# Nondisclosure with Conflicting Motives

Ying Gao*

November 2, 2020

**Abstract**

In this paper, I analyze communication through the disclosure of verifiable evidence when the receiver/decision maker is uncertain about where the sender's preferred action lies in relation to her own. In contrast to the known-preference case, fully informative communication is impossible: receivers cannot distinguish between senders of opposite preferences who pool by withholding their information when it is unfavorable. Two opposing patterns of partial disclosure emerge. When senders are biased relative to receivers, but just as state-sensitive, nondisclosure is driven by senders with extreme preferences, who choose to withhold slightly unfavorable evidence; but the reverse occurs when senders are state-insensitive.

## 1 Introduction

Why do people have differing beliefs about issues that appear, to an informed audience, to be all but settled with hard evidence? One explanation is that people with access to evidence don't always disclose it, but instead choose to lie by omission to influence others.

Folk wisdom, however, states that hiding information doesn't work if audiences anticipate what's being hidden. Grossman [1981] and Milgrom [1981] provide a classic argument that if the evidence-holder always wants to influence the receiver's action up, then only senders with the lowest signals will consider hiding them, and because of adverse selection, senders can't benefit from silence.

To bridge a gap between this result and the many examples of influential nondisclosure in the world, this paper explores the possibility that receivers *don't* know how the sender

*14.192 final paper. Thanks to Alex and Stephen for helpful advice and feedback.

wants to influence them: they are uncertain about the sender's preferences. There are a few reasons why this might be the case. First, a receiver may know the identity of the sender, but be uncertain about her preferences on a given issue, either because the issue or the sender herself is unfamiliar. Alternatively, a receiver may not know the sender's identity at all, only that she comes from a population of possible informants with different preferences.

My main observation is that, when the receiver doesn't know which *direction* the sender would like him to adjust his action towards, he will never be able to fully back out what the sender is hiding. This is because of *bidirectional pooling*: senders with opposite kinds of preferences and evidence on opposite sides of the status quo both end up withholding their information, each relying on the possibility of the other to prevent the receiver from catching on to the nature of the omission. The best the receiver can do in response is take an intermediate action that does well in expectation. In contrast, when the sender's desired direction of influence is known, full disclosure occurs.

Under bidirectional pooling equilibria, the sender discloses some signals, but not others. Their disclosure policies order signal values, and under some additional conditions, the set of sender-preference types, into two-sided spectra with increasingly influential signals and increasingly biased preference types surrounding a center signal and center type. Disclosure near the center differs from disclosure at the extremes. Strikingly, two opposite patterns of disclosure arise depending on how much more sensitive the receiver is to the state than the sender. When the receiver and the sender care about the state equally, a sender finds it worthwhile to share highly impactful information, but nudges receivers in the direction of her private bias by omitting details that slightly contradict it. On the other hand, when only the receiver cares about the state, the sender is reluctant to share big news, because she doesn't want the receiver to overadjust; instead, she tries to influence receivers to take her preferred action by disclosing minor evidence in its favor.

The comparative statics of disclosure from these two cases straddle debates about the importance of heterogeneous perspectives to informative communication. In the first case, greater differences in preferences discourage communication, while in the second, a diversity of perspectives is necessary for unconventional truths to get across.

These results give a framework by which the power to disseminate or withhold facts allows idiosyncratic preferences to affect public outcomes. It is applicable to several important real-world examples. One is media ownership: although the magnitude of the effect is disputed, studies agree that content put forth by media outlets changes with ownership in ways that are

consistent with pushing the owner's agenda (Gentzkow and Shapiro [2010], Baum and Zhukov [2018]). My results predict this, and suggest that, if in addition to promoting their bias, media companies also care about benefiting their readership or about citizens doing the right thing, then they will all report the most essential headlines, but differently biased companies may differ in their coverage by selectively skipping unfavorably partisan minutiae. On the other hand, lobbyists pursuing a specific issue in Washington may completely fail to disclose even important information about the issue to Congressmen, so long as it's not aligned with their goal – this is consistent with predictions when the informant is fundamentally state-agnostic.

This paper is laid out as follows. Section 2 gives the main framework of a disclosure model with sender preference uncertainty. Section 3 shows that with enough preference variation, bidirectional pooling equilibria occur. Section 4 discusses the form of the disclosure policy and comparative statics under two main cases. Section 5 works out an example, Section 6 considers the assumptions distinguishing my results from full disclosure, and Section 7 concludes.

## 1.1 Literature review

Models of disclosure typically involve senders who choose to disclose verifiable signals to receivers, who then act upon the information. Classically, in applications such as the quality-signalling problem, imperfect information between the parties lies along a single dimension, that is, the sender's private signal about some shared, payoff-relevant state. Because the sender always wants to influence the receiver's action up, there is adverse selection into withholding. Any pooling strategies unravel, leading to full disclosure in equilibrium (Milgrom [1981], Grossman [1981]). A general statement of these results, from Okuno-Fujiwara et al. [1990], is that whenever each sender's utility is always strictly increasing (i.e. positive monotone) in each receiver's beliefs about their signal, all state-relevant private information will be revealed.

Strands of the literature have pointed out the possibility of imperfect disclosure when receivers are uncertain about more than the sender's payoff-relevant signal. Dye [1985] observed that, if observers are uncertain about a manager's endowment of information, then those with unfavorable evidence can profitably pool with the uninformed. For the sake of comparison, we could frame the focus of this paper as uncertainty about *what kind* of sender holds useful information, under certainty that the information exists. Banerjee and

Somanathan [2001] consider voice in organizations, starting with a model of binary disclosure in which informants may have different priors about the promise of a project. However, in their model, all communication is unidirectional, since the verifiable signal itself is always good news. My paper departs from their binary state/single signal framework and focuses on settings with a continuum of signals and states, and bidirectional communication; however, it shares their focus on the effects of sender heterogeneity.

Also related to the idea of pooling under multidimensional sender heterogeneity are models of costly signaling with privately-known costs (Frankel and Kartik [2019], Esteban and Ray [2006]). There, an informed party with known preferences observes a natural state, as well as her private cost of distorting the decision-maker's perception of the state. Similarly to the sender-specific preferences in this paper, distortion costs are not directly payoff-relevant to the decision-maker, but they confound reports of the state, so that uncertainty about the state remains at the time that decisions are made.

Some more distantly related discussions of cheap talk and partial provability are nevertheless interesting in the context of this problem. Chakraborty and Harbaugh [2010] show that, when there are $N$ dimensions in a message, at most one dimension is payoff-relevant to a sender with fixed preferences, therefore he and the receiver can find an $N-1$ dimensional subspace of common interest on which they can communicate informatively in a cheap talk game. This paper considers something of an opposite case, in which the feasible message is restricted to a single dimension, while the sender's payoff depends on multidimensional information, and shows that this restricts the informativeness of communication. Among models of partial provability, Seidmann and Winter [1997] show that while generically equilibria are partially informative, full revelation occurs when each verifiable subset of types – analogous to a message in my setting – admits a worst-case type, which no other type in that subset wishes to masquerade as. In my setting, sufficient richness of sender preferences and compactness of the signal set precludes the existence of a worst-case type for the empty message.

# 2 Model

I focus on a simple disclosure model with one sender and one receiver, who both aim to maximize their expected utility under utility functions

$$u_s(\theta, x_s, a_r), \qquad u_r(\theta, a_r).$$

Payoffs for both players can be directly influenced only through the receiver's action, $a_r \in \mathbb{R}$.

The state of the world, $\theta \in \mathbb{R}$, is unknown, but both players know that its distribution is $f(\theta)$. Conditional on the state, a signal $m_s$ is drawn at the start of the game from the distribution $h(m_s|\theta)$.

**Assumption 2.1** *The marginal distribution of the signal, $\int_\theta h(m_s|\theta) \cdot f(\theta) d\theta$, is continuously supported on a bounded interval $[\underline{m}, \bar{m}]$.*

Finally, the sender has a privately-known preference type $x_s \sim g_s(x) \in [\underline{x}_s, \bar{x}_s]$, with commonly known distribution independent of $f(\theta)$.[1] For the main body of this paper I assume that $g_s(x)$ is not a degenerate (point) distribution, but I reconsider this possibility, and its relation to full disclosure results, in the last section.

**Assumption 2.2** *The receiver is uncertain about the sender's type: $g_s(x)$ is not supported on a single point.*

Though inconsequential in this one-shot model, the underlying idea is that $x_s$ is intrinsic and known well ahead of time, while $m_s$ is a signal drawn at the beginning of the game. To reflect this, I will refer to $x_s$ as simply the sender's *type*, and when referring to a particular sender, I mean a sender endowed with a type $x_s$. To avoid confusion, the sender's full set of private information $(x_s, m_s)$ (which is what "type" refers to elsewhere in the literature) will instead be called a *scenario* henceforth in this paper.

