigma) > L(\varnothing,\sigma)$) if and only if $\sigma > \sigma_l$ where σ_l solves

$$E_R\left[\omega\left|\varnothing\right.\right] - E_S\left[\omega\left|\sigma_l\right.\right] = E_S\left[\omega\left|\sigma_l\right.\right] - E_R\left[\omega\left|\sigma_l\right.\right]. \tag{C.142}$$

Finally, $L(\sigma,\sigma) > L(\varnothing,\sigma)$ in s_3. Hence, S prefers not to disclose in this interval. In sum, S prefers non-disclosure if and only if $\sigma \in (\sigma_l, \sigma_h)$ where σ_h lies at the boundary of s_3 and s_4, i.e., solves

$$E_R[\omega|\varnothing] = E_R\left[\omega\left|\sigma_h\right.\right]. \tag{C.143}$$

Case 2: $E_R\left[\omega\left|\widetilde{\sigma}\right.\right] > E_R[\omega|\varnothing]$.

Let us show that this is impossible in equilibrium. Assume by contradiction that $E_R\left[\omega\left|\widetilde{\sigma}\right.\right] > E_R[\omega|\varnothing]$ in equilibrium. Then, the signal space can be split into four adjacent non-empty intervals of signals s_1, s_2,s_3 and s_4 such that:

1) $E_S\left[\omega\left|\sigma\right.\right] < E_R\left[\omega\left|\sigma\right.\right] \leq E_R[\omega|\varnothing]$ for all $\sigma \in s_1$.

2) $E_S\left[\omega\left|\sigma\right.\right] \leq E_R\left[\omega\left|\varnothing\right.\right] < E_R[\omega|\sigma]$ for all $\sigma \in s_2$.

3) $E_R\left[\omega\left|\varnothing\right.\right] < E_S[\omega|\sigma] < E_R\left[\omega\left|\sigma\right.\right]$ for all $\sigma \in s_3$.

4) $E_R\left[\omega\left|\varnothing\right.\right] < E_R[\omega|\sigma] \leq E_S\left[\omega\left|\sigma\right.\right]$ for all $\sigma \in s_4$.

Clearly, $L(\sigma,\sigma) \leq L(\varnothing,\sigma)$ in both s_1 and s_4, i.e., the sender prefers to disclose σ in these intervals. Next, in s_2, S prefers non-disclosure as then $L(\sigma,\sigma) > L(\varnothing,\sigma)$. Finally, in s_3 S prefers non-disclosure if and only if $\sigma < \sigma_h$ where σ_h solves

$$E_R\left[\omega\left|\sigma_h\right.\right] - E_S\left[\omega\left|\sigma_h\right.\right] = E_S\left[\omega\left|\sigma_h\right.\right] - E_R\left[\omega\left|\varnothing\right.\right].$$

In sum, S prefers non-disclosure if and only if $\sigma \in (\sigma_l, \sigma_h)$ where σ_l lies at the boundary of s_1 and s_2, i.e., solves

$$E_R[\omega|\varnothing] = E_R\left[\omega\left|\sigma_l\right.\right]. \tag{C.144}$$

However, we now obtain a contradiction. First, observe that $\sigma_h < \widetilde{\sigma} < 0$. Indeed, all signals in s_3 (including σ_h) are lower than $\widetilde{\sigma}$ (the intersection point of $E_S[\omega|\sigma]$ and $E_R[\omega|\sigma]$) since $E_S[\omega|\sigma] < E_R\left[\omega\left|\sigma\right.\right]$ (i.e., $E_S[\omega|\sigma]$ has not yet crossed $E_R\left[\omega\left|\sigma\right.\right]$) in this interval. At the same time, $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} < 0$ as $\gamma_S > \gamma_R$ and $\mu_S > 0$ by assumption. Consequently, $\sigma_h < 0$. Then, given the above disclosure strategy (and $\mu_R = 0$ by

assumption), we have

$$\begin{aligned} E_R[\omega|\varnothing] &= \frac{\varphi(F_R(\sigma_h)-F_R(\sigma_l))}{\varphi(F_R(\sigma_h)-F_R(\sigma_l))+1-\varphi}\int_{\sigma_l}^{\sigma_h} E_R\left[\omega\,|\sigma\right]\frac{f_R(\sigma)}{F_R(\sigma_h)-F_R(\sigma_l)}d\sigma \\ &> \int_{\sigma_l}^{\sigma_h} E_R\left[\omega\,|\sigma\right]\frac{f_R(\sigma)}{F_R(\sigma_h)-F_R(\sigma_l)}d\sigma \\ &> E_R\left[\omega\,|\sigma_l\right], \end{aligned}$$

where the first inequality follows from $\sigma_h < 0$ as established above. We obtain a contradiction to (C.144).

Case 3: $E_R\left[\omega\,|\widetilde{\sigma}\right] = E_R[\omega|\varnothing]$.

Let us again show that this is impossible in equilibrium. Indeed, in this case the signal space can be split into two adjacent non-empty intervals of signals s_1 and s_2 such that:

1) $E_S\left[\omega\,|\sigma\right] < E_R\left[\omega\,|\sigma\right] < E_R[\omega|\varnothing]$ for all $\sigma \in s_1$.

2) $E_R\left[\omega\,|\varnothing\right] \leq E_R[\omega|\sigma] \leq E_S\left[\omega\,|\sigma\right]$ for all $\sigma \in s_2$.

Clearly, in both cases the distance between $E_S[\omega|\sigma]$ and $E_R[\omega|\sigma]$ is always smaller than between $E_S[\omega|\sigma]$ and $E_R[\omega|\varnothing]$, i.e., $L(\sigma,\sigma) \leq L(\varnothing,\sigma)$. Thus, in equilibrium S prefers to disclose all signals. Consequently, $E_R[\omega|\varnothing] = \mu_R = 0$. At the same time, $E_R\left[\omega\,|\widetilde{\sigma}\right] < 0$ that is a contradiction to the initial assumption of the case. The latter inequality holds since $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} < 0$ as $\gamma_S > \gamma_R$ and $\mu_S > 0$ by assumption.

Summarizing Cases 1-3, we obtain that any equilibrium is characterized by a non-disclosure interval (σ_l, σ_h) such that

$$E_R\left[\omega\,|\varnothing\right] - E_S\left[\omega\,|\sigma_l\right] = E_S\left[\omega\,|\sigma_l\right] - E_R\left[\omega\,|\sigma_l\right], \tag{C.145}$$

$$E_R[\omega|\varnothing] = E_R\left[\omega\,|\sigma_h\right]. \tag{C.146}$$

Given this disclosure strategy, the equilibrium value of $E_R[\omega|\varnothing]$ is given by

$$E_R[\omega|\varnothing] = \frac{\varphi}{\varphi(F_R(\sigma_h)-F_R(\sigma_l))+1-\varphi}\int_{\sigma_l}^{\sigma_h} E_R\left[\omega\,|\sigma\right] f_R(\sigma)d\sigma. \tag{C.147}$$

Claim 2. *A necessary and sufficient equilibrium condition is*

$$-\varphi\int_{\sigma_l}^{\sigma_h}(\sigma_h-\sigma)f_R(\sigma)d\sigma - \sigma_h(1-\varphi) = 0, \tag{C.148}$$

$$\sigma_l = b_1\sigma_h - b_2, \tag{C.149}$$

where

$$\begin{aligned} b_1 &= \frac{\gamma_R^2(\gamma_S^2+\gamma_\varepsilon^2)}{\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)}, \\ b_2 &= \frac{2\gamma_\varepsilon^2(\gamma_R^2+\gamma_\varepsilon^2)\mu_S}{\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)}. \end{aligned}$$

Proof. The result for the case of $\gamma_S = \gamma_R$ (in which case $b_1 = 1$ and $b_2 = \frac{2\gamma_\varepsilon^2}{\gamma_R^2}\mu_S$) has been established in the proof of Proposition 1 in Che and Kartik (2009), given that the sender's preferences in their model give rise to the same set of equilibria by Proposition 7. Let us extend the result for the case of $\gamma_S > \gamma_R$.

Note that the system of equations (C.145)-(C.147) represents the necessary and sufficient equilibrium condition (S selects optimal disclosure given R's beliefs and the latter are formed by Bayes rule). Let us simplify these conditions further. Plugging (C.146) into (C.147) implies

$$E_R\left[\omega\,|\sigma_h\right] = \frac{\varphi}{\varphi(F_R(\sigma_h)-F_R(\sigma_l))+1-\varphi}\int_{\sigma_l}^{\sigma_h} E_R\left[\omega\,|\sigma\right] f_R(\sigma)d\sigma.$$

Since $E_R\left[\omega\,|\sigma\right] = \left(\frac{1}{1+\frac{\gamma_\varepsilon^2}{\gamma_R^2}}\right)\sigma$ this further simplifies to

$$\sigma_h = \frac{\varphi}{\varphi(F_R(\sigma_h)-F_R(\sigma_l))+1-\varphi}\int_{\sigma_l}^{\sigma_h} \sigma f_R(\sigma)d\sigma. \tag{C.150}$$

In turn, this holds if and only if

$$-\varphi\int_{\sigma_l}^{\sigma_h}(\sigma_h-\sigma)f_R(\sigma)d\sigma - \sigma_h(1-\varphi) = 0. \tag{C.151}$$

Next, (C.145) and (C.146) together imply

$$E_R\left[\omega\,|\sigma_l\right] = 2E_S\left[\omega\,|\sigma_l\right] - E_R\left[\omega\,|\sigma_h\right]. \tag{C.152}$$

Substituting for $E_S\left[\omega\,|\sigma_i\right]$ and $E_R\left[\omega\,|\sigma_i\right]$ and simplifying yields

$$\sigma_l = b_1\sigma_h - b_2, \tag{C.153}$$

where

$$\begin{aligned} b_1 &= \frac{\gamma_R^2(\gamma_S^2+\gamma_\varepsilon^2)}{\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)}, \\ b_2 &= \frac{2\gamma_\varepsilon^2(\gamma_R^2+\gamma_\varepsilon^2)\mu_S}{\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)} \end{aligned}$$

Claim 3. $\sigma_h < 0$.

Proof. Assume by contradiction $\sigma_h \geq 0$. Then, since $\sigma_l = b_1\sigma_h - b_2 < \sigma_h$ (where the inequality is strict since $\mu_S > 0$ by assumption), we have

$$\begin{aligned} \sigma_h &> \frac{\varphi(F_R(\sigma_h)-F_R(\sigma_l))}{\varphi(F_R(\sigma_h)-F_R(\sigma_l))+1-\varphi}\int_{\sigma_l}^{\sigma_h}\sigma\frac{f_R(\sigma)}{F_R(\sigma_h)-F_R(\sigma_l)}d\sigma \\ &= \frac{\varphi}{\varphi(F_R(\sigma_h)-F_R(\sigma_l))+1-\varphi}\int_{\sigma_l}^{\sigma_h}\sigma f_R(\sigma)d\sigma \end{aligned}$$

that contradicts (C.150).

Claim 4. *There exists at least one equilibrium.*

Proof. Plugging (C.149) into (C.148) we obtain a sufficient and necessary equilibrium condition

$$\Lambda(\sigma_h) := -\varphi\int_{b_1\sigma_h-b_2}^{\sigma_h}(\sigma_h-\sigma)f_R(\sigma)d\sigma - \sigma_h(1-\varphi) = 0. \tag{C.154}$$

Let us show that at least one solution to the equation exists. We have

$$\lim_{\sigma_h\to-\infty}\Lambda(\sigma_h) = \lim_{\sigma_h\to-\infty}\left(-\varphi\sigma_h\int_{b_1\sigma_h-b_2}^{\sigma_h}f_R(\sigma)d\sigma - \sigma_h(1-\varphi) + \varphi\int_{b_1\sigma_h-b_2}^{\sigma_h}\sigma f_R(\sigma)d\sigma\right) = \infty \tag{C.155}$$

since for the last term in brackets it holds $\int_{b_1\sigma_h-b_2}^{\sigma_h}\sigma f_R(\sigma)d\sigma > \int_{-\infty}^{\sigma_h}\sigma f_R(\sigma)d\sigma$ (for $\sigma_h < 0$) while

$$\lim_{\sigma_h\to-\infty}\int_{-\infty}^{\sigma_h}\sigma f_R(\sigma)d\sigma = -\lim_{\sigma_h\to-\infty}\left((\gamma_R^2+\gamma_\varepsilon^2)f_R(\sigma_h)\right) = 0,$$

where the first equality follows by Lemma C.1.

Next, note that $\Lambda(0) < 0$. This together with (C.155) and continuity of $\Lambda(\sigma_h)$ implies existence of at least one solution to $\Lambda(\sigma_h) = 0$ by the intermediate function theorem.

Claim 5. *The equilibrium is unique.*

Let us show that the solution to the necessary and sufficient equilibrium condition $\Lambda(\sigma_h) = 0$ (that exists by Claim 4) is unique.

We have

$$\frac{\partial\Lambda(\sigma_h)}{\partial\sigma_h} = \varphi\left[b_1((1-b_1)\sigma_h + b_2)f_R(b_1\sigma_h - b_2) - \int_{b_1\sigma_h - b_2}^{\sigma_h} f_R(\sigma)d\sigma\right] - (1-\varphi). \quad \text{(C.156)}$$

Consider the derivative at the equilibrium value of σ_h denoted by σ_h^* so that $\Theta(\sigma_h^*) = 0$. The latter implies

$$\begin{aligned} -\varphi\int_{b_1\sigma_h^* - b_2}^{\sigma_h^*}(\sigma_h^* - \sigma)f_R(\sigma)d\sigma - \sigma_h^*(1-\varphi) &= 0 \\ \iff 1 - \varphi = -\frac{\varphi}{\sigma_h^*}\int_{b_1\sigma_h^* - b_2}^{\sigma_h^*}(\sigma_h^* - \sigma)f_R(\sigma)d\sigma, \end{aligned}$$

where $\sigma_h^* = 0$ is excluded by Claim 3. Substituting the right-hand side of the last equality for $1-\varphi$ in (C.156) we obtain (substituting also $\sigma_l^* = b_1\sigma_h^* - b_2$ by (C.149)):

$$\begin{aligned} \frac{\partial\Lambda(\sigma_h^*)}{\partial\sigma_h} &= \varphi\left[b_1(\sigma_h^* - \sigma_l^*)f_R(\sigma_l^*) - \int_{\sigma_l^*}^{\sigma_h^*} f_R(\sigma)d\sigma\right] + \frac{\varphi}{\sigma_h^*}\int_{\sigma_l^*}^{\sigma_h^*}(\sigma_h^* - \sigma)f_R(\sigma)d\sigma \\ &= \frac{\varphi}{\sigma_h^*}\left[b_1\sigma_h^*(\sigma_h^* - \sigma_l^*)f_R(\sigma_l^*) - \int_{\sigma_l^*}^{\sigma_h^*}\sigma f_R(\sigma)d\sigma\right] \quad \text{(C.157)} \end{aligned}$$

Let us show that the term in brackets is strictly positive. We have

$$\begin{aligned} & b_1\sigma_h^*(\sigma_h^* - \sigma_l^*)f_R(\sigma_l^*) - \int_{\sigma_l^*}^{\sigma_h^*}\sigma f_R(\sigma)d\sigma \\ > \ & \sigma_h^*(\sigma_h^* - \sigma_l^*)f_R(\sigma_l^*) - \int_{\sigma_l^*}^{\sigma_h^*}\sigma f_R(\sigma)d\sigma \\ > \ & \sigma_h^*(\sigma_h^* - \sigma_l^*)f_R(\sigma_l^*) - \int_{\sigma_l^*}^{\sigma_h^*}\sigma f_R(\sigma_l^*)d\sigma \\ = \ & 0.5f_R(\sigma_l^*)(\sigma_h^* - \sigma_l^*)^2 > 0, \end{aligned}$$

where the first inequality is due to $b_1 < 1$ (as $\gamma_S > \gamma_R$ by assumption), $\sigma_h^* < 0$ by Claim 3 and $\sigma_l^* = b_1\sigma_h - b_2 < \sigma_h$ (as $\mu_S > 0$ by assumption); and the second inequality is also due to $\sigma_h^* < 0$ and $\sigma_l = b_1\sigma_h - b_2 < \sigma_h$. Thus, the term in brackets on the right-hand side of (C.157) is strictly positive, which together with $\sigma_h^* < 0$ by Claim 3 implies $\frac{\partial\Lambda(\sigma_h^*)}{\partial\sigma_h} < 0$. Thus, $\Lambda(\sigma_h)$ strictly decreases at its roots, which by continuity of $\Lambda(\sigma_h)$ implies that it may have at most one root. Hence, the equilibrium (existing by Claim 4) is unique. ■

C.5.3 Proof of Proposition 8(b)

In what follows, we normalize $\mu_S = \mu_R = 0$ without loss of generality.

Claim 1. *There exists equilibrium with full disclosure.*

Proof. Assume S discloses all signals. Then, in case of non-disclosure $E_R[\omega|\varnothing] = 0$. At the same time, $E_S[\omega|\sigma]$ and $E_R[\omega|\sigma]$ intersect at the 0-disagreement signal $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} = 0$. Consequently, the signal space can be split into two adjacent non-empty intervals of signals s_1 and s_2 such that:

1) $E_S[\omega|\sigma] < E_R[\omega|\sigma] < E_R[\omega|\varnothing]$ for all $\sigma \in s_1$.

