Disliking to disagree

Florian Hoffmann[*] Kiryl Khalmetski[†] Mark T. Le Quement[‡]

PRELIMINARY AND INCOMPLETE

June 21, 2017

Abstract

Within a simple binary states model, we study strategic information transmission and acquisition between agents holding different priors and who are averse to perceived disagreement in posteriors. In a simple disclosure game, we find that for a given quality of information, full disclosure is feasible only if the receiver's (R) prior is close enough to one minus the sender's (S) prior. If full disclosure is not feasible, then S only fully discloses information congruent with the relative bias of the player whose prior is the most extreme. In a game of cheap talk with costly lying, we find that a more informative signal may lead to more noisy communication and less learning by R. Finally, in a game of voluntary joint exposure to a public signal with a random cost of participation, we find that 1) the player whose prior is most extreme is the most eager to participate, 2) the probability that both agree to participate is maximized when priors are symmetric around $\frac{1}{2}$ and not too extreme and 3) for symmetric priors and perfectly informative signals, the expected ex post disagreement may be locally decreasing in the difference in priors. Finally, we show that the assumed preferences arise endogenously within a simple game of collective decision making by compromise.

Keywords: cheap talk, disclosure, psychological games

JEL classification: D81, D83.

[*]University of Bonn. E-mail: fhoffmann@uni-bonn.de.
[†]University of Cologne. E-mail: kiryl.khalmetski@uni-koeln.de.
[‡]University of East Anglia. E-mail: M.Le-Quement@uea.ac.uk.

1 Introduction

People dislike to disagree. One potential reason is that someone who disagrees with you is someone who actually thereby *thinks that you are wrong*, which most would arguably deem an intrinsically unpleasant state of affairs.[1] Aversion to perceived disagreement in beliefs might instead arise because disagreement threatens the shared identity and cohesion of a group (Akerlof and Kranton, 2000). Disagreement in beliefs might also be disliked for more practical reasons, because it hinders the accomplishment of some valuable collective task, for example by leading different parties to strategically distort their positions. Political cultures in north-western Europe are known for assigning a high value to reaching consensus in negotiations (e.g. the Netherlands and the so-called Polder model of consensual politics, parliamentary politics and labour market negotiations in Scandinavian countries). We refer to Golman et al. (2016) who discuss a wide array of possible psychological and social-psychological mechanisms behind aversion to disagreement.

A second stylized fact is that prior disagreement in beliefs is very often an exogenously given fact of social interaction (Aumann, 1976; Che and Kartik, 2009). A main reason is that individuals have different personal histories involving exposure to different information.

Our analysis sheds light on some key implications for information sharing and acquisition of the above two stylized facts, namely the aversion to (perceived) disagreement in beliefs and the commonality of differences in priors. Our main underlying objective is to disentangle the mechanisms by which prior disagreement translates into posterior disagreement given such preferences. Can individuals who disagree in terms of priors share information, possibly even generate information via debate? Does preexisting disagreement generate more posterior disagreement? We primarily address two more specific questions. First, how completely will the private information held by a party be revealed, depending on the nature of that information (Is it verifiable? How conclusive is it?) and

[1] In psychology, the balance theory of Heider (1946, 1958) and the cognitive dissonance theory of Festinger (1957) posited that conflicts of opinions or attitudes among individuals may lead to psychological stress which they strive to reduce. See Matz and Wood (2005) and Gerber et al. (2012) for empirical evidence regarding psychological costs of revealed disagreement.

the extent of prior conflict between parties? Second, if groups can decide to expose themselves collectively to public information (for example through debates, reading the same book, watching a documentary movie together), how does prior conflict affect the likelihood that individuals will want to participate in such joint exposure?

Strategic information transmission with diverse priors has been examined in Che and Kartik (2009) and Kartik et al. (2015). A main contribution of our paper is to reexamine the above type of game with a new type of preferences, namely aversion to disagreement. We wish to emphasize that these preferences remain entirely unstudied and could as such also be fruitfully applied to other suitably chosen games, such as bargaining or voting games. A second contribution is the introduction of what we term *information exposure* as a new theoretical question. The problem is trivial within the context of most models but becomes interesting given our assumed preferences. A third contribution is our study of extended disclosure games featuring a stage of signaling of prior beliefs before the stage of actual information sharing or generation. The (unstudied) equivalent in a classical cheap talk model would be a game in which the sender, in a first stage, makes a cheap talk announcement concerning her unknown bias.

What we do We consider three types of games, which all build on the same basic features. There are two agents 1 and 2, the state is binary (0 or 1), and the two agents assign different publicly observed prior probabilities α_i, $i = 1, 2$, to the state being 0. A binary informative signal $\sigma \in \{0, 1\}$ known to be correct with probability p is (potentially) available and one or both agents need to make a decision concerning the release or the acquisition of this signal.

We first examine a classical disclosure game. One of the two agents (now called S) holds with positive probability an informative signal of known quality. She can either disclose the signal or omit to do so. Informative equilibrium scenarios involve respectively full disclosure, disclosure of only 0-signal or disclosure of only 1-signals.

The second game that we consider is a game of cheap talk with costly lying (almost cheap talk). Here, S is free to send any of the two available messages at each of her information sets. In equilibrium, she however bears a (psychological) cost of misleading R, i.e. of sending the message that induces the least accurate beliefs at a given information set. A fully revealing equilibrium generically does not exist if the cost of lying is low, in which case we consider instead noisy communication.

In the third and final basic setup that we consider, both agents can decide to jointly generate an informative public signal of known quality, participation being costly. The game captures a situation in which the parties can decide whether or not to have a conversation or watch a documentary movie together, the contents of the conversation or the movie being unpredictable.

Finally, we examine three important extensions in the last section of the paper. The first seeks to provide a foundation for the assumed preferences (aversion to perceived disagreement) within the context of a simple game of collective decision making. We wish to determine whether the expected payoff of participants is influenced by any pre-existing perceived disagreement. The second extension is an analysis of disclosure with continuous signals satisfying the MLRP property. We here address the same questions as in the binary signals environment and wish to test the robustness of our original findings. Our third extension is to study, within the first and the third of the above three setups, the case of prior uncertainty about at least one party's prior. In the extended game that we examine, the information disclosure or exposure stage is preceded by a stage of cheap talk communication during which information regarding priors can potentially be transmitted. This scenario appears empirically relevant: Strangers often test the water before initiating conversations that will involve sharing or generating sensitive information.

Findings Our results for the disclosure game are as follows. For a given quality of information and a given sender prior α_S, there is a most favorable opponent prior $\alpha_R^* = 1 - \alpha_S$ such that full disclosure is feasible only if R's prior is close enough to α_R^*. The higher the quality of information, the less close R's prior needs to be to α_R^* for full disclosure to be feasible. If α_R and α_S are located on different sides of $1 - \alpha_S$ and information quality is not high enough, then the only informative equilibrium is such that S only discloses a signal in line with R's relative prior bias. Instead, if α_R and α_S are located on the same side of $1 - \alpha_S$ and information quality is not high enough, then the only informative equilibrium is such that S only discloses a signal in line with her own relative prior bias. In other words, S only discloses information congruent with the relative bias of the player whose prior is the most extreme (i.e. closest to one of the boundaries of $[0,1]$). We discuss how these insights about partial revelation can be used as a starting point for a theory of echo-chambers.

Our results for the analysis of cheap talk with costly lying are as follows. A higher cost of lying and more aligned priors are as expected conducive to more informative communication. More surprisingly, for a low enough lying cost, there is a closed interval of signal precisions such that more precise information held by S leads to less informative communication. This indirect strategic effect can furthermore be strong enough to overrule the direct positive effect of more informative signals.

In our analysis of costly exposure, we obtain the following main results. First, we find that the player with the most extreme prior is the most eager to engage in joint exposure. Second, the probability that both players accept to engage in joint exposure is maximized by priors that are symmetric around $\frac{1}{2}$ as well as not too extreme. This establishes a sense in which diverse groups can be put to a productive use in terms of gathering information, if we take the perspective of an outside benevolent principal interested in making sure that agents discover the truth. Our third result establishes that for symmetric priors and perfectly informative signals, the expected ex post disagreement may be locally decreasing in the difference in priors. In other words, the positive effect of prior misalignment on the probability of exposure to consensus generating information (from an ex ante perspective) dominates the direct negative effect of higher ex ante disagreement.

We now describe the results obtained in our analysis of extensions. In our analysis of a simple game of collective decision making, we find that the expected payoff of participants is decreasing in perceived prior disagreement. If the game were preceded by a stage of information sharing or exposure, participants would thus be motivated to minimize perceived disagreement as assumed in our main analysis. The preferences assumed throughout our paper thus arise endogenously in such a context. Our analysis of disclosure with continuous signals yields insights that are similar to those obtained for the binary signals environment, thus showing that the findings are robust to the assumed information structure. Only information congruent with the relative prior bias of the most extreme player is fully disclosed. Finally, in the extended version of our disclosure game with signaling about priors, assuming that R is interested in maximizing the information disclosed by S, we find that if R's prior is taken from a uniform distribution over $[0,1]$ and S's prior is known to be $\frac{1}{2}$, there exists no equilibrium in which R truthfully announces her prior and the most informative disclosure equilibrium ensues in the subsequent subgame. In the extended version of the exposure game with signaling about priors, we find that under one-sided as well as two-sided uncertainty about priors, with priors drawn from a uniform distribution over $[0,1]$, there exists an equilibrium in which parties truthfully disclose whether their prior is above or below $\frac{1}{2}$.

Related literature An extensive body of research dating back to Crawford and Sobel (1982) studies information transmission between an informed sender and an uninformed receiver (see Sobel, 2013, for a review). Typically, these models involve an exogenous difference in preferences over the receiver's choices (i.e. a conflict of interest) between the sender and the receiver. As a result, the sender aims to manipulate the receiver's beliefs about the state of the world. In contrast, in our setting the sender manipulates the receiver's beliefs about her own beliefs, i.e., the second-order beliefs of the receiver, which ideally (from the perspective of the sender) should match the receiver's own first-order beliefs.

Mostly relevant for our study, Che and Kartik (2009) consider a verifiable disclosure game between a decision maker and an advisor. In the main specification of the model, both players have identical preferences over the decision maker's choice conditional on a given state realization, yet divergent prior beliefs. As a result, the players have different interim preferences over the decision maker's action (i.e. "opinions") conditional on learning the same signal (since they draw different inferences from the same signal, i.e. form different expectations of the state of the world given the same signal). In turn, this leads to concealment of signals in an intermediate range by the advisor.[2] Unlike in the

[2]The focus of Che and Kartik (2009) is also on how this structure of communication eventually affects the advisor's prior incentives to acquire information. Banerjee and Somanathan (2001) consider a related model where communication takes place between one receiver ("the leader") and multiple senders with heterogeneous priors (which also results in a difference in senders' opinions regarding the most suitable receiver action). Kartik et al. (2015) analyze how disagreement in beliefs between the receiver and a biased sender affects the latter's incentives to disclose or conceal information, and how this is affected by the presence of other senders.

model of Che and Kartik (2009), the sender in our setting is not interested in whether the receiver's posterior belief is correct (i.e., matches the expected state of the world), but instead cares about the receiver's *perception* of the ex-post disagreement in their posterior beliefs (which involves receiver's second-order beliefs as noted above). As a result, the sender may conceal signals even though they would lead to a lower deviation of the receiver's beliefs from what the sender believes is the true state (for instance, by concealing the signals in line with the sender's own prior bias).

Several papers feature some elements of endogenously arising preference for belief conformity. In particular, Ottaviani and Sørensen (2006a,b) consider settings where an expert has an incentive to signal a high quality of her signal to an evaluator who ultimately observes the actual state (i.e. the expert is concerned to achieve a good reputation w.r.t. to the quality of her information). This may lead the expert to bias her message in the direction of her prior belief in order to maximize the chance of predicting the actually realized state. This bias eventually hampers information transmission. Gentzkow and Shapiro (2006) study a similar mechanism where the sender is not always perfectly informed about the state of the world and wishes to be perceived as informed. This again leads to a distortion of the message in the direction of the receiver's prior.[3] Similarly, in our setup the sender sometimes prefers not to reveal signals which contradict the receiver's prior, yet for very different reasons related to mitigating the receiver's perception of ex-post disagreement (while the quality of the sender's signal is common knowledge in our setting).

A number of papers study the emergence of ex-post disagreement in beliefs between players who do not have strategic concerns to mitigate disagreement. As in our case, disagreement may result from different prior beliefs (Dixit and Weibull, 2007; Acemoglu

[3]Sobel (1985) and Morris (2001) study related sender-receiver games with an endogenous reputational concern of the sender for being perceived as unbiased, which also leads to distorted informativeness of communication. Prendergast (1993) in a principal-agent setting examines the agent's incentive to match the (noisy) information of the principal in his report. Bursztyn et al. (2017) consider a setting where a sender has to communicate his type to a receiver and has incentive to appear of the same type as the receiver. Bénabou (2012) shows that agents with anticipatory utility may converge to each other's wrong beliefs due to the dependence of one's payoffs on the actions of the others.

et al., 2007; Sethi and Yildiz, 2012), or different prior (privately observed) signals (Andreoni and Mylovanov, 2012). Notably, under certain conditions such disagreement in beliefs may also persist in the long run, i.e. asymptotically.[4]

Finally, our paper contributes to the growing body of literature on psychological game theory, which considers models where preferences directly incorporate beliefs (of arbitrary order) about others' strategies or beliefs (Geanakoplos et al., 1989; Battigalli and Dufwenberg, 2009). Many applied models in this field focus on preferences which depend on the interplay between beliefs and material payoffs, as in models of reciprocity (Rabin, 1993; Dufwenberg and Kirchsteiger, 2004) or guilt aversion (Battigalli and Dufwenberg, 2007). A related model of Ely et al. (2015) (although formally not belonging to the domain of psychological game theory) considers the behavior of a principal who wishes the beliefs of an agent to follow a specific path over time involving suspense or surprises.

We proceed as follows. Section 1 presents the basic setup that is common to all the games that we study. Section 2 considers strategic disclosure. Section 3 considers a game of cheap talk with a cost of lying. Section 4 considers strategic exposure. Section 5 considers three extensions.

2 Basic setup

The state is denoted ω and the state space is $\{0, 1\}$. There are two agents 1 and 2. Each agent i's prior belief that the state is 0 is denoted α_i. Unless explicitly specified otherwise, we shall assume that these two priors are publicly observed at the beginning of the game. We shall assume that at least one of the two agents is averse to being perceived by her opponent as disagreeing with the latter . We model the aversion as follows. Conditional on knowing that information I_j is available to her opponent j, agent i obtains the following utility:

$$-\left|E_j[E_i[\omega \,|I_i] \,\middle|I_j] - E_j[\omega \,|I_j]\right|. \tag{1}$$

[4]Several studies in network economics consider the effect of individual conformity to the belief/opinions of others on the overall polarization (i.e. disagreement) of beliefs in networks (Dandekar et al., 2013; Buechel et al., 2015; Golub and Jackson, 2012).

The expression decomposes into the following building blocks. Given information I_j, j's own belief about the state is $E_j[\omega \mid I_j]$. Furthermore, j's belief about i's belief regarding the state is given by $E_j[E_i[\omega \mid I_i] \mid I_j]$. Note that j, conditional on I_j, might not be entirely sure as to what information is held by i (i.e. might not be able to pin down I_i), but I_j, combined with knowledge of i's communication strategy (if i disclosed some information), might be very informative regarding I_i. Note that j is commonly known to update beliefs by Bayes' rule (using her prior distribution α_j). Also, if the information observed by j is determined by i (in the case of disclosure or cheap talk), then j is assumed to know the strategy used by i in equilibrium.

Our equilibrium concept throughout is Perfect Bayesian equilibrium. I.e. players' strategies are sequentially rational given their beliefs and others' equilibrium strategies. Second, beliefs are derived via Bayes' rule whenever possible.

We introduce some further notation. If $\alpha_i < \alpha_j, (>)$ we say that i is relatively prior-biased towards state 0 (1) relative to j. Furthermore, if $\alpha_i < \alpha_j, (>)$, we say that a 0-signal is congruent with i's relative prior bias. We say that player i is more extreme than player j if player i's prior is strictly closer to one of the boundaries of $[0, 1]$ than player j's prior, i.e. if $\min\{\alpha_i, 1 - \alpha_i\} < \min\{\alpha_j, 1 - \alpha_j\}$.

3 Strategic disclosure

In this section, we denote agent 1 by S and agent 2 by R. Agent S holds with probability $\varphi \in (0, 1)$ a privately observed signal $\sigma \in \{0, 1\}$. The signal is identical to the state with probability p, i.e. $P(\sigma = \omega) = p, \forall \omega$. Player S can decide to disclose the signal to R or not. R is entirely passive and simply observes S's signal if disclosed and subsequently updates beliefs. The objective function of S is given as in (1), i.e. S cares only about minimizing R's ex post perceived disagreement. The preferences of R are left unspecified as these are inconsequential, R's only role being to update beliefs according to Bayes' rule. In what follows, an equilibrium featuring full-disclosure is called a FD equilibrium. An equilibrium in which only 1-signals are disclosed is called a D1 equilibrium. An equilibrium in which only 0-signals are disclosed is called a D0 equilibrium.

The following proposition offers a complete characterization of pure strategy equilibria in our game.

Proposition 1 ***I. Assume that*** $\alpha_S > \alpha_R$.

a) If $\alpha_R < 1 - \alpha_S$ *(i.e. R is more extreme than S), there exists an FD equilibrium if and only if* $p > P_0(\alpha_S, \alpha_R)$*, where*

$$P_0(\alpha_S, \alpha_R) \equiv \frac{\alpha_S + \alpha_R - \alpha_S\alpha_R - 1}{\alpha_S + \alpha_R - 2\alpha_S\alpha_R - 1}$$

and where $\frac{\partial P_0(\alpha_S,\alpha_R)}{\partial \alpha_S} < 0$, $\frac{\partial P_0(\alpha_S,\alpha_R)}{\partial \alpha_R} < 0$*. If instead* $\alpha_R \geq 1 - \alpha_S$ *(i.e. R is less extreme than S), there exists an FD equilibrium if and only if* $p > P_1(\alpha_S, \alpha_R)$*, where*

$$P_1(\alpha_S, \alpha_R) \equiv \frac{-\alpha_S\alpha_R}{\alpha_S + \alpha_R - 2\alpha_S\alpha_R - 1}$$

and where $\frac{\partial P_1(\alpha_S,\alpha_R)}{\partial \alpha_S} > 0$ *and* $\frac{\partial P_1(\alpha_S,\alpha_R)}{\partial \alpha_R} > 0$.

b) If $\alpha_R < 1 - \alpha_S$ *and* $p < P_0(\alpha_S, \alpha_R)$ *there exists a D1 equilibrium and no D0 equilibrium. In other words, R is only shown evidence congruent with her relative prior bias.*

c) If $\alpha_R \geq 1 - \alpha_S$ *and* $p < P_1(\alpha_S, \alpha_R)$*, there exists a D0 equilibrium and no D1 equilibrium. In other words, R is only shown evidence congruent with* S'*s relative prior bias.*

II. Assume that $\alpha_R = \alpha_S$*. There exists an FD equilibrium.*

III. Assume that $\alpha_R > \alpha_S$.

a) If $\alpha_R < 1 - \alpha_S$ *(i.e. R is less extreme than S), there exists an FD equilibrium if and only if* $p < P_0(\alpha_S, \alpha_R)$*. If instead* $\alpha_R \geq 1 - \alpha_S$ *(i.e. R is more extreme than S), there exists an FD equilibrium if and only if* $p < P_1(\alpha_S, \alpha_R)$.

b) If $\alpha_R < 1 - \alpha_S$ *and* $p < P_0(\alpha_S, \alpha_R)$*, there exists a D1 equilibrium and no D0 equilibrium. In other words, R is only ever shown evidence congruent with* S'*s relative prior bias.*

c) If $\alpha_R \geq 1 - \alpha_S$ *and* $p < P_1(\alpha_S, \alpha_R)$*, there exists a D0 equilibrium and no D1 equilibrium. In other words, R is only ever shown evidence congruent with her relative prior bias.*

Proof: see in Appendix A.

Note that the results of Part III are the mirror image of to those of Part I. The following figure illustrates Proposition 1. We assume $\alpha_S = .7$. The function $P_0(.7, \alpha_R)$ is drawn in continuous, $P_1(.7, \alpha_R)$ is in dashed. There are four relevant areas to be considered in the figure. In the first area, below the downward sloping continuous curve, only a D1 equilibrium exists. In the second area, above the downward sloping continuous curve and the upwards sloping dashed curve, the full disclosure equilibrium exists. In the third area, below the upwards sloping dashed curve, only a D0 equilibrium exists. The fourth area corresponds to $\alpha_R = \alpha_S$, where a full disclosure equilibrium exists no matter p. Note finally that for $\alpha_R = 1 - \alpha_S$, there exists a truth-telling equilibrium for any p.