## 2.1 Timing and actions.

First, the sender's type is drawn, and she learns it. Then, nature draws a state and a signal conditional on it. The sender observes the signal and chooses whether to disclose the signal to the receiver, or to withhold it. I interpret the signal as a piece of hard evidence that can be passed on costlessly to the receiver, and means the same thing to both players. The sender's preference type, on the other hand, is not verifiable, and the sender cannot engage in cheap talk or in any other way influence the receiver's belief about it. Neither the sender nor the receiver has commitment power. Observing only what the sender has passed along, the receiver chooses an action. To summarize, the timing is as follows:

0. $x_s$, $\theta$, and $m_s$ are realized. The sender is given $x_s$, and observes $m_s$.

[1] Everything will extend to the case where $x_s$ has unbounded support, as well, but I have chosen to keep $x_s$ bounded here for ease of exposition.

1. The sender sends a message $\tilde{m}(x_s, m_s)$ to the receiver. She may choose between sending their signal as-is ($\tilde{m} = m_s$), or withholding it ($\tilde{m} = \emptyset$).

2. The receiver observes the message if there was one. He forms a Bayesian posterior on the state, which is $\beta(\theta|\emptyset)$ if he saw no message ($\tilde{m} = \emptyset$), and $\tilde{h}(\theta|m_s)$ if he saw a message $\tilde{m} = m_s$.

3. The receiver chooses his action $a_r(\tilde{m})$, and payoffs are realized.

## 2.2 Notation and assumptions.

Let $a^*_{r,i}(\cdot)$ denote an optimal choice of $a_r$ from the perspective of player $i$. That is,

$$a^*_{r,r}(\theta) \in \arg\max \mathbb{E}[u_r(\theta, a)] \qquad a^*_{r,s}(\theta, x_s) \in \arg\max \mathbb{E}[u_s(\theta, x_s, a)].$$

I often write a posterior $\beta$ as an argument in utility function $u$ or maximizer $a^*$, in place of $\theta$. It is shorthand for taking the expectation over $\theta$ given $\beta$, e.g.

$$u_r(\beta, a_r) = \mathbb{E}[u_r(\theta, a_r)|\beta] \qquad a^*_{r,r}(\beta) \in \arg\max_a \mathbb{E}[u_r(\theta, a)|\beta].$$

This notation is natural because the state of the world enters the game only through the expectations induced by the signal. Thus, the signal is a "sufficient statistic" with respect to the state and players' strategies, and it is abstractly without loss to take utility functions over realized signals, instead of those over states, as primitives of the model.

In order to impose necessary structure on the basic setup above, I assume that preferences are continuous, differentiable in the action, single-peaked, and ordered in $m$ and $x_s$.

**Assumption 2.3** ***Continuity and differentiability:***

*$u_s(\theta, x_s, a_r)$ and $u_r(\theta, a_r)$ are continuous in all arguments, and differentiable in $a_r$.*

**Assumption 2.4** ***Quasiconcavity with increasing peaks (QCIP):***

*$u_s(m_s, x_s, a_r)$ and $u_r(m_s, a_r)$ are strictly quasiconcave in $a_r$, with peaks $a^*_{r,s}(\beta, x_s)$ strictly increasing in $x_s$, and $a^*_{r,r}(\tilde{h}(\theta|m_s))$ strictly increasing over the family $m_s \in [\underline{m}, \bar{m}]$.*

QCIP means that the utility functions of sender and receiver are both single-peaked, and are ordered with increasing peaks over the possible beliefs induced by signals (for the receiver)

and preference types (for the sender), holding the other fixed. Single-peakedness is a common assumption, and increasing peaks is a standard way to order single-peaked functions. The order applies only to the peaks, and preferences are not necessarily well-ordered elsewhere. In particular, this condition does not necessarily imply single crossing differences (SCD), which says that for arbitrary actions $a' > a$, a "higher type" (with higher signal or preference type) will have relatively higher utility for $a'$ rather than $a$ whenever a lower type does. In fact, Quah and Strulovici [2009] observe that among single-peaked functions, increasing peaks is strictly weaker than SCD, and equivalent to the interval dominance order.

Following the notational discussion above, I assume QCIP directly on the signal-dependent expected utility functions. It can be replaced with an equivalent assumption over the original state-dependent utility functions $u_s(\theta, x_s, a_r)$ and $u_r(\theta, a_r)$ as long as:

1. $\tilde{h}(\theta|m_s)$ satisfies the monotone likelihood ratio property, which is sufficient to guarantee a strong set order on the optima.
2. Strict single-peakedness of $u_s$, $u_r$ is preserved when expectations are taken over $\tilde{h}(\theta|m_s)$, for all $m_s$.

The second condition can often be checked by hand. It is satisfied for a fairly inclusive range of common functional forms. Some useful categories of utility functions and signal structures satisfying (2) are:

- The signal perfectly conveys the state, $m_s = \theta$.
- Conditional distribution $\tilde{h}(\theta|m_s)$ is strictly single-peaked for all $m_s$, and $\theta$ is a shifter of the utility functions, i.e. for some increasing functions $\gamma_s$, $\gamma_r$,

$$u_s(\theta, x_s, a + \Delta) = u_s(\theta - \gamma_s(\Delta), x_s, a), \qquad u_r(\theta, a + \Delta) = u_r(\theta - \gamma_r(\Delta), a).$$

The solution concept I consider is a perfect Bayesian equilibrium (PBE) in pure strategies for the receiver. I will show later that such equilibria always exist here.

PBE is the concept used in most signalling games, including by Grossman and Milgrom for the unraveling result, and in the partial provability literature. In my model, any PBE is pinned down by a single object, which is the receiver's empty-message belief $\beta(\theta|\emptyset)$. The sender best-responds to a given empty-message posterior by choosing between inducing the action the receiver takes upon seeing the true signal, or the one induced by $\beta(\theta|\emptyset)$. Being in

a PBE requires that the empty-message posterior be consistent with the state distribution conditional on an empty message, induced by the sender's best response.

Imposing that the receiver play a pure strategy is usually without loss, since under a generic utility function and belief distribution, there will be a single action that maximizes their expected utility. Furthermore, QCIP ensures a single optimal action for the receiver when the signal is revealed to him. However, since my other assumptions will not rule out that there can be a tie for the receiver's expected utility maximizer under the no-message posterior, and my approach relies on the receiver choosing one specific action, in that case, I will force a pure action. This assumption is quite realistic in the direct application to single receivers, since most people don't consciously randomize. It may be less reasonable when the "receiver" stands for the aggregate of a large population, but even then, since at least one pure strategy equilibrium exists, and additional equilibria relying on randomization will be knife's-edge and difficult to sustain, it seems natural to focus on the former.

# 3 Bidirectional pooling

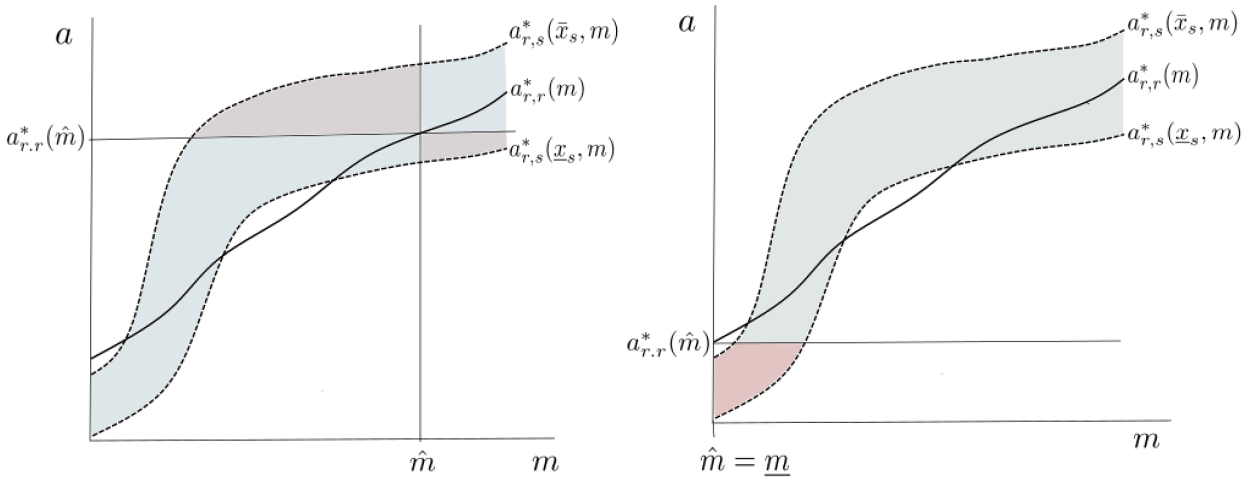
Does the sender always disclose her evidence? In this section, I argue that if the set of possible preferences contain ones that oppose each other under any beliefs for the receiver, then full disclosure never occurs. Formally, the key idea of uncertainty over opposing preferences is a combination of Assumption 2.1, which establishes type-uncertainty, and Assumption 3.1 below, which ensures that senders' preferences are sufficiently opposed to rule out full unraveling of nondisclosure.