2) $E_R[\omega|\varnothing] \leq E_R[\omega|\sigma] \leq E_S[\omega|\sigma]$ for all $\sigma \in s_2$.

Clearly, in both cases the distance between $E_S[\omega|\sigma]$ and $E_R[\omega|\sigma]$ is always smaller than between $E_S[\omega|\sigma]$ and $E_R[\omega|\varnothing]$. Thus, given $E_R[\omega|\varnothing] = 0$, S prefers to disclose all signals, that makes this strategy consistent with R's beliefs, and hence constitutes an equilibrium.

Claim 2. *There exists no other equilibrium except for full disclosure.*

Proof. Assume by contradiction that there exists an equilibrium where S does not disclose some signal σ.

Consider now separately three cases depending on the relation between $E_R[\omega|\widetilde{\sigma}]$ and $E_R[\omega|\varnothing]$. We will show that we come to contradiction in each case.

Case 1: $E_R[\omega|\widetilde{\sigma}] < E_R[\omega|\varnothing]$.

By the same argument as in the proof of Claim 1 of Proposition 8(i.a), it must hold that S does not disclose all signals in the non-empty interval (σ_l, σ_h) satisfying

$$E_R[\omega|\varnothing] - E_S[\omega|\sigma_l] = E_S[\omega|\sigma_l] - E_R[\omega|\sigma_l]. \tag{C.158}$$

$$E_R[\omega|\varnothing] = E_R[\omega|\sigma_h]. \tag{C.159}$$

Since $E_R[\omega|\widetilde{\sigma}] = 0$ as $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} = 0$, case assumption $E_R[\omega|\widetilde{\sigma}] < E_R[\omega|\varnothing]$ implies $E_R[\omega|\varnothing] > 0$ so that by (C.159)

$$E_R[\omega|\sigma_h] > 0 \Longrightarrow \sigma_h > 0.$$

Besides, one can show that $\sigma_l \geq 0$. Indeed, otherwise,

$$E_R[\omega|\sigma_l] > E_S[\omega|\sigma_l] > 2E_S[\omega|\sigma_l] > 2E_S[\omega|\sigma_l] - E_R[\omega|\varnothing]$$

that contradicts (C.158), where the first inequality is due to $\gamma_S > \gamma_R$ and $\mu_S = 0$ by

assumption, and the last inequality is due to $E_R[\omega|\varnothing] > E_R[\omega|\widetilde{\sigma}] = 0$ by assumption.

Then, since $\sigma_h > 0$ and $\sigma_l \geq 0$,

$$\begin{aligned} E_R[\omega|\varnothing] &= \frac{\varphi(F_R(\sigma_h) - F_R(\sigma_l))}{\varphi(F_R(\sigma_h) - F_R(\sigma_l)) + 1 - \varphi} \int_{\sigma_l}^{\sigma_h} E_R[\omega|\sigma] \frac{f_R(\sigma)}{F_R(\sigma_h) - F_R(\sigma_l)} d\sigma \\ &< \int_{\sigma_l}^{\sigma_h} E_R[\omega|\sigma] \frac{f_R(\sigma)}{F_R(\sigma_h) - F_R(\sigma_l)} d\sigma \\ &< E_R[\omega|\sigma_h] \end{aligned}$$

that contradicts (C.159).

Case 2: $E_R[\omega|\widetilde{\sigma}] > E_R[\omega|\varnothing]$.

By the same argument as in the proof of Claim 1 of Proposition 8(i.a), it must hold that S does not disclose all signals in the non-empty interval (σ_l, σ_h) satisfying

$$E_R[\omega|\sigma_h] - E_S[\omega|\sigma_h] = E_S[\omega|\sigma_h] - E_R[\omega|\varnothing]. \tag{C.160}$$

$$E_R[\omega|\varnothing] = E_R[\omega|\sigma_l]. \tag{C.161}$$

Since $E_R[\omega|\widetilde{\sigma}] = 0$ as $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} = 0$, case assumption $E_R[\omega|\widetilde{\sigma}] > E_R[\omega|\varnothing]$ implies $E_R[\omega|\varnothing] < 0$ so that by (C.161)

$$E_R[\omega|\sigma_l] < 0 \Longrightarrow \sigma_l < 0.$$

Besides, one can show that $\sigma_h \leq 0$. Indeed, otherwise,

$$E_R[\omega|\sigma_h] < E_S[\omega|\sigma_h] < 2E_S[\omega|\sigma_h] < 2E_S[\omega|\sigma_h] - E_R[\omega|\varnothing]$$

that contradicts (C.158), where the first inequality is due to $\gamma_S > \gamma_R$ and $\mu_S = 0$ by assumption, and the last inequality is due to $E_R[\omega|\varnothing] < E_R[\omega|\widetilde{\sigma}] = 0$ by assumption.

Then, since $\sigma_l < 0$ and $\sigma_h \leq 0$,

$$\begin{aligned} E_R[\omega|\varnothing] &= \frac{\varphi(F_R(\sigma_h) - F_R(\sigma_l))}{\varphi(F_R(\sigma_h) - F_R(\sigma_l)) + 1 - \varphi} \int_{\sigma_l}^{\sigma_h} E_R[\omega|\sigma] \frac{f_R(\sigma)}{F_R(\sigma_h) - F_R(\sigma_l)} d\sigma \\ &> \int_{\sigma_l}^{\sigma_h} E_R[\omega|\sigma] \frac{f_R(\sigma)}{F_R(\sigma_h) - F_R(\sigma_l)} d\sigma \\ &> E_R[\omega|\sigma_l] \end{aligned}$$

that contradicts (C.161).

Case 3: $E_R[\omega|\widetilde{\sigma}] = E_R[\omega|\varnothing]$.

By the same argument as in the proof of Claim 1 of Proposition 8(i.a), in this case S prefers to disclose all signals which contradicts the initial assumption.

Thus, we have come to contradiction in all possible parameter cases. ■

C.5.4 Proof of Proposition 9(a)

In what follows, we normalize μ_R to 0 without loss of generality.

Let $\gamma_S \geq \gamma_R$.

Claim 1. *R's expected utility strictly decreases in μ_S for $\mu_S > 0$.*

Step 1. Denote R's expected quadratic loss as

$$L_R = E_R[(a-\omega)^2].$$

Let $g_R(\omega)$ denote R's prior density function of ω, and $g_R(\omega|\sigma)$ denote R's posterior density function of ω conditional on σ. Then, given S's equilibrium disclosure strategy (to disclose if and only if $\sigma \notin (\sigma_l, \sigma_h)$) and R's optimal action conditional on disclosure $a_R(d) = E_R[\omega|d]$, we have

$$\begin{aligned}
L_R &= E_R[(a_R-\omega)^2] \\
&= E_R[(E_R[\omega|d]-\omega)^2] \\
&= \varphi \int_{\sigma_l}^{\sigma_h} \int_{-\infty}^{\infty} \left(E_R[\omega|\varnothing]-\omega\right)^2 g_R(\omega|\sigma) d\omega f_R(\sigma) d\sigma \\
&\quad + \varphi \int_{-\infty}^{\sigma_l} \int_{-\infty}^{\infty} \left(E_R[\omega|\sigma]-\omega\right)^2 g_R(\omega|\sigma) d\omega f_R(\sigma) d\sigma \\
&\quad + \varphi \int_{\sigma_h}^{\infty} \int_{-\infty}^{\infty} \left(E_R[\omega|\sigma]-\omega\right)^2 g_R(\omega|\sigma) d\omega f_R(\sigma) d\sigma \\
&\quad + (1-\varphi) \int_{-\infty}^{\infty} \left(E_R[\omega|\varnothing]-\omega\right)^2 g_R(\omega) d\omega \\
&= \varphi \int_{\sigma_l}^{\sigma_h} \int_{-\infty}^{\infty} \left(E_R[\omega|\varnothing]-\omega\right)^2 g_R(\omega|\sigma) d\omega f_R(\sigma) d\sigma \\
&\quad + \varphi \left(F_R(\sigma_l) + 1 - F_R(\sigma_h)\right) \frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}} \\
&\quad + (1-\varphi) \int_{-\infty}^{\infty} \left(E_R[\omega|\varnothing]-\omega\right)^2 g_R(\omega) d\omega,
\end{aligned}$$

where the second equality is due to $\int_{-\infty}^{\infty} \left(E_R[\omega|\sigma]-\omega\right)^2 g_R(\omega|\sigma) d\omega$ being the posterior variance of ω conditional on disclosed signal σ, which is constant at $\frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}}$ for any σ by

(6). This further simplifies to the following function of σ_l and σ_h (using Lemma C.1):

$$\begin{aligned} L_R(\sigma_l,\sigma_h) &= \varphi\left(F_R(\sigma_l)+1-F_R(\sigma_h)\right)\frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}} \\ &\quad +\varphi\begin{pmatrix} E_R[\omega|\varnothing,\sigma_l,\sigma_h]^2\left(F_R(\sigma_h)-F_R(\sigma_l)\right) \\ -2\gamma_R^2 E_R[\omega|\varnothing,\sigma_l,\sigma_h](f_R(\sigma_l)-f_R(\sigma_h)) \end{pmatrix} \\ &\quad +\varphi\left(\int_{\sigma_l}^{\sigma_h}\int_{-\infty}^{\infty}\omega^2 g_R(\omega|\sigma)d\omega f_R(\sigma)d\sigma\right) \\ &\quad +(1-\varphi)\left(E_R[\omega|\varnothing,\sigma_l,\sigma_h]^2+\int_{-\infty}^{\infty}\omega^2 g_R(\omega)d\omega\right). \end{aligned} \tag{C.162}$$

Taking the derivative of $L_R(\sigma_l,\sigma_h)$ with respect to μ_S and simplifying, we obtain:

$$\frac{dL_R(\sigma_l,\sigma_h)}{d\mu_S}=\varphi f_R(\sigma_l)\frac{\partial\sigma_l}{\partial\mu_S}\left(\frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}}-\int_{-\infty}^{\infty}(E_R[\omega|\varnothing]-\omega)^2 g_R(\omega|\sigma_l)d\omega\right). \tag{C.163}$$

In the following steps, we show that the right-hand side is strictly positive.

Step 2. Let us show that $\frac{\partial\sigma_l}{\partial\mu_S}<0$. By (C.149),

$$\frac{\partial\sigma_l}{\partial\mu_S}=b_1\frac{\partial\sigma_h}{\partial\mu_S}-\frac{\partial b_2}{\partial\mu_S} \tag{C.164}$$

Let us derive $\frac{\partial\sigma_h}{\partial\mu_S}$. Recall the equilibrium condition (C.154):

$$\Lambda(\sigma_h):=-\varphi\int_{b_1\sigma_h-b_2}^{\sigma_h}(\sigma_h-\sigma)f_R(\sigma)d\sigma-\sigma_h(1-\varphi)=0.$$

Then, by the implicit function theorem,

$$\frac{\partial\sigma_h}{\partial\mu_S}=-\frac{\partial\Lambda/\partial\mu_S}{\partial\Lambda/\partial\sigma_h}. \tag{C.165}$$

Taking the partial derivative $\partial\Lambda/\partial\mu_S$ and simplifying we obtain

$$\partial\Lambda/\partial\mu_S=-\varphi(\sigma_h-\sigma_l)f_R(\sigma_l)\frac{2\gamma_\varepsilon^2(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)}<0. \tag{C.166}$$

At the same time, in the proof of Claim 5 of the proof of Proposition 8, it has been shown that $\partial\Lambda/\partial\sigma_h<0$ at the equilibrium value of σ_h. This together with (C.165) and (C.166) implies that $\frac{\partial\sigma_h}{\partial\mu_S}<0$. Then, since $b_1=\frac{\gamma_R^2(\gamma_S^2+\gamma_\varepsilon^2)}{\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)}>0$ and $\frac{\partial b_2}{\partial\mu_S}=\frac{2\gamma_\varepsilon^2(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)}>0$ for $\gamma_S>\gamma_R$, (C.164) together with $\frac{\partial\sigma_h}{\partial\mu_S}<0$ yields

$$\frac{\partial\sigma_l}{\partial\mu_S}<0. \tag{C.167}$$

Step 3. Let us show that the term in brackets in (C.163) is strictly negative. Note that since $\frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}}$ is the posterior variance conditional on disclosure of any signal, we have

$$\frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}} = \int_{-\infty}^{\infty}(E_R[\omega|\sigma_l]-\omega)^2 g_R(\omega|\sigma_l)d\omega. \tag{C.168}$$

At the same time, $E_R[\omega|\sigma_l]$ is the unique solution to the first-order condition for minimizing the expected loss $\int_{-\infty}^{\infty}(a-\omega)^2 g_R(\omega|\sigma_l)d\omega$ with respect to a. Consequently,

$$\int_{-\infty}^{\infty}(E_R[\omega|\sigma_l]-\omega)^2 g_R(\omega|\sigma_l)d\omega < \int_{-\infty}^{\infty}(E_R[\omega|\varnothing]-\omega)^2 g_R(\omega|\sigma_l)d\omega \tag{C.169}$$

where the inequality is strict due to $E_R[\omega|\sigma_l] \neq E_R[\omega|\sigma_h] = E_R[\omega|\varnothing]$ by (C.146). (C.169) and (C.168) together imply

$$\frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}} - \int_{-\infty}^{\infty}(E_R[\omega|\varnothing]-\omega)^2 g_R(\omega|\sigma_l)d\omega < 0. \tag{C.170}$$

Step 4. (C.170) and (C.167) imply that the right-hand side of (C.163) is strictly positive so that $\frac{dL_R(\sigma_l,\sigma_h)}{d\mu_S} > 0$. Hence, R's expected utility is strictly decreasing in μ_S for $\mu_S > 0$.

Claim 2. *R's expected utility strictly increases in μ_S for $\mu_S < 0$.*

The proof follows by analogous arguments (as well as by symmetry considerations).

Claim 3. *For any μ_S', μ_S'' it holds $E[u_R|\mu_S'] \gtrless E[u_R|\mu_S'']$ if $|\mu_R - \mu_S'| \lesseqgtr |\mu_R - \mu_S''|$.*

Without loss of generality, assume $\mu_S' < \mu_S''$. The claim for the fixed order of μ_S and $\mu_R = 0$, i.e., for the cases $\mu_S' < \mu_S'' \leq 0$ and $0 \leq \mu_S' < \mu_S''$, holds by Claims 1 and 2, and the fact that R's utility is maximized at $\mu_S = 0$ by Proposition 8(i.b). The proof for the last case $\mu_S' < 0 < \mu_S''$ follows by the same argument as in Step 3 of the proof of Proposition 2. ■

C.5.5 Proof of Proposition 9(b)

In what follows, we normalize μ_R to 0 and assume $\mu_S > 0$ without loss of generality.

Let $\gamma_S \geq \gamma_R$.

Step 1. As before, denote R's expected quadratic loss as

$$L_R = E_R[(a-\omega)^2].$$

Let $g_R(\omega)$ denote R's prior density function of ω, and $g_R(\omega|\sigma)$ denote R's posterior density function of ω conditional on σ.

By (C.162),

$$\begin{aligned} L_R(\sigma_l,\sigma_h) &= \varphi\left(F_R(\sigma_l)+1-F_R(\sigma_h)\right)\frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}} \\ &\quad +\varphi\begin{pmatrix} E_R[\omega|\varnothing,\sigma_l,\sigma_h]^2\left(F_R(\sigma_h)-F_R(\sigma_l)\right) \\ -2\gamma_R^2 E_R[\omega|\varnothing,\sigma_l,\sigma_h](f_R(\sigma_l)-f_R(\sigma_h)) \end{pmatrix} \\ &\quad +\varphi\left(\int_{\sigma_l}^{\sigma_h}\int_{-\infty}^{\infty}\omega^2 g_R(\omega|\sigma)d\omega f_R(\sigma)d\sigma\right) \\ &\quad +(1-\varphi)\left(E_R[\omega|\varnothing,\sigma_l,\sigma_h]^2+\int_{-\infty}^{\infty}\omega^2 g_R(\omega)d\omega\right). \end{aligned} \tag{C.171}$$

Taking the derivative of $L_R(\sigma_l,\sigma_h)$ with respect to γ_S and simplifying, we obtain:

$$\frac{dL_R(\sigma_l,\sigma_h)}{d\gamma_S} = \varphi f_R(\sigma_l)\frac{\partial\sigma_l}{\partial\gamma_S}\left(\frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}}-\int_{-\infty}^{\infty}(E_R[\omega|\varnothing]-\omega)^2 g_R(\omega|\sigma_l)d\omega\right). \tag{C.172}$$

In the following steps, we show that the right-hand side is strictly negative.