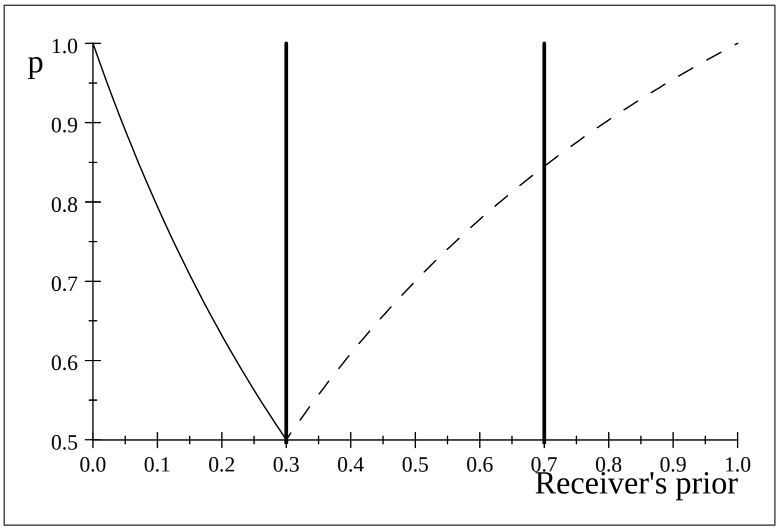

Figure 1

We summarize the qualitative implications of the above characterization in the following corollary of Proposition 1.

Corollary 1 *a) For $p \in \left(\frac{1}{2}, 1\right)$ and a given $\alpha_S \in (0,1)$, for α_R sufficiently high or low, R is exposed only to information congruent with her relative prior bias.*

b) For $p \in \left(\frac{1}{2}, 1\right)$ and a given $\alpha_S \in (0,1)$, a full disclosure equilibrium exists only if either 1) $\alpha_R = \alpha_S$ or 2) α_R is sufficiently close to $1 - \alpha_S$.

The above results follow immediately from our characterization. The argument behind b) is as follows. Let $\underline{p}(\alpha_S, \alpha_R)$ be the minimal value of p such that there exists a truth-telling equilibrium given α_S, α_R. The function $\underline{p}(\alpha_S, \alpha_R)$ is continuous and decreasing in

α_R for α_R such that $\alpha_R \neq \alpha_S$ and $\alpha_R < 1 - \alpha_S$. Instead, $\underline{p}(\alpha_S, \alpha_R)$ is continuous and increasing in α_R for α_R such that $\alpha_R \neq \alpha_S$ and $\alpha_R > 1 - \alpha_S$. Finally, $\underline{p}(\alpha_S, 1 - \alpha_S) = \underline{p}(\alpha_S, \alpha_S) = \frac{1}{2}$.

As a preliminary remark, note that a property which underlies all of our main results is that disclosure of information can lead to increased disagreement between individuals with differing priors. This is already known from Kartik et al. (2015). Subjects always update their prior in the same direction (i.e. no polarization arises) but with different intensities, i.e. update their respective beliefs to different extents. Furthermore, the signal that is susceptible to generate increased disagreement is the one that is congruent with the relative prior bias of the most extreme player.

Our first substantial finding is that more prior misalignment can generate more disclosure, as long as the misalignment is not too extreme. The intuition here is that close (but different priors) induce a strong status quo bias. Consider a putative full disclosure equilibrium and moderate signal quality. For one of the two possible signal realizations, disclosure of S's signal will lead to higher perceived disagreement than pretending to hold no signal and thus sticking to prior disagreement. The key is that prior disagreement being small, it constitutes an attractive default option for S that is difficult to improve upon. If instead, prior disagreement is relatively large, though not extreme, this is not true anymore. The status quo is unattractive and any disclosed signal realization will improve on it. Finally, if prior disagreement is very large, some signal realization increases disagreement despite the already large prior disagreement. The reason is that the very large prior disagreement causes different subjects to update with very different intensities. Our second main finding that a receiver who is more extreme than the sender that she faces is only exposed to information that confirms her relative prior bias, if the quality of information is not high enough. The intuition here is that the signals that generate an increase in disagreement are those contradicting the extreme party's beliefs, as already mentioned.

The above characterization provides an interesting starting point for a theory of echo-chambers, which we expand on in what follows. Consider the following simple dynamic simple model of random encounters. Suppose that R sequentially encounters senders numbered $1, 2, ... + \infty$ over a set of periods $t = 1, 2, ... + \infty$. The prior of R at the beginning of each period-t encounter is denoted α_R^{t-1} and includes all the information received over time until (including) $t-1$. Senders have priors denoted $\alpha_{S,i}$ for each sender, and each sender's prior is randomly drawn from a uniform distribution on $[0,1]$. Agents' priors are publicly observed at the beginning of an encounter. Each sender receives an independently drawn signal of commonly known quality p with probability φ. A sender privately knows whether she holds a signal. A sender i, if consulted, plays a static game with R, in which she maximizes the usual objective function given by:

$$-\left|E_R\left[E_{S,i}[\omega]\,|I_{S,i}\right] - E_R[\omega\,|I_{S,i}]\right|,$$

where $E_{S,i}[\omega]$ is sender i's expected value of the state and $I_{S,i}$ is the information put forward by sender i. Each period there is a probability τ that consultation is irreversibly terminated.

An interesting question to ask in this setup is now as follows: Suppose that R has at the beginning of the game a very extreme prior putting very high weight on state 1, i.e. α_R^0 is very low. Is R likely to be stuck with this extreme belief over a long period of time, even if the realized state is actually 0? How long would it take R, in expectation, to significantly shift her beliefs towards the truth? If it takes very long, it means that she is very unlikely to reach this state before she interrupts sampling.

Our equilibrium prediction for the disclosure subgames is as follows. If at t, α_R^{t-1} and the encountered sender's prior $\alpha_{S,i}$ are such that there exists a full disclosure equilibrium in the one-shot game, we assume full disclosure. If instead there only exists a D1- (or alternatively D0-) equilibrium, then we assume that this is played.

Suppose that R starts with a very low prior, i.e. α_R^0 very low and that p is relatively low. Given a uniform distribution of α_S^t in every period, the probability is roughly one that that the first sender encountered satisfies $\alpha_S^1 > \alpha_R^0$ and $\alpha_R^0 < 1-\alpha_S^1$. Accordingly, the equilibrium of the first disclosure game is with very high probability a D1-equilibrium. If the encounter leads to disclosure of a 1-signal, then α_R^1 will be even lower than α_R^0. If the encounter instead leads to no disclosure, then:

$$\alpha_R^1 = \frac{\alpha_R^0(\varphi p + (1-\varphi))}{\alpha_R^0(\varphi p + (1-\varphi)) + (1-\alpha_R^0)(\varphi(1-p) + (1-\varphi))},$$

which is larger than α_R^0. Indeed, no disclosure means that S either holds no signal or a signal indicating state 0. No disclosure thus provides some evidence in favor of 0. This evidence will be stronger, the higher p and the higher φ, the latter parameter affecting the probability that S actually holds a signal when not disclosing.

Consider a scenario (call it scenario K) characterized as follows. First, R samples n senders in a row. Second, all senders encountered either hold no signal or hold a 0-signal. Under suitably chosen parameter conditions (e.g. α_R^0 and p small enough), it is likely that given this scenario, for all t up till n, it holds true that $\alpha_R^{t-1} < \min\{\alpha_S^t, 1-\alpha_S^t\}$, so that R observes no disclosure n times in a row because each of the n stage game equilibria is a D1 equilibrium. The final belief of R in such a case will be:

$$k(\alpha_R^0, \varphi, p, n) = \frac{\alpha_R^0(\varphi p + (1-\varphi))^n}{\alpha_R^0(\varphi p + (1-\varphi))^n + (1-\alpha_R^0)(\varphi(1-p) + (1-\varphi))^n}.$$

If R's beliefs move very little after multiple encounters all featuring a D1 stage game equilibrium as well as no disclosure, R is effectively stuck with a low prior α_R^t over time because this low prior is self-confirming: The fact that it is low at the beginning of t makes it impossible to encounter information at t that significantly shifts its value. It is interesting to compare the quantity $k(\alpha_R^0, \varphi, p, n)$ to its expected counterpart in a putative equilibrium with full disclosure, conditional on scenario K realizing. This quantity is:

$$\widetilde{k}(\alpha_R^0, \varphi, p, n) = \sum_{k=0}^{n} \binom{n}{k} \varphi^k (1-\varphi)^{n-k} \frac{\alpha_R p^k}{\alpha_R p^k + (1-\alpha_R)(1-p)^k}.$$

Clearly, $\widetilde{k}(\alpha_R^0, \varphi, p, n)$ is larger than $k(\alpha_R^0, \varphi, p, n)$. In other words, partial disclosure of the D1-equilibrium type slows down learning in case the state is 0, as compared to full disclosure. The figure below shows numerical examples of $k(\alpha_R^0, \varphi, p, n)$. The function $k(.05, .2, .7, n)$ is continuous, $k(.05, .5, .7, n)$ is dashed and $k(.025, .5, .7, n)$ is thick continuous. The noteworthy aspect is that the function k can be very flat.

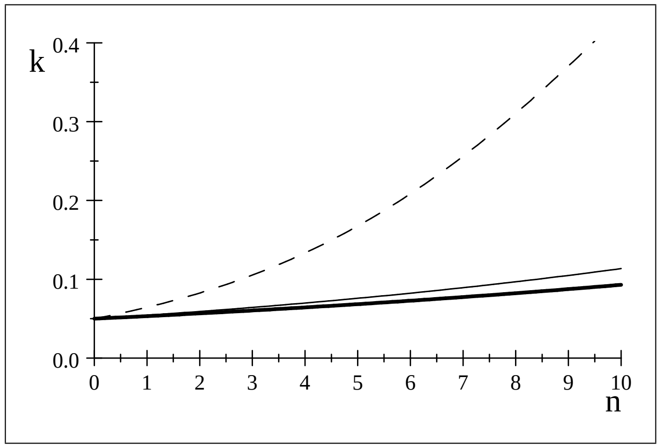

Figure 2

4 Almost cheap talk

We here consider a communication game featuring what could be termed *almost cheap-talk*. Player 1 (also called S) is known to hold a signal $\sigma \in \{0, 1\}$ with probability one, the signal being identical to the state with probability p. After observing the signal, S must send a message from the set $\{m_0, m_1\}$ which is observed by player 2 (also called R), who subsequently updates her beliefs. We assume a cost of lying which is determined as follows. In equilibrium, each message determines a conditional distribution over the signals (0 or 1) that S may be holding. If the equilibrium is informative, then the conditional distribution differs across messages. At a given information set of S (a given signal), in a given informative equilibrium, we define the *misleading message* as the message in the set $\{m_0, m_1\}$ which implies the lowest conditional probability of the signal actually held by S. We shall assume that at any information set, S bears an extra cost of $c > 0$ if she sends the misleading message. Besides this cost of lying or misleading, S's utility only depends on R's ex post perception of disagreement as in our analysis of disclosure.

We shall restrict ourselves to the special case of $\alpha_S = \frac{1}{2}$ and also assume w.l.o.g. that $\alpha_R < \frac{1}{2}$. As we know from our analysis of strategic disclosure, if $\alpha_R < 1 - \alpha_S$ and p is not high enough, S is reluctant (willing) to disclose a 0-signal (a 1-signal) because such a disclosure causes an ex post perceived disagreement that is higher (lower) than the ex ante perceived disagreement. Building on this intuition, we consider an equilibrium of the almost-cheap talk game in which S randomizes between m_0 and m_1 with probabilities $\{\beta, 1-\beta\}$ if and only if she holds a 0-signal. S instead sends m_1 with probability 1 if her signal is 1. We call such a putative equilibrium, if it exists, a *simple noisy equilibrium*. Note that in such an equilibrium, it is trivially true that m_1 (m_0) is the misleading message given signal 0 (1).

We state some preliminary observations before stating our main findings in the next Proposition. Given α_S, α_R, p, in a putative simple noisy communication equilibrium featuring the mixing probability β, the ex post perceived disagreement given message m_0 is:

$$t_0(\alpha_S, \alpha_R, p) = \left(\frac{\alpha_S p}{\alpha_S p + (1-\alpha_S)(1-p)} - \frac{\alpha_R p}{\alpha_R p + (1-\alpha_R)(1-p)} \right).$$

The ex post perceived disagreement given m_1 is instead given by:

$$t_1(\alpha_S, \alpha_R, p, \beta) = \left(\begin{array}{c} \left(\frac{\alpha_R(1-p)+(1-\alpha_R)p}{\alpha_R(1-p)+(1-\alpha_R)p+\alpha_R p(1-\beta)+(1-\alpha_R)(1-p)(1-\beta)} \right) \frac{\alpha_S(1-p)}{\alpha_S(1-p)+(1-\alpha_S)p} \\ + \left(1 - \frac{\alpha_R(1-p)+(1-\alpha_R)p}{\alpha_R(1-p)+(1-\alpha_R)p+\alpha_R p(1-\beta)+(1-\alpha_R)(1-p)(1-\beta)} \right) \frac{\alpha_S p}{\alpha_S p+(1-\alpha_S)(1-p)} \\ - \frac{\alpha_R((1-p)+p(1-\beta))}{\alpha_R((1-p)+p(1-\beta))+(1-\alpha_R)(p+(1-p)(1-\beta))} \end{array} \right).$$

The relative complexity of the second expression originates in the fact that R is uncertain about the signal held by S conditional on observing m_1, the latter being sent with positive probability given $\sigma = 1$ and $\sigma = 0$. We obtain the following characterization.

Proposition 2 *Let $\alpha_S = \frac{1}{2}$ and $\alpha_R < \frac{1}{2}$.*

a) There exists a simple noisy equilibrium if and only if

$$t_1\left(\frac{1}{2}, y, p, 0\right) + c \geq t_0\left(\frac{1}{2}, y, p\right) \geq t_1\left(\frac{1}{2}, y, p, 1\right) + c.$$

If there exists one, it is unique and characterized by the mixing probability $\beta^*(c, p, \alpha_R)$.

b) Given p, α_R, let $c' > c$ be such that there exists a simple noisy equilibrium under both c and c'. It holds true that $\beta^*(c', p, \alpha_R) > \beta^*(c, p, \alpha_R)$, i.e. a higher cost of lying implies more informative communication.

c) Given p, α_R, let $\alpha'_R > \alpha_R$ be such that there exists a simple noisy equilibrium under both α_R and α'_R. It holds true that $\beta^*(c, p, \alpha'_R) > \beta^*(c, p, \alpha_R)$, i.e. closer priors imply more informative communication.

d) Given α_R, if c is small enough, there exists a closed interval I of values of p such that 1) a simple noisy equilibrium exists if $p \in I$ and 2) $\beta^*(c, p, \alpha_R)$ is locally decreasing in p for $p \in I$ and 3) $I \subset \left(\frac{1}{2}, 1 - \alpha_R\right)$. Conditional on $p \in I$, an increase in signal precision thus implies more noisy communication.

Proof: See Appendix B.

Point a) simplifies the analysis by establishing uniqueness within the considered class of simple noisy equilibria. Points b) and c) establish intuitive comparative statics properties regarding the effect of changes in the lying cost and prior closeness. An increase in each of these variables positively affects the informativeness of communication. Point d) is more surprising: It states that more informative signals, on some closed interval of values of p, lead to more noisy communication. The intuition is that conditional on a given mixing probability β, as the signal strength p increases the sender payoff attached to m_1 increases relative to that attached to m_0. As a consequence, in order to preserve S's indifference between m_0 and m_1, the conditional distribution over signals implied by m_1 needs to assign more weight to signal 0, i.e. β needs to decrease.

The immediate question raised by Point d) is whether the direct beneficial effect of increased signal precision may be dominated by its negative strategic effect, so that an increase in p leads to less learning by R. We provide an example below which indicates that this can indeed be true. Define the following loss function, which constitutes one way of measuring the informativeness of S's communication:

$$U(c, p, \alpha_R) = -\sum_{m \in \{m_0, m_1\}} \sum_{\omega \in \{0,1\}} P(\omega, m) \left| E\left[\omega \,|m\right] - E[\omega] \right|,$$

where $P(\omega, m)$ is determined by S's communication strategy. $U()$ is the square root of the variance of R's equilibrium beliefs. The figure below illustrates our numerical example, which features $\alpha_R = .25$ and $c = .075$. The x-axis corresponds to p, the quality of the signal available to S. The continuous curve corresponds to the truth-telling probability

$\beta^*()$ while the dashed curve corresponds to the loss-function $U()$. The noteworthy aspect is that $U(.075.p, .25)$ is (very slightly) decreasing in in p if $p \in [.7, .8]$.

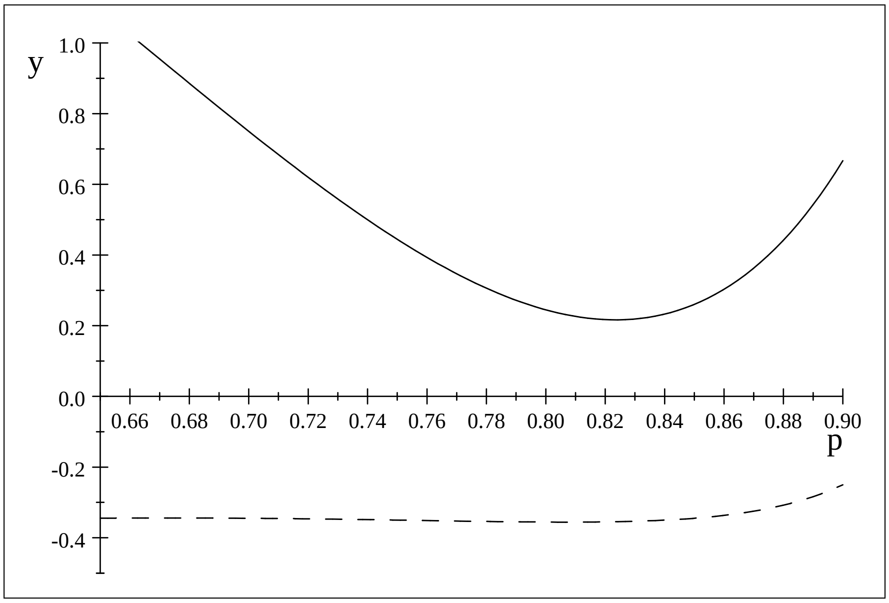

Figure 3

5 Strategic exposure

We here study the following simple game of voluntary and costly collective exposure to a public signal. Both players' utility function is given as in (1) minus a cost of participation. Both are averse to ex post perceived disagreement and bear an i.i.d. cost of observing the signal. In stage 1, both players decide whether or not to participate in a "conversation", each player's cost c_i of participating being drawn from a uniform distribution on $[0, 1]$. If both decide to participate, both players publicly observe a randomly drawn signal which is correct with probability p and each pays her private participation cost c_i. If at least one of the two players decides not to participate, the game ends. No signal is observed and no cost is paid. For notational simplicity, we shall denote player 1 by x and player 2 by y, where $\alpha_1 = x$ and $\alpha_2 = y$. Assume w.l.o.g that $x > y$. Note that there is no strategic aspect to the environment that we consider. Each player faces an individual decision problem and should simply agree to participate if and only if the expected reduction in perceived disagreement, conditional on joint observation of the signal, is larger than the private cost c_i of participating.

In studying the above problem, we implicitly take the stance of an outside principal interested in maximizing the amount of information acquired by a group of agents, for example through a debate. In order to formally complete the description of such a problem, one could add to the above described payoffs an extra component which depends on the absolute difference between an individually chosen ex post action $a_i \in [0,1]$ and the state ω. Given such a payoff component, each player will choose an action equal to her ex post expected value of the state, no matter how low the weight attributed to this component in the overall payoff function. In such a setup, one could assume that learning the truth and acting right is a marginal concern for agents while it is a first order concern for the outside principal, who thus wants to induce information acquisition at all cost.

We start by defining the following objects, which measure the difference in beliefs conditional on each possible public signal:

$$
\begin{aligned}
D_0(x,y,p) &= P(\omega = 0 \,|\sigma = j, x) - P(\omega = 0 \,|\sigma = j, y) \\
&\quad - \left(\frac{xp}{xp + (1-x)(1-p)} - \frac{yp}{yp + (1-y)(1-p)} \right),
\end{aligned}
$$

$$
\begin{aligned}
D_1(x,y,p) &= P(\omega = 0 \,|\sigma = 1, x) - P(\omega = 0 \,|\sigma = 1, y) \\
&= -\left(\frac{x(1-p)}{x(1-p) + (1-x)p} - \frac{y(1-p)}{y(1-p) + (1-y)p} \right)
\end{aligned}
$$

The expected difference in beliefs conditional on joint exposure to a signal of quality p, as computed by agent $z \in \{x, y\}$ is thus given by:

$$
\begin{aligned}
\Lambda^z(x,y,p) &= P(\sigma = 0 \,|z) D_0(x,y,p) + P(\sigma = 1 \,|z) D_1(x,y,p) \\
&= (zp + (1-z)(1-p))\, D_0(x,y,p) + (1 - (zp + (1-z)(1-p)))\, D_1(x,y,p)
\end{aligned}
$$

Note that $\Lambda^z(x, y, \frac{1}{2})$ is simply the status quo perceived disagreement (i.e. the difference in priors). It follows that the value of a public signal of quality p to player z is given as follows:

$$V^z(x, y, p) = \Lambda^z(x, y, p) - \Lambda^z(x, y, \frac{1}{2}).$$

Clearly, player z, $z \in x, y\}$, decides to participate in joint exposure if and only if $c_z \leq V^z(x, y, p)$. We obtain the following characterization.

