Consistency and Strategic Disclosure in Model-based Persuasion

Jun Hyun Ji*

July 20, 2026

Click here for the most recent version

Abstract

Persuasion often works by supplying a model through which observed data are interpreted. I study persuasion when the receiver, beyond requiring that the model fit the observed signals well, demands that it be consistent, treating the signals as independent and identical draws rather than reading meaning into patterns. I characterize the persuasion cost of consistency in terms of the entropy of the empirical distribution of the signals, so that the more the signals differ from one another, the weaker the persuasive power. I then consider a sender who privately observes the signals and discloses only part of them at the time of persuasion, before the receiver sees the rest. When the signals are not all alike, the sender can avoid the cost of consistency entirely and fully manipulate the receiver's beliefs. When they are all identical, disclosing more of them makes the signals easier to fit, but consistency limits how much this added evidence can work in the sender's favor. The optimal disclosure is either a single signal or all signals; a single signal is best only when the receiver initially reads the signal as sufficiently bad news.

*Department of Economics, University of Pittsburgh. Email: juj25@pitt.edu.

1. Introduction

Persuasion is often achieved by providing models that fit observed data well. In contexts ranging from politics to science, experts provide models to help audiences map observed data to underlying states of the world. As Schwartzstein and Sunderam (2021) argue, this model-based persuasion is ubiquitous, and its persuasive power depends on how surprising the data are to the audience: persuasion tends to succeed when the audience's own understanding of the world fits the data poorly, leaving them to seek out an expert's interpretation of what they have seen.

But fitting the data well does not by itself make an interpretation convincing. Fit can be achieved by appealing to patterns in the realized data—reading meaning into a trend, or into the particular order in which events unfolded—and this frequently invites skepticism. A financial advisor would lose credibility if she argued that market fundamentals shift erratically from day to day, just as a climate scientist would lose trust proposing a model in which physical laws vary by season. A regression whose R^2 is suspiciously high invites concerns about overfitting rather than confidence in the result. In each case the problem is not that the interpretation fails to explain the data; it is that the interpretation appears to exploit the data already in hand to manufacture a desired story, in a way that conflicts with what the audience believes must be true of how the data are generated in the first place.

I study one sharp form of this skepticism: the knowledge that the observations in the data are independent of one another. An audience that thinks this way grants the persuader no room to read meaning into trends or patterns. How difficult is it to persuade this skeptical audience compared to audiences that do not possess such knowledge? Can persuaders overcome this difficulty by holding private information about the data?

To address these questions, I formalize such an audience within the framework of model-based persuasion, where a sender seeks to influence a receiver's beliefs about unknown states by proposing an alternative model. A model describes how informative each observable history of signals is of unknown states. Formally, it specifies likelihoods of signals conditional on states. As in Schwartzstein and Sunderam (2021), a naive Bayesian receiver rejects the sender's proposed model unless it fits the observed history better than his own default model—unless the new worldview makes the evidence he has seen appear more likely than it did under his initial view. I refer to this requirement as *fit-dominance*. In addition, the sender is required to offer a consistent account that applies to every signal alike, rather than reading a different meaning into each part of the realized history. This amounts to requiring the sender's model to treat each signal as an independent draw from a state-dependent distribution. I refer to this requirement as *consistency*.

Once consistency is required, three questions become tractable. First, how costly is it? I measure the cost by the ratio between the maximum belief the sender could induce under fit-dominance alone and the maximum belief she can induce once consistency is also required. I characterize this ratio in terms of the Shannon entropy of the empirical distribution of the observed signals. Entropy measures how heterogeneous the observed history is: it is highest when each different signal appears equally often and lowest when they are all the same. The more heterogeneous the history, the harder it is to explain, and the greater the cost of consistency. For example, if some market data look purely random, a fund manager's best argument is that "pure randomness is a good sign." But this also means that when conditions are good, the data could have looked less random as well (e.g., flipping two fair coins is just as likely to result in two heads or two tails). In other words, proposing a model that generates random signals in favorable states cannot strongly support those states, weakening its persuasive power.