Step 2. Let us show that $\frac{\partial\sigma_l}{\partial\gamma_S}<0$. By (C.149),

$$\frac{\partial\sigma_l}{\partial\gamma_S} = b_1\frac{\partial\sigma_h}{\partial\gamma_S}+\frac{\partial b_1}{\partial\gamma_S}\sigma_h-\frac{\partial b_2}{\partial\gamma_S}. \tag{C.173}$$

Let us derive $\frac{\partial\sigma_h}{\partial\gamma_S}$. Recall the equilibrium condition (C.154):

$$\Lambda(\sigma_h) := -\varphi\int_{b_1\sigma_h-b_2}^{\sigma_h}(\sigma_h-\sigma)f_R(\sigma)d\sigma-\sigma_h(1-\varphi)=0.$$

Then, by the implicit function theorem,

$$\frac{\partial\sigma_h}{\partial\gamma_S} = -\frac{\partial\Lambda/\partial\gamma_S}{\partial\Lambda/\partial\sigma_h}. \tag{C.174}$$

Taking the partial derivative $\partial\Lambda/\partial\gamma_S$ and simplifying we obtain

$$\partial\Lambda/\partial\gamma_S = \varphi(\sigma_h-\sigma_l)f_R(\sigma_l)\begin{pmatrix} -\frac{4\gamma_S\gamma_R^2\gamma_\varepsilon^2(\gamma_R^2+\gamma_\varepsilon^2)}{(\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2))^2}\sigma_h \\ +\frac{4\gamma_S\gamma_\varepsilon^2(\gamma_R^2+\gamma_\varepsilon^2)(\gamma_R^2+2\gamma_\varepsilon^2)\mu_S}{(\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2))^2} \end{pmatrix} > 0, \tag{C.175}$$

where the inequality follows since $\sigma_h < 0$ by Claim 3 in the proof of Proposition 8. At the same time, in the proof of Claim 5 of the proof of Proposition 8, it has been shown that $\partial\Lambda/\partial\sigma_h < 0$ at the equilibrium value of σ_h. This together with (C.174) and (C.175) implies that $\frac{\partial\sigma_h}{\partial\gamma_S} > 0$. Then, since $b_1 = \frac{\gamma_R^2(\gamma_S^2+\gamma_\varepsilon^2)}{\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)} > 0$, $\frac{\partial\sigma_h}{\partial\gamma_S} > 0$, $\frac{\partial b_1}{\partial\gamma_S} = -\frac{4\gamma_S\gamma_R^2\gamma_\varepsilon^2(\gamma_R^2+\gamma_\varepsilon^2)}{(\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2))^2} < 0$, $\sigma_h < 0$ and $\frac{\partial b_2}{\partial\gamma_S} = -\frac{4\gamma_S\gamma_\varepsilon^2(\gamma_R^2+\gamma_\varepsilon^2)(\gamma_R^2+2\gamma_\varepsilon^2)\mu_S}{(\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2))^2} < 0$, (C.173) yields

$$\frac{\partial\sigma_l}{\partial\gamma_S} > 0. \tag{C.176}$$

Step 3. By the same argument as in Step 3 of the Proof of Proposition 9(a) we obtain

$$\frac{1}{\frac{1}{\gamma_R^2}+\frac{1}{\gamma_\varepsilon^2}} - \int_{-\infty}^{\infty}(E_R[\omega|\varnothing]-\omega)^2 g_R(\omega|\sigma_l)d\omega < 0. \tag{C.177}$$

Step 4. (C.177) and (C.176) imply that the right-hand side of (C.172) is strictly negative so that $\frac{dL_R(\sigma_l,\sigma_h)}{d\gamma_S} < 0$. Hence, R's expected utility is strictly increasing in γ_S for $\gamma_S \geq \gamma_R$. ■

C.5.6 Proof of Proposition 10

Let $\gamma_S \geq \gamma_R$.

In what follows, we show that for any given R's prior means denoted as μ_R' and μ_R'' and fixed R's prior variance, S prefers R with the prior mean closer to μ_S. Throughout the proof, we assume $\mu_R'' < \mu_R' \leq \mu_S$ (the proof for other possible constellations of prior means then follows by symmetry considerations).

Denote the actual disagreement expected ex ante by S by $E_S[\widehat{\Delta}] = E_S[|E_S[\omega|\sigma] - E_R[\omega|d]|]$. The claim is equivalent to showing that $E_S[\widehat{\Delta}]$ strictly decreases with μ_R for any $\mu_R \leq \mu_S$ (so that S prefers an R with a prior mean closer to μ_S).

Claim 1. *Let $\mu_S > \mu_R$. Denote σ_0 the value of σ such that $E_R[\omega|\varnothing] - E_S[\omega|\sigma_0] = 0$. Then, $\widetilde{\sigma} < \sigma_l \leq \sigma_0 < \sigma_h$.*

Proof. As shown in the proof of Claim 1 of Proposition 8, $\sigma_l \in s_2$ and $\sigma_h \in s_3$ where s_2 and s_3 are defined by

- $E_R[\omega|\sigma] < E_S[\omega|\sigma] \leq E_R[\omega|\varnothing]$ for all $\sigma \in s_2$.
- $E_R[\omega|\sigma] < E_R[\omega|\varnothing] < E_S[\omega|\sigma]$ for all $\sigma \in s_3$.

This implies that $E_S[\omega|\sigma_l] \leq E_R[\omega|\varnothing] < E_S[\omega|\sigma_h]$. Then, given that $E_R[\omega|\varnothing] = E_S[\omega|\sigma_0]$ by definition, we obtain $E_S[\omega|\sigma_l] \leq E_S[\omega|\sigma_0] < E_S[\omega|\sigma_h]$ that holds if and only

if $\sigma_l \leq \sigma_0 < \sigma_h$.

Next, for $\gamma_S \geq \gamma_R$, condition $E_R[\omega|\sigma] < E_S[\omega|\sigma]$ for all $\sigma \in s_2$ implies that $\sigma > \widetilde{\sigma}$ (the crossing point of $E_R[\omega|\sigma]$ and $E_S[\omega|\sigma]$) for all $\sigma \in s_2$, including σ_l. Thus, $\widetilde{\sigma} < \sigma_l$.

Claim 2. *Let $\mu_S > \mu_R$. Then, $\sigma_h \leq \mu_R < \mu_S$.*

Proof. Since $\mu_R < \mu_S$ by assumption, it is sufficient to show that $\sigma_h \leq \mu_R$. Assume by contradiction $\sigma_h > \mu_R$. Then, it should hold

$$\begin{aligned} E[\omega|\sigma_h] &> \frac{\varphi(F_R(\sigma_h) - F_R(\sigma_l))}{\varphi(F_R(\sigma_h) - F_R(\sigma_l)) + 1 - \varphi} \int_{\sigma_l}^{\sigma_h} E[\omega|\sigma] \frac{f_R(\sigma)}{F_R(\sigma_h) - F_R(\sigma_l)} d\sigma \\ &\quad + \frac{1-\varphi}{\varphi(F_R(\sigma_h) - F_R(\sigma_l)) + 1 - \varphi} \mu_R \\ &= E_R[\omega|\varnothing], \end{aligned}$$

that contradicts (C.141).

Claim 3. *Let $\mu_S > \mu_R$. Then, $\mu_S > E_R[\omega|\varnothing]$.*

Proof. Note that $E_R[\omega|\varnothing] < \mu_R$ in equilibrium, since $\sigma_l < \sigma_h \leq \mu_R$ by Claim 2, i.e. all non-disclosed signals are lower than μ_R. In turn, $\mu_S > \mu_R$ by assumption that together with $E_R[\omega|\varnothing] < \mu_R$ leads to the claim.

Claim 4. *$E_S[\widehat{\Delta}]$ continuously decreases in μ_R for $\mu_R < \mu_S$.*

Proof. Given the equilibrium disclosure strategy specified by 8, we have:

$$\begin{aligned} & E_S[\widehat{\Delta}] = E_S[|E_S[\omega|\sigma] - E_R[\omega|d]|] \\ = \; & \varphi(F_S(\sigma_h) - F_S(\sigma_l)) \left(\int_{\sigma_l}^{\sigma_h} |E_R[\omega|\varnothing] - E_S[\omega|\sigma]| \frac{f_S(\sigma)}{F_S(\sigma_h) - F_S(\sigma_l)} d\sigma \right) \\ & + \varphi(F_S(\sigma_l) + 1 - F_S(\sigma_h)) \begin{pmatrix} \frac{F_S(\sigma_l)}{F_S(\sigma_l) + 1 - F_S(\sigma_h)} \int_{-\infty}^{\sigma_l} \Delta(\sigma) \frac{f_S(\sigma)}{F_S(\sigma_l)} d\sigma \\ + \frac{1 - F_S(\sigma_h)}{F_S(\sigma_l) + 1 - F_S(\sigma_h)} \int_{\sigma_h}^{\infty} \Delta(\sigma) \frac{f_S(\sigma)}{1 - F_S(\sigma_h)} d\sigma \end{pmatrix} \\ & + (1-\varphi) |\mu_S - E_R[\omega|\varnothing]| \\ = \; & \varphi \left(\begin{array}{c} \int_{\sigma_l}^{\sigma_0} (E_R[\omega|\varnothing] - E_S[\omega|\sigma]) f_S(\sigma) d\sigma \\ + \int_{\sigma_0}^{\sigma_h} (E_S[\omega|\sigma] - E_R[\omega|\varnothing]) f_S(\sigma) d\sigma \end{array} + \begin{array}{c} \int_{-\infty}^{\widetilde{\sigma}} -(k_1\sigma + k_2) f_S(\sigma) d\sigma \\ + \int_{\widetilde{\sigma}}^{\sigma_l} (k_1\sigma + k_2) f_S(\sigma) d\sigma \\ + \int_{\sigma_h}^{\infty} (k_1\sigma + k_2) f_S(\sigma) d\sigma \end{array} \right) \quad \text{(C.178)} \\ & + (1-\varphi)(\mu_S - E_R[\omega|\varnothing]), \quad \text{(C.179)} \end{aligned}$$

where to derive absolute values we used Claim 1, Claim 3, and (C.18).

We need to show that $\frac{dE_S[\widehat{\Delta}]}{d\mu_R} < 0$ for any $\mu_R < \mu_S$. Taking the derivative and

simplifying we obtain

$$\begin{aligned}
&\frac{dE_S[\widehat{\Delta}]}{d\mu_R} \\
= \; & -\varphi\frac{\gamma_\varepsilon^2}{\gamma_\varepsilon^2+\gamma_R^2}\left(\int_{\sigma_h}^{\infty} f_S(\sigma)d\sigma - \int_{-\infty}^{\widetilde{\sigma}} f_S(\sigma)d\sigma + \int_{\widetilde{\sigma}}^{\sigma_l} f_S(\sigma)d\sigma\right) \\
& +\varphi\frac{dE_R[\omega|\varnothing]}{d\mu_R}\left(\int_{\sigma_l}^{\sigma_0} f_S(\sigma)d\sigma - \int_{\sigma_0}^{\sigma_h} f_S(\sigma)d\sigma\right) \\
& -(1-\varphi)\frac{dE_R[\omega|\varnothing]}{d\mu_R}. \qquad \text{(C.180)}
\end{aligned}$$

To show that the right-hand side is strictly negative we will prove three claims:

- Claim 4a: $\int_{\sigma_h}^{\infty} f_S(\sigma)d\sigma - \int_{-\infty}^{\widetilde{\sigma}} f_S(\sigma)d\sigma > 0$,
- Claim 4b: $\int_{\sigma_l}^{\sigma_0} f_S(\sigma)d\sigma - \int_{\sigma_0}^{\sigma_h} f_S(\sigma)d\sigma < 0$,
- Claim 4c: $\frac{dE_R[\omega|\varnothing]}{d\mu_R} > 0$.

Claim 4a. $\int_{\sigma_h}^{\infty} f_S(\sigma)d\sigma - \int_{-\infty}^{\widetilde{\sigma}} f_S(\sigma)d\sigma > 0$.

Proof. The claim follows from $\widetilde{\sigma} < \sigma_h < \mu_S$, where the first inequality is by Claim 1, and the second inequality is by Claim 2.

Claim 4b. $\int_{\sigma_l}^{\sigma_0} f_S(\sigma)d\sigma - \int_{\sigma_0}^{\sigma_h} f_S(\sigma)d\sigma < 0$.

Proof. Let us show that

$$\sigma_h - \sigma_0 - (\sigma_0 - \sigma_l) < 0 \qquad \text{(C.181)}$$

By definition of σ_0,

$$E_R[\omega|\varnothing] - E_S[\omega|\sigma_0] = 0.$$

Since $E_R[\omega|\varnothing] = E_R[\omega|\sigma_h]$ by (C.141), we have

$$E_R[\omega|\sigma_h] - E_S[\omega|\sigma_0] = 0.$$

Substituting expressions for $E_R[\omega|\sigma_h]$, $E_S[\omega|\sigma_0]$ and solving for σ_0 yields

$$\sigma_0 = \frac{1}{\gamma_S^2}\left(\frac{(\gamma_S^2+\gamma_\varepsilon^2)(\sigma_h\gamma_R^2+\gamma_\varepsilon^2\mu_R)}{\gamma_R^2+\gamma_\varepsilon^2} - \gamma_\varepsilon^2\mu_S\right). \qquad \text{(C.182)}$$

Substituting (C.182) for σ_0 and (C.149) for σ_l into (C.181) and simplifying we get

$$\sigma_h - \sigma_0 - (\sigma_0 - \sigma_l) = \frac{(\gamma_S^2-\gamma_R^2)\gamma_\varepsilon^2}{(\gamma_R^2+\gamma_\varepsilon^2)\gamma_S^2}(\sigma_h - \sigma_l) > 0. \qquad \text{(C.183)}$$

Note that $\sigma_l \leq \sigma_0 < \sigma_h < \mu_S$ by Claims 1 and 2 so that $f_S(\sigma)$ strictly increases on $[\sigma_l, \sigma_h]$. This together with (C.183) implies $\int_{\sigma_l}^{\sigma_0} f_S(\sigma)d\sigma - \int_{\sigma_0}^{\sigma_h} f_S(\sigma)d\sigma < 0$.

Claim 4c. $\frac{dE_R[\omega|\varnothing]}{d\mu_R} > 0$.

Proof. By (C.141),

$$\begin{aligned}\frac{dE_R[\omega|\varnothing]}{d\mu_R} &= \frac{dE_R[\sigma_h|\omega]}{d\mu_R} \\ &= \left(\frac{1}{1+\frac{\gamma_\varepsilon^2}{\gamma_R^2}}\right)\frac{d\sigma_h}{d\mu_R} + \left(\frac{1}{1+\frac{\gamma_R^2}{\gamma_\varepsilon^2}}\right). \end{aligned} \tag{C.184}$$

To compute $\frac{d\sigma_h}{d\mu_R}$ let us generalize the equilibrium conditions (C.148) and (C.149) for an arbitrary value of μ_R. Performing the same steps as in the proof of Claim 2 of the proof of Proposition 8 yet for general values of μ_R we obtain the following equilibrium condition:

$$\widetilde{\Lambda}(\sigma_h) := -\varphi \int_{\sigma_l}^{\sigma_h} (\sigma_h - \sigma)\, f_R(\sigma)d\sigma - (1-\varphi)\,(\sigma_h - \mu_R) = 0,$$

with σ_l given by

$$\sigma_l = b_1\sigma_h - \widetilde{b}_2, \tag{C.185}$$

where

$$\begin{aligned} b_1 &= \frac{\gamma_R^2(\gamma_S^2+\gamma_\varepsilon^2)}{\gamma_S^2\gamma_R^2 + \gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)}, \\ \widetilde{b}_2 &= \frac{2\gamma_\varepsilon^2((\gamma_R^2+\gamma_\varepsilon^2)\mu_S - (\gamma_S^2+\gamma_\varepsilon^2)\mu_R)}{\gamma_S^2\gamma_R^2 + \gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)} \end{aligned}$$

Then, by the implicit function theorem,

$$\frac{\partial\sigma_h}{\partial\mu_R} = -\frac{\partial\widetilde{\Lambda}/\partial\mu_R}{\partial\widetilde{\Lambda}/\partial\sigma_h}. \tag{C.186}$$

Noting that $\frac{\partial f_l}{\partial\mu_R} = -f_l\frac{\mu_R-\sigma_l}{\gamma_R^2+\gamma_\varepsilon^2}$ by the properties of the pdf of the normal distribution, we obtain

$$\begin{aligned}\partial\widetilde{\Lambda}/\partial\mu_R &= \varphi\frac{2\gamma_\varepsilon^2(\gamma_S^2+\gamma_\varepsilon^2)}{\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)}(\sigma_h-\sigma_l)f_R(\sigma_l) \\ &\quad +\varphi\int_{\sigma_l}^{\sigma_h}(\sigma_h-\sigma)f_R(\sigma)\frac{(\mu_R-\sigma)}{\gamma_R^2+\gamma_\varepsilon^2}d\sigma + (1-\varphi) \\ &> 0, \end{aligned} \tag{C.187}$$

where the inequality follows since $\mu_R - \sigma > 0$ for all $\sigma \in (\sigma_l, \sigma_h)$ by Claim 2.

At the same time, $\partial\widetilde{\Lambda}/\partial\sigma_h < 0$ by analogous derivations as in the proof of Claim 5 of the proof of Proposition 8. This together with (C.186) and (C.187) implies that $\frac{\partial\sigma_h}{\partial\mu_R} > 0$. Consequently, by (C.184), $\frac{dE_R[\omega|\varnothing]}{d\mu_R} > 0$.

Ultimately, Claims 4a, 4b and 4c together imply that the right-hand side of (C.180) is strictly negative so that $\frac{dE_S[\widehat{\Delta}]}{d\mu_R} < 0$.