**Assumption 3.1** ***Bidirectional sender bias (BSB):***

$$\min_{m_s} \left[a^*_{r,s}(\beta(\theta|m_s), \underline{x}_s)\right] < a^*_{r,r}(\underline{m}) \quad \textit{and} \quad a^*_{r,r}(\bar{m}) < \max_{m_s} \left[a^*_{r,s}(\beta(m_s), \bar{x}_s)\right].$$

In words, BSB means that the most extreme actions that could be optimal for the sender, over all type and signal realizations, are more extreme than the most extreme optimal actions for the receiver. It captures a strong notion of opposing biases between senders in different scenarios: the receiver is sure that no matter the action he plans to take, there are some scenarios in which the sender wishes it higher, and others in which the sender wishes it lower.

**Theorem 3.2** *Assume that $u_s$ and $u_r$ are continuous in all arguments, differentiable in $a_r$, and QCIP and BSB are satisfied. Then any signaling equilibrium features $a^*_{r,r}(\beta(\theta|\emptyset)) \in$*




(a) $a^*_{r,s}(\hat{m}, \underline{x}_s) < a^*_{r,r}(\hat{m}) < a^*_{r,s}(\hat{m}, \bar{x}_s)$ (Pooling)

(b) $a^*_{r,r}(\hat{m}) < a^*_{r,s}(\hat{m}, \underline{x}_s) < a^*_{r,s}(\hat{m}, \bar{x}_s)$ (Pooling)

Figure 1: Outcomes under disclosure and nondisclosure for different configurations of $\hat{m}$.
**Gray area** $= \{(a^*_{r,s}(x_s, m_s), m_s) : x_s \in [\underline{x}_s, \bar{x}_s], m_s \in [\underline{m}, \bar{m}]\}$.
**Red area** $= \{(a^*_{r,s}(x_s, m_s), m_s) : a^*_{r,r}(\hat{m}) \in (a^*_{r,s}(x_s, m_s), a^*_{r,r}(m_s))$ or $(a^*_{r,r}(m_s), a^*_{r,s}(x_s, m_s))\}$.
(a) In equilibrium, $\hat{m} \in (\underline{m}, \bar{m})$, and the sender's best response entails pooling towards $\hat{m}$ from either side. In particular, senders in the red regions will always withhold the signal. (b) Unlike in Grossman and Milgrom, there cannot be a corner posterior given the empty message. If $\hat{m} = \underline{m}$, then the sender would best respond by withholding some higher signals, violating belief consistency.

$\big(a^*_{r,r}(\tilde{h}(\theta|\underline{m})), a^*_{r,r}(\tilde{h}(\theta|\bar{m}))\big)$*, with a positive probability of withholding both "high" signals* $(a^*_{r,r}(\beta(\theta|\emptyset)) < a^*_{r,r}(\tilde{h}(\theta|m_s)))$ *and "low" signals* $(a^*_{r,r}(\beta(\theta|\emptyset)) > a^*_{r,r}(\tilde{h}(\theta|m_s)))$*. All sender types, except possibly one, withhold under some signal realizations.*

The only type of sender who may never find it worthwhile to withhold any signal is one who prefers the receiver to take action $a^*_{r,r}(\tilde{\beta}(\theta|\emptyset))$ exactly when their true signal would induce $r$ to take that action anyways.

For a full proof of the theorem, please see the appendix. Here, I will explain the intuition, which is simple. First, fixing the sender's strategy, the receiver's beliefs are also fixed, and his strategy is determined: he plays $a^*_{r,r}(\beta(\theta|\tilde{m}))$. There is a signal, $\hat{m} \in [\underline{m}, \bar{m}]$, such that the action taken by the receiver upon seeing the signal $\hat{m}$ is the same as the action taken under $\tilde{m} = \emptyset$: $a^*_{r,r}(\beta(\theta|\emptyset)) = a^*_{r,r}(\tilde{h}(\theta|\hat{m}))$. This signal functions as an endogenously determined "center", the benchmark to which impactful news will be contrasted. It is the status quo not because it represents receivers' prior beliefs, but because it represents the posterior under silence, which can be quite different.

Recall from our discussion of the model that $\hat{m}$ fixes the equilibrium. To see that any equilibrium will satisfy Theorem 3.2, observe that under single-peakedness, if $a^*_{r,s}(m_s, x_s) > a^*_{r,r}(\hat{m}) > a^*_{r,r}(m_s)$ or $a^*_{r,s}(m_s, x_s) < a^*_{r,r}(\hat{m}) < a^*_{r,r}(m_s)$, then the sender's strict optimal action is to withhold the signal. It will be helpful to refer to Figure 1, where such scenarios appear in red. They are given by the intersection between the 2nd & 4th quadrants of the plane centered on $(\hat{m}, a^*_{r,r}(\hat{m}))$ and the gray area representing the set of possible (signal, sender-optimal action) pairs. Importantly, the assumptions above don't suffice to pin down the *entire* set of scenarios under which nondisclosure is the sender's best move, but the region just described will constitute a strict subset of such scenarios.

It is not hard to see that under BSB, the red region must have positive measure over $g_s(x) \int_\theta h(m_s|\theta) f(\theta) d\theta$, and, more importantly, there is a spread over the true value of $m_s$ across the region. Technical assumption 2.1 prevents the receiver from taking on beliefs that have probability 0 ex ante and that almost every sender wants to avoid.[2] Thus, senders withhold signals with positive probability, and whenever they do, the receiver is uncertain about the true signal realization.

Single-peakedness also implies that whenever the receiver's belief is consistent and not a singleton, there are both types of senders who prefer to withhold signals above $\hat{m}$ and ones who withhold signals below. Observing that senders' optimal-action curves are functions bounded within the gray region of Figure 1, it's clear that the only type of sender who may never wish to withhold any signal is the one whose optimal action curve passes through $(\hat{m}, a^*_{r,r}(\hat{m}))$.

Finally, note that while $\hat{m}$ characterizes the equilibrium strategies of sender and receiver uniquely up to indifference, the equilibrium need not be unique, as there may be multiple equilibrium values of $\hat{m}$ that give rise to distinct strategy profiles. An equilibrium in pure strategies for the receiver does always exist, however: a simple intermediate value theorem argument, in conjunction with BSB, shows that a function that takes in $\hat{m}$ and outputs the implied posterior $\hat{m}$ from the sender's BR has a fixed point in $(\underline{m}, \bar{m})$.[3]

[2] If the domain of $m_s$ were unbounded, and the receiver's belief $b(m_s|\emptyset)$ allowed to be supported on its closure, then under families of preferences in which the receiver's optimal action varies unboundedly with the message and the sender's utility becomes unboundedly negative with distance from their optimal action, a point belief on $m_s = \infty$ or $m_s = -\infty$ would be self-sustaining, due to infinite losses from withholding any finite realization of $m_s$. Thus, fully informative equilibria are once again possible. Similar issues arise when $m_s$ lies in an open interval.

[3] Let $\hat{m}^{BR}(\cdot)$ be an operator taking in a hypothetical value of $\hat{m}$ and outputting the new value of $\hat{m}$ that would represent the receiver's no-message posterior after one round of best responding by the sender. Because the sender's best response is continuous in $\hat{m}$, and the receiver's posterior is continuous in the sender's

# 4 Who withholds information, and when?

The discussion above makes it clear that most types of senders disclose some signals, and withhold others. I now examine which signals each sender withholds, and how that affects the kinds of information that make it through to the receiver. The end result of these comparisons is a set of comparative statics over senders' propensity to communicate and signals' likelihoods of being transmitted, depending on their extremeness relative to special "central" signals and types.

## 4.1 Sender-type monotonicity of disclosure

Are senders more likely to hide impactful information the more it contradicts their bias? Intuition suggests so. By withholding their signal, the sender "corrects" a misalignment between their preferences and the receiver's by letting the receiver carry on with a belief that is slanted relative to the truth, from the sender's point of view. Senders with increasingly extreme low types should be willing to withhold an increasingly large set of signals higher than $\hat{m}$, and senders with increasingly high types should more often withhold signals lower than $\hat{m}$. For the rest of the paper, I will assume single crossing differences, under which this prediction is easy to verify:

**Assumption 4.1** ***Single crossing differences (SCD) in*** $x_s$***:*** *For all* $m_s$*,* $x_s < x_s'$*, and* $a_r < a_r'$*,*

$$\begin{aligned} u_s(\tilde{h}(\theta|m_s), x_s, a_r') - u_s(\tilde{h}(\theta|m_s), x_s, a_r) \geq (>)\ 0 \\ \implies u_s(\tilde{h}(\theta|m_s), x_s', a_r') - u_s(\tilde{h}(\theta|m_s), x_s', a_r) \geq (>)\ 0. \end{aligned} \tag{1}$$

Single crossing differences in $x_s$ means that if, between a lower action and a higher one, the utility of a sender of lower type is higher for the lower action than for the higher action, then the same is true of the higher type.