Second, when is this cost severe enough that persuasion becomes impossible? This happens when the receiver's default model already treats the signals as uninformative and matches the observed history—a far weaker condition than under fit-dominance alone, where the sender is powerless only if the default fits the history perfectly. Third, when is the cost severe enough that persuasion makes the sender worse off? If the receiver's own default model does not satisfy consistency, yet he still demands consistency from the sender, the best model the sender can offer may move his belief below where his default already left it. The sender then prefers to propose nothing at all.

Can persuaders use private information about the data to mitigate the cost of consistency? I next turn to a setting in which the timing of persuasion and the timing of the receiver's learning come apart. Experts often know more of the data than their audience does at the moment they offer an interpretation, which may give them an incentive to strategically choose which part of the data to reveal when proposing a model. To capture this, I let the sender privately observe the full history, disclose a subset of it while proposing a model, and let the receiver adopt or reject that model based on how well it fits the disclosed data alone. Only afterward does the receiver observe the full history and apply the adopted model to it. The receiver is naive in the sense that he does not re-evaluate the model after seeing the full history. The sender thus chooses which evidence to reveal before committing to an interpretation, anticipating that the receiver will later evaluate all the evidence through that interpretation.

This ability to frame which data enters the explanation changes the sender's power, and it does so in a way that depends on the heterogeneity of the history. Recall that a heterogeneous history is exactly what makes consistency costly. Yet whenever the history contains more

than one type of signal, the sender can move the receiver's posterior all the way to certainty in the favored state, regardless of his default model, by disclosing any subset that excludes at least one signal type. By doing so, she retains full flexibility in interpreting the excluded signals once they are eventually revealed, and thereby avoids the cost of consistency entirely. The combination of the sender's private observation of the history and her ability to frame which data enters the explanation can therefore severely weaken the disciplining role of both consistency and fit-dominance.

When the history instead consists only of identical signals, consistency carries no direct cost because there is no heterogeneity in the history. However, when the sender chooses how many of them to disclose, her problem is shaped by a tension between the two requirements she faces. Disclosing more signals makes the receiver's default model fit them less well, which gives the sender more flexibility to differentiate the states in her interpretation. But because the sender must explain every signal through the same fixed likelihood, disclosing a longer sequence forces her to raise the likelihood she assigns to the disclosed signals in every state, which reduces that flexibility. I show that this trade-off can be characterized tractably, and that the optimal disclosure is always at a corner: the sender discloses either a single signal or the entire history.

Which of the two corners is optimal depends on how the receiver's default model views the signal. Disclosing a single signal is optimal when the default model regards the signal as strong evidence for the unfavorable state (i.e., the receiver sees it as a bad sign). Intuitively, when the receiver's initial interpretation of the signal is sufficiently pessimistic, each additional signal he sees pushes his belief further toward the unfavorable state, thus revealing many of them only deepens the view the sender is trying to overturn. Disclosing just one keeps that effect to a minimum.

This paper contributes to a growing literature on model-based persuasion initiated by Schwartzstein and Sunderam (2021), in which a persuader influences beliefs not by supplying new data, but by supplying an interpretation of publicly observed data. Their sender faces a single requirement, fit-dominance, and I adopt it along with their assumption that the receiver evaluates a proposed model by its fit rather than by strategic inference about the sender's motives. My contribution is twofold. First, I introduce the consistency requirement and characterize its consequences for persuasion.[1] Second, I study the sender's disclosure problem, in which the timing of persuasion and the timing of the receiver's learning are separated, and characterize her optimal disclosure strategy.

[1] Schwartzstein and Sunderam (2021), in their online appendix, raise the idea of restricting the sender's flexibility by requiring the sender's model to relate different parts of the history to the state in the same way. In that sense, the consistency requirement studied here sits under that idea, and yields a tractable refinement.

Several papers extend Schwartzstein and Sunderam (2021) in other directions. Aina (2025) lets the sender commit to a menu of models before the history is realized; Ba et al. (2026) let the sender design the signals in addition to proposing a model, and endow the receiver with reasoning about the sender's motives; and Jain (2023) separates the roles of generating and interpreting the data across two competing senders. This paper takes a different direction: refining the receiver's knowledge of how the signals are generated, which the consistency requirement captures, and the sender's private information about the data, which gives rise to the disclosure problem.