Claim 5. *$E_S[\widehat{\Delta}]$ is continuous at $\mu_R = \mu_S$.*

Since in equilibrium $\sigma_l \leq \sigma_h$, by (C.185) it should hold (given that $1 - b_1 > 0$)

$$\begin{aligned} & b_1\sigma_h - \widetilde{b}_2 \leq \sigma_h \\ \iff \quad & \sigma_h \geq \frac{-\widetilde{b}_2}{1-b_1} = \frac{\mu_R(\gamma_S^2+\gamma_\varepsilon^2) - \mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_S^2-\gamma_R^2} \end{aligned} \tag{C.188}$$

At the same time, $\sigma_h \leq \mu_R$ by Claim 2. This together with (C.188) yields

$$\mu_R \geq \sigma_h \geq \frac{\mu_R(\gamma_S^2+\gamma_\varepsilon^2) - \mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_S^2-\gamma_R^2}.$$

This implies that as μ_R converges to μ_S, σ_h converges to μ_S. Then,

$$\begin{aligned} \lim_{\mu_R\to\mu_S}\sigma_l &= \lim_{\mu_R\to\mu_S}\left(b_1\sigma_h - \widetilde{b}_2\right) = \lim_{\mu_R\to\mu_S}\left(b_1\mu_S - \widetilde{b}_2\right) \\ &= \lim_{\mu_R\to\mu_S}\left(\frac{\gamma_R^2(\gamma_S^2+\gamma_\varepsilon^2)}{\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)}\mu_S - \frac{2\gamma_\varepsilon^2((\gamma_R^2+\gamma_\varepsilon^2)\mu_S - (\gamma_S^2+\gamma_\varepsilon^2)\mu_R)}{\gamma_S^2\gamma_R^2+\gamma_\varepsilon^2(2\gamma_S^2-\gamma_R^2)}\right) \\ &= \mu_S \end{aligned}$$

Thus, as μ_R converges to μ_S, both σ_l and σ_h continuously converge to μ_S. Then, by (C.178), $E_S[\widehat{\Delta}]$ continuously converges to the value of $E_S[\widehat{\Delta}]$ under full disclosure, which is its value under $\mu_S = \mu_R$ by Proposition 8. Hence, $E_S[\widehat{\Delta}]$ is continuous at $\mu_R = \mu_S$.

Claims 4 and 5 together imply that $E_S[\widehat{\Delta}]$ strictly decreases in μ_R for any $\mu_R \leq \mu_S$, i.e., for any $\mu_R^{'}$ and μ_R'' such that $\mu_R'' < \mu_R' \leq \mu_S$, S obtains a higher expected utility with μ_R'. The claim for other possible constellations of prior means then follows by symmetry considerations. ■

C.6 Information design

C.6.1 Preliminaries

Define the two functions

$$A(\eta) = \widetilde{\Delta}_R(\varnothing|\eta) - \Delta(\widetilde{\sigma}+\eta)$$

and

$$B(\eta) = \frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)} - \frac{f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)}{f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)}, \tag{C.189}$$

where

$$P_i^{\varnothing}(\eta) = \varphi F_i(\widetilde{\sigma}-\eta) + \varphi\left(1-F_i(\widetilde{\sigma}+\eta)\right) + (1-\varphi), \tag{C.190}$$

$$\begin{aligned}\widetilde{\Delta}_R(\varnothing) &= E_R[\Delta(\sigma)\,|d=\varnothing,\eta] \\ &= \left[\frac{\varphi\left[F_R(\widetilde{\sigma}-\eta)\right]}{\varphi\left[F_R(\widetilde{\sigma}-\eta)+(1-F_R(\widetilde{\sigma}+\eta))\right]+1-\varphi}\right] E_R[\Delta(\sigma)\,|\sigma<\widetilde{\sigma}-\eta] \\ &\quad+\left[\frac{\varphi\left[1-F_R(\widetilde{\sigma}+\eta)\right]}{\varphi\left[F_R(\widetilde{\sigma}-\eta)+(1-F_R(\widetilde{\sigma}+\eta))\right]+1-\varphi}\right] E_R[\Delta(\sigma)\,|\sigma>\widetilde{\sigma}+\eta] \\ &\quad+\left[\frac{(1-\varphi)}{\varphi\left[F_R(\widetilde{\sigma}-\eta)+(1-F_R(\widetilde{\sigma}+\eta))\right]+1-\varphi}\right]|\mu_S-\mu_R|\,.\end{aligned} \tag{C.191}$$

Denote also R's perceived disagreement expected ex ante by S conditional on commitment to a disclosure interval $[\widetilde{\sigma}-\eta,\widetilde{\sigma}+\eta]$ as $E_S[\widetilde{\Delta}_R|\eta]$, where

$$\begin{aligned}E_S[\widetilde{\Delta}_R|\eta] &= (1-P_S^{\varnothing}(\eta))\frac{1}{F_S(\widetilde{\sigma}+\eta)-F_S(\widetilde{\sigma}-\eta)}\int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta}\Delta(\sigma)f_S(\sigma)d\sigma \\ &\quad+P_S^{\varnothing}(\eta)\widetilde{\Delta}_R(\varnothing|\eta).\end{aligned} \tag{C.192}$$

Lemma C.12 *It holds true that*

$$\frac{\partial E_S[\widetilde{\Delta}_R|\eta]}{\partial\eta} = \begin{pmatrix}\varphi\left[f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)\right]\frac{P_S^{\varnothing}}{P_R^{\varnothing}} \\ -\varphi\left[f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)\right]\end{pmatrix}\left(\widetilde{\Delta}_R(\varnothing)-\Delta(\widetilde{\sigma}+\eta)\right),$$

so the sign of $\frac{\partial E_S[\widetilde{\Delta}_R|\eta]}{\partial\eta}$ is the same as the sign of $A(\eta)B(\eta)$. I.e., $\frac{\partial E_S[\widetilde{\Delta}_R|\eta]}{\partial\eta}$ is positive if $A(\eta)$ and $B(\eta)$ are either both positive or both negative. $\frac{\partial E_S[\widetilde{\Delta}_R|\eta]}{\partial\eta}$ is negative otherwise.

Proof. Taking the derivative of $E_S[\widetilde{\Delta}_R|\eta]$ we obtain

$$\begin{aligned}\frac{\partial E_S[\widetilde{\Delta}_R|\eta]}{\partial\eta} &= \left[-\varphi f_S(\widetilde{\sigma}-\eta)-\varphi f_S(\widetilde{\sigma}+\eta)\right]\widetilde{\Delta}_R(\varnothing) \\ &\quad+\left[(1-\varphi)+\varphi F_S(\widetilde{\sigma}-\eta)+\varphi\left(1-F_S(\widetilde{\sigma}+\eta)\right)\right]\frac{\partial\widetilde{\Delta}_R(\varnothing)}{\partial\eta} \\ &\quad+\varphi\left[f_S(\widetilde{\sigma}+\eta)\Delta(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)\Delta(\widetilde{\sigma}-\eta)\right].\end{aligned}$$

Noting that $\Delta(\widetilde{\sigma}+\eta) = \Delta(\widetilde{\sigma}-\eta)$ for any η, this implies

$$\begin{aligned}\frac{\partial E_S[\widetilde{\Delta}_R|\eta]}{\partial \eta} &= \left[(1-\varphi)+\varphi F_S(\widetilde{\sigma}-\eta)+\varphi\left(1-F_S(\widetilde{\sigma}+\eta)\right)\right]\frac{\partial \widetilde{\Delta}_R(\varnothing)}{\partial \eta} \\ &\quad -\varphi\left[f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)\right]\left[\widetilde{\Delta}_R(\varnothing)-\Delta(\widetilde{\sigma}+\eta)\right] \qquad \text{(C.193)}\end{aligned}$$

Next, taking the corresponding derivative of the right-hand side of (C.191) and simplifying we obtain

$$\frac{\partial \widetilde{\Delta}_R(\varnothing)}{\partial \eta} = \frac{\varphi\left[f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)\right]}{\varphi\left[F_R(\widetilde{\sigma}-\eta)+(1-F_R(\widetilde{\sigma}+\eta))\right]+1-\varphi}\left[\widetilde{\Delta}_R(\varnothing)-\Delta(\widetilde{\sigma}+\eta)\right]. \qquad \text{(C.194)}$$

Substituting this into (C.193) we get

$$\frac{\partial E_S[\widetilde{\Delta}_R|\eta]}{\partial \eta} = \begin{pmatrix} \varphi\left[f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)\right]\frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)} \\ -\varphi\left[f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)\right] \end{pmatrix}\left[\widetilde{\Delta}_R(\varnothing)-\Delta(\widetilde{\sigma}+\eta)\right]. \qquad \text{(C.195)}$$

Now, note that

$$\begin{aligned} & \varphi\left[f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)\right]\frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)} - \varphi\left[f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)\right] > 0 \\ \iff & \frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)} - \frac{f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)}{f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)} > 0 \\ \iff & B(\eta) > 0. \end{aligned}$$

Hence, by (C.195), the sign of $\frac{\partial E_S[\widetilde{\Delta}_R|\eta]}{\partial \eta}$ corresponds to the sign of $B(\eta)A(\eta)$. ■

Lemma C.13 *There is a unique η_A such that $A(\eta)$ is positive for $\eta < \eta_A$ and negative otherwise.*

Proof. Recall that

$$\Delta(\sigma) = \begin{cases} k_1\sigma + k_2 \text{ for } \sigma \geq \widetilde{\sigma}, \\ -k_1\sigma - k_2 \text{ for } \sigma < \widetilde{\sigma}, \end{cases} \qquad \text{(C.196)}$$

where

$$k_1 = \operatorname{sgn}(\gamma_S - \gamma_R)\left(\frac{\gamma_S^2}{\gamma_S^2+\gamma_\varepsilon^2} - \frac{\gamma_R^2}{\gamma_R^2+\gamma_\varepsilon^2}\right), \qquad \text{(C.197)}$$

$$k_2 = \operatorname{sgn}(\gamma_S - \gamma_R)\left(\frac{\gamma_\varepsilon^2\mu_S}{\gamma_S^2+\gamma_\varepsilon^2} - \frac{\gamma_\varepsilon^2\mu_R}{\gamma_R^2+\gamma_\varepsilon^2}\right). \qquad \text{(C.198)}$$

Then, by (C.194),

$$\begin{aligned}
&\frac{\partial\widetilde{\Delta}_R(\varnothing)}{\partial\eta} - \frac{\partial\Delta(\widetilde{\sigma}+\eta)}{\partial\eta} \\
= \;& \frac{\varphi\left[f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)\right]}{\left(\varphi\left[F_R(\widetilde{\sigma}-\eta)+(1-F_R(\widetilde{\sigma}+\eta))\right]+1-\varphi\right)}\left[\widetilde{\Delta}_R(\varnothing)-\Delta(\widetilde{\sigma}+\eta)\right] \\
& -\text{sgn}\,(\gamma_S-\gamma_R)\left(\frac{\gamma_S^2}{\gamma_S^2+\gamma_\varepsilon^2}-\frac{\gamma_R^2}{\gamma_R^2+\gamma_\varepsilon^2}\right)
\end{aligned}$$

So at the equilibrium value of η such that $\left[\widetilde{\Delta}_R(\varnothing)-\Delta(\widetilde{\sigma}+\eta)\right]=0$, the sign of the derivative is the sign of $\left(-\text{sgn}(\gamma_S-\gamma_R)\left(\frac{\gamma_S^2}{\gamma_S^2+\gamma_\varepsilon^2}-\frac{\gamma_R^2}{\gamma_R^2+\gamma_\varepsilon^2}\right)\right)$, which is always negative, i.e., $A(\eta)$ is decreasing at its root. Moreover, there must exist a unique equilibrium value of η by Proposition 1. ■

Lemma C.14 *a). If $\gamma_S<\gamma_R$, there is a unique η_B such that $B(\eta)$ is negative for $\eta<\eta_B$ and positive otherwise.*

b). If $\gamma_S>\gamma_R$, there is a unique η_B such that $B(\eta)$ is positive for $\eta<\eta_B$ and negative otherwise.

Proof. We have

$$B(\eta)=\frac{1-\varphi+\varphi\left[F_S(\widetilde{\sigma}-\eta)+(1-F_S(\widetilde{\sigma}+\eta))\right]}{1-\varphi+\varphi\left[F_R(\widetilde{\sigma}-\eta)+(1-F_R(\widetilde{\sigma}+\eta))\right]}-\frac{f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)}{f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)}.$$

Then,

$$\begin{aligned}
\frac{\partial B(\eta)}{\partial\eta} \;=\; & \frac{\varphi\left[-f_S(\widetilde{\sigma}-\eta)-f_S(\widetilde{\sigma}+\eta)\right]P_R^{\varnothing}(\eta)-P_S^{\varnothing}(\eta)\varphi\left[-f_R(\widetilde{\sigma}-\eta)-f_R(\widetilde{\sigma}+\eta)\right]}{\left(P_R^{\varnothing}(\eta)\right)^2} \\
& -\partial\left(\frac{f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)}{f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)}\right)/\partial\eta \qquad \text{(C.199)}
\end{aligned}$$

Note that

$$\frac{\varphi\left[-f_S(\widetilde{\sigma}-\eta)-f_S(\widetilde{\sigma}+\eta)\right]P_R^{\varnothing}(\eta)-P_S^{\varnothing}(\eta)\varphi\left[-f_R(\widetilde{\sigma}-\eta)-f_R(\widetilde{\sigma}+\eta)\right]}{\left(P_R^{\varnothing}(\eta)\right)^2}=0$$

is equivalent to

$$\frac{f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)}{f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)}=\frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)}.$$

So by (C.199) whenever $\frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)}-\frac{f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)}{f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)}=0\iff B(\eta)=0$, we have:

$$\left.\frac{\partial B(\eta)}{\partial\eta}\right|_{\eta\text{ s.t. }B(\eta)=0}=-\partial\left(\frac{f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)}{f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)}\right)/\partial\eta.$$

Now, by Lemma C.15 the right-hand side is positive if $\gamma_S < \gamma_R$ whereas it is negative if $\gamma_S > \gamma_R$, i.e., $B(\eta)$ is increasing (decreasing) at its root if $\gamma_S < \gamma_R$ ($\gamma_S > \gamma_R$). By continuity of $B(\eta)$ this implies that whenever the root exists it must be unique.

Let us show that the root of $B(\eta)$ exists. Let $\mu_R = 0$ without loss of generality. We have

$$\begin{aligned} & \lim_{\eta\to 0}\left(\frac{1-\varphi+\varphi\left[F_S(\widetilde{\sigma}-\eta)+(1-F_S(\widetilde{\sigma}+\eta))\right]}{1-\varphi+\varphi\left[F_R(\widetilde{\sigma}-\eta)+(1-F_R(\widetilde{\sigma}+\eta))\right]}-\frac{f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)}{f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)}\right) \\ = \; & 1-\frac{f_S(\widetilde{\sigma})}{f_R(\widetilde{\sigma})}=1-\frac{\sqrt{\gamma_R^2+\gamma_\varepsilon^2}}{\sqrt{\gamma_S^2+\gamma_\varepsilon^2}}e^{-\frac{\mu_S^2}{2(\gamma_S^2-\gamma_R^2)}}\gtrless 0 \text{ iff } \gamma_S\gtrless\gamma_R. \end{aligned} \tag{C.200}$$

At the same time, by (C.202),

$$\begin{aligned} & \lim_{\eta\to\infty}\left(\frac{1-\varphi+\varphi\left[F_S(\widetilde{\sigma}-\eta)+(1-F_S(\widetilde{\sigma}+\eta))\right]}{1-\varphi+\varphi\left[F_R(\widetilde{\sigma}-\eta)+(1-F_R(\widetilde{\sigma}+\eta))\right]}-\frac{f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)}{f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)}\right) \\ = \; & 1-\lim_{\eta\to\infty}\frac{f_S(\widetilde{\sigma}+\eta)+f_S(\widetilde{\sigma}-\eta)}{f_R(\widetilde{\sigma}-\eta)+f_R(\widetilde{\sigma}+\eta)} \\ = \; & 1-\frac{\sqrt{\gamma_R^2+\gamma_\varepsilon^2}}{\sqrt{\gamma_S^2+\gamma_\varepsilon^2}}\lim_{\eta\to\infty}\left(e^{\frac{\eta^2(\gamma_S^2-\gamma_R^2)}{2(\gamma_S^2+\gamma_\varepsilon^2)(\gamma_R^2+\gamma_\varepsilon^2)}-\frac{\mu_S^2}{2(\gamma_S^2-\gamma_R^2)}}\right)\lesseqgtr 0 \text{ if } \gamma_S\gtrless\gamma_R. \end{aligned} \tag{C.201}$$

This together with (C.200) and continuity of $B(\eta)$ implies by the intermediate value theorem that there exists η_B such that $B(\eta_B)=0$. ■

Lemma C.15 *$\partial\left(\frac{f_S(\widetilde{\sigma}+x)+f_S(\widetilde{\sigma}-x)}{f_R(\widetilde{\sigma}+x)+f_R(\widetilde{\sigma}-x)}\right)/\partial x > 0$ if $\gamma_S > \gamma_R$ and $\partial\left(\frac{f_S(\widetilde{\sigma}+x)+f_S(\widetilde{\sigma}-x)}{f_R(\widetilde{\sigma}+x)+f_R(\widetilde{\sigma}-x)}\right)/\partial x < 0$ if $\gamma_S < \gamma_R$.*

Proof. Let $\mu_R = 0$ without loss of generality. Plugging in explicit expressions for normal pdfs and $\widetilde{\sigma}$ and simplifying we obtain

$$\frac{f_S(\widetilde{\sigma}+x)+f_S(\widetilde{\sigma}-x)}{f_R(\widetilde{\sigma}+x)+f_R(\widetilde{\sigma}-x)}=\frac{\sqrt{\gamma_R^2+\gamma_\varepsilon^2}}{\sqrt{\gamma_S^2+\gamma_\varepsilon^2}}e^{\frac{x^2(\gamma_S^2-\gamma_R^2)}{2(\gamma_S^2+\gamma_\varepsilon^2)(\gamma_R^2+\gamma_\varepsilon^2)}-\frac{\mu_S^2}{2(\gamma_S^2-\gamma_R^2)}} \tag{C.202}$$

so that

$$\partial\left(\frac{f_S(\widetilde{\sigma}+x)+f_S(\widetilde{\sigma}-x)}{f_R(\widetilde{\sigma}+x)+f_R(\widetilde{\sigma}-x)}\right)/\partial x=\frac{x(\gamma_S^2-\gamma_R^2)}{(\gamma_S^2+\gamma_\varepsilon^2)^{3/2}\sqrt{\gamma_R^2+\gamma_\varepsilon^2}}e^{\frac{x^2(\gamma_S^2-\gamma_R^2)}{2(\gamma_S^2+\gamma_\varepsilon^2)(\gamma_R^2+\gamma_\varepsilon^2)}-\frac{\mu_S^2}{2(\gamma_S^2-\gamma_R^2)}}. \tag{C.203}$$

The claim follows directly. ■

Lemma C.16 *If S can either commit to FD or not commit, she prefers to commit (not commit) if $\gamma_S < \gamma_R$ ($\gamma_S > \gamma_R$). S is indifferent between FD and no commitment if $\gamma_S = \gamma_R$.*

Proof.