Proposition 3 *I. For any x, y, p, $V^x(x, y, p) > 0$ and $V^y(x, y, p) > 0$.*

II. Let $y = 1 - x$.

a) $V^x(x, 1 - x, p) = V^y(x, 1 - x, p)$ for all $x \geq \frac{1}{2}, p$.

b) $V^x(x, 1 - x, p)$ is concave and single peaked in x with a maximum $x^* \in \left(\frac{1}{2}, 1\right)$, for a given p.

III. Let $y = \frac{1}{2}$ and $x > \frac{1}{2}$.

a) $V^y(x, \frac{1}{2}, p)$ is increasing in x for a given p. $V^x(x, \frac{1}{2}, p)$ is concave and single peaked in x with a maximum $x^* \in \left(\frac{1}{2}, 1\right)$.

b) $V^y(x, \frac{1}{2}, p) = \frac{V^y(x, 1-x, p)}{2}$ for all x, p.

c) $V^x(x, \frac{1}{2}, p) > V^y(x, \frac{1}{2}, p)$ for any $x > \frac{1}{2}$ and p.

IV. Let $x > y$.

a) Given x, $V^x(x, y, p)$ is increasing in y for y close enough to 0 and instead decreasing in y for y close enough to x. $V^y(x, y, p)$ is decreasing in y.

b) $V^x(x, y, p) \lesseqgtr V^y(x, y, p)$ if $y \lesseqgtr 1 - x$.

c) $V^x(x, y, p)$ and $V^y(x, y, p)$ are strictly increasing in p, for given x, y.

d) The prior pair $\{x, y\}$ that maximizes $\min\{V^x(x, y, p), V^y(x, y, p)\}$ is given by $\{x^*, 1 - x^*\}$ where $x^* = \arg\max_{x > \frac{1}{2}} V^x(x, 1 - x, p)$.

Proof: see in Appendix C.

Point I simply establishes that the gross value (i.e. ignoring the participation cost) of observing the public signal is always positive for both agents. Parts II, III and IV consider three different cases defined by different conditions on the players' priors. Point II examines the special case of priors that are symmetric around $\frac{1}{2}$. Result II.a) states that both players derive the same value from observing the signal. II.b) shows that this value is maximized for player priors that are neither too radical nor too moderate.

Point III considers the case where one player has a uniform prior while the other's prior is biased towards state 1. III.a) states that the value of the signal for the unbiased player is increasing in the misalignment of prior beliefs. Instead, the value of information to the biased player is concave and single peaked in the degree of prior belief misalignment. The value of information for the unbiased player bears a simple relation to the value of information in case her bias equals $1-x$, where x is the prior of the other (biased) player. It equals exactly half of the latter value. Finally, the biased player always assigns a higher value to the signal than the unbiased player.

Point IV considers the general case $x > y$. IV.b) establishes that the most extreme player derives the highest value from joint exposure. IV.c) shows that a more informative signal, as expected, increases the value of observing the signal. IV.d) shows that the preference constellation that maximizes the probability that a conversation will take place (by maximizing the minimal value of exposure across players) is given by $\{x^*, 1-x^*\}$ where x^* maximizes $V^x(x, 1-x, p)$. In other words, in order to maximize the probability of a conversation, one should pick two individuals who have symmetric priors, these priors being neither too moderate nor too radical.

The above characterization raises the question of whether an increase in prior disagreement can lead to a decrease in expected posterior disagreement, by way of its effect on players' willingness to participate. In what follows, we provide a simple numerical example and a formal result that show that this can indeed be the case given a high signal precision. Assume that the individual cost of participation is i.i.d. and uniformly distributed on the interval $[0, \bar{c}]$, where $\bar{c} > 0$. Given commonly known x, y, p, (assuming w.l.o.g. that $x > y$) the probability that both players accept to participate is given by:

$$\rho(x,y,p,\bar{c}) = \left(\frac{V^z(x,y,p)}{\bar{c}}\right)\left(\frac{V^y(x,y,p)}{\bar{c}}\right).$$

It follows that the expected ex post perceived disagreement in our game, from the perspective of player $z \in \{x, y\}$, is given by:

$$\Gamma(z,x,y,p,\bar{c}) = \rho(x,y,p,\bar{c})\Lambda^z(x,y,p) + (1-\rho(x,y,p,\bar{c}))(x-y).$$

Note that given our previous Proposition, it holds true that

$$\Gamma\left(x, x, 1-x, p, \bar{c}\right) = \Gamma\left(1-x, x, 1-x, p, \bar{c}\right).$$

We now provide a simple numerical example. For simplicity, we consider the case of $y = 1 - x$. In Figure 4, we set $p = .9$ and $\bar{c} = .25$. The horizontal axis corresponds to variable x and prior disagreement thus increases as x increases starting from $\frac{1}{2}$. The expected disagreement $\Gamma\left(x, x, 1-x, .9, .25\right) = \Gamma\left(1-x, x, 1-x, .9, .25\right)$ is mapped in thick continuous. The marginal value of exposure $V^x(x, 1-x, p) = V^y(x, 1-x, p)$ is shown in dashed. For any $x \leq .71$, $\left(\frac{V^z(x,1-x,p)}{\bar{c}}\right) \leq 1$ so that $\rho(x, 1-x, .9, .25)$ is indeed a probability. The noteworthy feature is that $\Gamma\left(x, x, 1-x, .9, .25\right)$ is decreasing in x for $x \geq 65$. I.e. higher prior disagreement implies lower expected ex post disagreement.

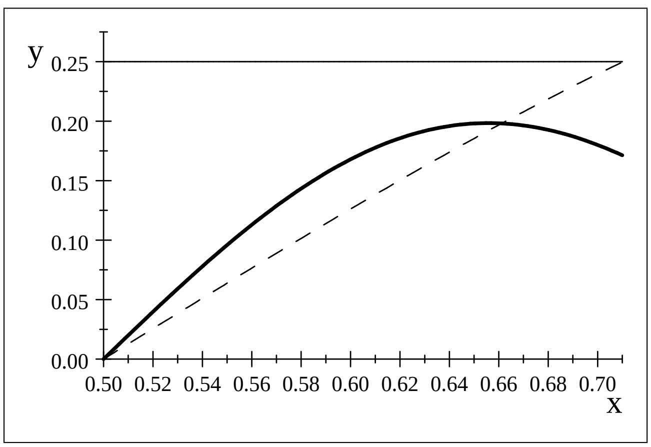

Figure 4

We now state a formal result for the special case of $y = 1 - x$ and $p = 1$, i.e. a perfectly informative signal.

Proposition 4 *Let $x > \frac{1}{2}$, $y = 1 - x$ and $p = 1$. For every $\bar{c} \in (0, 1]$, there exist $\bar{x} = \frac{1+\bar{c}}{2}$ and $x^* = \frac{1+\frac{\sqrt{3}}{3}\bar{c}}{2} \in (0, \bar{x})$ such that 1) $\rho(x, x, 1-x, 1, \bar{c}) \in [0, 1]$ if $x \leq \bar{x}$ and 2) $\Gamma(x, x, x, 1-x, 1, \bar{c})$ is increasing in x for $x < x^*$ and decreasing in x for $x > x^*$.*

Proof: see in Appendix D.

The next figure illustrates the above result. We set $\bar{c} = 1$. The expected disagreement $\Gamma(x, x, 1-x, 1, 1)$ corresponds to the thick continuous curve. The marginal value of exposure $V^x(x, 1-x, 1) = V^y(x, 1-x, 1)$ corresponds to the dashed curve. Note that $\Gamma(x, x, 1-x, 1, 1)$ peaks at $x^* = 0.788\,68$.

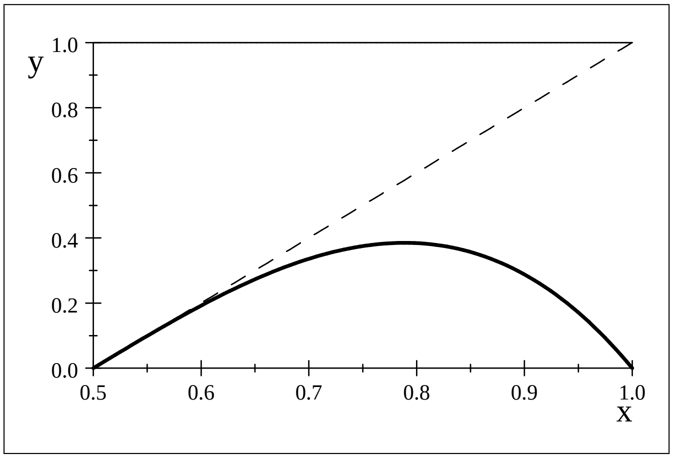

Figure 5

6 Extensions

6.1 An instrumental foundation of preferences for belief conformity

Agents may be averse to perceived disagreement in beliefs for instrumental reasons, because it hinders the achievement of their practical goals in a strategic context. They may for example be engaged in a multi-stage game in which a stage of disclosure is followed by a stage of collective decision making that requires each to provide input, the nature of that input being affected by perceived disagreement. We here analyze the second stage of such a two stage game which posits decision making by compromise. Two parties (S and R) each simultaneously propose a policy and the collective choice is a combination of the proposals, an agent's ideal policy being her expected value of a binary state $\omega \in \{0, 1\}$. While S's actual beliefs about the state are potentially uncertain to R at the beginning of the voting substage, R's beliefs are assumed publicly known. We show that S's payoff in equilibrium is decreasing in R's perception of disagreement in beliefs at the beginning of the policy proposal stage, as it encourages R to strategically distort her proposal.

Each of two parties $i \in \{S, R\}$ submits a proposal $x_i \in \mathbb{R}$ (e.g. a draft of a law). The final policy x that is implemented is a compromise, i.e. is a convex combination of the two proposals (this could alternatively be framed as voting for a particular position and the average is implemented). For concreteness we assume that $x = \frac{1}{2}(x_S + x_R)$. Each party has a preferred policy given by $\beta_i \in [0,1]$, where β_i is the parties' posterior belief about a binary state of the world $\omega \in \{0,1\}$. Further, we suppose that given β_i each party has a cost of submitting a untruthful proposal, the cost function being $c(\beta_i, x) = \frac{1}{2}(\beta_S - x_S)^2$. In other words, a moderate party is intrinsically reluctant to submit an extreme proposal just to get its way in negotiations, for example due to reputational concerns. β_R is publicly known. Instead, β_S is potentially only known to S and R's expectation of β_S is denoted $E_R[\beta_S]$. We now solve the policy-choice subgame. S's problem is:

$$\min_{x_S} \left\{ \left(\beta_S - \frac{x_S + x_R}{2} \right)^2 + \frac{1}{2}(\beta_S - x_S)^2 \right\}$$

which implies $x_S = \frac{4\beta_S - x_R}{3}$. Similarly, R solves

$$\min_{x_R} \left\{ E_R \left[\left(\beta_R - \frac{x_S + x_R}{2} \right)^2 \right] + \frac{1}{2}(\beta_R - x_R)^2 \right\}$$

implying $x_R = \frac{4\beta_R - E_R[x_S]}{3}$. It follows that in equilibrium, we have

$$\begin{aligned} x_S &= \frac{8\beta_S - 3\beta_R + E_R[\beta_S]}{6}, \\ x_R &= \frac{3\beta_R - E_R[\beta_S]}{2}, \end{aligned}$$

Plugging the above quantities into S's payoff function, we may conclude that S obtains the following payoff:

$$-\frac{3}{72}\left(2(\beta_S - \beta_R) + (E_R[\beta_S] - \beta_R)\right)^2.$$

S's payoff in the policy-choice game is thus affected by R's perception of disagreement $(E_R[\beta_S] - \beta_R)$, though it also depends on actual disagreement $(\beta_S - \beta_R)$. If the policy choice game is preceded by a stage of disclosure as analyzed in section 2, one can use backward induction to solve for S's equilibrium disclosure choice. Note that as compared to the preferences assumed in section, S now not only wants to reduce perceived ex-post disagreement but also wants to reduce actual disagreement.

6.2 Disclosure with continuous signals

As in section 2, S holds an informative signal with probability $\varphi \in (0,1)$. Instead of considering a binary signal, we here assume that S's signal is drawn from $[\underline{s}, \overline{s}]$. Given state $\omega \in \{l, h\}$, with $h = 1 > l = 0$, the signal is distributed according to $F(s|\omega)$. For simplicity, assume that F is absolutely continuous, i.e. admits a density $f(s|\omega)$ with full support ($f(s|\omega) > 0$ for all $s \in S$ and $\omega \in \{l, h\}$) and is ordered in the sense of the MLRP, i.e. $\frac{d}{ds}\frac{f(s|h)}{f(s|l)} > 0$. Denote by α_i, $i \in \{S, R\}$, S's and R's prior belief that $\omega = l$. Upon learning s, the updated belief of i is

$$\alpha_i(s) = \frac{\alpha_i f(s|l)}{\alpha_i f(s|l) + (1-\alpha_i) f(s|h)} = \frac{\alpha_i}{\alpha_i + (1-\alpha_i)\frac{f(s|h)}{f(s|l)}},$$

which is decreasing in s. We assume that extreme signals $\underline{s}$ and $\overline{s}$ are perfectly revealing, i.e., $\frac{f(s|h)}{f(s|l)} = 0$ for $s = \underline{s}$ and $= \infty$ for $s = \overline{s}$. Note that there exists a threshold signal $\tilde{s} \in (\underline{s}, \overline{s})$ such that $\alpha_i(s) \lesseqqgtr \alpha_i$ for $s \gtreqqless \tilde{s}$. The threshold satisfies $f(\tilde{s}|h) = f(\tilde{s}|l)$. We say that signal $s > \tilde{s}$ is congruent with j's relative prior bias if $\alpha_j > \alpha_i$.

We state a preliminary lemma and two propositions.

Lemma 1 *If $\alpha_S \neq \alpha_R$, then $\Delta(s) := |\alpha_S(s) - \alpha_R(s)|$ satisfies the following. i) $\Delta(\underline{s}) = \Delta(\overline{s}) = 0$. ii) There exists $\widehat{s}$ such that $\Delta(s)$ is increasing in s for all $s < \widehat{s}$ and decreasing in s for all $s > \widehat{s}$. iii) $\widetilde{s} < (>) \widehat{s}$ if the player with the lower prior is less (more) extreme.*

Proof: See in the Appendix E.

Proposition 5 *Assume that $\alpha_S > \alpha_R$. Then there exist two thresholds $s_1 \leq s_2$ such that the informed sender discloses s if and only if $s \leq s_1$ or $s \geq s_2$. If $\alpha_R = 1 - \alpha_S$ then $s_1 = s_2$, i.e. there is full disclosure, else $s_1 < s_2$, i.e. a non-empty set of signals $s \in (s_1, s_2)$ is not disclosed. If*

$\alpha_R < 1 - \alpha_S$, i.e. R is more extreme than S, then $s_1 < s_2 < \widetilde{s}$, i.e. all signals congruent with R's relative prior bias are disclosed. If $\alpha_R > 1 - \alpha_S$, i.e. R is less extreme than S, then $\widetilde{s} < s_1 < s_2$, i.e. all signals congruent with S's relative prior bias are disclosed.

Proof: See in the Appendix E.

The above proposition exhibits the same fundamental qualitative features as arising in the case of binary signals. Moderate prior misalignment is better than small or large prior misalignment in terms of ensuring disclosure. More specifically, given $\alpha_S > \alpha_R$ the most favorable prior of R in terms of disclosure is $1 - \alpha_S$. For any other prior of R, a closed set of signals is not disclosed. The second main finding is that signals that are congruent with the relative bias of the most extreme player are fully disclosed whereas signals that contradict the latter are disclosed only partially. If R is more extreme than S, S fully discloses signals that are congruent with R's relative bias whereas she imperfectly discloses signals that contradict R's relative bias. If instead R is less extreme than S, S fully discloses signals that contradict R's relative bias whereas she imperfectly discloses signals that are congruent with R's relative bias.

For our next result, we keep α_R fixed and examine how S's communication changes as α_S increases and moves away from α_R. This involves clarifying how the non-disclosure interval changes as a function of α_S.

Proposition 6 *Assume that $\alpha_S > \alpha_R$. i) s_2 is increasing in α_S. ii) s_1 is increasing in α_S.*

Proof: See in Appendix E.

Let $\alpha_S^{''} > \alpha_S' > \alpha_R$. Denote the event of no disclosure by ND. The result states that the conditional distribution of s given $(ND, \alpha_S = \alpha_S^{''})$ first order stochastically dominates the conditional distribution of s given $(ND, \alpha_S = \alpha_S^{'})$. Note that for $\alpha_S = 1 - \alpha_R$, $s_1 = s_2 = \tilde{s}$, where $\tilde{s}$ is the uninformative signal. Let $\widetilde{F}(s)$ denote the unconditional distribution of signals given prior α_R, which is endowed with pdf $\widetilde{f}(s)$. Define the average informativeness of signals conditional on no-disclosure as follows:

$$\left| \left(\frac{1}{\widetilde{F}(s_2) - \widetilde{F}(s_1)} \int_{s_1}^{s_2} \frac{f(s\,|h)}{f(s\,|l)} \widetilde{f}(s) ds \right) - 1 \right|$$

We may conclude the following. The average informativeness of non-disclosed signals decreases as α_S increases conditional on $\alpha_S < 1 - \alpha_R$. Instead, it increases as α_S increases conditional on $\alpha_S > 1 - \alpha_R$. In other words, as α_S moves away from $1 - \alpha_R$, S omits increasingly informative signals on average. Another way of formulating our result is as follows. Denote by $\widetilde{\alpha}_R(ND)$ the posterior probability assigned by R to $\omega = l$ conditional on ND in equilibrium. It holds true that $|\widetilde{\alpha}_R(ND) - \alpha_R|$ decreases as α_S increases conditional on $\alpha_S < 1 - \alpha_R$. Instead, $|\widetilde{\alpha}_R(ND) - \alpha_R|$ increases as α_S increases conditional on $\alpha_S > 1 - \alpha_R$.

As a final remark, note that the above does not allow us to conclude that the ex ante informativeness of S's communication decreases as α_S moves away from $1 - \alpha_R$.

6.3 Signaling priors

Disclosure and signaling of priors We here consider an extended version of the game of strategic disclosure analyzed in section 2. It features the following three new elements. First, while S's prior is publicly observed, R's prior is now instead privately observed by the latter and drawn from a publicly known distribution. Second, while S is interested in minimizing perceived disagreement as in the original disclosure game that we considered, R's utility function is now explicitly defined and given by $-(a - \omega)^2$, where $a \in [0, 1]$ is the action chosen by R after S's (potential) disclosure. Third, before the disclosure stage, there is a stage during which R can send a cheap talk message from the message set $[0, 1]$ to S. In equilibrium, this message could be informative regarding R's prior.

Does there exist an equilibrium in which R truthfully reveals her prior, so that the subsequent disclosure subgame features no asymmetric information concerning R's prior? The first point of our next Proposition establishes R's preference across partially informative experiments of the D1- and D0 types. We slightly abuse notation and simply speak of R preferring either D0 or D1 communication. Point b) draws the negative implication concerning the existence of an equilibrium featuring truthful revelation of her prior by R.

Proposition 7 *a) Given $\alpha_R < \frac{1}{2}$, R strictly prefers D0-communication over D1-communication. Given $\alpha_R > \frac{1}{2}$, R's preference over these is reversed. Given $\alpha_R = \frac{1}{2}$, R is indifferent between the* *two.*

b) Suppose that the set of possible priors of R contains only two types of priors; priors such that the only informative equilibrium in the disclosure subgame is a D1-equilibrium and priors such that the only informative equilibrium is a D0-equilibrium. There exists no equilibrium of the extended disclosure game in which R truthfully discloses her bias and the implied partial disclosure equilibrium ensues.

Proof: See in the Appendix F.

Note that a more general version of point b) would read as follows. Consider any distribution of R's prior that places positive probability on prior values α_R and α'_R such that the latter two priors imply different most informative equilibria in the ensuing disclosure subgame (i.e. D0, D1 or full disclosure). There exists no equilibrium in which R truthfully discloses her prior and the most informative equilibrium is played in the ensuing disclosure subgame. Note that if the distribution of R's prior contains a value such that full disclosure is feasible, any R type would trivially want to announce this prior value in a putative equilibrium featuring truthful announcement of priors.