**Proposition 4.2** *If, in addition to the assumptions of Theorem 3.2,* $u_s$ *satisfies SCD, then the propensity to withhold signals in order to induce a given slant is monotone in sender type: for all* $m_s$ *such that* $a^*_{r,r}(\tilde{h}(\theta|m_s)) > (<) a^*_{r,r}(\beta(\theta|\emptyset))$*, whenever a sender of type* $x_s$ *chooses to withhold* $m_s$*, so do all senders of type* $x_s' < (>) x_s$*.*

strategy, $\hat{m}^{BR}$ is continuous in $\hat{m}$. By the argument used to prove Theorem 3.2, under BSB $\hat{m}^{BR}(\underline{m}) - \underline{m} > 0$ and $\hat{m}^{BR}(\bar{m}) - \bar{m} < 0$; therefore, $\hat{m}^{BR}$ has at least one fixed point in $[\underline{m}, \bar{m}]$.

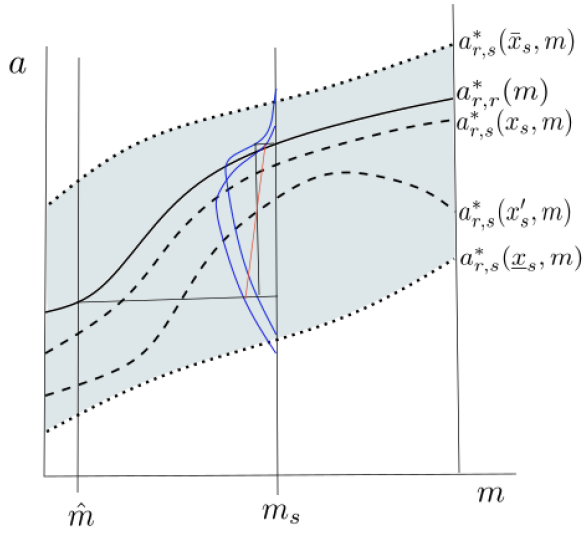



Figure 2: SCD guarantees that if a type $x_s$ is at least indifferent between $\hat{m}$ and $m_s > \hat{m}$, then a type $x'_s < x_s$ will certainly prefer $\hat{m}$, and thus withholds $m_s$ for sure.

**Proof** Defining $\hat{m}$ as in the proof of the previous theorem, observe that for all $m_s < \hat{m}$, SCD in $x_s$ directly implies that if a type $x_s$ prefers $\hat{m}$ to $m_s$, then a type $x'_s > x_s$ does as well, and similarly for $m_s > \hat{m}$ and $x'_s < x_s$. ∎

Proposition 4.2 states that under single crossing differences, the order on senders' types perfectly captures the (weak) inclusion order on both the set of signals $m_s < \hat{m}$ that they benefit from withholding, and the set of signals $m_s > \hat{m}$ that they benefit from disclosing.

Without single crossing differences, there exist counterexamples to this proposition. The reason is that, even if under a given signal $m_s$ type $x'_s$ has a preferred action closer to $a^*_{r,r}(\tilde{h}(\theta|\hat{m}))$ and further from $a^*_{r,r}(\tilde{h}(\theta|m_s))$ than type $x_s$, a change in the shape of the rest of the curve may mean that type $x'_s$ gets *greater* utility than type $x_s$ from disclosing $m_s$, and less from withholding.

## 4.2 Disclosure policies by type

I now look at disclosure choices within types. Given $\hat{m}$, a full characterization of the disclosure policy for each type is possible.

A few more definitions will be helpful. In particular, since there is a 1-to-1 mapping between disclosed signals and receiver-optimal actions, it will be useful to view the sender's

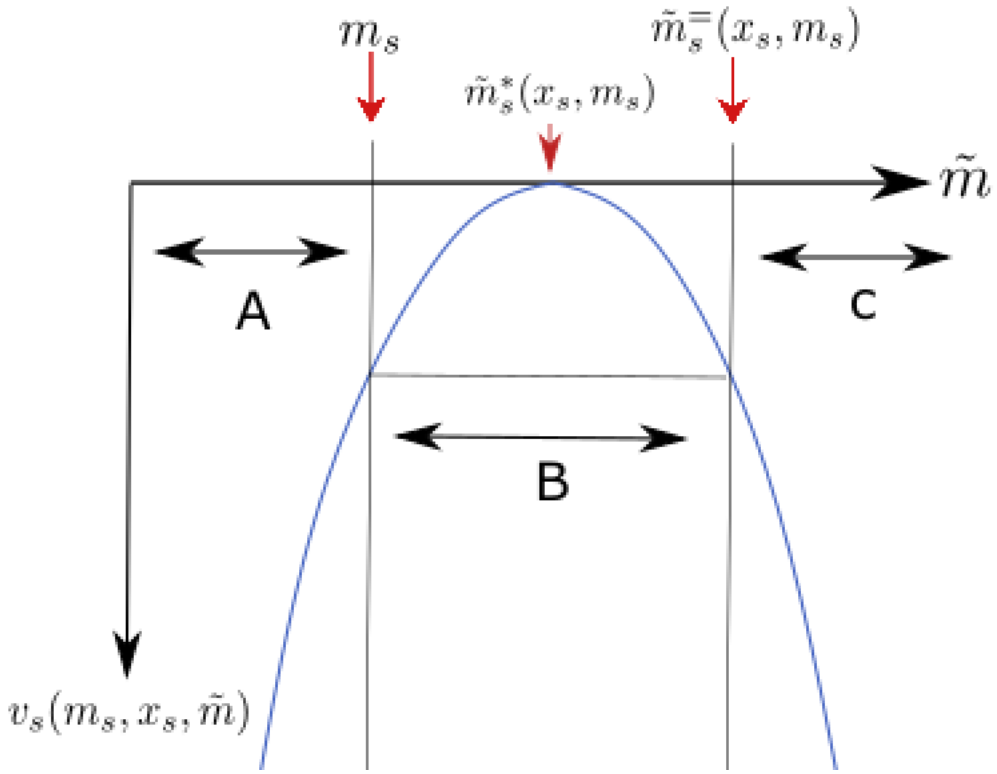



Figure 3: The value of obfuscating evidence in a given scenario $(x_s, m_s)$, when $\hat{m}$ lies in one of 3 regions.
**A**: Withholding influences beliefs in the wrong direction $\Rightarrow$ disclosure.
**B**: Profitable nondisclosure.
**C**: Withholding overcorrects in the direction of bias $\Rightarrow$ disclosure.

problem as hypothetically maximizing their utility over all possible messages, subject to the disclosure constraint that only $\hat{m}$ and $m_s$ are actually feasible. Taking as given the receiver's strategy, in the first period the sender chooses a message as if maximizing directly over $\tilde{m}$ the utility function

$$v_s(m_s, x_s, \tilde{m}) := u_s(m_s, x_s, a^*_{r,r}(\beta(\theta|\tilde{m}))).$$

The function $v_s$ will take on the properties of $u_s$; in particular, it is single-peaked in $\tilde{m}$. Furthermore, if the sender's choice of message was unrestricted, there would be a unique sender-optimal message

$$\tilde{m}^*_s(x_s, m_s) := \arg\max_{\tilde{m}} v_s(m_s, x_s, \tilde{m}).$$

Finally, each sender has a "breakeven message" as a function of true signal $m_s$.

**Definition** A *breakeven message* $m_s^=(m_s, x_s)$ for the sender is the furthest-away alternative signal that, if sent, would allow the sender to receive at least the same utility as disclosing their true signal:

$$m_s^=(m_s, x_s) = \begin{cases} \min(m \in [\underline{m}, \bar{m}] : v_s(m_s, x_s, m) \geq v_s(m_s, x_s, m_s)) & \text{if } m_s > \tilde{m}_s^*(x_s, m_s) \\ \max(m \in [\underline{m}, \bar{m}] : v_s(m_s, x_s, m) \geq v_s(m_s, x_s, m_s)) & \text{if } m_s < \tilde{m}_s^*(x_s, m_s) \\ m_s & \text{if } m_s = \tilde{m}_s^*(x_s, m_s) \end{cases}$$

Single-peakedness of $v_s(m_s, x_s, \cdot)$ implies that an alternative hypothetical message is preferred to $m_s$ if and only if it lies between $m_s$ and $m_s^=(m_s, x_s)$. Figure 3 shows why: when $\hat{m}$ lies towards $\tilde{m}_s^*(x_s, m_s)$ relative to $m_s$, nondisclosure directionally favors the sender's bias, but if it is to the other side of $\tilde{m}_s^=(x_s, m_s)$, then the omission goes too far.