Step 1. Let $\gamma_S = \gamma_R$. Then, by Proposition 1.ii, the equilibrium under no commitment is full disclosure. Hence, S is indifferent between no commitment and commitment to full disclosure.

Step 2. Let $\gamma_S \neq \gamma_R$. Without loss of generality, assume $\mu_S \geq \mu_R = 0$.
As before, denote the equilibrium disclosure interval as $[\sigma_l, \sigma_h]$, and R's perceived disagreement conditional on non-disclosure as $\widetilde{\Delta}_R(\varnothing)$. Denote also

$$P_S^{\varnothing} = \Pr_S[d = \varnothing] = \varphi F_S(\sigma_l) + \varphi(1 - F_S(\sigma_h)) + 1 - \varphi$$

(the probability of non-disclosure from S's ex ante perspective). Then, R's perceived disagreement expected ex ante by S conditional on the equilibrium disclosure strategy is:

$$\begin{aligned} E_S^{eq}[\widetilde{\Delta}_R] &= (1 - P_S^{\varnothing}) \frac{1}{F_S(\sigma_h) - F_S(\sigma_l)} \int_{\sigma_l}^{\sigma_h} \Delta(\sigma) f_S(\sigma) d\sigma + P_S^{\varnothing} \widetilde{\Delta}_R(\varnothing) \quad \text{(C.204)} \\ &= \varphi \int_{\sigma_l}^{\sigma_h} \Delta(\sigma) f_S(\sigma) d\sigma + P_S^{\varnothing} \widetilde{\Delta}_R(\varnothing). \quad \text{(C.205)} \end{aligned}$$

At the same time, R's perceived disagreement expected ex ante by S under commitment to full disclosure is:

$$\begin{aligned} E_S^{FD}[\widetilde{\Delta}_R] &= \varphi \int_{-\infty}^{\infty} \Delta(\sigma) f_S(\sigma) d\sigma + (1 - \varphi)\mu \\ &= \varphi \int_{\sigma_l}^{\sigma_h} \Delta(\sigma) f_S(\sigma) d\sigma + \varphi \left(\int_{-\infty}^{\sigma_l} \Delta(\sigma) f_S(\sigma) d\sigma + \int_{\sigma_h}^{\infty} \Delta(\sigma) f_S(\sigma) d\sigma \right) + (1 - \varphi)\mu. \end{aligned}$$

Taking the difference we thus obtain:

$$\begin{aligned} & E_S^{FD}[\widetilde{\Delta}_R] - E_S^{eq}[\widetilde{\Delta}_R] \\ = \; & \varphi \left(\int_{-\infty}^{\sigma_l} \Delta(\sigma) f_S(\sigma) d\sigma + \int_{\sigma_h}^{\infty} \Delta(\sigma) f_S(\sigma) d\sigma \right) + (1 - \varphi)\mu - P_S^{\varnothing} \widetilde{\Delta}_R(\varnothing) \quad \text{(C.206)} \end{aligned}$$

Since in equilibrium $\widetilde{\Delta}_R(\varnothing) = \Delta(\sigma_h)$ by Lemma C.3, the latter equality is equivalent to:

$$\begin{aligned} & E_S^{FD}[\widetilde{\Delta}_R] - E_S^{eq}[\widetilde{\Delta}_R] \\ = \; & \varphi \left(\int_{-\infty}^{\sigma_l} \Delta(\sigma) f_S(\sigma) d\sigma + \int_{\sigma_h}^{\infty} \Delta(\sigma) f_S(\sigma) d\sigma \right) + (1 - \varphi)\mu - P_S^{\varnothing} \Delta(\sigma_h) \\ = \; & \left(\begin{array}{c} -\varphi k_1 F_S(\sigma_l) E_S[\sigma | \sigma \leq \sigma_l] - \varphi F_S(\sigma_l) k_2 + \varphi k_1 (1 - F_S(\sigma_h)) E_S[\sigma | \sigma \geq \sigma_h] \\ + \varphi(1 - F_S(\sigma_h)) k_2 + (1 - \varphi)\Delta_0 \end{array} \right) \\ & - (\varphi(F_S(\sigma_l) + 1 - F_S(\sigma_h)) + 1 - \varphi)(k_1 \sigma_h + k_2). \quad \text{(C.207)} \end{aligned}$$

Substituting expressions for $E_S[\sigma|\sigma \leq \sigma_l]$ and $E_S[\sigma|\sigma \geq \sigma_h]$ from Lemma C.1, and using the fact that $\Delta(\sigma_l) = \Delta(\sigma_h) \Longleftrightarrow -k_1\sigma_l - k_2 = k_1\sigma_h + k_2$, the latter equality simplifies to

$$E_S^{FD}[\widetilde{\Delta}_R] - E_S^{eq}[\widetilde{\Delta}_R] = k_1\tau_S(\mu_S, \gamma_S, k_1, k_2, \sigma_l, \sigma_h, \Delta_0), \tag{C.208}$$

where

$$\begin{aligned} & \tau_S(\mu_S, \gamma_S, k_1, k_2, \sigma_l, \sigma_h, \Delta_0) \\ = \; & \varphi \left(\begin{array}{c} F_S(\sigma_l)(\sigma_l - \mu_S) + (\gamma_S^2 + \gamma_\varepsilon^2)(f_S(\sigma_l) + f_S(\sigma_h)) \\ +(1 - F_S(\sigma_h))(\mu_S - \sigma_h) \end{array} \right) - (1 - \varphi) \left(\sigma_h + \frac{k_2 - \Delta_0}{k_1} \right). \end{aligned}$$

We can show the following claims regarding the properties of $\tau_S(\mu_S, \gamma_S, k_1, k_2, \sigma_l, \sigma_h, \Delta_0)$.

Claim 1. $\tau_S(\mu_R, \gamma_R, k_1, k_2, \sigma_l, \sigma_h, \Delta_0) = 0$.

Proof. Follows by Lemma C.4.

Claim 2. $\frac{\partial \tau_S(\mu_S, \gamma_S, k_1, k_2, \sigma_l, \sigma_h, \Delta_0)}{\partial \mu_S} \geq (\leq) 0$ if $\gamma_S > (<)\gamma_R$.

Proof: By the properties of the pdf and cdf of the normal distribution,

$$\frac{\partial F_S(\sigma_i)}{\partial \mu_S} = -f_S(\sigma_i), \tag{C.209}$$

$$\frac{\partial f_S(\sigma_i)}{\partial \mu_S} = -f_S(\sigma_i)\frac{\mu_S - \sigma_i}{\gamma_S^2 + \gamma_\varepsilon^2}. \tag{C.210}$$

Then,

$$\begin{aligned} & \frac{\partial \tau_S(\mu_S, \gamma_S, k_1, k_2, \sigma_l, \sigma_h, \Delta_0)}{\partial \mu_S} \\ = \; & \varphi(-f_S(\sigma_l)(\sigma_l - \mu_S) - F_S(\sigma_l) - f_S(\sigma_l)(\mu_S - \sigma_l) - f_S(\sigma_h)(\mu_S - \sigma_h) \\ & +(1 - F_S(\sigma_h)) + f_S(\sigma_h)(\mu_S - \sigma_h)) \\ = \; & \varphi(1 - F_S(\sigma_l) - F_S(\sigma_h)). \end{aligned} \tag{C.211}$$

Let $\gamma_S > \gamma_R$. Then,

$$F_S(\sigma_h) = F_S(\widetilde{\sigma} + \eta^*) \leq F_S(\mu_S + \eta^*) = 1 - F_S(\mu_S - \eta^*) \leq 1 - F_S(\widetilde{\sigma} - \eta^*) = 1 - F_S(\sigma_l),$$

where the inequalities are due to $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2 + \gamma_\varepsilon^2)}{\gamma_R^2 - \gamma_S^2} \leq 0 \leq \mu_S$ for $\gamma_S > \gamma_R$, and the second equality follows by the symmetry of f_S around μ_S. By the analogous argument, $F_S(\sigma_h) \geq$

$1 - F_S(\sigma_l)$ if $\gamma_S < \gamma_R$. Thus, we obtain

$$1 - F_S(\sigma_l) - F_S(\sigma_h) \geq (\leq) 0 \text{ if } \gamma_S > (<)\gamma_R,$$

which together with (C.211) leads to the claim.

Claim 3. $\frac{\partial \tau_S(\mu_S,\gamma_S,k_1,k_2,\sigma_l,\sigma_h,\Delta_0)}{\partial \gamma_S} > 0$.

Proof: By the properties of the pdf and cdf of the normal distribution,

$$\frac{\partial F_S(\sigma_i)}{\partial \gamma_S} = -f_S(\sigma_i)\frac{\gamma_S}{\gamma_S^2+\gamma_\varepsilon^2}(\sigma_i - \mu_S), \tag{C.212}$$

$$\frac{\partial f_S(\sigma_i)}{\partial \gamma_S} = -f_S(\sigma_i)\frac{\gamma_S(\gamma_S^2+\gamma_\varepsilon^2-(\sigma_i-\mu_S)^2)}{(\gamma_S^2+\gamma_\varepsilon^2)^2}. \tag{C.213}$$

Then,

$$\begin{aligned}
&\frac{\partial \tau_S(\mu_S,\gamma_S,k_1,k_2,\sigma_l,\sigma_h,\Delta_0)}{\partial \gamma_S} \\
= \;& -\varphi f_S(\sigma_l)\frac{\gamma_S}{\gamma_S^2+\gamma_\varepsilon^2}(\sigma_l-\mu_S)^2 + 2\varphi\gamma_S(f_S(\sigma_l)+f_S(\sigma_h)) \\
& -\varphi f_S(\sigma_l)\frac{\gamma_S(\gamma_S^2+\gamma_\varepsilon^2-(\mu_S-\sigma_l)^2)}{\gamma_S^2+\gamma_\varepsilon^2} - \varphi f_S(\sigma_h)\frac{\gamma_S(\gamma_S^2+\gamma_\varepsilon^2-(\mu_S-\sigma_h)^2)}{\gamma_S^2+\gamma_\varepsilon^2} \\
& -\varphi f_S(\sigma_h)\frac{\gamma_S}{\gamma_S^2+\gamma_\varepsilon^2}(\sigma_h-\mu_S)^2 \\
= \;& \varphi\gamma_S\left(f_S(\sigma_l)+f_S(\sigma_h)\right) > 0.
\end{aligned}$$

The above claims imply that for any $\mu_S \geq \mu_R$ and $\gamma_S > \gamma_R$ it holds

$$\tau_S(\mu_S,\gamma_S,k_1,k_2,\sigma_l,\sigma_h,\Delta_0) > \tau_S(\mu_R,\gamma_R,k_1,k_2,\sigma_l,\sigma_h,\Delta_0) = 0,$$

where the inequality is by Claims 2 and 3, and the equality is by Claim 1. At the same time, by the same claims, for any $\mu_S \geq \mu_R$ and $\gamma_S < \gamma_R$ it holds

$$\tau_S(\mu_S,\gamma_S,k_1,k_2,\sigma_l,\sigma_h,\Delta_0) < \tau_S(\mu_R,\gamma_R,k_1,k_2,\sigma_l,\sigma_h,\Delta_0) = 0.$$

Since $E_S^{FD}[\widetilde{\Delta}_R] - E_S^{eq}[\widetilde{\Delta}_R] = k_1\tau_S(\mu_S,\gamma_S,k_1,k_2,\sigma_l,\sigma_h,\Delta_0)$ by (C.208), we obtain that $E_S^{FD}[\widetilde{\Delta}_R] - E_S^{eq}[\widetilde{\Delta}_R] > (<)0$ if $\gamma_S > (<)\gamma_R$, i.e., S prefers to commit to full disclosure if and only if $\gamma_S < \gamma_R$. ■

Lemma C.17 *If S can either commit to ND or not commit, she prefers to commit (not commit) if $\gamma_S > \gamma_R$ ($\gamma_S < \gamma_R$). S is indifferent between ND and no commitment if $\gamma_S = \gamma_R$.*

Proof. In what follows, we let $\mu_S \geq \mu_R = 0$ without loss of generality. Denote R's perceived disagreement expected ex ante by S conditional on equilibrium disclosure and ND as $E_S^{eq}[\widetilde{\Delta}_R]$ and $E_S^{ND}[\widetilde{\Delta}_R]$, respectively. Denote also the equilibrium probability of non-disclosure from player i's ex ante perspective as

$$P_i^{\varnothing} = \Pr_i[d = \varnothing] = \varphi F_i(\sigma_l) + \varphi(1 - F_i(\sigma_h)) + 1 - \varphi$$

for $i = S, R$.

Case 1: $\gamma_S > \gamma_R$.

Claim 1. *A sufficient condition for $E_S^{eq}[\widetilde{\Delta}_R] - E_S^{ND}[\widetilde{\Delta}_R] > 0$ is $\varsigma(x)$ is strictly increasing in x* on $(0, \eta)$, *where*

$$\varsigma(x) := \frac{F_S(\widetilde{\sigma} + x) - F_S(\widetilde{\sigma} - x)}{F_R(\widetilde{\sigma} + x) - F_R(\widetilde{\sigma} - x)}$$

Proof. R's perceived disagreement expected ex ante by S conditional on the equilibrium disclosure strategy is:

$$E_S^{eq}[\widetilde{\Delta}_R] = (1 - P_S^{\varnothing}) \int_{\sigma_l}^{\sigma_h} \Delta(\sigma) \frac{f_S(\sigma)}{F_S(\sigma_h) - F_S(\sigma_l)} d\sigma + P_S^{\varnothing} \widetilde{\Delta}_R(\varnothing). \tag{C.214}$$

At the same time, R's perceived disagreement expected ex ante by S under commitment to no disclosure is:

$$\begin{aligned} E_S^{ND}[\widetilde{\Delta}_R] &= \varphi \int_{-\infty}^{\infty} \Delta(\sigma) f_R(\sigma) d\sigma + (1 - \varphi)\mu \\ &= \varphi \left(\int_{\sigma_l}^{\sigma_h} \Delta(\sigma) f_R(\sigma) d\sigma + \begin{array}{c} \int_{-\infty}^{\sigma_l} \Delta(\sigma) f_R(\sigma) d\sigma \\ + \int_{\sigma_h}^{\infty} \Delta(\sigma) f_R(\sigma) d\sigma \end{array} \right) + (1 - \varphi)\mu \\ &= (1 - P_R^{\varnothing}) \int_{\sigma_l}^{\sigma_h} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\sigma_h) - F_R(\sigma_l)} d\sigma + P_R^{\varnothing} \widetilde{\Delta}_R(\varnothing). \end{aligned}$$

Then, we have

$$
\begin{aligned}
& E_S^{eq}[\widetilde{\Delta}_R] - E_S^{ND}[\widetilde{\Delta}_R] \\
= \; & (1-P_S^{\varnothing}) \int_{\sigma_l}^{\sigma_h} \Delta(\sigma) \frac{f_S(\sigma)}{F_S(\sigma_h)-F_S(\sigma_l)} d\sigma + P_S^{\varnothing} \widetilde{\Delta}_R(\varnothing) \\
& -(1-P_R^{\varnothing}) \int_{\sigma_l}^{\sigma_h} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\sigma_h)-F_R(\sigma_l)} d\sigma - P_R^{\varnothing} \widetilde{\Delta}_R(\varnothing) \\
\geq \; & (1-P_S^{\varnothing}) \int_{\sigma_l}^{\sigma_h} \Delta(\sigma) \frac{f_S(\sigma)}{F_S(\sigma_h)-F_S(\sigma_l)} d\sigma + P_S^{\varnothing} \widetilde{\Delta}_R(\varnothing) \\
& -(1-P_S^{\varnothing}) \int_{\sigma_l}^{\sigma_h} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\sigma_h)-F_R(\sigma_l)} d\sigma - P_S^{\varnothing} \widetilde{\Delta}_R(\varnothing) \\
= \; & (1-P_S^{\varnothing}) \left(\int_{\sigma_l}^{\sigma_h} \Delta(\sigma) \frac{f_S(\sigma)}{F_S(\sigma_h)-F_S(\sigma_l)} d\sigma - \int_{\sigma_l}^{\sigma_h} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\sigma_h)-F_R(\sigma_l)} d\sigma \right),
\end{aligned}
$$

where the inequality is due to $P_R^{\varnothing} < P_S^{\varnothing}$ by Lemma C.11 and $\widetilde{\Delta}_R(\varnothing) = \Delta(\sigma_h) > \int_{\sigma_l}^{\sigma_h} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\sigma_h)-F_R(\sigma_l)} d\sigma$ by Lemma C.3. Hence, to prove that $E_S^{eq}[\widetilde{\Delta}_R] - E_S^{ND}[\widetilde{\Delta}_R] > 0$ under $\gamma_S > \gamma_R$ it is sufficient to show that

$$
\int_{\sigma_l}^{\sigma_h} \Delta(\sigma) \frac{f_S(\sigma)}{F_S(\sigma_h)-F_S(\sigma_l)} d\sigma > \int_{\sigma_l}^{\sigma_h} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\sigma_h)-F_R(\sigma_l)} d\sigma. \tag{C.215}
$$