Exposure and signaling of priors We now reconsider the case of strategic exposure analyzed in section 4, extending the setup by adding a stage of signaling about priors. We make the simplifying assumption that $c_1 = c_2 = 0$ so that both players always decide to participate in stage 2. In point a) we assume that S's prior is known to equal $\frac{1}{2}$, while α_R is privately known to R and drawn from a uniform distribution over $[0, 1]$. The game is given as follows. In stage 1, S can send a cheap talk message from $\{m_1, m_2\}$. In stage 2, the two parties simultaneously observe a public signal. In point b) we consider a generalized version in which both players' prior is privately known and drawn from a uniform distribution on $[0, 1]$. In stage 1, both players can send a simultaneous cheap talk message $m \in \{m_1, m_2\}$that might yield information about their bias. In stage 2, the two parties simultaneously observe a public signal.

In both cases considered, can a player credibly communicate information about her prior before the joint observation stage? The answer is positive in both cases. We obtain the following result:

Proposition 8 *a) Let* $\alpha_R = \frac{1}{2}$*. There exists an equilibrium such that S truthfully announces whether her prior is above or below* $\frac{1}{2}$ *before the public signal stage.*

b) Let the priors of both players be unknown and uniformly distributed on $[0, 1]$*. There exists an equilibrium in which each player truthfully announces whether her prior is above or below* $\frac{1}{2}$ *before the public signal stage.*

Proof: See in the Appendices G and H.

7 Appendix A: Disclosure (proof of Proposition 1)

7.1 Point I.a)

Step 0 Assume a putative full-disclosure equilibrium. We prove a sequence of substatements which together yield the statement of Point a).

Step 1 *Suppose that S holds a 0-signal and discloses it. The difference between ex ante and ex post perceived disagreement given disclosure of a 0-signal is given by:*

$$u_0(x,y,p) = (x-y) - \left(\frac{xp}{xp+(1-x)(1-p)} - \frac{yp}{yp+(1-y)(1-p)} \right).$$

If $y < 1-x$, it holds true that $u_0(x,y,p) < 0$ if and only if $p < P_0(x,y)$, where $P_0(x,y) \in \left(\frac{1}{2},1\right)$. If $y \geq 1-x$, it holds true that $u_0(x,y,p) \geq 0$ for any $p \geq \frac{1}{2}$.

Proof:

Note that

$$u_0(x,y,p) = (2p-1)\frac{(x-y)(p+x+y-px-py-xy+2pxy-1)}{(p+x-2px-1)(p+y-2py-1)}.$$

We solve

$$p+x+y-px-py-xy+2pxy-1=0$$

and obtain

$$p = P_0(x,y) \equiv \frac{x+y-xy-1}{x+y-2xy-1}.$$

Now, note that $P_0(x,y)$ is a decreasing function of y. Solving $P_0(x,y) = \frac{1}{2}$, yields $y = 1-x$. In other we have $P_0(x,y) \geq \frac{1}{2}$ if and only if $y \leq 1-x$.

Step 2 *Suppose that S holds a 1-signal. The difference between ex ante and ex post perceived disagreement given disclosure of a 1-signal is given by:*

$$u_1(x,y,p) = (x-y) - \left(\frac{x(1-p)}{x(1-p)+(1-x)p} - \frac{y(1-p)}{y(1-p)+(1-y)p} \right).$$

If $y \leq 1-x$, it holds true that $u_1(x,y,p) \geq 0$ for any $p \geq \frac{1}{2}$. If $y > 1-x$, it holds true that $u_1(x,y,p) < 0$ if and only if $p < P_1(x,y)$, where $P_1(x,y) \in \left(\frac{1}{2},1\right)$.

Proof: Note that

$$u_1(x,y,p) = -(2p-1)\frac{x-y}{(p+x-2px)(p+y-2py)}(px-p+py+xy-2pxy)$$

and that

$$px-p+py+xy-2pxy=0.$$

yields the solution:

$$p = P_1(x,y) \equiv -x\frac{y}{x+y-2xy-1}.$$

Now, note that $P_1(x,y)$ is an increasing function of y. Solving $P_1(x,y) = \frac{1}{2}$, yields $y = 1-x$. In other we have $P_1(x,y) \leq \frac{1}{2}$ if $y \leq 1-x$ while instead $P_1(x,y) > \frac{1}{2}$ if $y > 1-x$.

Step 3 Note the following:

$$\begin{aligned}
\frac{\partial P_0(\alpha_S,\alpha_R)}{\partial \alpha_S} &= \alpha_R\frac{\alpha_R-1}{(\alpha_S+\alpha_R-2\alpha_S\alpha_R-1)^2} < 0,\\
\frac{\partial P_0(\alpha_S,\alpha_R)}{\partial \alpha_R} &= \alpha_S\frac{\alpha_S-1}{(\alpha_S+\alpha_R-2\alpha_S\alpha_R-1)^2} < 0,\\
\frac{\partial P_1(\alpha_S,\alpha_R)}{\partial \alpha_S} &= -\alpha_R\frac{\alpha_R-1}{(\alpha_S+\alpha_R-2\alpha_S\alpha_R-1)^2} > 0,\\
\frac{\partial P_1(\alpha_S,\alpha_R)}{\partial \alpha_R} &= -\alpha_S\frac{\alpha_S-1}{(\alpha_S+\alpha_R-2\alpha_S\alpha_R-1)^2} > 0.
\end{aligned}$$

■

7.2 Point I.b) (D1 equilibrium with only 1-signals shown)

Step 0 Assume a putative D1 equilibrium. We prove a sequence of substatements which together yield the statement of Point I.b).

Step 1 After a 0-signal, S should prefer no disclosure to disclosure. So we examine

$$\begin{aligned}
&f_0(x,y,p,\varphi) \equiv \\
&\left(\begin{array}{c}\left(\frac{\varphi(yp+(1-y)(1-p))}{\varphi(yp+(1-y)(1-p))+(1-\varphi)}\right)\left(\frac{xp}{xp+(1-x)(1-p)}\right)\\ +\left(\frac{(1-\varphi)}{y\varphi p+(1-y)\varphi(1-p)+(1-\varphi)}\right)x\end{array}\right)\\
&-\left(\frac{y(\varphi p+1-\varphi)}{y(\varphi p+1-\varphi)+(1-y)(\varphi(1-p)+1-\varphi)}\right)\\
&-\left(\frac{xp}{xp+(1-x)(1-p)}-\frac{yp}{yp+(1-y)(1-p)}\right),
\end{aligned}$$

which simplifies to

$$(\varphi-1)(2p-1)\frac{(x-y)(p+x+y-px-py-xy+2pxy-1)}{(p+x-2px-1)(p+y-2py-1)(p\varphi+y\varphi-2py\varphi-1)}.$$

We need to solve for

$$f_0(x,y,p,\varphi)=0.$$

The solution is given by the following value of p:

$$P_0 \equiv \frac{x+y-xy-1}{x+y-2xy-1}.$$

Now, note that $P_0(x,y)$ is a decreasing function of y. Solving $P_0(x,y)=\frac{1}{2}$, yields $y=1-x$. In other we have $P_0(x,y)\geq\frac{1}{2}$ if and only if $y\leq 1-x$.

Step 2 After a 1-signal, S should prefer disclosing to not disclosing. So we study:

$$\begin{aligned}
&f_1(x,y,p,\varphi) \equiv \\
&\left(\begin{array}{c}\left(\frac{\varphi(yp+(1-y)(1-p))}{\varphi(yp+(1-y)(1-p))+(1-\varphi)}\right)\left(\frac{xp}{xp+(1-x)(1-p)}\right)\\ +\left(\frac{(1-\varphi)}{y\varphi p+(1-y)\varphi(1-p)+(1-\varphi)}\right)x\end{array}\right)\\
&-\left(\frac{y(\varphi p+1-\varphi)}{y(\varphi p+1-\varphi)+(1-y)(\varphi(1-p)+1-\varphi)}\right)\\
&-\left(\frac{x(1-p)}{x(1-p)+(1-x)p}-\frac{y(1-p)}{y(1-p)+(1-y)p}\right).
\end{aligned}$$

The argument is in two steps. Examine first the following object

$$\begin{aligned}&\widetilde{f}_1(x,y,p,\varphi)\\ \equiv\ &\left(\begin{array}{c}\left(\frac{\varphi(yp+(1-y)(1-p))}{\varphi(yp+(1-y)(1-p))+(1-\varphi)}\right)\left(\frac{xp}{xp+(1-x)(1-p)}\right)+\left(\frac{(1-\varphi)}{y\varphi p+(1-y)\varphi(1-p)+(1-\varphi)}\right)x\\ -\left(\frac{y(\varphi p+1-\varphi)}{y(\varphi p+1-\varphi)+(1-y)(\varphi(1-p)+1-\varphi)}\right)\end{array}\right)\\ &-(x-y)\,,\end{aligned}$$

which simplifies to

$$-\varphi\,(2p-1)\,\frac{(x-y)\,(p+x+y-px-py-xy+2pxy-1)}{(p+x-2px-1)\,(p\varphi+y\varphi-2py\varphi-1)}.$$

Note that this expression is positive for any $p \leq P_0 = \frac{x+y-xy-1}{x+y-2xy-1}$. Note also that $P_0(x,y)$ is a decreasing function of y. Solving $P_0(x,y) = \frac{1}{2}$, yields $y = 1-x$. We thus have $P_0(x,y) \geq \frac{1}{2}$ if and only if $y \leq 1-x$. Note finally that given $y \leq 1-x$,

$$(x-y) > \frac{x(1-p)}{x(1-p)+(1-x)p} - \frac{y(1-p)}{y(1-p)+(1-y)p}.$$

We may thus conclude that a fortiori, for any $p \leq P_0$ it also holds true that $f_1(x,y,p,\varphi) > 0$, implying that after a 1-signal, S should prefer disclosing to not disclosing.■

7.3 Point I.c) (D0 equilibrium with only 0-signals shown)

Step 0 Assume a putative D0 equilibrium. We prove a sequence of substatements which together yield the statement of Point I.c).

Step 1 After a 0-signal, S should prefer to disclose. So we check:

$$\begin{aligned}&g_0(x,y,p,\varphi)\\ \equiv\ &\left(\begin{array}{c}\left(\frac{\varphi(y(1-p)+(1-y)p)}{\varphi(y(1-p)+(1-y)p)+(1-\varphi)}\right)\left(\frac{x(1-p)}{x(1-p)+(1-x)p}\right)+\left(\frac{(1-\varphi)}{\varphi(y(1-p)+(1-y)p)+(1-\varphi)}\right)x\\ -\left(\frac{y(\varphi(1-p)+1-\varphi)}{y(\varphi(1-p)+1-\varphi)+(1-y)(\varphi p+1-\varphi)}\right)\end{array}\right)\\ &-\left(\frac{xp}{xp+(1-x)(1-p)}-\frac{yp}{yp+(1-y)(1-p)}\right).\end{aligned}$$

Here, the argument is in two steps. Note the following function:

$$r(x,y,p,\varphi) \equiv \begin{pmatrix} \left(\frac{\varphi(y(1-p)+(1-y)p)}{\varphi(y(1-p)+(1-y)p)+(1-\varphi)}\right)\left(\frac{x(1-p)}{x(1-p)+(1-x)p}\right)+\left(\frac{(1-\varphi)}{\varphi(y(1-p)+(1-y)p)+(1-\varphi)}\right)x \\ -\left(\frac{y(\varphi(1-p)+1-\varphi)}{y(\varphi(1-p)+1-\varphi)+(1-y)(\varphi p+1-\varphi)}\right) \end{pmatrix} - (x-y)\,.$$

Note the following simplification:

$$r(x,y,p,\varphi) = \varphi\,(2p-1)\,\frac{x-y}{p+x-2px}\,\frac{px-p+py+xy-2pxy}{p\varphi-\varphi+y\varphi-2py\varphi+1}$$

Note that $r(x,y,p,\varphi) \geq 0$ for any $p \leq P_1(x,y)$, recalling that $P_1(x,y) = -x\frac{y}{x+y-2xy-1}$. Recall also that $P_1(x,y) > \frac{1}{2}$ if $y > 1-x$. Now, simply note that given $y \geq 1-x$,

$$\frac{xp}{xp+(1-x)(1-p)} - \frac{yp}{yp+(1-y)(1-p)} < (x-y)\,.$$

We may conclude that a fortiori for any $p \leq P_1(x,y)$, it holds true that $g_0(x,y,p,\varphi) \geq 0$, implying that after a 0-signal, S prefers to disclose.

Step 2 After a 1-signal, S should prefer to omit disclosing. We examine:

$$\begin{aligned} & g_1(x,y,p,\varphi) \\ \equiv & \begin{pmatrix} \left(\frac{\varphi(y(1-p)+(1-y)p)}{\varphi(y(1-p)+(1-y)p)+(1-\varphi)}\right)\left(\frac{x(1-p)}{x(1-p)+(1-x)p}\right)+\left(\frac{(1-\varphi)}{\varphi(y(1-p)+(1-y)p)+(1-\varphi)}\right)x \\ -\left(\frac{y(\varphi(1-p)+1-\varphi)}{y(\varphi(1-p)+1-\varphi)+(1-y)(\varphi p+1-\varphi)}\right) \end{pmatrix} \\ & -\left(\frac{x(1-p)}{x(1-p)+(1-x)p} - \frac{y(1-p)}{y(1-p)+(1-y)p}\right). \end{aligned}$$

Note that $g_1(x,y,p,\varphi)$ simplifies to

$$(\varphi-1)\,(2p-1)\,\frac{x-y}{(p+x-2px)\,(p+y-2py)}\,\frac{px-p+py+xy-2pxy}{p\varphi-\varphi+y\varphi-2py\varphi+1}.$$

Now, simply note that $g_1(x,y,p,\varphi) \leq 0$ for for any $p \leq P_1(x,y)$.■

7.4 Proof of Part III

Proofs are omitted but identical to proofs for Part I. In order to study the full disclosure equilibrium, we need to study the following two objects:

$$l_0(x,y,p) = (y-x) - \left(\frac{yp}{yp+(1-y)(1-p)} - \frac{xp}{xp+(1-x)(1-p)}\right),$$

$$l_1(x,y,p) = (y-x) - \left(\frac{y(1-p)}{y(1-p)+(1-y)p} - \frac{x(1-p)}{x(1-p)+(1-x)p}\right).$$

Solving $l_0(x,y,p) = 0$ yields the solution $\frac{x+y-xy-1}{x+y-2xy-1}$. Solving $l_1(x,y,p) = 0$ yields the solution $-x\frac{y}{x+y-2xy-1}$.■

8 Appendix B: Almost Cheap Talk (proof of Proposition 2)

Step 1 The considered communication strategy is incentive compatible for S if and only if:

$$t_1\left(\frac{1}{2},y,p,\beta\right) + c = t_0(\frac{1}{2},y,p)$$

and

$$t_1\left(\frac{1}{2},y,p,\beta\right) \leq t_0(\frac{1}{2},y,p) + c.$$

The first condition is necessary and sufficient to ensures S's willingness to mix when holding a 0-signal, while the second condition is necessary and sufficient to ensure S's willingness to send m_1 with probability one when holding a 1-signal. Note that the first condition implies the second. Now, simply note that

$$\frac{\partial t_1(\frac{1}{2},y,p,\beta)}{\partial\beta} = p\,(2y-1)^2\,\frac{2p^2-3p+1}{(p\beta-\beta+y\beta-2py\beta+1)^2},$$

which is always negative. Imposing

$$t_1\left(\frac{1}{2},y,p,0\right) + c \geq t_0\left(\frac{1}{2},y,p\right) \geq t_1\left(\frac{1}{2},y,p,1\right) + c \tag{2}$$

thus ensures that a single crossing condition holds. I.e., there exists a unique $\beta \in (0,1)$ such that $t_1\left(\frac{1}{2},y,p,\beta\right) + c = t_0(\frac{1}{2},y,p)$. Note that (2) is equivalent to:

$$t_0(\frac{1}{2},y,p) - t_1(\frac{1}{2},y,p,0) \leq c \leq t_0(\frac{1}{2},y,p) - t_1(\frac{1}{2},y,p,1).$$

Now, developing the LHS and RHS expressions in the above double inequality, the latter in turn rewrites as follows:

$$\begin{aligned} v_1(p,y) &\equiv p\,(2y-1)^2 \frac{\left(2p^2-3p+1\right)}{p+y-2py-1} \leq c \\ &\leq v_2(p,y) \equiv -p\,(2y-1)^2 \frac{\left(2p^2-3p+1\right)}{\left(-4p^2y^2+4p^2y-p^2+4py^2-4py+p-y^2+y\right)}. \end{aligned}$$

Now, the following three properties can be easily shown. First, $v_1(\frac{1}{2},y) = v_2(\frac{1}{2},y) = 0$. Second, $v_1(1,y) = v_2(1,y) = 0$. Finally, for any $p \in \left(\frac{1}{2},1\right)$, it holds true that $v_2(p,y) > v_1(p,y) > 0$. Finally, note that $t_1\left(\frac{1}{2},y,p,\beta\right) + c = t_0(\frac{1}{2},y,p)$ yields a unique solution given by

$$\begin{aligned} &\beta^*(c,p,y) \\ \equiv\ & \frac{\begin{pmatrix} c+p+4p\alpha_R^2+12p^2\alpha_R-8p^3\alpha_R-12p^2\alpha_R^2 \\ +8p^3\alpha_R^2-cp-c\alpha_R-4p\alpha_R-3p^2+2p^3+2cp\alpha_R \end{pmatrix}}{c+cp^2+c\alpha_R^2-2cp-2c\alpha_R-4cp\alpha_R^2-4cp^2\alpha_R+4cp^2\alpha_R^2+6cp\alpha_R}. \end{aligned}$$

Step 2 This proves Point b) and thus examines the effect of changes in c. We here directly use the formula obtained for the equilibrium value of β, i.e. $\beta^*(c,p,y)$. Note that:

$$\frac{\partial \beta^*(c,p,y)}{\partial c} = -\frac{1}{c^2} p\,(2y-1)^2 \frac{2p^2-3p+1}{(p+y-2py-1)^2}.$$

which is always positive.

Step 3 This proves point c) and thus examines the effect of changes in y. Note that:

$$\frac{\partial \beta^*(c,p,y)}{\partial y} = \frac{1}{c} \frac{(2p-1)\left(c-2p-4p^2y-cp-cy+4py+2p^2+2cpy\right)}{(p+y-2py-1)^3}.$$

Note first that $\frac{1}{c}\frac{(2p-1)}{(p+y-2py-1)^3}$ is trivially always negative. The sign of $\frac{\partial \beta^*(c,p,y)}{\partial y}$ thus depends on the sign of the following expression:

$$g(c,p,y) \equiv c-2p-4p^2y-cp-cy+4py+2p^2+2cpy,$$

which we now focus on. Note first that

$$\frac{\partial g(c,p,y)}{\partial c} = 2py-y-p+1$$

which is always positive. Note second that

$$g(v_2(p,y),p,y) = -p\,(2y-1)\,\frac{p-1}{p+y-2py} < 0.$$

It follows that given p, y, $g(c,p,y)$ is positive for any $c \in [\,v_1(p,y), v_2(p,y)]$. Note also that

$$\frac{\partial g(c,p,y)}{\partial y} = 4p - 4p^2 + 2cp - c > 0.$$

It follows from the above that if y is such that $\left.\frac{\partial \beta^*(c,p,y)}{\partial y}\right|_{y=y'} > 0$, then $\left.\frac{\partial \beta^*(c,p,y)}{\partial y}\right|_{y=y''} > 0$ for any $y'' > y'$ (call this Fact 1). We may now conclude. Thus, consider y', p and c such that there exists a simple mixed equilibrium. Now, consider y'' strictly larger but sufficiently close to y'. It follows from Fact 1 that $\beta^*(c,p,y'') > \beta^*(c,p,y')$.