Therefore, fixing $\hat{m}$, the sender's strategy, up to indifference at the boundaries, is

$$\tilde{m}(m_s, x_s) = \begin{cases} \emptyset & \text{if } m_s \leq m_s^=(m_s, x_s) \text{ and } \hat{m} \in [m_s, m_s^=(m_s, x_s)] \\ & \quad \text{or } m_s \geq m_s^=(m_s, x_s) \text{ and } \hat{m} \in [m_s^=(m_s, x_s), m_s] \\ m_s & \text{otherwise.} \end{cases} \tag{2}$$

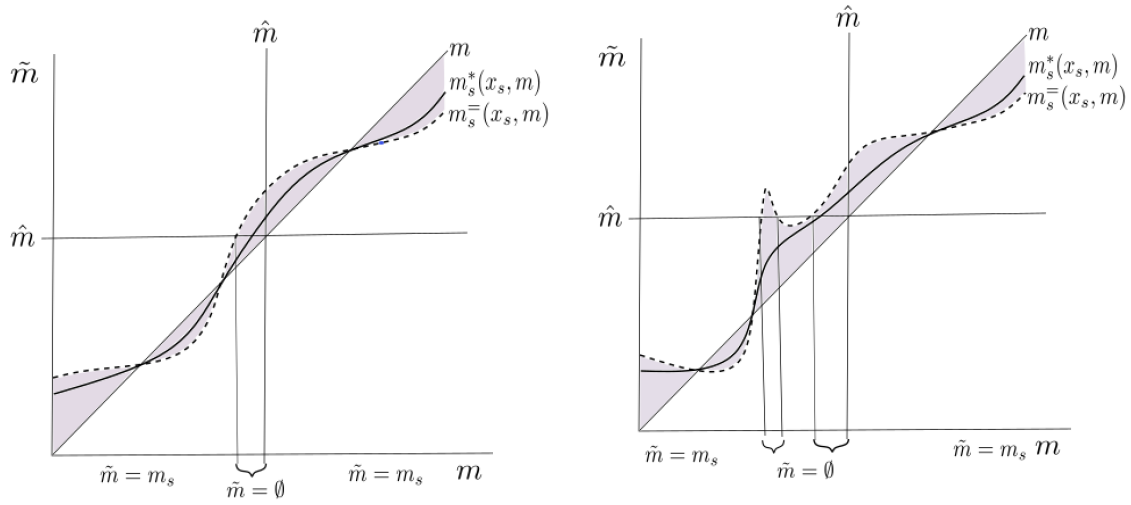
Figures 4a and 4b give examples of this concept. Fixing a sender, the intersection of $\tilde{m} = \hat{m}$ with the purple region between the true and breakeven messages gives the set of signals the sender will withhold.

## 4.3 Monotone breakeven message

Only when the breakeven message is well-behaved will there be crisp predictions about what kinds of signals senders disclose. The multi-segmented policy of Figure 4b occurs because the breakeven message in that case is nonmonotone, which can occur when the shape of the sender's utility function or the amount of misalignment with the receiver is highly state-dependent. On the other hand, when the breakeven message is monotone, strategies involving a single interval of nondisclosure will be the only possible outcome.

**Proposition 4.3 (Montone breakeven message (MBM))** *In addition to previous assumptions, if for all $x_s$, $\tilde{m}_s^=(m_s, x_s)$ is strictly increasing in $m_s$, then*

- *There is $\hat{x}$ such that the sender of type $\hat{x}$ is indifferent between disclosing or withholding*




(a) Monotonicity of the breakeven message implies an interval withholding strategy.

(b) An optimal strategy for the sender when the breakeven message is not monotone in $m_s$.

Figure 4: Breakeven messages plotted against true signals.

*$\hat{m}$ and discloses everything else.*

- *As $x_s$ increases from $\hat{x}$, senders withhold an increasing interval of signals $[m^*, \hat{m}]$, and as $x_s$ decreases from $\hat{x}$, senders withhold an increasing interval of signals $[\hat{m}, m^*]$.*

*On the other hand, if for all $x_s$, $\tilde{m}_s^=(m_s, x_s)$ is strictly decreasing in $m_s$, then*

- *There is $\hat{x}$ such that the sender of type $\hat{x}$ is indifferent between disclosing or withholding $\hat{m}$ and withholds everything else.*
- *As $x_s$ increases from $\hat{x}$, senders disclose an increasing interval of signals $[m^*, \hat{m}]$, and as $x_s$ decreases from $\hat{x}$, senders disclose an increasing interval of signals $[\hat{m}, m^*]$.*

Figure 4a illustrates the concept. Intuitively, positive MBM means that the threshold for overshooting the sender's bias is increasing in the realized signal, while negative MBM implies that it is decreasing. Loosely speaking, the sender's and receiver's interests are relatively aligned under positive MBM, whereas under negative MBM they can be quite misaligned. Neither implies, nor is implied by single crossing differences or QCIP.

Both the positive and negative monotone cases fit some simple cases. Positive monotone breakeven messages tend to occur when senders are misaligned due to simple bias. Formally, I define a simple bias setting to be one in which the sender's utility function is simply shifted

relative to the receiver's by a bias function $\omega$ increasing in $x_s$:

$$u_s(\theta, x_s, a_r) = u_r(\theta, a_r - \omega(x_s)).$$

The breakeven message will always be positive monotone when preference misalignment takes the form of simple bias and:

1. $u_s$ is symmetric about $a^*_{r,s}(m_s, x_s)$, or
2. $m_s$ shifts $u_s$ and $u_r$ together: there is a single increasing function $\gamma$ such that

$$u_s(m_s, x_s, a + \Delta) = u_s(m_s - \gamma(\Delta), x_s, a), \qquad u_r(m_s, a + \Delta) = u_r(m_s - \gamma(\Delta), a).$$

Bias, symmetry, and signals as shifters are all common in models of policy targeting and principal-agent models of delegation with communication.

On the other hand, an important class of problems for which the breakeven message is negative monotone is that in which the sender's preferences are completely state-independent. That is, senders could be completely dogmatic, or the state could be payoff-relevant only to receivers, even though the sender cares about the realized outcome. Examples include lobbyists and interest groups, or strict ideologues.

Section 6 gives an example of positive MBM in a situation with pure bias, symmetry, and a state-matching motive, as well as an example with negative MBM when the sender is very insensitive to the state, relative to the receiver.

## 4.4 Comparative statics under MBM

Having established settings in which positive and negative MBM are reasonable assumptions, I turn to highlight some comparative statics of the probability of disclosure when breakeven messages are monotone. Observe that signals and types are very much bidirectional, with the "center" of each bidirectional spectrum naturally defined by $\hat{m}$ and $\hat{x}$, respectively. A rough symmetry of scenarios across the center means that the most interesting comparative statics will be about *extremeness*, or distance from the center, rather than about high vs. low signals or types. In what follows, for the sake of conciseness I assume that indifferent senders always choose disclosure.[4]

[4]This does not change the set of equilibria, nor their properties

The first set of comparative statics concern the communicativeness of senders. There is exactly one central type $\hat{x}$, who can be considered a pure centrist. Under positive MBM, this type's interests are aligned enough with the receiver's to want to disclose everything (even though their preferences and the receiver's generally differ). With a negative MBM, the centrist's interests are not particularly aligned with the receiver's, nor do they have a particular interest in championing a cause; thus, they never disclose anything. As a corollary of Prop. 4.2, the total probability of disclosure is monotone with distance from $x_s$ to either side:

**Corollary 4.4** *Under positive MBM, the sender's total probability of disclosing a signal is quasiconcave in $x_s$ and maximized at $\hat{x}$.*

*Under negative MBM, the sender's total probability of disclosing a signal is quasiconvex in $x_s$ and minimized at $\hat{x}$.*

So, in cases where senders' and receivers' misalignment approximates a simple bias, extreme senders are, on the whole, less likely to disclose a message. Fixing receivers' beliefs, a greater spread in the distribution of sender types decreases the amount of communication. On the other hand, when senders' preferences are less state-sensitive than the audience's, only extreme types are willing to share influential information, and more dispersed preferences lead to a greater flow of information.

The contrast between these two cases relates to a debate about the positive or negative impacts of diverse values. A common thought is that polarization can jam communication. Many issues for the receiver, such as decreased trust or lack of common ground, contribute, but the positive MBM outcome supports reluctance to communicate on the extreme sender's side as another possible factor. The pattern under negative MBM, on the other hand, echoes an argument in favor of pluralism made by Banerjee and Somanathan [2001].[5] There, a greater variety in types is associated with more communication, because it takes an agreeably biased sender to support the transmission of any major news – strong supporters are necessary to bring things to light.