The paper also relates to Burkovskaya and Starkov (2026), who study a sender aiming to convince a receiver that one variable causes another. Their sender privately observes data on multiple variables, discloses a subset of them, and proposes a directed acyclic graph as an interpretation of the causal relationships among the variables. As in this paper, the sender privately observes the data and discloses only part of it before an interpretation is fixed, but the two settings differ in several respects. Their sender seeks to establish a causal link, whereas the sender here seeks to move the receiver's belief about a fixed, unknown state; and there are essentially two variables throughout this paper, the history and the state; hence the sender discloses a subset of the observed history rather than disclosing new variables.

More broadly, the paper relates to the two classical branches of the literature on strategic communication: the disclosure of verifiable information, in which a sender chooses what to reveal from data she cannot fabricate (Grossman and Hart, 1980; Milgrom, 1981), and Bayesian persuasion, in which a sender designs the process that generates a signal (Kamenica and Gentzkow, 2011). Model-based persuasion differs from both, since the sender supplies an interpretation of already-realized data rather than withholding realizations or designing the signal process. The disclosure problem studied here shares with the first branch the sender's choice of which observations to disclose, while retaining the interpretive model-proposal problem of model-based persuasion.

The remainder of the paper is organized as follows. Section 2 sets up the model. Section 3 characterizes the cost of consistency when the data are public. Section 4 analyzes the sender's disclosure problem when she privately observes the data. Proofs, if omitted, are collected in the appendix.

2. Consistency in Model-Based Persuasion

A sender (she) seeks to influence a receiver's (he) beliefs by proposing a model to interpret publicly observed signals. The receiver adopts the proposed model if it satisfies two conditions: it must fit the observed signals well, and it must be consistent. In this section, I

specify these conditions and discuss their implications.

2.1. Setup

Formally, there is an unknown state of the world $\omega \in \Omega = \{G, B\}$. The receiver holds a prior belief $\mu_0 \in \text{int}(\Delta\Omega)$, known to the sender. There is a history $h = (s_1, \ldots, s_N) \in S^N$ of N signals drawn from the set S of signals. For each $s \in S$, let $h(s) = \sum_{i=1}^{N} \mathbb{1}\{s_i = s\}$ denote the number of times signal s appears in the history. A *model* specifies, for each state, a probability distribution over histories. Formally, a model is an element $p \in (\Delta S^N)^{|\Omega|}$. Thus, for each $\omega \in \Omega$, the model assigns a distribution $p(\cdot \mid \omega) \in \Delta S^N$ over histories. Define the *h-fit* of model p by $\Pr(h \mid p) := \sum_{\omega \in \Omega} \mu_0(\omega) p(h \mid \omega)$, which is the likelihood of h under model p. Given a model p and the observed history h, let the posterior belief induced by model p be $\mu_h(\cdot \mid p) \in \Delta\Omega$. After the model selection, the receiver behaves as a Bayesian and thus $\mu_h(\omega \mid p) = \frac{\mu_0(\omega) p(h|\omega)}{\Pr(h|p)}$ for each ω.

The receiver's payoff function $U^R(a, \omega)$ depends on his action $a \in A$ and the state ω. When he adopts model p and observes the history h, he chooses an action that maximizes his expected payoff using the posterior belief $\mu_h(\cdot \mid p)$: that is,

$$a(h; p) \in \arg\max_{a \in A} \mathbb{E}_{\mu_h(\cdot|p)} \left[U^R(a, \omega) \right].$$

For simplicity, I assume that $A = [0, 1]$ and that the optimal action is $a(h; p) = \mu_h(G \mid p)$ for all h, p, and hence I do not specify the function U^R explicitly. The sender's payoff function U^S is state-independent and strictly increasing in the receiver's action: e.g., $U^S(a) = a$ and thus, $U^S(a(h; p)) = \mu_h(G \mid p)$. That is, her objective is to choose a model p that maximizes the posterior belief about state G, where p is chosen from a compact set $\mathbb{M}$. When $\mathbb{M} = (\Delta S^N)^{|\Omega|}$, the sender faces no restriction on model choice.

Figure 1 summarizes the timeline of the game, which proceeds in two stages. In the *observation stage*, the history $h = (s_1, \ldots, s_N)$ is publicly observed by both the sender and the receiver. In the *persuasion and learning stage*, the sender proposes a model p, the receiver updates his belief using the proposed model, and chooses an action accordingly.