Step 4 This proves point d) and thus examines the effect of changes in p. Note first that $\beta^*(c,\frac{1}{2},y) = 2$ and $\beta^*(c,1,y) = \frac{1}{y}$. Second, note that

$$\beta^*(c,1-y,y) = \frac{1}{4cy-4cy^2}\left(8y^3 - 12y^2 + 6y + 2c - 1\right).$$

Now note that $8y^3 - 12y^2 + 6y - 1 < 0$ (>0) for $y < (>)\frac{1}{2}$ and equal 0 for $y = \frac{1}{2}$. It follows that given $y < \frac{1}{2}$, if c is small enough, $\beta^*(c,1-y,y)$ is negative. We may conclude. For any given $y < \frac{1}{2}$, for c small enough, we have $\beta^*(c,\frac{1}{2},y) > 1$ and $\beta^*(c,1-y,y) < 0$. Given that $\beta^*(c,p,y)$ is continuous in p, it follows that for any given y, for c small enough, there must be an interval of values of p such that $\beta^*(c,p,y) \in (0,1)$ and is decreasing in p.■

9 Appendix C: Exposure (proof of Proposition 3)

Step 1 This proves Point I. We here obtain the following expressions for the change in belief differences as a function of the public signal. Given a 0-signal, we have:

$$u_0(x,y,p) = (x-y) - \left(\frac{xp}{xp+(1-x)(1-p)} - \frac{yp}{yp+(1-y)(1-p)}\right).$$

Given a 1-signal, we instead have:

$$u_1(x,y,p) = (x-y) - \left(\frac{x(1-p)}{x(1-p)+(1-x)p} - \frac{y(1-p)}{y(1-p)+(1-y)p}\right).$$

The expected change in difference in beliefs for the agent with prior x is given by:

$$A(x,y,p) = (xp + (1-x)(1-p))\, u_0(x,y,p) + (1 - (xp + (1-x)(1-p)))\, u_1(x,y,p).$$

The expected change in difference in beliefs for the agent with prior y is given by:

$$B(x,y,p) = (yp + (1-y)(1-p))\, u_0(x,y,p) + (1 - (yp + (1-y)(1-p)))\, u_1(x,y,p).$$

Note that these expressions simplify. It holds true that:

$$A(x,y,p) = -y\,(2p-1)^2\,(x-y)\,\frac{y-1}{-4p^2y^2 + 4p^2y - p^2 + 4py^2 - 4py + p - y^2 + y}$$

and

$$B(x,y,p) = -x\,(2p-1)^2\,(x-y)\,\frac{x-1}{-4p^2x^2 + 4p^2x - p^2 + 4px^2 - 4px + p - x^2 + x}.$$

Now, note that it can be shown trivially that these expressions are always positive no matter the values of x, y and p.

Step 2 Define the following objects, corresponding to the difference in prior beliefs after each type of signal.

$$M(x,y,p) = -\left(\frac{xp}{xp + (1-x)(1-p)} - \frac{yp}{yp + (1-y)(1-p)}\right),$$

$$N(x,y,p) = -\left(\frac{x(1-p)}{x(1-p) + (1-x)p} - \frac{y(1-p)}{y(1-p) + (1-y)p}\right).$$

The expected ex post difference in beliefs for the agent with prior x, given that the public signal is observed, is given by:

$$\Pi(x,y,p) = (xp + (1-x)(1-p))\, M(x,y,p) + (1 - (xp + (1-x)(1-p)))\, N(x,y,p).$$

The expected difference in beliefs for the agent with prior y, given that the public signal is observed, is given by:

$$\Lambda(x,y,p) = (yp + (1-y)(1-p))\, M(x,y,p) + (1 - (yp + (1-y)(1-p)))\, N(x,y,p).$$

Step 3 This proves Point II.a). Let us assume that $y = 1 - x$. Note that:

$$\begin{aligned}\Pi(x, 1-x, p) - \Pi(x, 1-x, \frac{1}{2}) &= \Lambda(x, 1-x, p) - \Lambda(x, 1-x, \frac{1}{2}) \\ &= \frac{-x\,(2p-1)^2\left(2x^2 - 3x + 1\right)}{-4p^2x^2 + 4p^2x - p^2 + 4px^2 - 4px + p - x^2 + x},\end{aligned}$$

which is always positive (this was already proven more generally earlier).

Step 4 This proves Point II.b). Note that

$$\begin{aligned}&\frac{\partial\left(\Pi(x, 1-x, p) - \Pi(x, 1-x, \frac{1}{2})\right)}{\partial x} \\ = &\frac{(2p-1)^2}{\left(-4p^2x^2 + 4p^2x - p^2 + 4px^2 - 4px + p - x^2 + x\right)^2} \\ &\left(\begin{array}{c} 8p^2x^4 - 16p^2x^3 + 14p^2x^2 - 6p^2x + p^2 - 8px^4 \\ +16px^3 - 14px^2 + 6px - p + 2x^4 - 4x^3 + 2x^2 \end{array}\right)\end{aligned}$$

We now consider the function:

$$\begin{aligned}&T(p, x) \\ \equiv\ &8p^2x^4 - 16p^2x^3 + 14p^2x^2 - 6p^2x + p^2 - 8px^4 \\ &+16px^3 - 14px^2 + 6px - p + 2x^4 - 4x^3 + 2x^2.\end{aligned}$$

and show that there is some x^* such that $T(p, x) > 0$ for $x < x^*$ and $T(p, x) < 0$ for $x > x^*$. To show this, we show that $T(p, x)$ is monotonically decreasing in x as well as positive for $x = \frac{1}{2}$ and negative for $x = 1$. Note first that that

$$\begin{aligned}&\frac{\partial T(p, x)}{\partial x} \\ = \ &32p^2x^3 - 48p^2x^2 + 28p^2x - 6p^2 - 32px^3 + 48px^2 - 28px + 6p + 8x^3 - 12x^2 + 4x.\end{aligned}$$

Now, note that $\frac{\partial T(p,x)}{\partial x} = 0$ has the following solutions:

$$\left\{\frac{1}{2}, \frac{1}{4p-2}\left(2p + \sqrt{2p - 2p^2 + 1} - 1\right), -\frac{1}{4p-2}\left(-2p + \sqrt{2p - 2p^2 + 1} + 1\right)\right\},$$

where the second is always larger than 1 for $p \in \left[\frac{1}{2}, 1\right]$ and the smaller is always smaller than 0 for $p \in \left[\frac{1}{2}, 1\right]$. This implies that $\frac{\partial T(p,x)}{\partial x}$ always has the same sign for any $p \in$

$\left[\frac{1}{2},1\right]$. Note finally that $T(p,\frac{1}{2})=\frac{1}{8}$ while instead $T(p,1)=p\,(p-1)$. Unfortunately, one cannot identify an explicit solution to $T(p,x)=0$. To see that $\Pi(x,1-x,.p)-\Pi(x,1-x,\frac{1}{2})$ is concave in x, note that:

$$\frac{\partial^2\left(\Pi(x,1-x,p)-\Pi(x,1-x,\frac{1}{2})\right)}{\partial^2 x}$$
$$2p\,(2p-1)^2\,(2x-1)\,\frac{(p-1)\left(4p^2x^2-4p^2x+p^2-4px^2+4px-p+x^2-x+1\right)}{\left(-4p^2x^2+4p^2x-p^2+4px^2-4px+p-x^2+x\right)^3},$$

which is easily shown to be negative for any x,p. The figure below illustrates $\Pi(x,1-x,p)-\Pi(x,1-x,\frac{1}{2})$ for two different values of p ($p=.75$ in continuous, $p=.85$ in dashed).

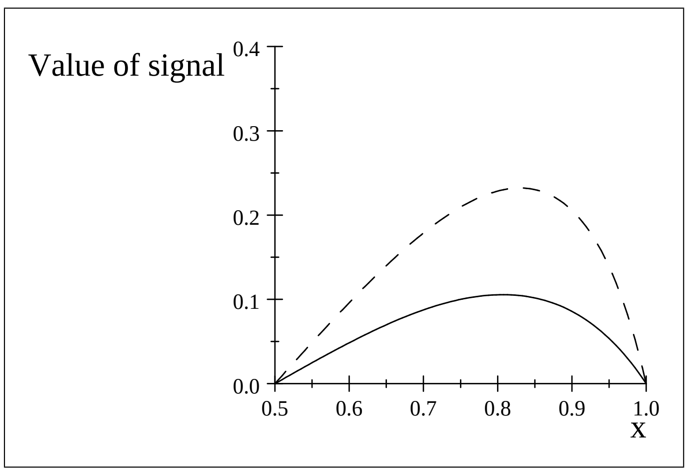

Figure A

We see that the value of the public signal is single-peaked and concave in x, for a given p. We also see that the value of the public signal is increasing in p, for a given x.

Step 5 This proves Point III. Assume now $x>\frac{1}{2}$ and $y=\frac{1}{2}$. Note that

$$\Pi(x,\frac{1}{2},p)-\Pi(x,\frac{1}{2},\frac{1}{2})=\frac{1}{2}\,(2p-1)^2\,(2x-1)\,,$$

which is monotonically increasing in x. On the other hand,

$$\Lambda\left(x,\frac{1}{2},p\right)-\Lambda\left(x,\frac{1}{2},\frac{1}{2}\right)=-\frac{1}{2}x\,(2p-1)^2\,\frac{2x^2-3x+1}{-4p^2x^2+4p^2x-p^2+4px^2-4px+p-x^2+x}$$

, which in turn equals $\frac{\Lambda(x,1-x,p)-\Lambda(x,1-x,\frac{1}{2})}{2}$. Note also that:

$$\begin{aligned}
&\Pi\left(x,\frac{1}{2},p\right)-\Pi\left(x,\frac{1}{2},\frac{1}{2}\right)-\left[\Lambda\left(x,\frac{1}{2},p\right)-\Lambda\left(x,\frac{1}{2},\frac{1}{2}\right)\right]\\
=&\ -\frac{1}{2}p\left(2p-1\right)^2\left(2x-1\right)^3\frac{p-1}{-4p^2x^2+4p^2x-p^2+4px^2-4px+p-x^2+x},
\end{aligned}$$

which is always positive. The figure below illustrates $\Pi(x,\frac{1}{2},p)-\Pi(x,\frac{1}{2},\frac{1}{2})$ in continuous and $\Lambda(x,\frac{1}{2},p)-\Lambda(x,\frac{1}{2},\frac{1}{2})$ in dashed. We set $p=.75$.

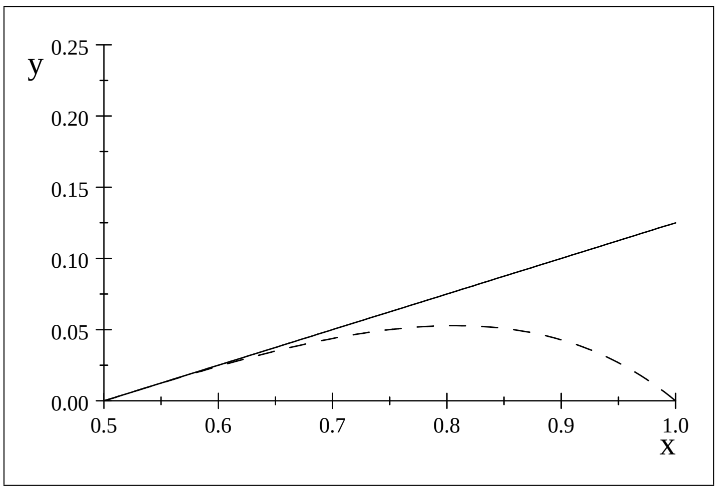

Figure B

Step 6 This step proves Point IV.a). Note that

$$\begin{aligned}
\frac{\partial\left(\Pi(x,y,p)-\Pi(x,y,\frac{1}{2})\right)}{\partial y} =&\ -\frac{(2p-1)^2}{\left(-4p^2y^2+4p^2y-p^2+4py^2-4py+p-y^2+y\right)^2}\\
&\left(\begin{array}{c} p^2x-7py^2-2p^2y+8py^3-4py^4+7p^2y^2\\ -8p^2y^3+4p^2y^4-px+2py+y^2-2y^3+y^4-2p^2xy+2pxy\end{array}\right).
\end{aligned}$$

Note that $-\frac{(2p-1)^2}{\left(-4p^2y^2+4p^2y-p^2+4py^2-4py+p-y^2+y\right)^2}$ is always negative and continuous in y. Now, define

$$\begin{aligned}
h(p,x,y) =&\ p^2x-7py^2-2p^2y+8py^3-4py^4+7p^2y^2-8p^2y^3\\
&+4p^2y^4-px+2py+y^2-2y^3+y^4-2p^2xy+2pxy.
\end{aligned}$$

Note the following facts. First, $h(p,x,0) = px(p-1)$ which is trivially negative. Second,

$$h(p,x,x) = -x(x-1)\left(-4p^2x^2 + 4p^2x - p^2 + 4px^2 - 4px + p - x^2 + x\right).$$

which is trivially always positive. Given the continuity of $h(p,x,y)$ in y, we may conclude that $h(p,x,y)$ is negative for y close enough to 0 and positive for y close enough to x. It follows immediately that $\frac{\partial\left(\Pi(x,y,p)-\Pi(x,y,\frac{1}{2})\right)}{\partial y}$ is positive for y close enough to 0 and negative for y close enough to x. Note, as a side remark, that we are not able to establish that $\frac{\partial h(p,x,y)}{\partial y}$ is always positive for all $y < x$, which would have allowed us to establish a single crossing condition and obtain a stronger result concerning $\frac{\partial\left(\Pi(x,y,p)-\Pi(x,y,\frac{1}{2})\right)}{\partial y}$. Note finally that

$$\frac{\partial\left(\Lambda(x,y,p) - \Lambda(x,y,\frac{1}{2})\right)}{\partial y} = x(2p-1)^2 \frac{x-1}{-4p^2x^2 + 4p^2x - p^2 + 4px^2 - 4px + p - x^2 + x},$$

which is trivially always negative.

Step 7 This proves Point IV.b). Note that

$$\begin{aligned} & \Pi(x,y,p) - \Pi(x,y,\frac{1}{2}) - \left(\Lambda(x,y,p) - \Lambda(x,y,\frac{1}{2})\right) \\ = & -p(2p-1)^2(x-y)^2(p-1) \\ & \frac{x+y-1}{\begin{pmatrix} (-4p^2x^2 + 4p^2x - p^2 + 4px^2 - 4px + p - x^2 + x) \\ (-4p^2y^2 + 4p^2y - p^2 + 4py^2 - 4py + p - y^2 + y) \end{pmatrix}} \end{aligned}$$

which is trivially positive for $y < 1-x$, equal 0 for $y = 1-x$ and negative for $y > 1-x$.

Step 8 This proves Point IV.c). Note simply that:

$$\begin{aligned} \frac{\partial\left(\Pi(x,y,p) - \Pi(x,y,\frac{1}{2})\right)}{\partial p} = & -y(2p-1)(x-y) \\ & \frac{y-1}{(-4p^2y^2 + 4p^2y - p^2 + 4py^2 - 4py + p - y^2 + y)^2} \end{aligned}$$

and

$$\frac{\partial\left(\Lambda(x,y,p)-\Lambda(x,y,\frac{1}{2})\right)}{\partial p} = -x\,(2p-1)\,(x-y)$$
$$\frac{x-1}{\left(-4p^2x^2+4p^2x-p^2+4px^2-4px+p-x^2+x\right)^2}.$$

Both expressions are trivially positive. The figure below shows $\Pi(.8,s,.75)-\Pi(.8,s,\frac{1}{2})$ in continuous and $\Lambda(.8,s,.75)-\Lambda(.8,s,\frac{1}{2})$ in dashed. We assume different values of p given by respectively $p=.6$ or $p=.75$ or $p=.9$.

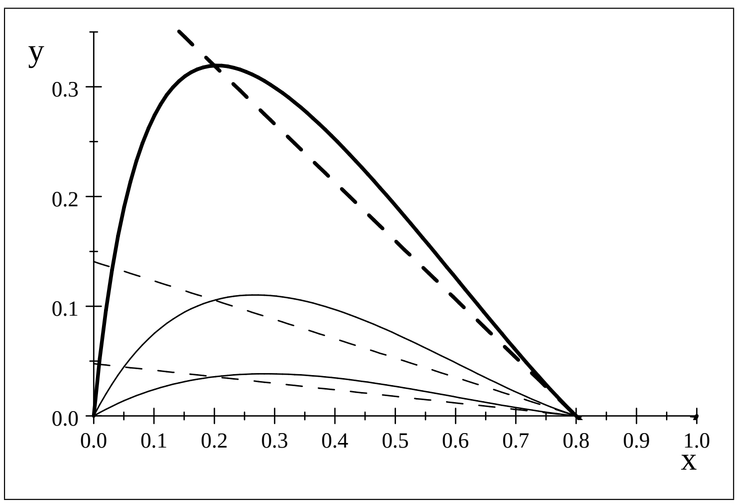

Figure C

Step 9 This proves Point IV.d): Note that for any given x, the y that maximizes

$$\min\{V^x(x,y,p),V^y(x,y,p)\}$$

is given by $y^* = 1-x$. Now, we furthermore know from previous points that first, $V^y(x,1-x,p)=V^x(x,1-x,p)$ and second, $V^x(x,1-x,p)$ is single peaked in x.■

10 Appendix D: Exposure with perfectly informative signals (proof of Proposition 4)

Step 1 Note that

$$\Gamma(x, x, 1-x, 1, c) = -\frac{1}{c^2}(2x-1)\left(-c^2+4x^2-4x+1\right).$$

and $V^x(x, x, 1-x, 1) = 2x-1$.

Step 2 Note first that $2x-1=c$ yields $x=\frac{1}{2}c+\frac{1}{2}$. Second, $\Gamma(x, x, 1-x, 1, c)=0$ also yields $x=\frac{1}{2}c+\frac{1}{2}$. Third, the equality $\frac{\partial(\Gamma(x,x,1-x,1,c))}{\partial x}=0$ yields the solution $x^*=\frac{1}{6}\sqrt{3}c+\frac{1}{2}$. The function $\Gamma(x, x, 1-x, 1, c)$ is increasing in x below this value and decreasing in x above. Fourth, note that $\frac{1}{6}\sqrt{3}c+\frac{1}{2}<(\frac{1}{2}c+\frac{1}{2})$.$\blacksquare$

11 Appendix E: Disclosure with continuous signals

11.1 Proof of Lemma 1

Step 1 i) is obvious. To show ii) we show that there is a unique s such that $\frac{d}{ds}(\alpha_S(s)-\alpha_R(s))=0$ (this assumes strict MLRP). Note that

$$\begin{aligned}\frac{d}{ds}(\alpha_S(s)-\alpha_R(s)) &= \frac{d}{ds}\left(\frac{\alpha_S}{\alpha_S+(1-\alpha_S)\frac{f(s|h)}{f(s|l)}}-\frac{\alpha_R}{\alpha_R+(1-\alpha_R)\frac{f(s|h)}{f(s|l)}}\right)\\ &= \left(\frac{\alpha_R(1-\alpha_R)}{\left(\alpha_R+(1-\alpha_R)\frac{f(s|h)}{f(s|l)}\right)^2}-\frac{\alpha_S(1-\alpha_S)}{\left(\alpha_S+(1-\alpha_S)\frac{f(s|h)}{f(s|l)}\right)^2}\right)\frac{d}{ds}\frac{f(s|h)}{f(s|l)}.\end{aligned}$$

So consider the solution to

$$\alpha_R(1-\alpha_R)\left(\alpha_S+(1-\alpha_S)\frac{f(s|h)}{f(s|l)}\right)^2=\alpha_S(1-\alpha_S)\left(\alpha_R+(1-\alpha_R)\frac{f(s|h)}{f(s|l)}\right)^2.$$

Both sides are increasing in s, but we claim that they increase at different rates. To see this, note that

$$\begin{aligned} & \frac{d}{ds}\alpha_R\left(1-\alpha_R\right)\left(\alpha_S+\left(1-\alpha_S\right)\frac{f(s|h)}{f(s|l)}\right)^2 \\ = \;& 2\alpha_R\left(1-\alpha_R\right)\left(1-\alpha_S\right)\left(\alpha_S+\left(1-\alpha_S\right)\frac{f(s|h)}{f(s|l)}\right)\frac{d}{ds}\frac{f(s|h)}{f(s|l)}, \\ & \frac{d}{ds}\alpha_S\left(1-\alpha_S\right)\left(\alpha_R+\left(1-\alpha_R\right)\frac{f(s|h)}{f(s|l)}\right)^2 \\ = \;& 2\alpha_S\left(1-\alpha_R\right)\left(1-\alpha_S\right)\left(\alpha_R+\left(1-\alpha_R\right)\frac{f(s|h)}{f(s|l)}\right)\frac{d}{ds}\frac{f(s|h)}{f(s|l)}. \end{aligned}$$

The result follows by assumption as

$$\begin{aligned} & 2\alpha_R\left(1-\alpha_R\right)\left(1-\alpha_S\right)\left(\alpha_S+\left(1-\alpha_S\right)\frac{f(s|h)}{f(s|l)}\right)\frac{d}{ds}\frac{f(s|h)}{f(s|l)} \\ \gtreqless \;& 2\alpha_S\left(1-\alpha_R\right)\left(1-\alpha_S\right)\left(\alpha_R+\left(1-\alpha_R\right)\frac{f(s|h)}{f(s|l)}\right)\frac{d}{ds}\frac{f(s|h)}{f(s|l)}, \end{aligned}$$

which is equivalent to

$$\alpha_R\alpha_S+\alpha_R\left(1-\alpha_S\right)\frac{f(s|h)}{f(s|l)} \gtreqless \alpha_R\alpha_S+\alpha_S\left(1-\alpha_R\right)\frac{f(s|h)}{f(s|l)}$$

which is equivalent to $\alpha_R \gtreqless \alpha_S$. Hence, $\widehat{s}$ must be unique. Existence follows from continuity and (i) together with $\Delta(\widetilde{s}) = |\alpha_S - \alpha_R| > 0$.