It matters not just that different senders can be more or less forthcoming, but also that some signals may be disclosed more often than others. Abstracting away from the realiza-

[5]Banerjee and Somanathan's verifiable communication model actually satisfies conditions for a *positive* monotone breakeven message, but their signals are unidirectional by design, and for them, an extreme sender is one who has a private desire to support a project, which corresponds in my model to a higher, not a more misaligned, type.

tion of senders' types, differential transmissibility of signals directly determines welfare and outcomes on the receiver side. A corollary of Proposition 4.3 is that:

**Corollary 4.5** *Under positive MBM, a signal's total probability of being disclosed is quasi-convex in $m_s$ except at $\hat{m}$, where it is* 1. *As $m_s \uparrow \hat{m}$ or $m_s \downarrow \hat{m}$, the probability of disclosure decreases.*

*Under negative MBM, a signal's total probability of being disclosed is quasiconcave in $m_s$ and maximized at $\hat{m}$.*

More extreme signals are more transmissible under positive MBM because senders disclose all bias-favoring signals. An interpretation of this is that bigger news is more likely to travel because everyone agrees on the importance of publicizing it, whereas biased sources are all too happy to sway audiences by fudging the small stuff. The contrasting prediction under negative MBM is that extreme signals are less likely to be transmitted through disclosure. They are too influential, and must be withheld for fear of causing the receiver to overreact.

I emphasize transmissibility through disclosure because I believe it has consequences for more extended models that feature disclosure as a subgame: for example, when a chain of different senders is required to deliver evidence to its final recipient, its effect may be amplified, and all information that isn't sufficiently impactful, or any information that's too outrageous, may be dropped in transmission. Again, both possibilities are plausible.

The preceding discussion of conditions for positive and negative monotone breakeven messages suggests that one distinguishing factor is the sender's state-sensitivity relative to the receiver. With a specific setting in mind, this distinction can help settle the debate both about the transmissibility of extreme signals and about extreme senders' communicativeness.

# 5 Example: policy platforms.

Following the discussion of equilibrium policies, I provide an illustrative example. Suppose that a receiver follows a member of the press (sender) on Twitter, and would, in an upcoming election, like to support one of a continuum of government spending policies, which range from very contractionary (-1) to very expansionary (1). There is a factor, $\theta \sim U[-1, 1]$, that influences the optimal level of government spending, but it is unknown to both agents. The sender is affected by what the receiver does through its impact on the outcome of the election. In addition, each agent may be in a position to benefit privately from government

programs, or to suffer from higher taxes, so I model their utilities as a quadratic loss function

$$u_s(\theta, x_s, a_r) = -(a_r - \theta - x_s)^2, \qquad u_r(\theta, a_r) = -(a_r - c\theta)^2.$$

with the sender's private preference parameter $x_s$ uniform on $[-1, 1]$. Finally, suppose that the sender has access to briefing with information relevant to $\theta$ that can be summarized as a signal $m_s \sim U[\theta - \epsilon, \theta + \epsilon]$, and may disclose it verifiably.

The utility functions in this example satisfy continuity, differentiability, and BSB; the distributions satisfy Assumptions 2.1 and 2.2, and both jointly satisfy QCIP and SCD. Thus, there will be bidirectional pooling and an ordering of disclosure in sender types. When $c < 2$, the breakeven message satisfies positive MBM, and when $c > 2$, it satisfies negative MBM.

I solve for the full disclosure policy to illustrate the outstanding characteristics of each case, starting with a description of the best response of the receiver and sender when the other side's strategy is fixed.

**Receiver's decision**: $a^*_{r,r}(\tilde{m}) = c\mu(\theta|\tilde{m})$, where the mean posterior message is either, if $\tilde{m} = m_s$,

$$\mu(\theta|m_s) = \begin{cases} m_s, & m_s \in [-1+\epsilon, 1-\epsilon] \\ \frac{1+m_s-\epsilon}{2}, & m_s \in [1-\epsilon, 1+\epsilon] \\ \frac{-1+m_s+\epsilon}{2}, & m_s \in [-1-\epsilon, -1+\epsilon] \end{cases}$$

or, if the message is empty, $\mu(\theta|\emptyset) = \int_{\underline{x}}^{\bar{x}} \int_{\underline{m}}^{\bar{m}} 1_{\tilde{m}(x_s,m_s)=\emptyset} \mu(\theta|m_s) dm_s dx_s$.

**Sender's decision**: Fix the mean posterior upon seeing nothing, $\mu(\theta|\emptyset)$. Then

$$\tilde{m}(x_s, m_s) = \begin{cases} \emptyset & \text{if } x_s > 0 \text{ and } \mu(\theta|\emptyset) - \mu(\theta|m_s) \in [0, \frac{2x_s}{c}] \\ & \quad \text{or } x_s < 0 \text{ and } \mu(\theta|\emptyset) - \mu(\theta|m_s) \in [\frac{2x_s}{c}, 0] \\ m_s & \text{else.} \end{cases}$$

The belief $\mu(\theta|\emptyset)$ is determined in equilibrium, and fully characterizes the equilibrium actions through the above best responses.

**Claim 5.1** *For all $c$, the unique equilibrium consistent with a Bayesian receiver is given by the above actions and $\mu(\theta|\emptyset) = 0$.*

Figure 5 shows when the sender chooses to disclose or withhold signals when $c = 1$. This is

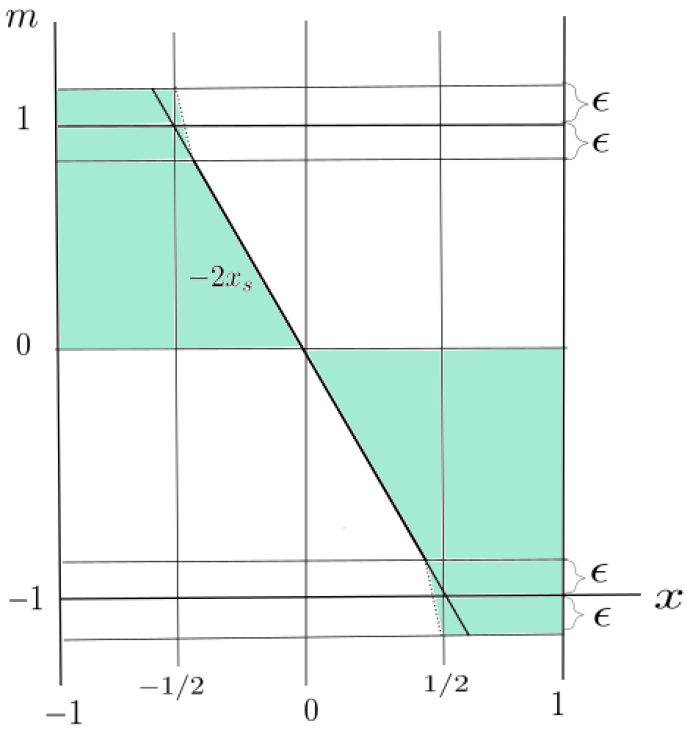



Figure 5: $c = 1$

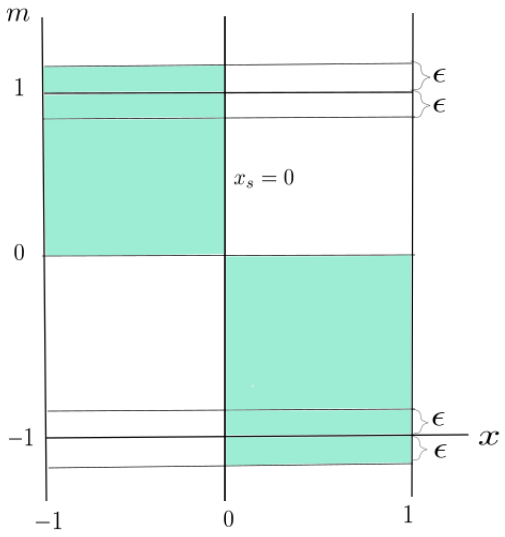



Figure 6: $c = 2$

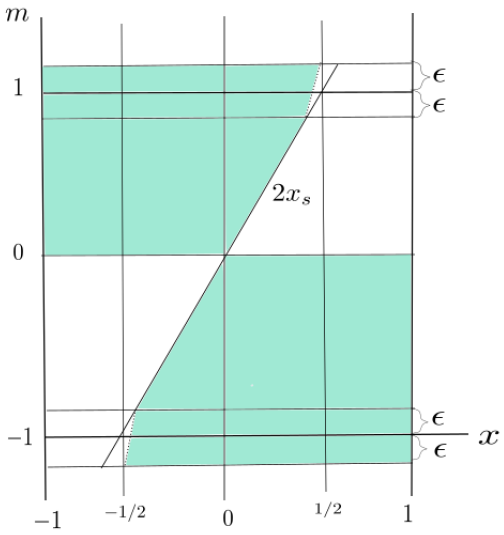



Figure 7: $c = 3$

a "pure bias" setting in which senders and receivers are equally sensitive to the state, but the sender's preferences are offset from the receiver's based on private preferences. Withholding occurs in the shaded area.