In what follows, we prove this inequality. Denote $\delta(\sigma) = |\sigma - \widetilde{\sigma}|$ (the distance of signal σ from the 0-disagreement signal). Then, for $i = S, R$ we can rewrite

$$
\int_{\sigma_l}^{\sigma_h} \Delta(\sigma) \frac{f_i(\sigma)}{F_i(\sigma_h)-F_i(\sigma_l)} d\sigma = \int_0^{\eta} \Delta(\widetilde{\sigma}+\delta) dG_i(\delta),
$$

where $G_i(\delta)$ is the cdf of $\delta(\sigma)$ conditional on $\sigma \in [\sigma_l, \sigma_h]$ from the perspective of i. Then, (C.215) holds if and only if

$$
\int_0^{\eta} \Delta(\widetilde{\sigma}+\delta) dG_S(\delta) > \int_0^{\eta} \Delta(\widetilde{\sigma}+\delta) dG_R(\delta).
$$

Since $\Delta(\widetilde{\sigma}+\delta)$ is increasing in δ for $\delta > 0$, a sufficient condition for the latter inequality is that the distribution of δ from the perspective of S first-order stochastically dominates the distribution of δ from the perspective of R, i.e.

$$
G_S(\delta) < G_R(\delta) \text{ for any } \delta \in (0, \eta). \tag{C.216}
$$

We have

$$
\begin{aligned}
G_i(x) &= \Pr[\delta(\sigma) \leq x | \sigma \in [\sigma_l, \sigma_h]] \\
&= \Pr[\sigma \in [\widetilde{\sigma}-x, \widetilde{\sigma}+x] | \sigma \in [\sigma_l, \sigma_h]] \\
&= \frac{F_i(\widetilde{\sigma}+x) - F_i(\widetilde{\sigma}-x)}{F_i(\widetilde{\sigma}+\eta) - F_i(\widetilde{\sigma}-\eta)}.
\end{aligned}
$$

Thus, to prove (C.216) we need to show that for any $\delta \in (0, \eta)$,

$$\begin{aligned} & \frac{F_S(\widetilde{\sigma}+\delta) - F_S(\widetilde{\sigma}-\delta)}{F_S(\widetilde{\sigma}+\eta) - F_S(\widetilde{\sigma}-\eta)} < \frac{F_R(\widetilde{\sigma}+\delta) - F_R(\widetilde{\sigma}-\delta)}{F_R(\widetilde{\sigma}+\eta) - F_R(\widetilde{\sigma}-\eta)} \\ \iff & \frac{F_S(\widetilde{\sigma}+\delta) - F_S(\widetilde{\sigma}-\delta)}{F_R(\widetilde{\sigma}+\delta) - F_R(\widetilde{\sigma}-\delta)} < \frac{F_S(\widetilde{\sigma}+\eta) - F_S(\widetilde{\sigma}-\eta)}{F_R(\widetilde{\sigma}+\eta) - F_R(\widetilde{\sigma}-\eta)}. \end{aligned}$$

A sufficient condition for this to hold is that function

$$\varsigma(x) := \frac{F_S(\widetilde{\sigma}+x) - F_S(\widetilde{\sigma}-x)}{F_R(\widetilde{\sigma}+x) - F_R(\widetilde{\sigma}-x)}$$

is strictly increasing in x on $(0, \eta)$.

Claim 2: *$\varsigma(x)$ is strictly increasing in x on $(0, \eta)$.*

Proof. We have

$$\frac{\partial \varsigma(x)}{\partial x} = \frac{1}{(F_R(\widetilde{\sigma}+x) - F_R(\widetilde{\sigma}-x))^2} \begin{pmatrix} (f_S(\widetilde{\sigma}+x) + f_S(\widetilde{\sigma}-x))(F_R(\widetilde{\sigma}+x) - F_R(\widetilde{\sigma}-x)) \\ -(F_S(\widetilde{\sigma}+x) - F_S(\widetilde{\sigma}-x))(f_R(\widetilde{\sigma}+x) + f_R(\widetilde{\sigma}-x)) \end{pmatrix}.$$

Consequently,

$$\frac{\partial \varsigma(x)}{\partial x} > 0 \text{ if and only if } h(x) > 0 \tag{C.217}$$

where

$$h(x) := \frac{f_S(\widetilde{\sigma}+x) + f_S(\widetilde{\sigma}-x)}{f_R(\widetilde{\sigma}+x) + f_R(\widetilde{\sigma}-x)} - \frac{F_S(\widetilde{\sigma}+x) - F_S(\widetilde{\sigma}-x)}{F_R(\widetilde{\sigma}+x) - F_R(\widetilde{\sigma}-x)}.$$

Let us show that $h(x) > 0$ for any $x \in (0, \eta)$. Assume by contradiction that there exists $x' \in (0, \eta)$ such that $h(x') \leq 0$. Then, there can be only two cases:

- *Case 1*: $\frac{\partial h(x')}{\partial x} \leq 0$.

  Then,

  $$\begin{aligned} & \frac{\partial h(x')}{\partial x} \leq 0 \\ \iff & \partial \left(\frac{f_S(\widetilde{\sigma}+x') + f_S(\widetilde{\sigma}-x')}{f_R(\widetilde{\sigma}+x') + f_R(\widetilde{\sigma}-x')} \right) / \partial x - \partial \left(\frac{F_S(\widetilde{\sigma}+x') - F_S(\widetilde{\sigma}-x')}{F_R(\widetilde{\sigma}+x') - F_R(\widetilde{\sigma}-x')} \right) / \partial x \leq 0 \\ \iff & \partial \left(\frac{f_S(\widetilde{\sigma}+x') + f_S(\widetilde{\sigma}-x')}{f_R(\widetilde{\sigma}+x') + f_R(\widetilde{\sigma}-x')} \right) / \partial x - \frac{\partial \varsigma(x')}{\partial x} \leq 0. \end{aligned}$$

  We obtain a contradiction, since $\partial \left(\frac{f_S(\widetilde{\sigma}+x') + f_S(\widetilde{\sigma}-x')}{f_R(\widetilde{\sigma}+x') + f_R(\widetilde{\sigma}-x')} \right) / \partial x > 0$ by Lemma C.15, while $\frac{\partial \varsigma(x')}{\partial x} \leq 0$ due to $h(x') \leq 0$ and (C.217).

- *Case 2*: $\frac{\partial h(x')}{\partial x} > 0$.

Note that

$$\begin{aligned}\lim_{x\to 0} h(x) &= \lim_{x\to 0}\frac{f_S(\widetilde{\sigma}+x)+f_S(\widetilde{\sigma}-x)}{f_R(\widetilde{\sigma}+x)+f_R(\widetilde{\sigma}-x)} - \lim_{x\to 0}\frac{F_S(\widetilde{\sigma}+x)-F_S(\widetilde{\sigma}-x)}{F_R(\widetilde{\sigma}+x)-F_R(\widetilde{\sigma}-x)}\\ &= \lim_{x\to 0}\frac{f_S(\widetilde{\sigma}+x)+f_S(\widetilde{\sigma}-x)}{f_R(\widetilde{\sigma}+x)+f_R(\widetilde{\sigma}-x)} - \lim_{x\to 0}\frac{f_S(\widetilde{\sigma}+x)+f_S(\widetilde{\sigma}-x)}{f_R(\widetilde{\sigma}+x)+f_R(\widetilde{\sigma}-x)}\\ &= 0,\end{aligned}$$

where the second equality is by L'Hopital's rule. Then, by continuity of $h(x)$ at any $x > 0$ there must exist $x'' \in (0, x')$ such that both $h(x'') \leq 0$ and $\frac{\partial h(x'')}{\partial x} \leq 0$. The latter again leads to contradiction by the same argument as for x' in Case 1.

Thus, $h(x) > 0$ for any $x \in (0, \eta)$ that together with (C.217) leads to the claim.

Claims 1 and 2 together imply that $E_S^{eq}[\widetilde{\Delta}_R] - E_S^{ND}[\widetilde{\Delta}_R] > 0$ if $\gamma_S > \gamma_R$, i.e., S prefers to commit to no disclosure in this case.

Case 2: $\gamma_S < \gamma_R$.

The proof proceed analogously and is omitted.

Case 3: $\gamma_S = \gamma_R$.

In this case, (C.18) implies that at any signal σ the disagreement function is equal to the constant $\frac{\gamma_\varepsilon^2 \mu_S}{\gamma_R^2+\gamma_\varepsilon^2}$. Hence, commitment to no disclosure results in the following expected perceived disagreement for S:

$$E_S[\widetilde{\Delta}_R|\gamma_S = \gamma_R] = (1-\varphi)\mu_S + \varphi\frac{\gamma_\varepsilon^2\mu_S}{\gamma_R^2+\gamma_\varepsilon^2}. \tag{C.218}$$

Yet, this is also the expected perceived disagreement after full disclosure that is the equilibrium outcome under $\gamma_S = \gamma_R$ by Proposition 1.ii. Hence, S is indifferent between no commitment (resulting in full disclosure) and commitment to no disclosure. ■

C.6.2 Proof of Proposition 11

As before, denote R's perceived disagreement expected ex ante by S conditional on equilibrium disclosure, FD and ND as $E_S^{eq}[\widetilde{\Delta}_R]$, $E_S^{FD}[\widetilde{\Delta}_R]$ and $E_S^{ND}[\widetilde{\Delta}_R]$, respectively.

Let $\gamma_S < \gamma_R$. Then, $E_S^{ND}[\widetilde{\Delta}_R] > E_S^{eq}[\widetilde{\Delta}_R] > E_S^{FD}[\widetilde{\Delta}_R]$, where the first inequality is by C.17 and the second one is by C.16. Hence, S prefers commitment to FD out of the three

possible commitment options in this case. Analogously, if $\gamma_S > \gamma_R$, then by the same lemmas $E_S^{ND}[\widetilde{\Delta}_R] < E_S^{eq}[\widetilde{\Delta}_R] < E_S^{FD}[\widetilde{\Delta}_R]$, i.e., commitment to ND is preferred. Finally, by the same lemmas we have indifference $E_S^{ND}[\widetilde{\Delta}_R] = E_S^{eq}[\widetilde{\Delta}_R] = E_S^{FD}[\widetilde{\Delta}_R]$ if $\gamma_S = \gamma_R$. ■

C.6.3 Proof of Proposition 12

Let $\gamma_S \neq \gamma_R$. By Lemma C.12, the derivative of $E_S[\widetilde{\Delta}_R|\eta]$ can switch sign only at the roots of $A(\eta)$ and $B(\eta)$. These roots, η_A and η_B, exist and are unique by Lemmas C.13 and C.14. Hence, $E_S[\widetilde{\Delta}_R|\eta]$ has only two local extrema at η_A and η_B (if $\eta_A = \eta_B$ then no local extrema exist). Note that η_A is excluded as a possible solution since it is preferred by either full disclosure (FD) or no disclosure (ND) by Proposition 11. Consequently, $E_S[\widetilde{\Delta}_R|\eta]$ is minimized either at η_B or a corner solution (ND or FD). If $\gamma_S > \gamma_R$, then the only possible corner solution is ND (since S prefers η_A over FD by Lemma C.16). If $\gamma_S < \gamma_R$, then the only possible corner solution is FD (since S prefers η_A over ND by Lemma C.17).[18]

Finally, the root condition for η_B may be rewritten as:

$$
\begin{aligned}
& B(\eta_B) = 0 \\
\iff \quad & \frac{1-\varphi+\varphi\left[F_S(\widetilde{\sigma}-\eta_B)+(1-F_S(\widetilde{\sigma}+\eta_B))\right]}{1-\varphi+\varphi\left[F_R(\widetilde{\sigma}-\eta_B)+(1-F_R(\widetilde{\sigma}+\eta_B))\right]} - \frac{f_S(\widetilde{\sigma}+\eta_B)+f_S(\widetilde{\sigma}-\eta_B)}{f_R(\widetilde{\sigma}-\eta_B)+f_R(\widetilde{\sigma}+\eta_B)} = 0 \\
\iff \quad & \frac{f_S(\widetilde{\sigma}+\eta_B)+f_S(\widetilde{\sigma}-\eta_B)}{1-\varphi+\varphi\left[F_S(\widetilde{\sigma}-\eta_B)+(1-F_S(\widetilde{\sigma}+\eta_B))\right]} = \frac{f_R(\widetilde{\sigma}-\eta_B)+f_R(\widetilde{\sigma}+\eta_B)}{1-\varphi+\varphi\left[F_R(\widetilde{\sigma}-\eta_B)+(1-F_R(\widetilde{\sigma}+\eta_B))\right]} \\
\iff \quad & \frac{P'_{\varnothing,S}(\eta_B)}{P_{\varnothing,S}(\eta_B)} = \frac{P'_{\varnothing,R}(\eta_B)}{P_{\varnothing,R}(\eta_B)},
\end{aligned}
$$

where $P_{\varnothing,i}(x) = \varphi F_i(\widetilde{\sigma}-x) + \varphi(1-F_i(\widetilde{\sigma}+x)) + 1 - \varphi$ for $i = S, R$. ■

C.6.4 Proof of Proposition 13(a)

Let $\gamma_S < \gamma_R$. By Proposition 12 there can be only two possible solutions to the optimal commitment problem: $\widehat{\eta}$ defined by (14) and FD. Denote R's perceived disagreement expected ex ante by S conditional on disclosure interval $[\widetilde{\sigma}-\widehat{\eta},\widetilde{\sigma}+\widehat{\eta}]$ as $E_S[\widetilde{\Delta}_R|\widehat{\eta}]$. Denote R's perceived disagreement expected ex ante by S conditional on FD as $E_S^{FD}[\widetilde{\Delta}_R]$.

[18] One can construct numerical examples showing that both η_B and ND are possible as optimal commitments if $\gamma_S > \gamma_R$, and both η_B and FD are possible as optimal commitments if $\gamma_S < \gamma_R$, depending on parameter values.

Let us show that if $|\mu_S - \mu_R|$ is sufficiently small, then $E_S[\widetilde{\Delta}_R|\widehat{\eta}] > E_S^{FD}[\widetilde{\Delta}_R]$.

Claim 1. *$E_S[\widetilde{\Delta}_R|\widehat{\eta}] > E_S^{FD}[\widetilde{\Delta}_R]$ if and only if $\widetilde{\Delta}_R(\varnothing|\widehat{\eta}) - \widetilde{\Delta}_S(\varnothing|\widehat{\eta}) > 0$, where*

$$\begin{aligned}\widetilde{\Delta}_i(\varnothing|\widehat{\eta}) &= \left[\frac{\varphi\left[F_i(\widetilde{\sigma}-\widehat{\eta})\right]}{\varphi\left[F_i(\widetilde{\sigma}-\widehat{\eta})+(1-F_i(\widetilde{\sigma}+\widehat{\eta}))\right]+1-\varphi}\right] E_i[\Delta(\sigma)\,|\sigma<\widetilde{\sigma}-\widehat{\eta}] \\ &+\left[\frac{\varphi\left[1-F_i(\widetilde{\sigma}+\widehat{\eta})\right]}{\varphi\left[F_i(\widetilde{\sigma}-\widehat{\eta})+(1-F_i(\widetilde{\sigma}+\widehat{\eta}))\right]+1-\varphi}\right] E_i[\Delta(\sigma)\,|\sigma>\widetilde{\sigma}+\widehat{\eta}] \\ &+\left[\frac{(1-\varphi)}{\varphi\left[F_i(\widetilde{\sigma}-\widehat{\eta})+(1-F_i(\widetilde{\sigma}+\widehat{\eta}))\right]+1-\varphi}\right] |\mu_S-\mu_R|\,. \qquad \text{(C.219)}\end{aligned}$$

for $i = S, R$.