Step 2 To show (iii), define $\alpha_{\max} = \max\{\alpha_S, \alpha_R\}$ and $\alpha_{\min} = \min\{\alpha_S, \alpha_R\}$ such that $\Delta(s) = \alpha_{\max}(s) - \alpha_{\min}(s)$. We then have:

$$\begin{aligned} \frac{d}{ds}\Delta(\widetilde{s}) \;&=\; \left(\frac{\alpha_{\min}\left(1-\alpha_{\min}\right)}{\left(\alpha_{\min}+\left(1-\alpha_{\min}\right)\right)^2}-\frac{\alpha_{\max}\left(1-\alpha_{\max}\right)}{\left(\alpha_{\max}+\left(1-\alpha_{\max}\right)\right)^2}\right)\frac{d}{ds}\frac{f(\widetilde{s}|h)}{f(\widetilde{s}|l)} \gtreqless 0 \\ & \iff \alpha_{\min}\left(1-\alpha_{\min}\right) \gtreqless \alpha_{\max}\left(1-\alpha_{\max}\right), \end{aligned}$$

and $\widetilde{s} < (>)\widehat{s}$ if $\alpha_{\min}$ is less extreme than $\alpha_{\max}$.■

11.2 Proof of Proposition 5

Step 1 Denote the set of signals by Ψ. Denote the (sub)set of the set of signals Ψ that is being disclosed by Ψ^d and the complement by Ψ^{nd}. From R's point of view, S does not disclose an observed signal with probability

$$\Pr_R(s \in \Psi^{nd}) = \alpha_R \int_{\Psi^{nd}} f(s|l)ds + (1-\alpha_R) \int_{\Psi^{nd}} f(s|h)ds.$$

Hence, when S does not disclose, R's posterior is

$$\begin{aligned}\alpha_R(nd) &= \frac{\varphi}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}\int_{\Psi^{nd}}\left(\alpha_R f(s|l)+(1-\alpha_R)f(s|h)\right)\alpha_R(s)ds \\ &\quad +\frac{(1-\varphi)}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}\alpha_R \\ &= \frac{\varphi\int_{\Psi^{nd}} f(s|l)ds+(1-\varphi)}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}\alpha_R.\end{aligned}$$

Similarly, R's belief about S's posterior in this case is

$$\begin{aligned}\alpha_{RS}(nd) &= \frac{\varphi}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}\int_{\Psi^{nd}}\left(\alpha_R f(s|l)+(1-\alpha_R)f(s|h)\right)\alpha_S(s)ds \\ &\quad +\frac{(1-\varphi)}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}\alpha_S.\end{aligned}$$

Step 2 The discrepancy in beliefs upon disclosure is given by $|\alpha_S(s)-\alpha_R(s)|$ and at s_1 and s_2 the sender must be indifferent between disclosure and non-disclosure. Hence, we require

$$|\alpha_S(s)-\alpha_R(s)| = |\alpha_{RS}(nd)-\alpha_R(nd)| \text{ for } s=s_1,s_2, \tag{3}$$

which directly implies from the preceding Lemma that $s_1 \leq \widehat{s} \leq s_2$, strictly if $s_1 < s_2$. Further, it then follows directly from

$$\begin{aligned}&\alpha_{RS}(nd)-\alpha_R(nd) \\ =\ &\frac{\varphi}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}\int_{s_1}^{s_2}\left(\alpha_R f(s|l)+(1-\alpha_R)f(s|h)\right)\left(\alpha_S(s)-\alpha_R(s)\right)ds \\ &+\frac{(1-\varphi)}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}\left(\alpha_S-\alpha_R\right)\end{aligned}$$

together with (3) that under the optimal disclosure rule we must have $\alpha_S(s_1)-\alpha_R(s_1) = \alpha_S(s_2)-\alpha_R(s_2) = \alpha_S(nd)-\alpha_R(nd) > \alpha_S-\alpha_R$, implying that the uninformative signal $\widetilde{s}$ is always disclosed. Thus, if $\widetilde{s}=\widehat{s}$, then there is full disclosure conditional on an available signal.

Step 3 It remains to be shown that for $\widetilde{s} \neq \widehat{s}$, we have $s_1 < \widehat{s} < s_2$. We will argue by contradiction. Suppose that $\widetilde{s} \neq \widehat{s}$ and that there is an equilibrium with full disclosure conditional on an available signal. In such an equilibrium, disclosing $s = \widehat{s}$ leads to the a perceived disagreement of $\alpha_S(\widehat{s}) - \alpha_R(\widehat{s})$ which, from $\widetilde{s} \neq \widehat{s}$, is strictly greater than the perceived disagreement without disclosure, the latter being given by $|\alpha_{RS}(nd) - \alpha_R(nd)| = \alpha_S - \alpha_R$. Hence, disclosure of $s = \widehat{s}$ cannot be optimal. The remaining results then follow from the preceding Lemma.■

11.3 Proof of Proposition 6

Outline of the proof Steps 1-3 introduce notation as well as two preliminary properties. Steps 4-7 prove Point i) while remaining steps prove Point ii).

Step 1 For notational convenience, let us denote in what follows the priors as α_k^0, $k = \{S, R\}$. Let $\alpha_{RS}^{ND}(s_1, s_2)$ denote R's belief about S's posterior given no disclosure, the non-disclosure interval being given by (s_1, s_2). Similarly, let $\alpha_R^{ND}(s_1, s_2)$ denote R's posterior given no disclosure and the non-disclosure interval (s_1, s_2). We define the following normalized perceived disagreement functions:

$$\begin{aligned}\widetilde{\Delta}^{ND}(s_1, s_2) &\equiv \alpha_{RS}^{ND}(s_1, s_2) - \alpha_R^{ND}(s_1, s_2) - (\alpha_S^0 - \alpha_R^0),\\ \widetilde{\Delta}(s) &\equiv \alpha_S(s) - \alpha_R(s) - (\alpha_S^0 - \alpha_R^0).\end{aligned}$$

Given that R assigns probability α_R to state l $(= 0)$, she attributes the following unconditional pdf to signals:

$$\widetilde{f}(s) \equiv \alpha_R f(s|l) + (1 - \alpha_R) f(s|h).$$

Finally, define $LR(s) \equiv \frac{f(s|h)}{f(s|l)}$.

Step 2 It holds true that $\frac{\partial \widetilde{\Delta}^{ND}(s_1,s_2)}{\partial s_1} = \frac{\partial \widetilde{\Delta}^{ND}(s_1,s_2)}{\partial s_2} = 0$. Call this Property A. The proof is as follows.

$$
\begin{aligned}
&\widetilde{\Delta}^{ND}(s_1,s_2) = \frac{\varphi}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}\int_{s_1}^{s_2}\left(\alpha_S(s)-\alpha_R(s)\right)\widetilde{f}(s)ds \\
&+\frac{(1-\varphi)}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}\left(\alpha_S^0-\alpha_R^0\right)-\left(\alpha_S^0-\alpha_R^0\right) \\
= \;& \frac{\varphi}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}\left(\int_{s_1}^{s_2}\left(\alpha_S(s)-\alpha_R(s)\right)\widetilde{f}(s)ds-\Pr_R(s\in\Psi^{nd})\left(\alpha_S^0-\alpha_R^0\right)\right) \\
= \;& \frac{\varphi}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})} \\
&\left(\int_{s_1}^{s_2}\left(\alpha_S(s)-\alpha_R(s)\right)\widetilde{f}(s)ds-\Pr_R(s\in\Psi^{nd})\frac{1}{\Pr_R(s\in\Psi^{nd})}\int_{s1}^{s2}\left(\alpha_S^0-\alpha_R^0\right)\widetilde{f}(s)ds\right) \\
= \;& \frac{\varphi}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}\int_{s_1}^{s_2}(\alpha_S(s)-\alpha_R(s)-\left(\alpha_S^0-\alpha_R^0\right))\widetilde{f}(s)ds. \qquad (4)
\end{aligned}
$$

Denote $\eta(s_1)=\frac{\varphi}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}$ so that

$$
\begin{aligned}
\eta'(s_1) &= -\left(\frac{\varphi}{(1-\varphi)+\varphi\Pr_R(s\in\Psi^{nd})}\right)^2\frac{\partial\Pr_R(s\in\Psi^{nd})}{\partial s_1} \\
&= \eta^2\widetilde{f}(s_1).
\end{aligned}
$$

Then, taking the derivative of the whole term for $\widetilde{\Delta}^{ND}(s_1^*,s_2)$ given in (4) we obtain

$$
\begin{aligned}
\frac{\partial\widetilde{\Delta}^{ND}(s_1,s_2)}{\partial s_1} &= \eta'(s_1)\int_{s_1}^{s_2}(\alpha_S(s)-\alpha_R(s)-\left(\alpha_S^0-\alpha_R^0\right))\widetilde{f}(s)ds \\
&\quad -\eta(s_1)(\alpha_S(s_1)-\alpha_R(s_1)-\left(\alpha_S^0-\alpha_R^0\right))\widetilde{f}(s_1) \\
&= \eta(s_1)^2\widetilde{f}(s_1)\int_{s_1}^{s_2}(\alpha_S(s)-\alpha_R(s)-\left(\alpha_S^0-\alpha_R^0\right))\widetilde{f}(s)ds \\
&\quad -\eta(s_1)(\alpha_S(s_1)-\alpha_R(s_1)-\left(\alpha_S^0-\alpha_R^0\right))\widetilde{f}(s_1) \\
&= \eta(s_1)\widetilde{f}(s_1)\begin{pmatrix}\eta(s_1)\int_{s_1}^{s_2}(\alpha_S(s)-\alpha_R(s)-\left(\alpha_S^0-\alpha_R^0\right))\widetilde{f}(s)ds \\ -(\alpha_S(s_1)-\alpha_R(s_1)-\left(\alpha_S^0-\alpha_R^0\right))\end{pmatrix} \\
&= \eta(s_1)\widetilde{f}(s_1)\left(\widetilde{\Delta}^{ND}(s_1,s_2)-\widetilde{\Delta}(s_1)\right).
\end{aligned}
$$

At the same time, the term in brackets is 0 by (3). Thus,

$$
\frac{\partial\widetilde{\Delta}^{ND}(s_1,s_2)}{\partial s_1}=0.
$$

The claim for $\frac{\partial \widetilde{\Delta}^{ND}(s_1,s_2)}{\partial s_2}$ proceeds symmetrically.

Step 3 It holds true that for any $s > s_1$, $\frac{\frac{\partial \alpha_S(s)}{\partial \alpha_S^0}-1}{\widetilde{\Delta}(s)} > \frac{\frac{\partial \alpha_S(s_1)}{\partial \alpha_S^0}-1}{\widetilde{\Delta}(s_1)}$. Call this Property B. The proof is as follows. The sufficient condition for the claim is that function $\delta(s)$ defined by

$$\delta(s) \equiv \frac{\frac{\partial \alpha_S(s)}{\partial \alpha_S^0} - 1}{\widetilde{\Delta}(s)} \tag{5}$$

is increasing. We have

$$\frac{\partial \delta(s)}{\partial s} = \frac{\frac{\partial \alpha_S^2(s)}{\partial \alpha_S^0 \partial s}\widetilde{\Delta}(s) - \widetilde{\Delta}'(s)\left(\frac{\partial \alpha_S(s)}{\partial \alpha_S^0} - 1\right)}{\widetilde{\Delta}^2(s)}. \tag{6}$$

We thus need to show that the numerator is positive. Taking the corresponding derivatives and simplifying we obtain

$$\frac{\partial \alpha_S^2(s)}{\partial \alpha_S^0 \partial s}\widetilde{\Delta}(s) - \widetilde{\Delta}'(s)\left(\frac{\partial \alpha_S(s)}{\partial \alpha_S^0} - 1\right) \tag{7}$$

$$= \frac{((\alpha_R^0 - \alpha_S^0)^2(1-LR(s))^2(\alpha_R^0\left(\alpha_S^0\right)^2 + (1-\alpha_R^0)(1-\alpha_S^0)^2 LR^2(s))LR'(s)}{\left(\alpha_R^0 + LR(s) - \alpha_R^0 LR(s)\right)^2(\alpha_S^0 + LR(s) - \alpha_S^0 LR(s))^4} > 0. \tag{8}$$

This implies that $\frac{\partial \delta(s)}{\partial s}$ given by (6) is positive, which concludes the proof of Property B.

Step 4 Point i) is directly proved in steps 4-7. Recall that i) states that s_2 is increasing in α_S^0 for $\alpha_S^0 > \alpha_R^0$. Note first that in the proof of our first Proposition on disclosure with continuous signals, we have established the equilibrium condition

$$\alpha_S(s_1) - \alpha_R(s_1) = \alpha_S(s_2) - \alpha_R(s_2). \tag{9}$$

This implicitly defines s_1 as a function of s_2 and α_S^0, and respectively s_2 as a function of s_1 and α_S^0. Denote the resulting implicit functions as $s_1^*(s_2)$ and $s_2^*(s_1)$. Let us define the following functions:

$$\gamma_1(s_1, \alpha_S^0) \equiv \widetilde{\Delta}^{ND}(s_1, s_2^*(s_1)) - \widetilde{\Delta}(s_1), \tag{10}$$

$$\gamma_2(s_2, \alpha_S^0) \equiv \widetilde{\Delta}^{ND}(s_1^*(s_2), s_2) - \widetilde{\Delta}(s_2). \tag{11}$$

Note that in equilibrium

$$\begin{aligned}\gamma_1(s_1,\alpha_S^0) &= 0, && (12)\\ \gamma_2(s_2,\alpha_S^0) &= 0. && (13)\end{aligned}$$

Then, by the implicit function theorem

$$\frac{\partial s_2}{\partial \alpha_S^0} = -\frac{\partial \gamma_2/\partial \alpha_S^0}{\partial \gamma_2/\partial s_2}. \tag{14}$$

In the next three steps, we establish three claims pertaining to the signs of the three derivatives contained in (14). The first two claims establish the signs of the RHS derivatives and the final claim draws the implication that $\frac{\partial s_2}{\partial \alpha_S^0} > 0$.

Step 5 It holds true that $\frac{\partial \gamma_2}{\partial s_2} > 0$. Call this claim 1. The proof is as follows. We have

$$\begin{aligned}\frac{\partial \gamma_2}{\partial s_2} &= \frac{d\widetilde{\Delta}^{ND}(s_1^*(s_2),s_2)}{ds_2} - \frac{\partial \widetilde{\Delta}(s_2)}{\partial s_2}\\ &= \frac{\partial \widetilde{\Delta}^{ND}(s_1^*,s_2)}{\partial s_1^*}\frac{\partial s_1^*(s_2)}{\partial s_2} + \frac{\partial \widetilde{\Delta}^{ND}(s_1^*(s_2),s_2)}{\partial s_2} - \frac{\partial \widetilde{\Delta}(s_2)}{\partial s_2}\\ &= -\frac{\partial \widetilde{\Delta}(s_2)}{\partial s_2}, && (15)\end{aligned}$$

where the last equality follows by Property A (see Step 2). At the same time, $\frac{\partial \widetilde{\Delta}(s_2)}{\partial s_2} < 0$ since $s_2 > \widehat{s}$ (which, in turn, straightforwardly follows from Lemma 1). Hence, we may conclude that $\frac{\partial \gamma_2}{\partial s_2} > 0$.

Step 6 It holds true that $\frac{\partial \gamma_2}{\partial \alpha_S^0} < 0$. Call this claim 2. The proof is as follows. First, let us derive the following result:

$$\frac{\partial \alpha_S(s)}{\partial \alpha_S^0} - \frac{\partial \alpha_S(s_2)}{\partial \alpha_S^0} < 0 \text{ for any } s < s_2.$$

Taking the corresponding derivatives and simplifying we obtain

$$\begin{aligned}&\frac{\partial \alpha_S(s)}{\partial \alpha_S^0} - \frac{\partial \alpha_S(s_2)}{\partial \alpha_S^0} && (16)\\ =\ &\frac{(LR(s_2)-LR(s))}{\left(\alpha_S^0 + LR(s) - \alpha_S^0 LR(s)\right)^2\left(\alpha_S^0 + LR(s) - \alpha_S^0 LR(s)\right)^2} && (17)\\ &\times\left(\left(1-\alpha_S^0\right)^2 LR(s_1)LR(s_2) - \left(\alpha_S^0\right)^2\right). && (18)\end{aligned}$$

The fraction is clearly positive. Consider the remaining term. From the equilibrium condition $\widetilde{\Delta}(s_2) = \widetilde{\Delta}(s_1)$ we obtain

$$\begin{aligned} \widetilde{\Delta}(s_2) &= \widetilde{\Delta}(s_1), \\ \alpha_S(s_2) - \alpha_R(s_2) &= \alpha_S(s_1) - \alpha_R(s_1), \\ LR(s_2) &= \frac{\alpha_S^0 \alpha_R^0}{(1-\alpha_S^0)(1-\alpha_R^0) LR(s_1)}. \end{aligned}$$

Substituting this in the considered term and simplifying we obtain

$$\begin{aligned} & \left(1-\alpha_S^0\right)^2 LR(s_1) LR(s_2) - \left(\alpha_S^0\right)^2 \\ = & -\frac{\alpha_S^0(\alpha_S^0 - \alpha_R^0)}{1-\alpha_R^0} < 0. \end{aligned}$$

Getting back to (16), we eventually obtain

$$\frac{\partial \alpha_S(s)}{\partial \alpha_S^0} - \frac{\partial \alpha_S(s_2)}{\partial \alpha_S^0} < 0 \text{ for any } s < s_2. \tag{19}$$

Next, we have

$$\begin{aligned} \frac{\partial \gamma_2}{\partial \alpha_S^0} &= \frac{\partial \widetilde{\Delta}^{ND}(s_1^*, s_2)}{\partial s_1^*} \frac{\partial s_1^*}{\partial \alpha_S^0} + \frac{\partial \widetilde{\Delta}^{ND}(s_1^*, s_2)}{\partial \alpha_S^0} - \frac{\partial \widetilde{\Delta}(s_2)}{\partial \alpha_S^0} \\ &= \frac{\partial \widetilde{\Delta}^{ND}(s_1^*, s_2)}{\partial \alpha_S^0} - \frac{\partial \widetilde{\Delta}(s_2)}{\partial \alpha_S^0}, \end{aligned} \tag{20}$$

where the last equality is by Property A (see Step 2). Further, given the expression for $\widetilde{\Delta}^{ND}$ in (4), we have:

$$\frac{\partial \widetilde{\Delta}^{ND}(s_1^*, s_2)}{\partial \alpha_S^0} - \frac{\partial \widetilde{\Delta}(s_2)}{\partial \alpha_S^0} = \frac{\varphi}{(1-\varphi) + \varphi \Pr_R(s \in \Psi^{nd})} \int_{s_1}^{s_2} \left(\frac{\partial \alpha_S(s)}{\partial \alpha_S^0} - 1 \right) \widetilde{f}(s) ds - \left(\frac{\partial \alpha_S(s_2)}{\partial \alpha_S^0} - 1 \right).$$

At the same time, since $\frac{\partial \alpha_S(s_2)}{\partial \alpha_S^0} - 1 > \frac{\partial \alpha_S(s)}{\partial \alpha_S^0} - 1$ for any s by (20), it is trivially true that the whole expression is negative (since the first term is a weighted average of $\frac{\partial \alpha_S(s)}{\partial \alpha_S^0} - 1$ for $s \in (s_1, s_2)$). Hence, we may conclude that $\frac{\partial \gamma_2}{\partial \alpha_S^0} < 0$.

Step 7 It holds true that $\frac{\partial s_2}{\partial \alpha_S^0} > 0$. This final claim follows from (14), Claim 1 and Claim 2.