In contrast, when $c > 1$ the sender is more moderate relative to the receiver, introducing a new dimension of misalignment between the players that varies with the magnitude of the signal $m_s$. The case $c = 2$ is liminal, and disclosure policies are particularly simple here: senders disclose everything aligned with their bias, and hide everything opposed.

When the sender is much less sensitive to the state than the receiver, i.e. $c > 2$, we approach the case where, relative to the receiver, the sender is almost state-agnostic: extreme signals are rarely revealed, and only the most extreme senders are willing to disclose them.

We can contrast this outcome with those possible when the receiver is certain about how the sender wishes to influence him:

**Claim 5.2** *If, in this example, the sign of the sender's preference type $x_s$ is known, then in the unique equilibrium when $0 < c < 2$, the sender's messages are fully separated according to the realized signal.*

Knowledge of sender type alone does not guarantee full disclosure, and in particular, Claim 5.2 is not true if $c > 2$: as Claim 6.2 in the following section will show, positive MBM is a sufficient condition to guarantee full disclosure under sender-type certainty, but it turns out negative MBM is not.[6]

# 6 Full separation with unidirectionality or certainty

The main way in which my model departs from the literature is by assuming that senders' preferences are both uncertain and bidirectional. Plenty of models assume the opposite, and obtain full disclosure. A well-known result allowing some variation in sender preferences by type is the monotonicity theorem of Okuno-Fujiwara et al. [1990], which states that full disclosure is the unique possible outcome whenever senders' utilities are monotone in the receiver's beliefs over the entire space of scenarios. Though intuitive, this theorem is not a particularly good fit to settings in which senders' preferences have a single peak in the interior of the action space. More applicable is the idea from Seidmann and Winter

[6]Nor is negative MBM sufficent to get complete nondisclosure. Partial disclosure strategies are possible under negative monotonicity.

[1997] that full disclosure is a possible outcome if and only if each possible message admits a different "worst-case" scenario, which senders in no other scenario would like to pretend to be. Their logic applied here shows that if BSB fails dramatically, in that the sender-optimal action is either always greater or always less than the receiver-optimal action in any scenario, then full disclosure is the *unique* possible outcome.

**Claim 6.1** *If for all $m_s$, $x_s$, $a^*_{r,r}(\tilde{h}(m_s)) < (a^*_{r,s}(\tilde{h}(m_s), x_s))$, then $\hat{m} = \underline{m}$, and in the unique equilibrium the sender reveals fully reveals his signal by choosing $\tilde{m} = m_s$ under all signal realizations. A similar argument holds when $a^*_{r,r}(\tilde{h}(m_s)) > \max_{x_s}(a^*_{r,s}(\tilde{h}(m_s), x_s))$.*

What happens if, instead, BSB is satisfied, but the receiver knows exactly what the sender's preferences are as a function of the signal? Because misalignment between the sender's and receiver's preferences may be state-dependent, this is not as straightforward as ensuring the direction of bias is known, and the literature does not seem to have touched on it. Nevertheless, under positive MBM, I can show that in stark contrast to the outcome under type uncertainty, if the receiver has full knowledge that the sender is of preference type $x_s$, all equilibria are fully revealing, and there exists at least one.

**Claim 6.2** *If $x_s$ is known to the receiver, and the breakeven message is positive monotone, then an equilibrium exists, and in any equilibrium $m_s$ is fully revealed.*

# 7 Conclusion

Can a sender with unknown objectives and access to hard evidence influence others through their choice to disclose or withhold evidence? The answer depends on how the audience updates their beliefs under nondisclosure. This paper shows that when a sender could potentially have either of two opposing biases, receivers can't fully back out the sender's evidence or their identity. Therefore, relative to a symmetric information benchmark, ownership of evidence benefits the sender by allowing her to withhold some unfavorable news. In cases where senders and receivers are similarly sensitive to the state, but senders have a state-independent bias, strong signals tend to be revealed, whereas weak ones are often hidden, especially by heavily biased sources. Alternatively, when senders are agnostic to the state, they avoid disclosing strong signals unless also strongly biased.

A couple of extensions of this model are straightforward. I have considered preference uncertainty, and Dye [1985] considers imperfect disclosure under uncertainty about information endowments. If there is uncertainty about *both*, then imperfect disclosure will still

occur, and nondisclosing senders will pool with both informed senders under opposite scenarios, and the uninformed. The set of equilibria will differ from that without uncertainty in informedness, but the form of equilibrium and comparative statics will follow what I have outlined in this paper: the intuition is that information endowment uncertainty changes the posterior under silence, but it doesn't change informed senders' best responses conditional on the center.

Another potential addition is receiver-preference uncertainty. Among other things, allowing receivers' preferences to vary is natural in an anonymous setting where neither the sender nor the receiver's identity is known. Passing articles over the internet is a nice example of this. When receivers' preferences aggregate into a utility function satisfying the single-receiver conditions, my conclusions carry over immediately. In some work omitted here, I show that when receiver types are well-ordered and each type's expected utility depends on the state only through its expectation, pooling is also guaranteed under similar conditions.

Directions for future work include evaluating the impact of certain assumptions. I have assumed no cheap talk about the sender's preferences, but in practice communication about preferences may sometimes be possible, and may alter disclosure. In addition, signals may be divisible, allowing the sender some freedom in the degree of disclosure beyond a binary choice. I've also assumed that senders' preference types are payoff-irrelevant to the receiver, but they could instead be thought of as a payoff-relevant signal that cannot be disclosed in a game of multidimensional communication. Finally, although many instances of disclosure, such as those in a courtroom, or regarding the viability of a short-term opportunity, are well approximated by a one-shot game, some relationships between informants and audiences are long-lived. In these repeated games, persistent uncertainty about preferences may be less sustainable, but it would be interesting to explore the possibility that senders may still exercise some power by building a reputation.

# References

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

M. Baum and Y. Zhukov. Media ownership and news coverage of international conflict. *Political Communication*, 2018.

A. Chakraborty and R. Harbaugh. Persuasion by cheap talk. *American Economic Review*, 100(5), 2010.

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

J. Esteban and D. Ray. Inequality, lobbying, and resource allocation. *American Economic Review*, 96(1), 2006.

A. Frankel and N. Kartik. Muddled information. *Journal of Political Economy*, 127(4), 2019.

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

S. Grossman. The informational role of warranties and private disclosure about product quality. *Journal of Law and Economics*, 24:461–489, 1981.

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

M. Okuno-Fujiwara, A. Postlewaite, and K. Suzumara. Strategic information revelation. *Review of Economic Studies*, 57:24–47, 1990.

J. K.-H. Quah and B. Strulovici. Comparative statics, informativeness, and the interval dominance order. *Econometrica*, 77(6):1949–1992, 2009.

Daniel J Seidmann and Eyal Winter. Strategic information transmission with verifiable messages. *Econometrica*, 65(1):163–169, 1997.

# 8 Appendix

**Proof of Theorem 3.2** First, since $u_s(\theta, x_s, a_r)$ and $u_r(\theta, a_r)$ are continuous in all arguments, as is $\tilde{h}(\theta|m_s)$, the expected utilities $u_s(\tilde{h}(\theta|m_s), x_s, a_r)$ and $u_r(\tilde{h}(\theta|m_s), a_r)$ are also continuous in all arguments. Since they are single-peaked, the peaks $a^*_{r,s}(\tilde{h}(\theta|m_s), x_s)$ and $a^*_{r,r}(\tilde{h}(\theta|m_s))$ are continuous in $m_s$ and $x_s$ as well.