Proof. We have

$$\begin{aligned}&E_S[\widetilde{\Delta}_R|\widehat{\eta}] \\ =\;& (1-P_S^{\varnothing}(\widehat{\eta}))\frac{1}{F_S(\widetilde{\sigma}+\widehat{\eta})-F_S(\widetilde{\sigma}-\widehat{\eta})}\int_{\widetilde{\sigma}-\widehat{\eta}}^{\widetilde{\sigma}+\widehat{\eta}}\Delta(\sigma)f_S(\sigma)d\sigma + P_S^{\varnothing}(\widehat{\eta})\widetilde{\Delta}_R(\varnothing|\widehat{\eta}) \\ =\;& \varphi\int_{\widetilde{\sigma}-\widehat{\eta}}^{\widetilde{\sigma}+\widehat{\eta}}\Delta(\sigma)f_S(\sigma)d\sigma + P_S^{\varnothing}(\widehat{\eta})\widetilde{\Delta}_R(\varnothing|\widehat{\eta}),\end{aligned}$$

where $P_S^{\varnothing}(\widehat{\eta})$ is the ex ante probability of non-disclosure from S's perspective, i.e.

$$P_S^{\varnothing}(\widehat{\eta}) = \varphi F_S(\widetilde{\sigma}-\widehat{\eta}) + \varphi(1-F_S(\widetilde{\sigma}+\widehat{\eta})) + 1 - \varphi.$$

At the same time,

$$\begin{aligned}&E_S^{FD}[\widetilde{\Delta}_R] \\ =\;& \varphi\int_{-\infty}^{\infty}\Delta(\sigma)f_S(\sigma)d\sigma + (1-\varphi)\,|\mu_S-\mu_R| \\ =\;& \varphi\int_{\widetilde{\sigma}-\widehat{\eta}}^{\widetilde{\sigma}+\widehat{\eta}}\Delta(\sigma)f_S(\sigma)d\sigma + \varphi\left(\int_{-\infty}^{\widetilde{\sigma}-\widehat{\eta}}\Delta(\sigma)f_S(\sigma)d\sigma + \int_{\widetilde{\sigma}+\widehat{\eta}}^{\infty}\Delta(\sigma)f_S(\sigma)d\sigma\right) \\ &+(1-\varphi)\,|\mu_S-\mu_R|\,.\end{aligned}$$

Taking the difference we thus obtain:

$$\begin{aligned}&E_S[\widetilde{\Delta}_R|\widehat{\eta}] - E_S^{FD}[\widetilde{\Delta}_R] \\ =\;& P_S^{\varnothing}(\widehat{\eta})(\widetilde{\Delta}_R(\varnothing|\widehat{\eta}) - \varphi\left(\int_{-\infty}^{\widetilde{\sigma}-\widehat{\eta}}\Delta(\sigma)f_S(\sigma)d\sigma + \int_{\widetilde{\sigma}+\widehat{\eta}}^{\infty}\Delta(\sigma)f_S(\sigma)d\sigma\right) \\ &-(1-\varphi)\,|\mu_S-\mu_R| \\ =\;& P_S^{\varnothing}(\widehat{\eta})\left[\widetilde{\Delta}_R(\varnothing|\widehat{\eta}) - \widetilde{\Delta}_S(\varnothing|\widehat{\eta})\right], \qquad \text{(C.220)}\end{aligned}$$

which leads to the claim.

Claim 2. *Let* $\mu_S = \mu_R$. *Then,* $\widetilde{\Delta}_R(\varnothing|\widehat{\eta}) - \widetilde{\Delta}_S(\varnothing|\widehat{\eta}) > 0$.

Proof. Let $\mu_S = \mu_R = \mu$. Then, $\widetilde{\sigma} = \frac{\mu(\gamma_R^2+\gamma_\varepsilon^2)-\mu(\gamma_S^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} = \mu$. Hence, the disclosure interval $[\widetilde{\sigma} - \widehat{\eta},\widetilde{\sigma} + \widehat{\eta}]$ is symmetric around μ. Consequently, for $i = S, R$, it holds $1 - F_i(\widetilde{\sigma} + \widehat{\eta}) = F_i(\widetilde{\sigma} - \widehat{\eta})$. Besides, given that $\Delta(\sigma)$ is also symmetric around $\widetilde{\sigma} = \mu$, it is true that $E_i[\Delta(\sigma)\,|\sigma < \widetilde{\sigma} - \widehat{\eta}] = E_i[\Delta(\sigma)\,|\sigma > \widetilde{\sigma} + \widehat{\eta}]$. Then, by (C.219),

$$\widetilde{\Delta}_R(\varnothing|\widehat{\eta}) = E_R[\Delta(\sigma)\,|\sigma > \widetilde{\sigma} + \widehat{\eta}]\frac{2\varphi F_R(\mu - \widehat{\eta})}{2\varphi F_R(\mu - \widehat{\eta}) + 1 - \varphi}, \tag{C.221}$$

$$\widetilde{\Delta}_S(\varnothing|\widehat{\eta}) = E_S[\Delta(\sigma)\,|\sigma > \widetilde{\sigma} + \widehat{\eta}]\frac{2\varphi F_S(\mu - \widehat{\eta})}{2\varphi F_S(\mu - \widehat{\eta}) + 1 - \varphi}. \tag{C.222}$$

Consider both components on the right-hand sides of (C.221) and (C.222). First, for $i = S, R$ by (C.18) we have

$$E_i[\Delta(\sigma)\,|\sigma > \widetilde{\sigma} + \widehat{\eta}] = k_1 E_i[\sigma\,|\sigma > \widetilde{\sigma} + \widehat{\eta}] + k_2.$$

Since for normal distributions $E_i[\sigma\,|\sigma > \widetilde{\sigma} + \widehat{\eta}]$ is increasing in γ_i (see, e.g., Theorem 2 in Horrace, 2015), and $\gamma_S < \gamma_R$ by assumption, we then obtain

$$E_R[\Delta(\sigma)\,|\sigma > \widetilde{\sigma} + \widehat{\eta}] > E_S[\Delta(\sigma)\,|\sigma > \widetilde{\sigma} + \widehat{\eta}]. \tag{C.223}$$

Next, it holds

$$\begin{aligned} & F_R(\mu - \widehat{\eta}) > F_S(\mu - \widehat{\eta}) \\ \iff \quad & \frac{2\varphi F_R(\mu - \widehat{\eta})}{2\varphi F_R(\mu - \widehat{\eta}) + 1 - \varphi} > \frac{2\varphi F_S(\mu - \widehat{\eta})}{2\varphi F_S(\mu - \widehat{\eta}) + 1 - \varphi}, \end{aligned} \tag{C.224}$$

where the first inequality is due to the fact that $\gamma_S < \gamma_R$ by assumption, while for normal cdfs it holds $\frac{\partial F_i(\sigma)}{\partial \gamma_i} = -f_i(\sigma)\frac{\gamma_i}{\gamma_i^2+\gamma_\varepsilon^2}(\sigma - \mu)$ so that $\frac{\partial F_i(\mu-\widehat{\eta})}{\partial \gamma_i} = -f_i(\mu - \widehat{\eta})\frac{\gamma_i}{\gamma_i^2+\gamma_\varepsilon^2}(-\widehat{\eta}) > 0$.

(C.221)-(C.224) together imply $\widetilde{\Delta}_R(\varnothing|\widehat{\eta}) > \widetilde{\Delta}_S(\varnothing|\widehat{\eta})$.

Claim 3. *There exists* μ_S^* *such that for any* $\mu_S \in \{\mu_S' : |\mu_S' - \mu_R| \leq |\mu_S^* - \mu_R|\}$ *it holds* $\widetilde{\Delta}_R(\varnothing|\widehat{\eta}) - \widetilde{\Delta}_S(\varnothing|\widehat{\eta}) > 0$.

Proof. Follows by Claim 2 and continuity of $\widetilde{\Delta}_R(\varnothing|\widehat{\eta}) - \widetilde{\Delta}_S(\varnothing|\widehat{\eta})$ in μ_S (in particular, $\widehat{\eta}$ is continuous in μ_S by the implicit function theorem and (14)).

Claims 1 and 3 jointly imply the claim of Proposition 13(a). ■

C.6.5 Proof of Proposition 13(b)

Let $\gamma_S > \gamma_R$. Let $\mu_S \geq \mu_R = 0$ without loss of generality. Denote $P_i^{\varnothing}(\eta) = \varphi F_i(\widetilde{\sigma} - \eta) + \varphi(1 - F_i(\widetilde{\sigma} + \eta)) + 1 - \varphi$ for $i = S, R$.

Claim 1. *If $\gamma_S > \gamma_R$, then $P_S^{\varnothing}(\eta) > P_R^{\varnothing}(\eta)$ for any $\eta > 0$.*

Proof. We have

$$\frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)} = \frac{1 - \varphi + \varphi\left[F_S(\widetilde{\sigma} - \eta) + (1 - F_S(\widetilde{\sigma} + \eta))\right]}{1 - \varphi + \varphi\left[F_R(\widetilde{\sigma} - \eta) + (1 - F_R(\widetilde{\sigma} + \eta))\right]}.$$

Then,

$$\begin{aligned} & \partial \frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)} / \partial \eta \\ = \; & \frac{\varphi\left[-f_S(\widetilde{\sigma} - \eta) - f_S(\widetilde{\sigma} + \eta)\right] P_R^{\varnothing}(\eta) - P_S^{\varnothing}(\eta)\varphi\left[-f_R(\widetilde{\sigma} - \eta) - f_R(\widetilde{\sigma} + \eta)\right]}{\left(P_R^{\varnothing}(\eta)\right)^2}. \end{aligned}$$

Consequently, $\partial \frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)} / \partial \eta \geq 0$ if and only if $B(\eta) \geq 0$ defined in (C.189). By Lemma C.14, this implies that $\frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)}$ increases in η for $\eta \in (0, \eta_B]$ and decreases for $\eta \in (\eta_B, \infty)$. Together with the fact that $\lim_{\eta \to 0} \frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)} = \lim_{\eta \to \infty} \frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)} = 1$, and continuity of $\frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)}$, this implies that $\frac{P_S^{\varnothing}(\eta)}{P_R^{\varnothing}(\eta)} > 1$ for any $\eta > 0$.

Claim 2. *There exists μ_S^* such that $\mu_S > E_R[\Delta(\sigma)\, \sigma \in [\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]]$ for any $\eta > 0$ and $\mu_S \geq \mu_S^*$.*

Proof. For any $\eta > 0$,

$$\begin{aligned} & \mu_S - E_R[\Delta(\sigma) \,|\sigma \in [\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]] \\ \geq \; & \mu_S - E_R[\Delta(\sigma) \,|\sigma \in [-\infty, \infty]] \\ = \; & \mu_S - \int_{-\infty}^{\infty} \Delta(\sigma) f_R(\sigma) d\sigma, \end{aligned} \tag{C.225}$$

where the inequality follows since $E_R[\Delta(\sigma) \,|\sigma \in [\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]]$ is increasing in η as $\Delta(\sigma)$

increases in $|\sigma - \widetilde{\sigma}|$. At the same time, for the right-hand side we have

$$\begin{aligned}
& \mu_S - \int_{-\infty}^{\infty} \Delta(\sigma) f_R(\sigma) d\sigma \\
= \; & \mu_S - \int_{-\infty}^{\infty} |k_1\sigma + k_2| f_R(\sigma) d\sigma \\
= \; & \mu_S + k_1 \int_{-\infty}^{\widetilde{\sigma}} \sigma f_R(\sigma) d\sigma + k_2 F_R(\widetilde{\sigma}) - k_1 \int_{\widetilde{\sigma}}^{\infty} \sigma f_R(\sigma) d\sigma - k_2(1 - F_R(\widetilde{\sigma})) \\
= \; & \mu_S \left(1 - \frac{\gamma_\varepsilon^2}{\gamma_S^2 + \gamma_\varepsilon^2} \left(1 - 2F_R(\widetilde{\sigma})\right)\right) \\
& + \left(\frac{\gamma_S^2}{\gamma_S^2 + \gamma_\varepsilon^2} - \frac{\gamma_R^2}{\gamma_R^2 + \gamma_\varepsilon^2}\right) \left(\int_{-\infty}^{\widetilde{\sigma}} \sigma f_R(\sigma) d\sigma - \int_{\widetilde{\sigma}}^{\infty} \sigma f_R(\sigma) d\sigma\right). \qquad \text{(C.226)}
\end{aligned}$$

Since for $\gamma_S > \gamma_R$, $\lim_{\mu_S \to \infty} F_R(\widetilde{\sigma}) = \lim_{\mu_S \to \infty} F_R\left(\frac{\mu_S(\gamma_R^2 + \gamma_\varepsilon^2)}{\gamma_R^2 - \gamma_S^2}\right) = 0$, the right-hand side of (C.226) converges to ∞ if $\mu_S \to \infty$. Hence, $\lim_{\mu_S \to \infty} \left(\mu_S - \int_{-\infty}^{\infty} \Delta(\sigma) f_R(\sigma) d\sigma\right) = \infty$. Then, by (C.225), there exists μ_S^* such that $\mu_S > E_R[\Delta(\sigma) | \sigma \in [\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]]$ for any $\eta > 0$ and $\mu_S > \mu_S^*$.

Claim 3. *There exists μ_S^* such that $\widetilde{\Delta}_R(\varnothing|\eta) > \int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\widetilde{\sigma}+\eta) - F_R(\widetilde{\sigma}-\eta)} d\sigma$ for any $\eta > 0$ and $\mu_S > \mu_S^*$.*

Proof. Note that

$$\begin{aligned}
& \widetilde{\Delta}_R(\varnothing|\eta) - \int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\widetilde{\sigma} + \eta) - F_R(\widetilde{\sigma} - \eta)} d\sigma \\
= \; & \widetilde{\Delta}_R(\varnothing|\eta) - E_R[\Delta(\sigma) | \sigma \in [\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]], \qquad \text{(C.227)}
\end{aligned}$$

while

$$\begin{aligned}
\widetilde{\Delta}_R(\varnothing|\eta) \; = \; & \left[\frac{\varphi\left[F_R(\widetilde{\sigma} - \eta)\right]}{\varphi\left[F_R(\widetilde{\sigma} - \eta) + (1 - F_R(\widetilde{\sigma} + \eta))\right] + 1 - \varphi}\right] E_R[\Delta(\sigma) | \sigma < \widetilde{\sigma} - \eta] \\
& + \left[\frac{\varphi\left[1 - F_R(\widetilde{\sigma} + \eta)\right]}{\varphi\left[F_R(\widetilde{\sigma} - \eta) + (1 - F_R(\widetilde{\sigma} + \eta))\right] + 1 - \varphi}\right] E_R[\Delta(\sigma) | \sigma > \widetilde{\sigma} + \eta] \\
& + \left[\frac{(1 - \varphi)}{\varphi\left[F_R(\widetilde{\sigma} - \eta) + (1 - F_R(\widetilde{\sigma} + \eta))\right] + 1 - \varphi}\right] \mu_S.
\end{aligned}$$

Thus, $\widetilde{\Delta}_R(\varnothing|\eta)$ is the weighted average of three components, the first two of them being larger than $E_R[\Delta(\sigma) | \sigma \in [\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]]$ for any $\eta > 0$ since $\Delta(\sigma)$ increases in $|\sigma - \widetilde{\sigma}|$. At the same time, if μ_S is sufficiently large, then the third component, i.e., μ_S is also larger than $E_R[\Delta(\sigma) | \sigma \in [\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]]$ for any $\eta > 0$ by Claim 2. Thus, for sufficiently large μ_S, $\widetilde{\Delta}_R(\varnothing|\eta)$ is larger than $E_R[\Delta(\sigma) | \sigma \in [\widetilde{\sigma} - \eta, \widetilde{\sigma} + \eta]]$ for any $\eta > 0$. This together

with (C.227) leads to the claim.

Claim 4. *Let $E_S[\widetilde{\Delta}_R|\eta]$ be R's perceived disagreement expected ex ante by S conditional on commitment to disclosure interval $[\widetilde{\sigma}-\eta, \widetilde{\sigma}+\eta]$. Then, there exists μ_S^* such that $E_S[\widetilde{\Delta}_R|\eta] - E_S^{ND}[\widetilde{\Delta}_R] > 0$ for any $\eta > 0$ and $\mu_S > \mu_S^*$.*

Proof. Assume μ_S is sufficiently large such that $\widetilde{\Delta}_R(\varnothing) > \int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\widetilde{\sigma}+\eta)-F_R(\widetilde{\sigma}-\eta)} d\sigma$ for any $\eta > 0$ (see Claim 3). Then, for any $\eta > 0$,

$$
\begin{aligned}
& E_S[\widetilde{\Delta}_R|\eta] - E_S^{ND}[\widetilde{\Delta}_R] \\
= \; & (1-P_S^{\varnothing}) \int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta} \Delta(\sigma) \frac{f_S(\sigma)}{F_S(\widetilde{\sigma}+\eta)-F_S(\widetilde{\sigma}-\eta)} d\sigma + P_S^{\varnothing} \widetilde{\Delta}_R(\varnothing) \\
& -(1-P_R^{\varnothing}) \int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\widetilde{\sigma}+\eta)-F_R(\widetilde{\sigma}-\eta)} d\sigma - P_R^{\varnothing} \widetilde{\Delta}_R(\varnothing) \\
\geq \; & (1-P_S^{\varnothing}) \int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta} \Delta(\sigma) \frac{f_S(\sigma)}{F_S(\widetilde{\sigma}+\eta)-F_S(\widetilde{\sigma}-\eta)} d\sigma + P_S^{\varnothing} \widetilde{\Delta}_R(\varnothing) \\
& -(1-P_S^{\varnothing}) \int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\widetilde{\sigma}+\eta)-F_R(\widetilde{\sigma}-\eta)} d\sigma - P_S^{\varnothing} \widetilde{\Delta}_R(\varnothing) \\
= \; & (1-P_S^{\varnothing}) \left(\int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta} \Delta(\sigma) \frac{f_S(\sigma)}{F_S(\widetilde{\sigma}+\eta)-F_S(\widetilde{\sigma}-\eta)} d\sigma - \int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\widetilde{\sigma}+\eta)-F_R(\widetilde{\sigma}-\eta)} d\sigma \right),
\end{aligned}
$$

where the inequality follows due to $P_R^{\varnothing} < P_S^{\varnothing}$ by Claim 1 and $\widetilde{\Delta}_R(\varnothing) > \int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\widetilde{\sigma}+\eta)-F_R(\widetilde{\sigma}-\eta)} d\sigma$ by Claim 3. Hence, to prove that $E_S[\widetilde{\Delta}_R|\eta] - E_S^{ND}[\widetilde{\Delta}_R] > 0$ for any $\eta > 0$ under $\gamma_S > \gamma_R$ and sufficiently large μ_S, it is sufficient to show that

$$
\int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta} \Delta(\sigma) \frac{f_S(\sigma)}{F_S(\widetilde{\sigma}+\eta)-F_S(\widetilde{\sigma}-\eta)} d\sigma > \int_{\widetilde{\sigma}-\eta}^{\widetilde{\sigma}+\eta} \Delta(\sigma) \frac{f_R(\sigma)}{F_R(\widetilde{\sigma}+\eta)-F_R(\widetilde{\sigma}-\eta)} d\sigma. \quad \text{(C.228)}
$$

The rest of the proof follows by the same steps as the proof of Lemma C.17 following (C.215).