Step 8 We now prove Point ii) of the Proposition. Recall that ii) states that s_1 is increasing in α_S^0 for $\alpha_S^0 > \alpha_R^0$. Analogously to the previous case, the implicit function theorem yields

$$\frac{\partial s_1}{\partial \alpha_S^0} = -\frac{\partial \gamma_1 / \partial \alpha_S^0}{\partial \gamma_1 / \partial s_1}. \tag{21}$$

Using the same steps as in (15), we obtain

$$\frac{\partial \gamma_1}{\partial s_1} = -\frac{\partial \widetilde{\Delta}(s_1)}{\partial s_1} < 0, \tag{22}$$

where the inequality follows from $s_1 < \widehat{s}$ (which, in turn, straightforwardly follows from Lemma 1). At the same time, analogously to (20),

$$\begin{aligned} \frac{\partial \gamma_1}{\partial \alpha_S^0} &= \frac{\partial \widetilde{\Delta}^{ND}(s_1, s_2^*)}{\partial \alpha_S^0} - \frac{\partial \widetilde{\Delta}(s_1)}{\partial \alpha_S^0} \\ &= \frac{\varphi}{(1-\varphi) + \varphi \Pr_R(s \in \Psi^{nd})} \int_{s_1}^{s_2} \frac{\partial \widetilde{\Delta}(s)}{\partial \alpha_S^0} \widetilde{f}(s) ds - \frac{\partial \widetilde{\Delta}(s_1)}{\partial \alpha_S^0}. \end{aligned} \tag{23}$$

Let us show that this expression is positive. From the equilibrium condition $\gamma_1 = 0$ we obtain

$$\frac{\varphi}{(1-\varphi) + \varphi \Pr_R(s \in \Psi^{nd})} = \widetilde{\Delta}(s_1) \frac{1}{\int_{s_1}^{s_2} \widetilde{\Delta}(s) \widetilde{f}(s) ds}.$$

Hence,

$$\begin{aligned} &\frac{\varphi}{(1-\varphi) + \varphi \Pr_R(s \in \Psi^{nd})} \int_{s_1}^{s_2} \frac{\partial \widetilde{\Delta}(s)}{\partial \alpha_S^0} \widetilde{f}(s) ds \\ &= \widetilde{\Delta}(s_1) \frac{\int_{s_1}^{s_2} \frac{\partial \widetilde{\Delta}(s)}{\partial \alpha_S^0} \widetilde{f}(s) ds}{\int_{s_1}^{s_2} \widetilde{\Delta}(s) \widetilde{f}(s) ds}. \end{aligned} \tag{24}$$

At the same time, by Property B (see Step 3) we have for any $s > s_1$

$$\frac{\frac{\partial \widetilde{\Delta}(s_1)}{\partial \alpha_S^0}}{\widetilde{\Delta}(s_1)} < \frac{\frac{\partial \widetilde{\Delta}(s)}{\partial \alpha_S^0}}{\widetilde{\Delta}(s)}$$

which in turn rewrites as

$$\widetilde{\Delta}(s) < \frac{\frac{\partial \widetilde{\Delta}(s)}{\partial \alpha_S^0} \widetilde{\Delta}(s_1)}{\frac{\partial \widetilde{\Delta}(s_1)}{\partial \alpha_S^0}}.$$

Consequently,

$$\begin{aligned}
\widetilde{\Delta}(s_1)\frac{\int_{s_1}^{s_2}\frac{\partial\widetilde{\Delta}(s)}{\partial\alpha_S^0}\widetilde{f}(s)ds}{\int_{s_1}^{s_2}\widetilde{\Delta}(s)\widetilde{f}(s)ds} &> \widetilde{\Delta}(s_1)\frac{\int_{s_1}^{s_2}\frac{\partial\widetilde{\Delta}(s)}{\partial\alpha_S^0}\widetilde{f}(s)ds}{\int_{s_1}^{s_2}\frac{\frac{\partial\widetilde{\Delta}(s)}{\partial\alpha_S^0}\widetilde{\Delta}(s_1)}{\frac{\partial\widetilde{\Delta}(s_1)}{\partial\alpha_S^0}}\widetilde{f}(s)ds} \\
&= \widetilde{\Delta}(s_1)\frac{\int_{s_1}^{s_2}\frac{\partial\widetilde{\Delta}(s)}{\partial\alpha_S^0}\widetilde{f}(s)ds}{\frac{\widetilde{\Delta}(s_1)}{\frac{\partial\widetilde{\Delta}(s_1)}{\partial\alpha_S^0}}\int_{s_1}^{s_2}\frac{\partial\widetilde{\Delta}(s)}{\partial\alpha_S^0}\widetilde{f}(s)ds} = \frac{\partial\widetilde{\Delta}(s_1)}{\partial\alpha_S^0}. \qquad (25)
\end{aligned}$$

Eventually, we obtain

$$\begin{aligned}
\frac{\partial\gamma_1}{\partial\alpha_S^0} &= \frac{\varphi}{(1-\varphi)+\varphi\,\mathrm{Pr}_R(s\in\Psi^{nd})}\int_{s_1}^{s_2}\frac{\partial\widetilde{\Delta}(s)}{\partial\alpha_S^0}\widetilde{f}(s)ds - \frac{\partial\widetilde{\Delta}(s_1)}{\partial\alpha_S^0} \\
&= \widetilde{\Delta}(s_1)\frac{\int_{s_1}^{s_2}\frac{\partial\widetilde{\Delta}(s)}{\partial\alpha_S^0}\widetilde{f}(s)ds}{\int_{s_1}^{s_2}\widetilde{\Delta}(s)\widetilde{f}(s)ds} - \frac{\partial\widetilde{\Delta}(s_1)}{\partial\alpha_S^0} \\
&> \frac{\partial\widetilde{\Delta}(s_1)}{\partial\alpha_S^0} - \frac{\partial\widetilde{\Delta}(s_1)}{\partial\alpha_S^0} = 0, \qquad (26)
\end{aligned}$$

where the first equality follows from (23), the second equality follows from (24), and the inequality follows from (25). This together with (21) and (22) implies $\frac{\partial s_1}{\partial\alpha_S^0} > 0$.

12 Appendix F: Strategic disclosure and signaling priors (proof of Proposition 7)

Step 1 Note first that the expected payoff of R, given beliefs defined by a distribution $\{\alpha, 1-\alpha\}$ over $\{0,1\}$, is given by minus the variance of ω, which is given by $P(\omega = 0)P(\omega = 1) = \alpha(1-\alpha)$.

Step 2 We first consider the experiment in which only 1-signals are disclosed (denoted E_1) . Let us denote by $\sigma = \varnothing$ the event in which no signal is received by S. Denote by $\Pi^R(1, E_1)$ ($\Pi^R(\varnothing, E_1)$) the expected payoff of R given disclosure of a 1-signal (no signal).

The expected utility of R conditional on facing an E1 experiment is given by:

$$P(\sigma=1)\Pi^R(1,E_1)+P(\sigma=\varnothing)\Pi^R(\varnothing,E_1).$$

Note first that if R's prior distribution of the state is α, then

$$P(\sigma=1)=\varphi\left(\alpha(1-p)+(1-\alpha)p\right)$$

and

$$P(\sigma=\varnothing)=\varphi\left(\alpha p+(1-\alpha)(1-p)\right)+(1-\varphi).$$

Note next that the posterior distribution of the state after a 1-signal is given by:

$$\frac{\alpha(1-p)}{\alpha(1-p)+(1-\alpha)p},1-\frac{\alpha(1-p)}{\alpha(1-p)+(1-\alpha)p}.$$

It follows that

$$\begin{aligned}\Pi^R(1,E_1) &= -\left(\frac{\alpha(1-p)}{\alpha(1-p)+(1-\alpha)p}\right)\left(1-\frac{\alpha(1-p)}{\alpha(1-p)+(1-\alpha)p}\right)\\ &= -p\alpha\left(\alpha-1\right)\frac{p-1}{\left(p+\alpha-2p\alpha\right)^2}.\end{aligned}$$

Note that the posterior distribution of the state after no disclosure is given by:

$$P(\omega=0\,|\varnothing,\ E_1)\equiv\left(\frac{\alpha\left(\varphi p+(1-\varphi)\right)}{\alpha\left(\varphi p+(1-\varphi)\right)+(1-\alpha)\left(\varphi(1-p)+(1-\varphi)\right)}\right),1-P(\omega=0\,|\varnothing,\ E_1).$$

It follows that

$$\begin{aligned}&\Pi^R(\varnothing,E_1)\\ =\;& -\left(P(\omega=0\,|\varnothing,\ E_1)\right)\left(1-P(\omega=0\,|\varnothing,\ E_1)\right)\\ =\;& \alpha\frac{\alpha-1}{\left(p\varphi+\alpha\varphi-2p\alpha\varphi-1\right)^2}\left(p^2\varphi^2-p\varphi^2+\varphi-1\right).\end{aligned}$$

We may now conclude and define the value of the experiment E_1 for R if the latter has prior α :

$$
\begin{aligned}
& \Pi_1^R(\alpha,\varphi,p) \\
= \; & -\varphi\left(\alpha(1-p)+(1-\alpha)p\right)\left(\frac{\alpha(1-p)}{\alpha(1-p)+(1-\alpha)p}\right)\left(1-\frac{\alpha(1-p)}{\alpha(1-p)+(1-\alpha)p}\right) \\
& -(1-\varphi\left(\alpha(1-p)+(1-\alpha)p\right)) \\
& \times\left(\begin{array}{c}\left(\frac{\alpha(\varphi p+(1-\varphi))}{\alpha(\varphi p+(1-\varphi))+(1-\alpha)(\varphi(1-p)+(1-\varphi))}\right) \\ \left(1-\left(\frac{\alpha(\varphi p+(1-\varphi))}{\alpha(\varphi p+(1-\varphi))+(1-\alpha)(\varphi(1-p)+(1-\varphi))}\right)\right)\end{array}\right).
\end{aligned}
$$

Note that the expression simplifies significantly to the following:

$$
\Pi_1^R(\alpha,\varphi,p)=-\alpha\frac{(\alpha-1)\left(p+\alpha-2p\alpha-\alpha\varphi-p^2\varphi+2p\alpha\varphi\right)}{(p+\alpha-2p\alpha)\left(p\varphi+\alpha\varphi-2p\alpha\varphi-1\right)}.
$$

Step 3 We first consider the experiment in which only 0-signals are disclosed (denoted E_0) . Denote by $\Pi^R(0,E_0)$ ($\Pi^R(\varnothing,E_0)$) the expected payoff of R given disclosure of a 0-signal (no signal). The expected utility of R conditional on facing an E0 experiment is given by:

$$
P(\sigma=0)\Pi^R(0,E_0)+P(\sigma=\varnothing)\Pi^R(\varnothing,E_0)
$$

Note first that if R's prior distribution of the state is α, then

$$
P(\sigma=0)=\varphi\left(\alpha p+(1-\alpha)(1-p)\right)
$$

and

$$
P(\sigma=\varnothing)=\varphi+(1-\varphi)\alpha(1-p)+(1-\alpha)p.
$$

Note that the posterior distribution of the state after a 0-signal is given by:

$$
\frac{\alpha p}{\alpha p+(1-\alpha)(1-p)},1-\frac{\alpha p}{\alpha p+(1-\alpha)(1-p)}
$$

It follows that:

$$
\begin{aligned}
\Pi^R(0,E_0) &= -\left(\frac{\alpha p}{\alpha p+(1-\alpha)(1-p)}\right)\left(1-\frac{\alpha p}{\alpha p+(1-\alpha)(1-p)}\right) \\
&= p\alpha\left(\alpha-1\right)\frac{p-1}{(p+\alpha-2p\alpha-1)^2}.
\end{aligned}
$$

Note that the posterior distribution of the state after no disclosure is given by:

$$P(\omega = 0\,|\varnothing, E_0\,) \equiv \left(\frac{\alpha\,(\varphi(1-p)+(1-\varphi))}{\alpha\,(\varphi(1-p)+(1-\varphi))+(1-\alpha)(\varphi p+(1-\varphi))}\right), 1-P(\omega = 0\,|\varnothing, E_0\,).$$

It follows that:

$$\begin{aligned}
&\Pi^R(\varnothing, E_0) \\
= \; & -\left(\frac{\alpha\,(\varphi(1-p)+(1-\varphi))}{\alpha\,(\varphi(1-p)+(1-\varphi))+(1-\alpha)(\varphi p+(1-\varphi))}\right) \\
& \left(1-\left(\frac{\alpha\,(\varphi(1-p)+(1-\varphi))}{\alpha\,(\varphi(1-p)+(1-\varphi))+(1-\alpha)(\varphi p+(1-\varphi))}\right)\right) \\
= \; & -\alpha\,(\alpha-1)\,(p\varphi-1)\,\frac{p\varphi-\varphi+1}{(p\varphi-\varphi+\alpha\varphi-2p\alpha\varphi+1)^2}.
\end{aligned}$$

We are now ready to conclude and define the value of the experiment E_0 to R if her prior is α :

$$\begin{aligned}
&\Pi_0^R(\alpha, \varphi, p) \\
= \; & -\varphi\,(\alpha p+(1-\alpha)(1-p))\left(\frac{\alpha p}{\alpha p+(1-\alpha)(1-p)}\right)\left(1-\frac{\alpha p}{\alpha p+(1-\alpha)(1-p)}\right) \\
& -(1-\varphi\,(\alpha p+(1-\alpha)(1-p))) \\
& \times\left(\begin{array}{c} \left(\frac{\alpha(\varphi(1-p)+(1-\varphi))}{\alpha(\varphi(1-p)+(1-\varphi))+(1-\alpha)(\varphi p+(1-\varphi))}\right) \\ \left(1-\left(\frac{\alpha(\varphi(1-p)+(1-\varphi))}{\alpha(\varphi(1-p)+(1-\varphi))+(1-\alpha)(\varphi p+(1-\varphi))}\right)\right)\end{array}\right).
\end{aligned}$$

which simplifies significantly to:

$$\Pi_0^R(\alpha, \varphi, p) = \alpha\frac{(\alpha-1)\left(p+\alpha+\varphi-2p\alpha-2p\varphi-\alpha\varphi+p^2\varphi+2p\alpha\varphi-1\right)}{(p+\alpha-2p\alpha-1)\,(p\varphi-\varphi+\alpha\varphi-2p\alpha\varphi+1)}.$$

Step 4 We may now wrap up. Using the obtained formulas, we have:

$$\begin{aligned}
&\Pi_0^R(\alpha, \varphi, p) - \Pi_1^R(\alpha, \varphi, p) \\
= \; & -\alpha^2\varphi\,(\alpha-1)^2\,(\varphi-1)\,(2p-1)^3 \\
& \frac{2\alpha-1}{\begin{array}{c}\left(-4p^2\alpha^2+4p^2\alpha-p^2+4p\alpha^2-4p\alpha+p-\alpha^2+\alpha\right) \\ \left(4p^2\alpha^2\varphi^2-4p^2\alpha\varphi^2+p^2\varphi^2-4p\alpha^2\varphi^2+4p\alpha\varphi^2-p\varphi^2+\alpha^2\varphi^2-\alpha\varphi^2+\varphi-1\right)\end{array}}.
\end{aligned}$$

Now, simply not that for any p, φ, α,

$$\left(-4p^2\alpha^2 + 4p^2\alpha - p^2 + 4p\alpha^2 - 4p\alpha + p - \alpha^2 + \alpha\right)$$

has the same sign. Similarly, for any p, φ, α,

$$\left(4p^2\alpha^2\varphi^2 - 4p^2\alpha\varphi^2 + p^2\varphi^2 - 4p\alpha^2\varphi^2 + 4p\alpha\varphi^2 - p\varphi^2 + \alpha^2\varphi^2 - \alpha\varphi^2 + \varphi - 1\right)$$

has the same sign. The argument is as follows. Solving

$$-4p^2\alpha^2 + 4p^2\alpha - p^2 + 4p\alpha^2 - 4p\alpha + p - \alpha^2 + \alpha = 0$$

for α yields the solutions $\frac{p}{2p-1}$ and $\frac{1}{2p-1}(p-1)$, the first of which is < 0 for any p and the second of which is > 1 for any p. Similarly, solving

$$4p^2\alpha^2\varphi^2 - 4p^2\alpha\varphi^2 + p^2\varphi^2 - 4p\alpha^2\varphi^2 + 4p\alpha\varphi^2 - p\varphi^2 + \alpha^2\varphi^2 - \alpha\varphi^2 + \varphi - 1 = 0$$

for α yields the solutions $-\frac{1}{\varphi-2p\varphi}(-\varphi + p\varphi + 1)$ and $-\frac{p\varphi-1}{\varphi-2p\varphi}$. Define the functions:

$$t_1(p, \varphi) = -\frac{1}{\varphi - 2p\varphi}(-\varphi + p\varphi + 1)$$

$$t_2(p, \varphi) = -\frac{p\varphi - 1}{\varphi - 2p\varphi}$$

Note that $t_1(p, \varphi) > 1$ for any p, φ. To see this, note that $\frac{\partial t_1(p,\varphi)}{\partial \varphi} = -\frac{1}{\varphi^2(2p-1)}$ and that $t_1(p, 1) = \frac{p}{2p-1} > 1$. Note on the other hand that $t_2(p, \varphi) < 0$ for any p, φ. To see this, note that $\frac{\partial t_2(p,\varphi)}{\partial \varphi} = \frac{1}{\varphi^2(2p-1)}$ and that $t_2(p, 1) = \frac{1}{2p-1}(p-1) < 0$. We may conclude. It follows that the sign of $\Pi_0^R(\alpha, \varphi, p) - \Pi_1^R(\alpha, \varphi, p)$ is determined by the sign of $2\alpha - 1$. It holds true that $\Pi_0^R(\alpha, \varphi, p) - \Pi_1^R(\alpha, \varphi, p) > 0$ if $\alpha < \frac{1}{2}$ and $\Pi_0^R(\alpha, \varphi, p) - \Pi_1^R(\alpha, \varphi, p) < 0$ if $\alpha > \frac{1}{2}$. The figure below shows $\Pi_0^R(\alpha, .75, .85) - \Pi_1^R(\alpha, .75, .85)$ in continuous and

$\Pi_0^R(\alpha, .75, .65) - \Pi_1^R(\alpha, .75, .65)$ in dashed.

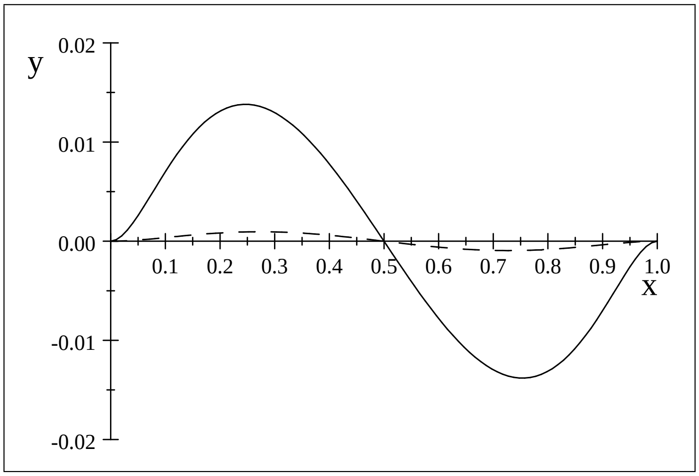

Figure D

Step 5 This proves Point b). Assume that the set of possible values of α_R (denoted Σ) is such that any possible value yields either the D0 or the D1 equilibrium in the disclosure subgame. Assume an equilibrium with truth-telling about her prior by R. Given our characterization appearing in Proposition 1, it must be that the set of priors Σ divides into two distinct sets of priors, Σ^- and Σ^+ such that 1) Σ^- contains all values such that $\alpha_R < 1 - \alpha_S$ and Σ^+ all values such that $\alpha_R > 1 - \alpha_S$ and 2) all priors in Σ^- yield equilibrium D1 in the disclosure subgame while all priors in Σ^+ yield equilibrium D0 in the disclosure subgame. Assume that in a putative truth-telling equilibrium, all members of Σ^- have an incentive to truth-tell. If this is the case, given Point a), it must be that all elements of Σ^- are larger than $\frac{1}{2}$ since they must prefer D1 over D0. This in turn implies that all priors in Σ^+ are larger than $\frac{1}{2}$, implying that prior types belonging to Σ^+ prefer D1 over D0 and will thus want to deviate to announcing that they belong to Σ^-. This yields a contradiction and thus proves that there cannot be an equilibrium in which R truthfully reveals her prior. ■

13 Appendix G: Strategic exposure and signaling priors with one-sided uncertainty (proof of Proposition 8(a))

Step 0 We here consider the case of public signal exposure followed by a stage in which both players simultaneously announce their priors. Recall that we assume that $\alpha_R = \frac{1}{2}$.