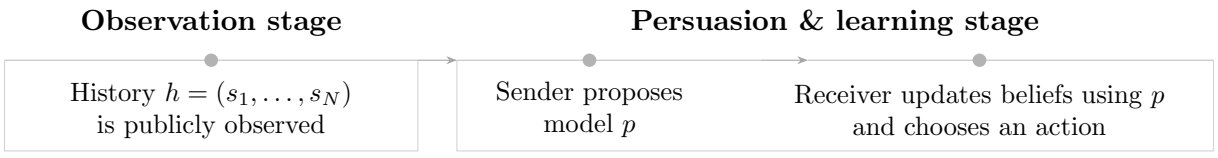


Figure 1: Timeline of the baseline model-based persuasion.

2.2. Fit-dominance Constraint

The receiver adopts the sender's proposed model only if it satisfies two conditions. First, the *fit-dominance* (FD) constraint requires that the proposed model fits the observed history well. As a benchmark for how well it fits the history, the receiver is endowed with a *default model* $p_d \in (\Delta S^N)^{|\Omega|}$, which is known to the sender. Formally, the h-fit of the proposed model must be weakly higher than that of the default model:

$$\Pr(h \mid p) \geq \Pr(h \mid p_d).$$

Otherwise, the receiver sticks to the default model and chooses $a(h; p_d)$.

The FD constraint is introduced by Schwartzstein and Sunderam (2021). As in their framework, the receiver has prior interpretations of the world that the sender cannot directly manipulate, evaluates models by their fit to the realized data, and, conditional on the selected model, updates by Bayes' rule. I also maintain their behavioral assumption that the receiver does not generate new models himself and does not adjust for the sender's incentives or for the sender's flexibility in choosing a model after observing the data. Thus, the sender's persuasive power comes from supplying an interpretation of the data that the receiver is willing to entertain.

2.3. Consistency

The second condition is that the sender can only propose models under which the signals $s_1, \ldots, s_N$ are independent and identically distributed (IID). This is called the *IID constraint*. Formally let $\mathbb{M}_{iid} \subset (\Delta S^N)^{|\Omega|}$ denote the set of such models, and let $(\Delta S)^{|\Omega|}$ denote the collection of state-contingent signal distributions, where each $\pi \in (\Delta S)^{|\Omega|}$ is called an *IID model*. Then, $p \in \mathbb{M}_{iid}$ if and only if there exists $\pi \in (\Delta S)^{|\Omega|}$ such that for each ω,

$$p(h \mid \omega) = \prod_{s \in S} \pi(s \mid \omega)^{h(s)}.$$

The imposition of the IID constraint departs from Schwartzstein and Sunderam (2021).[2] It requires the sender to treat the observed signals as independent draws from a fixed distribution—one that may depend on the state but not on when or in what order the signals arrived. This means the sender cannot assign likelihoods to each history directly, which rules out models that explain the history by appealing to trends, momentum, or any other form of cross-signal dependence.

The IID constraint can be interpreted as a parsimonious way to capture the receiver's prior knowledge that the signals are IID, or his demand for a consistent principle governing how each signal arises. In many applied settings, explanations that rely on path-specific or highly contingent features of the realized history are viewed as ad hoc and therefore unconvincing. An expert who argues that each individual realization occurred for a case-specific reason might lead the receiver to infer that the expert lacks knowledge of the true data-generating process.

I assume henceforth that the receiver's default model p_d is itself consistent, i.e., $p_d \in \mathbb{M}_{iid}$, and denote the corresponding IID model by π_d. If the receiver demands an IID model from the sender, it is natural that his own default interpretation of the data satisfies the same standard. I also assume the IID default model to be non-degenerate: i.e., $\pi_d(s \mid \omega) \in (0,1)$ for each $\omega \in \Omega$ and $s \in S$.

Under the two constraints discussed above, the sender's problem is:

$$p^*(h, p_d) \in \arg\max_{p \in \mathbb{M}} \mu_h(G \mid p) \text{ subject to } \underbrace{\Pr(h \mid p) \geq \Pr(h \mid p_d)}_{\text{FD}} \text{ and } \underbrace{p \in \mathbb{M}_{iid}}_{\text{IID}}.$$

Note that the sender essentially faces a decision problem, not a game. The receiver's strategic aspects are fully reduced to the sender's constraints. Any sophistication of the receiver may lead to a cheap talk environment whose nature depends on the structure of information asymmetry that is not specified in this paper (e.g., whether the sender knows the true model that generated the history, the true state, or both, and what the receiver can infer from the sender's proposed model). In this paper, the true model that generated the history influences neither the sender's model choice nor the receiver's decision to adopt it.