Claim 4 finally implies that under sufficiently large μ_S, S prefers to commit to ND over any finite disclosure interval $[\widetilde{\sigma}-\eta, \widetilde{\sigma}+\eta]$. ■

C.7 Receiver disagreement aversion

C.7.1 Proof of Proposition 14

In what follows, we show that for any given S's prior means denoted as μ_S' and μ_S'' and fixed S's prior variance, R prefers S with the prior mean closer to μ_R.

Throughout the proof, we normalize μ_R to 0 without loss of generality and assume $\mu_R \leq \max\{\mu'_S,\mu''_S\}$ (the proof for other possible constellations of prior means then follows by symmetry considerations). In this case, the claim is equivalent to showing that S's perceived disagreement expected ex ante by R $E_R[\widetilde{\Delta}_S(d)]$ strictly increases with μ_S for any $\mu_S \geq \mu_R$ (so that R prefers an S with a prior mean closer to μ_R). In turn, here we separately consider three possible parameter cases: $\gamma_S > \gamma_R$, $\gamma_S < \gamma_R$ and $\gamma_S = \gamma_R$.

Case 1: $\gamma_S > \gamma_R$.
By definition, S's ex post perceived disagreement is

$$\widetilde{\Delta}_S(d) = E_S[\Delta(\sigma)\,|d] = \Delta(\sigma),$$

as there is no uncertainty for S regarding her own signal $\sigma \in \mathbb{R} \cup \varnothing$. Hence, S's perceived disagreement expected ex ante by R is

$$\begin{aligned} & E_R[\widetilde{\Delta}_S(d)] \\ = \;& E_R[\Delta(\sigma)] \\ = \;& \varphi \int_{-\infty}^{\infty} \Delta(\sigma) f_R(\sigma) d\sigma + (1-\varphi)\mu_S. \end{aligned} \tag{C.229}$$

Thus, S's perceived disagreement expected ex ante by R does not depend on the equilibrium disclosure interval.

We need to show that $\frac{dE_R[\widetilde{\Delta}_S]}{d\mu_S} > 0$ for any $\mu_S \geq \mu_R$. We have

$$\begin{aligned} & E_R[\widetilde{\Delta}_S] \\ = \;& \varphi \int_{-\infty}^{\infty} \Delta(\sigma) f_R(\sigma) d\sigma + (1-\varphi)\mu_S \\ = \;& \varphi \left(-k_1 \int_{-\infty}^{\widetilde{\sigma}} \sigma f_R(\sigma) d\sigma - k_2 F_R(\widetilde{\sigma}) \right) \\ & + \varphi \left(k_1 \int_{\widetilde{\sigma}}^{\infty} \sigma f_R(\sigma) d\sigma + k_2 (1 - F_R(\widetilde{\sigma})) \right) \\ & + (1-\varphi)\mu_S \\ = \;& \frac{\gamma_S^2(\mu_S(1-\varphi) + 2\varphi f_R(\widetilde{\sigma})\gamma_\varepsilon^2) + \gamma_\varepsilon^2(\mu_S(1 - 2\varphi F_R(\widetilde{\sigma})) - 2\varphi f_R(\widetilde{\sigma})\gamma_R^2)}{\gamma_S^2 + \gamma_\varepsilon^2}, \end{aligned} \tag{C.230}$$

where the last equality is obtained after substituting $\int_{\widetilde{\sigma}}^{\infty} \sigma f_R(\sigma) d\sigma = -(\gamma_R^2 + \gamma_\varepsilon^2) f_R(\widetilde{\sigma})$ and $\int_{\widetilde{\sigma}}^{\infty} \sigma f_R(\sigma) d\sigma = (\gamma_R^2 + \gamma_\varepsilon^2) f_R(\widetilde{\sigma})$ by Lemma C.1, as well as explicit expressions for k_1 and k_2 from (C.19) and (C.20), and simplifying. Taking the derivative of the right-hand side with respect to μ_S (noting that $\partial f_R(\sigma)/\partial\sigma = -f_R(\sigma)\frac{\sigma}{\gamma_R^2+\gamma_\varepsilon^2}$) and simplifying yields:

$$\frac{dE_R[\widetilde{\Delta}_S]}{d\mu_S} = \frac{(1-\varphi)\gamma_S^2 + \gamma_\varepsilon^2(1-2\varphi F_R(\widetilde{\sigma}))}{\gamma_S^2+\gamma_\varepsilon^2} > 0, \tag{C.231}$$

where the inequality follows from $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} \leq 0 = \mu_R \iff F_R(\widetilde{\sigma}) \leq 0.5$ since $\gamma_S > \gamma_R$. Thus, R prefers an S with a smaller (i.e. closer) mean in the considered parameter case.

Case 2: $\gamma_S < \gamma_R$.

By the same computations (with different values of k_1 and k_2 by (C.19) and (C.20)), we obtain

$$E_R[\widetilde{\Delta}_S] = \frac{\gamma_S^2(\mu_S(1-\varphi) - 2\varphi f_R(\widetilde{\sigma})\gamma_\varepsilon^2) + \gamma_\varepsilon^2(\mu_S(1-2\varphi(1-F_R(\widetilde{\sigma}))) + 2\varphi f_R(\widetilde{\sigma})\gamma_R^2)}{\gamma_S^2+\gamma_\varepsilon^2} \tag{C.232}$$

so that

$$\frac{dE_R[\widetilde{\Delta}_S]}{d\mu_S} = \frac{(1-\varphi)\gamma_S^2 + \gamma_\varepsilon^2(1-2\varphi(1-F_R(\widetilde{\sigma})))}{\gamma_S^2+\gamma_\varepsilon^2} > 0, \tag{C.233}$$

where the inequality follows from $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} \geq 0 = \mu_R \iff F_R(\widetilde{\sigma}) \geq 0.5$ since $\gamma_S < \gamma_R$. Thus, R again prefers an S with a smaller (i.e., closer) mean in the considered parameter case.

Case 3: $\gamma_S = \gamma_R$.

In this case, (C.18) implies that at any signal σ the disagreement function is equal to the constant $\frac{\gamma_\varepsilon^2\mu_S}{\gamma_R^2+\gamma_\varepsilon^2}$. Moreover, by Proposition 1.ii full disclosure is the only equilibrium (so that S's perceived disagreement conditional on non-disclosure is equal to the prior disagreement $|\mu_S - \mu_R| = \mu_S$). Hence,

$$E_R[\widetilde{\Delta}_S|\gamma_S = \gamma_R] = (1-\varphi)\mu_S + \varphi\frac{\gamma_\varepsilon^2\mu_S}{\gamma_R^2+\gamma_\varepsilon^2} \tag{C.234}$$

and

$$\frac{dE_R[\widetilde{\Delta}_S|\gamma_S = \gamma_R]}{d\mu_S} = (1-\dot{\phi}) + \varphi\frac{\gamma_\varepsilon^2}{\gamma_R^2+\gamma_\varepsilon^2} > 0.$$

Thus, R again prefers an S with a smaller (i.e., closer) mean in the considered parameter case. ■

C.7.2 Proof of Proposition 15

a) Let $\mu_S = \mu_R$. Recall that $E_R[\widetilde{\Delta}_S] = E_R[\Delta(\sigma)]$ by (16). Hence, $E_R[\widetilde{\Delta}_S] = 0$ if $\gamma_S = \gamma_R$, as then $\Delta(\sigma) = 0$ for any $\sigma \in \mathbb{R} \cup \varnothing$. At the same time, $E_R[\widetilde{\Delta}_S] = E_R[\Delta(\sigma)] > 0$ if $\gamma_S \neq \gamma_R$, as then $\Delta(\sigma) > 0$ for any $\sigma \in \mathbb{R}$ except for $\sigma = \widetilde{\sigma}$. Consequently, $E_R[\widetilde{\Delta}_S]$ is minimized at $\gamma_S = \gamma_R$.

b) Without loss of generality, throughout the proof we assume $\mu_S > \mu_R$ and normalize μ_R to 0.

Claim 1. *$E_R[\widetilde{\Delta}_S]$ strictly decreases in γ_S if $\gamma_S < \gamma_R$.*

Proof. Let $\gamma_S < \gamma_R$. Then, by (C.232), we have

$$E_R[\widetilde{\Delta}_S] = \frac{\gamma_S^2(\mu_S(1-\varphi) - 2\varphi f_R(\widetilde{\sigma})\gamma_\varepsilon^2) + \gamma_\varepsilon^2(\mu_S(1-2\varphi(1-F_R(\widetilde{\sigma}))) + 2\varphi f_R(\widetilde{\sigma})\gamma_R^2)}{\gamma_S^2 + \gamma_\varepsilon^2}. \tag{C.235}$$

Taking the derivative with respect to γ_S (noting that $\partial f_R(\sigma)/\partial\sigma = -f_R(\sigma)\frac{\sigma}{\gamma_R^2+\gamma_\varepsilon^2}$) and simplifying yields

$$\frac{dE_R[\widetilde{\Delta}_S]}{d\gamma_S} = -\frac{2\gamma_S\gamma_\varepsilon^2\varphi(2f_R(\widetilde{\sigma})(\gamma_R^2+\gamma_\varepsilon^2) + (2F_R(\widetilde{\sigma})-1)\mu_S)}{(\gamma_S^2+\gamma_\varepsilon^2)^2} < 0,$$

where the inequality follows from $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} > 0 = \mu_R \iff F_R(\widetilde{\sigma}) > 0.5$ since $\gamma_S < \gamma_R$.

Claim 2. *$E_R[\widetilde{\Delta}_S]$ is continuous in γ_S at $\gamma_S = \gamma_R$.*

Proof. Note that

$$\begin{aligned}\lim_{\gamma_S\to\gamma_R^-}\widetilde{\sigma} &= \lim_{\gamma_S\to\gamma_R-}\frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} = \infty,\\ \lim_{\gamma_S\to\gamma_R^+}\widetilde{\sigma} &= \lim_{\gamma_S\to\gamma_R+}\frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} = -\infty.\end{aligned}$$

Then, $\lim_{\gamma_S\to\gamma_R^-} f_R(\widetilde{\sigma}) = 0$ and $\lim_{\gamma_S\to\gamma_R^-} F_R(\widetilde{\sigma}) = 1$ so that by (C.232),

$$\begin{aligned}&\lim_{\gamma_S\to\gamma_R^-} E_R[\widetilde{\Delta}_S]\\ =\ &\lim_{\gamma_S\to\gamma_R^-}\frac{\gamma_S^2(\mu_S(1-\varphi) - 2\varphi f_R(\widetilde{\sigma})\gamma_\varepsilon^2) + \gamma_\varepsilon^2(\mu_S(1-2\varphi(1-F_R(\widetilde{\sigma}))) + 2\varphi f_R(\widetilde{\sigma})\gamma_R^2)}{\gamma_S^2+\gamma_\varepsilon^2}\\ =\ &\frac{\gamma_S^2\mu_S(1-\varphi)+\gamma_\varepsilon^2\mu_S}{\gamma_S^2+\gamma_\varepsilon^2}.\end{aligned}$$

At the same time, $\lim_{\gamma_S \to \gamma_R^+} f_R(\widetilde{\sigma}) = 0$ and $\lim_{\gamma_S \to \gamma_R^+} F_R(\widetilde{\sigma}) = 0$ so that by (C.230),

$$\begin{aligned} & \lim_{\gamma_S \to \gamma_R^+} E_R[\widetilde{\Delta}_S] \\ = & \lim_{\gamma_S \to \gamma_R^+} \frac{\gamma_S^2(\mu_S(1-\varphi) + 2\varphi f_R(\widetilde{\sigma})\gamma_\varepsilon^2) + \gamma_\varepsilon^2(\mu_S(1-2\varphi F_R(\widetilde{\sigma})) - 2\varphi f_R(\widetilde{\sigma})\gamma_R^2)}{\gamma_S^2 + \gamma_\varepsilon^2} \\ = & \frac{\gamma_S^2 \mu_S(1-\varphi) + \gamma_\varepsilon^2 \mu_S}{\gamma_S^2 + \gamma_\varepsilon^2}. \end{aligned}$$

Thus, $\lim_{\gamma_S \to \gamma_R^-} E_R[\widetilde{\Delta}_S] = \lim_{\gamma_S \to \gamma_R^+} E_R[\widetilde{\Delta}_S]$ that leads to the claim.

Claim 3. *There exists $\gamma_S^* > \gamma_R$ such that $E_R[\widetilde{\Delta}_S]$ decreases in γ_S on (γ_R, γ_S^*) and increases in γ_S on $[\gamma_S^*, \infty)$.*

Proof. Let $\gamma_S > \gamma_R$. Then, by (C.230)

$$E_R[\widetilde{\Delta}_S] = \frac{\gamma_S^2(\mu_S(1-\varphi) + 2\varphi f_R(\widetilde{\sigma})\gamma_\varepsilon^2) + \gamma_\varepsilon^2(\mu_S(1-2\varphi F_R(\widetilde{\sigma})) - 2\varphi f_R(\widetilde{\sigma})\gamma_R^2)}{\gamma_S^2 + \gamma_\varepsilon^2}.$$

Taking the derivative with respect to γ_S (noting that $\partial f_R(\sigma)/\partial\sigma = -f_R(\sigma)\frac{\sigma}{\gamma_R^2+\gamma_\varepsilon^2}$) and simplifying yields

$$\frac{dE_R[\widetilde{\Delta}_S]}{d\gamma_S} = \frac{2\gamma_S\gamma_\varepsilon^2\varphi(2f_R(\widetilde{\sigma})(\gamma_R^2+\gamma_\varepsilon^2) + (2F_R(\widetilde{\sigma})-1)\mu_S)}{(\gamma_S^2+\gamma_\varepsilon^2)^2}. \tag{C.236}$$

Note that the sign of $\frac{dE_R[\widetilde{\Delta}_S]}{d\gamma_S}$ coincides with the sign of

$$K(\gamma_S) := 2f_R(\widetilde{\sigma})(\gamma_R^2+\gamma_\varepsilon^2) + (2F_R(\widetilde{\sigma})-1)\mu_S.$$

Note also that:

- $K(\gamma_S) < 0$ for γ_S sufficiently close to γ_R (as $\lim_{\gamma_S \to \gamma_R^+} \widetilde{\sigma} = \lim_{\gamma_S \to \gamma_{R^+}} \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} = -\infty$ so that $\lim_{\gamma_S \to \gamma_R^+} K(\gamma_S) = -\mu_S < 0$).
- $K(\gamma_S) > 0$ for γ_S sufficiently large (as $\lim_{\gamma_S \to \infty} \widetilde{\sigma} = \lim_{\gamma_S \to \infty} \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2} = 0 = \mu_R$ so that $\lim_{\gamma_S \to \infty} K(\gamma_S) = 2f_R(\mu_R)(\gamma_R^2+\gamma_\varepsilon^2) > 0$).
- $K(\gamma_S)$ increases in γ_S on (γ_R, ∞) since then $\widetilde{\sigma} = \frac{\mu_S(\gamma_R^2+\gamma_\varepsilon^2)}{\gamma_R^2-\gamma_S^2}$ increases in γ_S from $-\infty$ to 0, while both $f_R(\widetilde{\sigma})$ and $F_R(\widetilde{\sigma})$ increase in $\widetilde{\sigma}$ for $\widetilde{\sigma} \leq 0 = \mu_R$.

Then, by the intermediate value theorem there must exist unique $\gamma_S^* > \gamma_R$ such that $K(\gamma_S) < 0$ for $\gamma_S < \gamma_S^*$, $K(\gamma_S) = 0$ for $\gamma_S = \gamma_S^*$, and $K(\gamma_S) > 0$ for $\gamma_S > \gamma_S^*$. By (C.236)

the same holds for $\frac{dE_R[\widetilde{\Delta}_S]}{d\gamma_S}$, i.e., $E_R[\widetilde{\Delta}_S]$ decreases in γ_S on (γ_R, γ_S^*) and increases in γ_S on $[\gamma_S^*, \infty)$.

Claims 1-3 together lead to claim (b) of the proposition. ■