Step 1 Consider an equilibrium in which in stage 1, two messages are sent with positive probability, namely m_1 and m_2. There is a threshold α' s.t. S sends m_1 if her prior is below α' and m_2 otherwise. The conditional belief regarding α_S given each of these messages is given as follows. Message m_1 implies $\alpha_S \sim U[0, \alpha']$ while instead m_2 implies $\alpha_S \sim U[\alpha', 1]$. Given that her true prior is α_S, the expected utility of S after sending m_1 is

$$
\begin{aligned}
& U_S(m_1, \alpha_S, \alpha_R, \alpha', p) \\
= \; & -\Pr(\sigma = 0 \,|\alpha_S) E_S E_R \left| \frac{\alpha_S p}{\alpha_S p + (1-\alpha_S)(1-p)} - \frac{\alpha_R p}{\alpha_R p + (1-\alpha_R)(1-p)} \right| \\
-\Pr(\sigma = \; & 1 \,|\alpha_S) E_S E_R \left| \frac{\alpha_S (1-p)}{\alpha_S(1-p) + (1-\alpha_S) p} - \frac{\alpha_R (1-p)}{\alpha_R(1-p) + (1-\alpha_R) p} \right| \\
= \; & -(\alpha_S p + (1-p)(1-\alpha_S)) \int_0^{\alpha'} \left| \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} - \frac{\alpha_R p}{\alpha_R p + (1-\alpha_R)(1-p)} \right| \frac{1}{\alpha'} d\alpha \\
& -(p(1-\alpha_S) + (1-p)\alpha_S) \int_0^{\alpha'} \left| \frac{\alpha (1-p)}{\alpha(1-p) + (1-\alpha) p} - \frac{\alpha_R (1-p)}{\alpha_R(1-p) + (1-\alpha_R) p} \right| \frac{1}{\alpha'} d\alpha.
\end{aligned}
$$

Instead, her expected utility after sending m_2 is:

$$
\begin{aligned}
& U_S(m_2, \alpha_S, \alpha_R, \alpha', p) \\
= \; & -(\alpha_S p + (1-p)(1-\alpha_S)) \int_{\alpha'}^{1} \left| \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} - \frac{\alpha_R p}{\alpha_R p + (1-\alpha_R)(1-p)} \right| \frac{1}{1-\alpha'} d\alpha \\
& -(p(1-\alpha_S) + (1-p)\alpha_S) \int_{\alpha'}^{1} \left| \frac{\alpha (1-p)}{\alpha(1-p) + (1-\alpha) p} - \frac{\alpha_R (1-p)}{\alpha_R(1-p) + (1-\alpha_R) p} \right| \frac{1}{1-\alpha'} d\alpha.
\end{aligned}
$$

Step 2 The equilibrium characterized by α' exists if and only if the following is true:

$$\gamma(\alpha_S, \alpha_R, \alpha', p) = U_S(m_1, \alpha_S, \alpha_R, \alpha', p) - U_S(m_2, \alpha_S, \alpha_R, \alpha', p) \geq 0 \text{ iff } \alpha_S \leq \alpha'.$$

Given that it holds true that $\gamma(\alpha^S, \frac{1}{2}, \alpha', p)$ is strictly decreasing in α_S, the above condition is satisfied if and only if

$$\gamma(\alpha', \alpha_R, \alpha', p) = 0.$$

Step 3 Let us set $\alpha_R = \frac{1}{2}$. Clearly, $U_S(m_1, \alpha', \frac{1}{2}, \alpha')$ reduces to:

$$= -(\alpha' p + (1-p)(1-\alpha')) \int_0^{\alpha'} \left| \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} - p \right| \frac{1}{\alpha'} d\alpha$$
$$-(p(1-\alpha') + (1-p)\alpha') \int_0^{\alpha'} \left| \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} - (1-p) \right| \frac{1}{\alpha'} d\alpha$$

while instead $U_S(m_2, \alpha', \frac{1}{2}, \alpha')$ reduces to:

$$U_S(m_2, \alpha_S, \alpha_R, \alpha')$$
$$= -(\alpha' p + (1-p)(1-\alpha')) \int_{\alpha'}^{1} \left| \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} - p \right| \frac{1}{1-\alpha'} d\alpha$$
$$-(p(1-\alpha') + (1-p)\alpha') \int_{\alpha'}^{1} \left| \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} - (1-p) \right| \frac{1}{1-\alpha'} d\alpha.$$

Now, note that $U_S(m_1, \alpha', \frac{1}{2}, \alpha', p) = U_S(m_2, 1-\alpha', \frac{1}{2}, 1-\alpha', p)$ while $U_S(m_2, \alpha', \frac{1}{2}, \alpha', p) = U_S(m_1, 1-\alpha', \frac{1}{2}, 1-\alpha', p)$. It follows trivially that $\gamma(\alpha', \frac{1}{2}, \alpha', p) = 0$. This concludes the proof of the Proposition. We make two extra observations based on numerical examples and worth exploring more with an eye to obtain more general results. First, for p very high, keeping $\alpha_R = \frac{1}{2}$, there are multiple values of α' such that $\gamma(\alpha', \frac{1}{2}, \alpha', p) = 0$, implying the existence of multiple equilibria featuring a binary partition. Second, for α_R excessively far from $\frac{1}{2}$, there appears to not always exist a value of α' such that $\gamma(\alpha', \alpha_R, \alpha', p) = 0$.■

14 Appendix H: Strategic exposure and signaling priors with two-sided uncertainty (proof of Proposition 8(b))

Step 0 We consider the incentives of player S here, the incentives of R being identical. Given that S and R's priors are α_S and α_R, that the equilibrium threshold characterizing S's communication is α' and assuming that signal quality is p, the expected utility of S from sending m is denoted $U_S(m_1, \alpha_S, \alpha_R, \alpha', p)$. Define

$$\widetilde{\gamma}(\alpha_S, \alpha_R, \alpha', p) \equiv \int_0^1 U_S(m_1, \alpha', \alpha_R, \alpha', p) - U_S(m_2, \alpha', \alpha_R, \alpha', p) d\alpha_R.$$

Our proof is built in three overall steps. First, we note that the threshold α' is incentive compatible for S if and only if $\widetilde{\gamma}(\alpha', \alpha_R, \alpha', p) = 0$. Second, we prove that $\widetilde{\gamma}(\frac{1}{2}, \alpha_R, \frac{1}{2}, p) = 0$.

Finally, we prove that $\widetilde{\gamma}(\alpha_S, \alpha_R, \frac{1}{2}, p)$ is strictly decreasing in α_S. These three properties, taken together, prove our result.

Step 1 The threshold α' is incentive compatible for S if and only if

$$\int_0^1 U_S(m_1, \alpha_S, \alpha_R, \alpha', p) - U_S(m_2, \alpha_S, \alpha_R, \alpha', p) d\alpha_R \geq 0$$

if $\alpha_S < \alpha'$ while the relation is inversed if $\alpha_S > \alpha'$. Now, given that the above function is continuous and decreasing in α_S, it follows that $\widetilde{\gamma}(\alpha', \alpha_R, \alpha', p) = 0$ is a necessary and sufficient condition for α' to be incentive compatible for S.

Step 2 Suppose that S knows that the other player has with equal probability priors α or $1-\alpha$, assuming w.l.o.g. that $\alpha < \frac{1}{2}$. Now, note the following relations:

$$U_S(m_1, \alpha, 1-\alpha_R, \alpha, p) = U_S(m_2, 1-\alpha, \alpha_R, 1-\alpha, p) \tag{27}$$

and

$$U_S(m_2, \alpha, 1-\alpha_R, \alpha, p) = U_S(m_1, 1-\alpha, \alpha_R, 1-\alpha, p). \tag{28}$$

So let us consider the following expression:

$$\begin{aligned}\Delta(\alpha, \alpha_R, p) &= \frac{1}{2}\left[U_S(m_1, \alpha, \alpha_R, \alpha, p) - U_S(m_2, \alpha, \alpha_R, \alpha, p)\right] \\ &+ \frac{1}{2}\left[U_S(m_1, \alpha, 1-\alpha_R, \alpha, p) - U_S(m_2, \alpha, 1-\alpha_R, \alpha, p)\right].\end{aligned}$$

Using the above defined equalities (27) and (28), we may rewrite:

$$\begin{aligned}\Delta(\alpha, \alpha_R, p) &= \frac{1}{2}\left[U_S(m_1, \alpha, \alpha_R, \alpha, p) - U_S(m_2, \alpha, \alpha_R, \alpha, p)\right] \\ &+ \frac{1}{2}\left[U_S(m_2, 1-\alpha, \alpha_R, 1-\alpha, p) - U_S(m_1, 1-\alpha, \alpha_R, 1-\alpha, p)\right].\end{aligned}$$

Now, let us set $\alpha = \frac{1}{2}$. We obtain:

$$\begin{aligned}\Delta\left(\frac{1}{2}, \alpha_R, p\right) &= \frac{1}{2}\left[U_S(m_1, \frac{1}{2}, \alpha_R, \frac{1}{2}, p) - U_S(m_2, \frac{1}{2}, \alpha_R, \frac{1}{2}, p)\right] \\ &+ \frac{1}{2}\left[U_S(m_2, \frac{1}{2}, \alpha_R, \frac{1}{2}, p) - U_S(m_1, \frac{1}{2}, \alpha_R, \frac{1}{2}, p)\right],\end{aligned}$$

which equals 0 by definition.

Step 3 If player S's belief about the possible prior of her opponent is given by a uniform distribution on $[0,1]$, then they can also be described by a uniform distribution over pairs $\{\alpha, 1-\alpha\}$, for $\alpha \in \left[0, \frac{1}{2}\right]$. We may thus trivially write:

$$\begin{aligned} & \int_0^1 U_S(m_1, \frac{1}{2}, \alpha_R, \frac{1}{2}, p) - U_S(m_2, \frac{1}{2}, \alpha_R, \frac{1}{2}, p) d\alpha_R \\ = & \int_0^{\frac{1}{2}} \Delta\left(\frac{1}{2}, \alpha_R, p\right) d\alpha_R \end{aligned}$$

which thus equals 0 given that $\Delta\left(\frac{1}{2}, \alpha_R, p\right) = 0$ for any α_R.

Step 4 We now simply want to show that

$$\int_0^1 U_S(m_1, \alpha_S, \alpha_R, \frac{1}{2}, p) - U_S(m_2, \alpha_S, \alpha_R, \frac{1}{2}, p) d\alpha_R$$

is strictly decreasing in α_S. A sufficient condition for the above to be true is to show that for any given α_S, α_R, it holds true that $U_S(m_1, \alpha_S, \alpha_R, \frac{1}{2}, p) - U_S(m_2, \alpha_S, \alpha_R, \frac{1}{2}, p)$ is strictly decreasing in α_S. To see this, note that Leibniz' rule states that:

$$\begin{aligned} & \frac{\partial\left(\int_0^1 U_S(m_1, \alpha_S, \alpha_R, \frac{1}{2}, p) - U_S(m_2, \alpha_S, \alpha_R, \frac{1}{2}, p) d\alpha_R\right)}{\partial \alpha_S} \\ = & \int_0^1 \frac{\partial\left(U_S(m_1, \alpha_S, \alpha_R, \frac{1}{2}, p) - U_S(m_2, \alpha_S, \alpha_R, \frac{1}{2}, p)\right)}{\partial \alpha_S} d\alpha_R. \end{aligned}$$

Note that the following relation also holds:

$$U_S(m_2, \alpha_S, \alpha_R, \frac{1}{2}, p) = U_S(m_1, 1-\alpha_S, 1-\alpha_R, \frac{1}{2}, p).$$

We thus need to show now that $U_S(m_1, \alpha_S, \alpha_R, \alpha', p) - U_S(m_1, 1-\alpha_S, 1-\alpha_R, \alpha', p)$ is decreasing in α_S for any α_R. We have

$$\begin{aligned} & U_S(m_1, \alpha_S, \alpha_R, \alpha', p) - U_S(m_1, 1-\alpha_S, 1-\alpha_R, \alpha', p) \\ = & -(\alpha_S p + (1-p)(1-\alpha_S))\phi_1 - (p(1-\alpha_S) + (1-p)\alpha_S)\phi_2 \\ & +((1-\alpha_S)p + (1-p)\alpha_S)\phi_3 + (p\alpha_S + (1-p)(1-\alpha_S))\phi_4, \end{aligned}$$

where

$$\begin{aligned}
\phi_1 &= \int_0^{0.5} \left| \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} - \frac{\alpha_R p}{\alpha_R p + (1-\alpha_R)(1-p)} \right| \frac{1}{0.5} d\alpha, \\
\phi_2 &= \int_0^{0.5} \left| \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} - \frac{\alpha_R(1-p)}{\alpha_R(1-p) + (1-\alpha_R)p} \right| \frac{1}{0.5} d\alpha, \\
\phi_3 &= \int_0^{0.5} \left| \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} - \frac{(1-\alpha_R)p}{(1-\alpha_R)p + \alpha_R(1-p)} \right| \frac{1}{0.5} d\alpha, \\
\phi_4 &= \int_0^{0.5} \left| \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} - \frac{(1-\alpha_R)(1-p)}{(1-\alpha_R)(1-p) + \alpha_R p} \right| \frac{1}{0.5} d\alpha.
\end{aligned}$$

Note first that in general:

$$\begin{aligned}
& \frac{\partial \begin{pmatrix} (-(xp + (1-p)(1-x))\phi_1 - (p(1-x) + (1-p)x)\phi_2 \\ +((1-x)p + (1-p)x)\phi_3 + (px + (1-p)(1-x))\phi_4) \end{pmatrix}}{\partial x} \\
= \; & -(2p-1)\left(\phi_1 - \phi_2 + \phi_3 - \phi_4\right).
\end{aligned}$$

So in order to show that $U_S(m_1, \alpha_S, \alpha_R, \alpha', p) - U_S(m_1, 1-\alpha_S, 1-\alpha_R, \alpha', p)$ is decreasing in α_S for any given $\alpha_R \in [0,1]$, we need to show that for any α_R :

$$\begin{aligned}
& \int_0^{0.5} \left| \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} - \frac{\alpha_R p}{\alpha_R p + (1-\alpha_R)(1-p)} \right| \frac{1}{0.5} d\alpha \\
& - \int_0^{0.5} \left| \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} - \frac{\alpha_R(1-p)}{\alpha_R(1-p) + (1-\alpha_R)p} \right| \frac{1}{0.5} d\alpha \\
& + \int_0^{0.5} \left| \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} - \frac{(1-\alpha_R)p}{(1-\alpha_R)p + \alpha_R(1-p)} \right| \frac{1}{0.5} d\alpha \\
& - \int_0^{0.5} \left| \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} - \frac{(1-\alpha_R)(1-p)}{(1-\alpha_R)(1-p) + \alpha_R p} \right| \frac{1}{0.5} d\alpha \\
> \; & 0.
\end{aligned}$$

Call this inequality TS. Define

$$\begin{aligned} & \Gamma(\alpha_S, \alpha_R, \alpha, p) \\ \equiv \; & \left| \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} - \frac{\alpha_R p}{\alpha_R p + (1-\alpha_R)(1-p)} \right| \\ & - \left| \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} - \frac{\alpha_R(1-p)}{\alpha_R(1-p) + (1-\alpha_R)p} \right| \\ & + \left| \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} - \frac{(1-\alpha_R)p}{(1-\alpha_R)p + \alpha_R(1-p)} \right| \\ & - \left| \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} - \frac{(1-\alpha_R)(1-p)}{(1-\alpha_R)(1-p) + \alpha_R p} \right|. \end{aligned}$$

A sufficient condition for inequality TS is that for any $\alpha < 0.5$ and any $\alpha_R \in [0,1]$, $\Gamma(\alpha_S, \alpha_R, \alpha, p) \geq 0$, with a strict inequality at least for some α. We divide the analysis into four cases and show that the above inequality holds in each case. The four cases are given as follows. Case A: $\alpha < \alpha_R < 0.5$. Case B: $\alpha < 1 - \alpha_R < 0.5$. Case C: $\alpha_R < \alpha < 0.5$. Case D: $0 < 1 - \alpha_R < \alpha < 0.5$.

We first analyze cases A or B: Given that either $\alpha < \alpha_R < 0.5$ or $\alpha < 1 - \alpha_R < 0.5$ we have (using the fact that $\frac{\alpha p}{\alpha p + (1-\alpha)(1-p)}$ is increasing in α to eliminate absolute values):

$$\begin{aligned} & \Gamma(\alpha_S, \alpha_R, \alpha, p) \\ = \; & -\frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} + \frac{\alpha_R p}{\alpha_R p + (1-\alpha_R)(1-p)} \\ & + \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} - \frac{\alpha_R(1-p)}{\alpha_R(1-p) + (1-\alpha_R)p} \\ & - \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} + \frac{(1-\alpha_R)p}{(1-\alpha_R)p + \alpha_R(1-p)} \\ & + \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} - \frac{(1-\alpha_R)(1-p)}{(1-\alpha_R)(1-p) + \alpha_R p} \\ = \; & \frac{2(\alpha_R - \alpha)(1 - \alpha - \alpha_R)(1-p)(2p-1)p}{(1 - p + \alpha(2p-1))(1 - p + \alpha_R(2p-1))(\alpha + p(1-2\alpha))(\alpha_R + p(1-2\alpha_R))}. \end{aligned}$$

The above last expression is positive given the assumptions made on parameters.

We now examine case C. Here, given that $\alpha_R < \alpha < 0.5$, we have:

$$\begin{aligned}
& \Gamma(\alpha_S, \alpha_R, \alpha, p) \\
= & \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} - \frac{\alpha_R p}{\alpha_R p + (1-\alpha_R)(1-p)} \\
& - \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} + \frac{\alpha_R(1-p)}{\alpha_R(1-p) + (1-\alpha_R)p} \\
& - \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} + \frac{(1-\alpha_R)p}{(1-\alpha_R)p + \alpha_R(1-p)} \\
& + \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} - \frac{(1-\alpha_R)(1-p)}{(1-\alpha_R)(1-p) + \alpha_R p} \\
= & \ 0.
\end{aligned}$$

We now finally examine case D. Here, given that $0 < 1 - \alpha_R < \alpha < 0.5$, we have:

$$\begin{aligned}
& \Gamma(\alpha_S, \alpha_R, \alpha, p) \\
= & - \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} + \frac{\alpha_R p}{\alpha_R p + (1-\alpha_R)(1-p)} \\
& + \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} - \frac{\alpha_R(1-p)}{\alpha_R(1-p) + (1-\alpha_R)p} \\
& + \frac{\alpha p}{\alpha p + (1-\alpha)(1-p)} - \frac{(1-\alpha_R)p}{(1-\alpha_R)p + \alpha_R(1-p)} \\
& - \frac{\alpha(1-p)}{\alpha(1-p) + (1-\alpha)p} + \frac{(1-\alpha_R)(1-p)}{(1-\alpha_R)(1-p) + \alpha_R p} \\
= & \ 0.
\end{aligned}$$

■

References

Acemoglu, D., V. Chernozhukov, M. Yildiz, et al. (2007). Learning and disagreement in an uncertain world. Technical report, Collegio Carlo Alberto.

Akerlof, G. A. and R. E. Kranton (2000). Economics and identity. *The Quarterly Journal of Economics 115*(3), 715–753.

Andreoni, J. and T. Mylovanov (2012). Diverging opinions. *American Economic Journal: Microeconomics 4*(1), 209–232.

Aumann, R. J. (1976). Agreeing to disagree. *The annals of statistics*, 1236–1239.

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

Battigalli, P. and M. Dufwenberg (2007). Guilt in games. *The American economic review 97*(2), 170–176.

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

Bénabou, R. (2012). Groupthink: Collective delusions in organizations and markets. *The Review of Economic Studies*, rds030.

Buechel, B., T. Hellmann, and S. Klößner (2015). Opinion dynamics and wisdom under conformity. *Journal of Economic Dynamics and Control 52*, 240–257.

Bursztyn, L., G. Egorov, and S. Fiorin (2017). From extreme to mainstream: How social norms unravel. Technical report.

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

Crawford, V. P. and J. Sobel (1982). Strategic information transmission. *Econometrica: Journal of the Econometric Society*, 1431–1451.

Dandekar, P., A. Goel, and D. T. Lee (2013). Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences 110*(15), 5791–5796.

Dixit, A. K. and J. W. Weibull (2007). Political polarization. *Proceedings of the National Academy of Sciences 104*(18), 7351–7356.

Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and economic behavior 47*(2), 268–298.

Ely, J., A. Frankel, and E. Kamenica (2015). Suspense and surprise. *Journal of Political Economy 123*(1), 215–260.

Festinger, L. (1957). A theory of cognitive dissonance. *Evanston, Il: Row, Peterson*.

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989). Psychological games and sequential rationality. *Games and Economic Behavior 1*, 60–79.

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

Gerber, A. S., G. A. Huber, D. Doherty, and C. M. Dowling (2012). Disagreement and the avoidance of political discussion: Aggregate relationships and differences across personality traits. *American Journal of Political Science 56*(4), 849–874.

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.

Golub, B. and M. O. Jackson (2012). How homophily affects the speed of learning and best-response dynamics. *The Quarterly Journal of Economics 127*(3), 1287–1338.

Heider, F. (1946). Attitudes and cognitive organization. *The Journal of psychology 21*(1), 107–112.

Heider, F. (1958). The psychology of interpersonal relations.

Kartik, N., F. X. Lee, and W. Suen (2015). Does competition promote disclosure? Technical report.

Matz, D. C. and W. Wood (2005). Cognitive dissonance in groups: the consequences of disagreement. *Journal of personality and social psychology 88*(1), 22.

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

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.

Prendergast, C. (1993). A theory of "yes men". *The American Economic Review*, 757–770.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review 83*, 1281–1302.

Sethi, R. and M. Yildiz (2012). Public disagreement. *American Economic Journal. Microeconomics 4*(3), 57.

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

Sobel, J. (2013). Giving and receiving advice. In M. Acemoglu, D. Arellano and E. Dekel (Eds.), *Advances in Economics and Econometrics: Tenth World Congress*, pp. 305–341. New York: Cambridge University Press.