**Lemma 8.1** *Suppose that $b(m_s|\emptyset)$ is supported on a subset of $[c, d]$. Then $a^*_{r,r}(\beta(\theta|c)) \leq a^*_{r,r}(\beta(\theta|\emptyset)) \leq a^*_{r,r}(\beta(\theta|d))$, with strict inequality if $b(m_s|\emptyset)$ is not a point mass.*

**Proof** Consider the derivative of the receiver's expected utility with respect to their action under message $\emptyset$ given that their belief is $b(m_s|\emptyset)$. At $a^*_{r,r}(\beta(\theta|\emptyset)) = a^*_{r,r}(\tilde{h}(\theta|\hat{m}))$, the

derivative should be 0. But, for any $c' < c$ or $d' > d$,

$$\frac{\partial}{\partial a_r} u_r(\beta(\theta|\emptyset), a_r)\big|_{a^*_{r,r}(\tilde{h}(\theta|c'))} = \int_{m_s} \frac{\partial}{\partial a_r} u_r(\tilde{h}(\theta|m_s), a_r)|_{a^*_{r,r}(\tilde{h}(\theta|c'))} b(m_s|\emptyset) dm_s > 0,$$

$$\frac{\partial}{\partial a_r} u_r(\beta(\theta|\emptyset), a_r)\big|_{a^*_{r,r}(\tilde{h}(\theta|d'))} = \int_{m_s} \frac{\partial}{\partial a_r} u_r(\tilde{h}(\theta|m_s), a_r)|_{a^*_{r,r}(\tilde{h}(\theta|d'))} b(m_s|\emptyset) dm_s < 0,$$

and this is true even for $c' = c$ and $d' = d$ if $\beta(\theta|\emptyset)$ is not a point distribution. ∎

Fix a putative posterior, $\beta(\theta|\emptyset)$, for the receiver given the empty message. Observe that there is an posterior on $m_s$ given $\emptyset$, which can be denoted $b(m_s|\emptyset)$, with support on a subset of $[\underline{m}, \bar{m}]$, and

$$\beta(\theta|\emptyset) = \int_{m_s} \tilde{h}(\theta|m_s) b(m_s|\emptyset) dm_s.$$

Then $a^*_{r,r}(\beta(\theta|\emptyset)) \in [a^*_{r,r}(\tilde{h}(\theta|\underline{m})), a^*_{r,r}(\tilde{h}(\theta|\bar{m}))]$ by Lemma 8.1. The crucial step in the proof is to observe that, by intermediate value theorem, there exists $\hat{m} \in [\underline{m}, \bar{m}]$ such that $a^*_{r,r}(\beta(\theta|\emptyset)) = a^*_{r,r}(\tilde{h}(\theta|\hat{m}))$.

The following steps show that in equilibrium $\hat{m}$ must induce senders to pool via nondisclosure.

1. Fixing $\hat{m}$, define the sets

$$\bar{A}(\hat{m}) := \{(x_s, m_s) : a^*_{r,s}(\tilde{h}(\theta|m_s), x_s) > a^*_{r,r}(\tilde{h}(\theta|\hat{m}))\},$$

$$\underline{A}(\hat{m}) := \{(x_s, m_s) : a^*_{r,s}(\tilde{h}(\theta|m_s), x_s) < a^*_{r,r}(\tilde{h}(\theta|\hat{m})).$$

2. Then define

$$\bar{B}(\hat{m}) := \bar{A}(\hat{m})) \bigcap \{(x_s, m_s) : m_s < \hat{m}\},$$

$$\underline{B}(\hat{m}) := \underline{A}(\hat{m})) \bigcap \{(x_s, m_s) : m_s > \hat{m}\}.$$

   Both are the intersection of open sets, thus also open.

3. If $(x_s, m_s) \in \bar{B}(\hat{m}) \bigcup \underline{B}(\hat{m})$, then by single-peakedness the sender prefers $a^*_{r,r}(\tilde{h}(\theta|\hat{m}))$ to $a^*_{r,r}(\tilde{h}(\theta|m_s))$, and would like to withhold $m_s$ instead of disclosing it.

4. Whenever $\bar{B}(\hat{m}) \bigcup \bar{B}(\underline{m})$ is nonempty, $\{m_s : (x_s, m_s) \in \bar{B}(\hat{m}) \bigcup \underline{B}(\hat{m})\}$ must therefore be in the support of $b(m_s|\emptyset)$, and $b(m_s|\emptyset)$ has positive measure over it.

5. Because $g(x_s)$ is not a point distribution, there always exists either $x_s$ such that $a^*_{r,s}(\tilde{h}(\theta|\hat{m}), x_s) > a^*_{r,r}(\tilde{h}(\theta|\hat{m}))$ or $x_s$ such that $a^*_{r,s}(\tilde{h}(\theta|\hat{m}), x_s) < a^*_{r,r}(\tilde{h}(\theta|\hat{m}))$.

   If $\hat{m} \in (\underline{m}, \bar{m})$, then the former implies that $\bar{B}(\hat{m})$ is nonempty, and the latter implies $\underline{B}(\hat{m})$ is nonempty.

6. Alternatively, if $\hat{m} = \bar{m}$ or $\hat{m} = \underline{m}$, then BSB implies that $\bar{B}(\hat{m})$ or $\underline{B}(\hat{m})$ are nonempty, respectively.

This suffices to show that all equilibria must feature pooling: there is a positive measure of signals which senders, depending on their type, have a positive probability of withholding.

As an addendum to point 6, observe that if the receiver's belief is represented by $\hat{m} = \bar{m}$, then the sender's BR induces her to withhold some signals $m_s < \hat{m}$, but no signals $m_s > \hat{m}$ (since such signals do not exist), which is inconsistent with the original belief. A similar observation holds if $\hat{m} = \underline{m}$. Thus, in equilibrium it must be that $\hat{m} \in (\underline{m}, \bar{m})$. By Lemma 8.1, since $b(m_s|\emptyset)$ is not a singleton, it places positive probability on elements to either side of $\hat{m}$.

Finally, given $\hat{m}$, there is at most one sender type that satisfies $a^*_{r,s}(\tilde{h}(\theta|\hat{m}), x_s) = a^*_{r,r}(\tilde{h}(\theta|\hat{m}))$. Any type $x_s$ with $a^*_{r,s}(\tilde{h}(\theta|\hat{m}), x_s) > a^*_{r,r}(\tilde{h}(\theta|\hat{m}))$ will withhold for some set of signals $m_s < \hat{m}$, and any $x_s$ with $a^*_{r,s}(\tilde{h}(\theta|\hat{m}), x_s) < a^*_{r,r}(\tilde{h}(\theta|\hat{m}))$ will withhold for some $m_s > \hat{m}$. ∎

**Proof of Claim 6.1** Consider the case in which $a^*_{r,r}(\tilde{h}(m_s)) < a^*_{r,s}(\tilde{h}(m_s), \underline{x}_s)$ for all $m_s$. Whenever $\hat{m} < m_s$, the sender prefers to send $m_s$, because $a^*_{r,r}(\tilde{h}(\hat{m})) < a^*_{r,r}(\tilde{h}(m_s)) < a^*_{r,s}(\tilde{h}(m_s), \underline{x}_s)$. However, whenever $\hat{m} \neq \underline{m}$, for every type $x_s$ there exists a positive measure of $m_s < \hat{m}$ such that $a^*_{r,r}(\tilde{h}(m_s)) < a^*_{r,r}(\tilde{h}(\hat{m})) < a^*_{r,s}(\tilde{h}(m_s), \underline{x}_s)$. Therefore, if $\hat{m} \neq \underline{m}$, all withholding occurs for $m_s < \hat{m}$, which gives a contradiction.

When $\hat{m} = \underline{m}$, once again if $m_s > \hat{m}$, no sender wishes to withhold. Thus, senders may only withhold if $m_s = \hat{m}$, which is consistent with $\hat{m}$ representing the receiver's posterior, and results in full separation of signals.

**Proof of Claim 6.2** This claim follows directly from the fact that, under MBM, whenever the sender withholds, he withholds an interval of signals to one side or the other of $\hat{m}$. The sender cannot pool with other sender types, so if the sender attempts to withhold a nonempty interval of signals, the receiver should, as part of their best response, nontrivially update $\hat{m}$. Thus, $\hat{m}$ represents a fixed point of the two players' strategies only if the sender

never withholds, which may be the case either when $\hat{m} \in \{\underline{m}, \bar{m}\}$ or when $m_s^=(\hat{m}, x_s) = \hat{m}$.[7]

**Proof of Prop. 4.3** Since utilities are single-peaked, the sender weakly prefers to withhold whenever $\hat{m}$ is inside $[m_s^=(m_s, x_s), m_s]$ or $[m_s, m_s^=(m_s, x_s)]$.

Fix $x_s$. The breakeven message $m_s^=(m_s, x_s)$ is strictly increasing in $m_s$, so there is a single signal $M$ at which $m_s^=(x_s, M) = \hat{m}$. For all $m_s < M$, $m_s^=(m_s, x_s) < \hat{m}$; for all $m_s > M$, $m_s^=(m_s, x_s) > \hat{m}$. Thus, either $M < \hat{m}$, and $\hat{m} \in [m_s, m^=(x_s, m_s)]$ iff $m_s \in [M, \hat{m}]$; or, $M > \hat{m}$, and then $\hat{m} \in [m^=(x_s, m_s), m_s]$ iff $m_s \in [\hat{m}, M]$.

[7]Visually, this is to say that $\hat{m}$ must be either on one of the boundaries or at one of the nodes at which $m_s$ and $m_s^=(m_s, x_s)$ intersect in Figure 4a.