[2] The IID constraint relates to a refinement suggested by Schwartzstein and Sunderam (2021) in their online appendix, which they leave for future work. They suggest *internal consistency*: the receiver may demand that certain components of the history relate to the state in the same way. For example, consider a history $h = (s_1, s_2, s_3, s_4)$ and a receiver who knows that $p(s_1 \mid s_2, \omega) = p(s_1 \mid s_3, \omega)$ for all $\omega \in \Omega$: the likelihood of s_1 in every state must be the same whether the model conditions on s_2 or s_3. The IID constraint goes strictly further in two ways. First, it is universal: rather than applying to selected pairs, it requires every signal to relate to the state through the same mechanism. Second, it imposes independence: $p(s_1 \mid s_2, s_3, \omega) = \pi(s_1 \mid \omega)$. That is, the distribution of s_1 given ω does not depend on s_2 or s_3 at all. In this sense, the IID constraint provides a tractable formalization of their consistency refinement.

3. Cost of Consistency

3.1. Persuading without consistency requirement

I first characterize the benchmark case: the extent to which the sender can influence the receiver's beliefs under the FD constraint alone.

Proposition 1 (Schwartzstein and Sunderam (2021)'s Proposition 1)**.** *Given h, p_d, μ_0, let $\mu^{\text{FD}}(G)$ denote the maximum posterior about state G achievable under the FD constraint. Then*

$$\mu^{FD}(G) = \min\left\{1, \frac{\mu_0(G)}{\Pr(h \mid p_d)}\right\}.$$

This restates the benchmark result of Schwartzstein and Sunderam (2021). It shows that persuasive power is governed by how well the realized history is explained by the receiver's default model. When $\Pr(h \mid p_d)$ is small, the history is unlikely under the default model, and the receiver is therefore more willing to adopt an alternative model. In this sense, the sender's flexibility increases with the degree of surprise in the history. Formally, the term $\frac{1}{\Pr(h|p_d)}$ captures the level of surprise: a lower default likelihood corresponds to greater scope for persuasion. This effect is moderated by the prior $\mu_0(G)$. For a given level of surprise, a higher prior belief in state G allows the sender to induce a higher posterior. In the extreme, if $\Pr(h \mid p_d) \leq \mu_0(G)$, the sender can fully persuade the receiver and induce belief 100% in state G.

The result is driven by the structure of the sender's optimal model. In particular, the sender chooses a model $\bar{p}$ such that

$$\bar{p}(h \mid G) = 1, \qquad \bar{p}(h \mid B) = \max\left\{0, \frac{\Pr(h \mid p_d) - \mu_0(G)}{\mu_0(B)}\right\}.$$

Under this model, the history h is assigned probability one in state G, while its probability in state B is set as low as possible subject to the FD constraint. Absent this constraint, the sender would set $p(h \mid B) = 0$, thereby arguing that the observed history is impossible under state B. This induces posterior belief 100% on G. The FD constraint limits how far the sender can reduce $p(h \mid B)$. Specifically, the sender must ensure that the overall likelihood of h under the proposed model remains exactly the same as under the default model, i.e., $\Pr(h \mid \bar{p}) = \Pr(h \mid p_d)$. This requirement implies a lower bound on $p(h \mid B)$, given by $\frac{\Pr(h|p_d)-\mu_0(G)}{\mu_0(B)}$, which binds whenever full persuasion is not feasible. When $\mu_0(G) \geq \Pr(h \mid p_d)$, this lower bound is nonpositive, and the sender can set $p(h \mid B) = 0$. In this case, the prior is sufficiently favorable that the sender can present the realized history as conclusive evidence for state G, achieving full persuasion.

In contrast, the FD constraint places no restriction on $p(h \mid G)$. The sender can always set $p(h \mid G) = 1$, regardless of the realized history. Thus, any history can be made perfectly consistent with the good state: the sender can argue that the observed history—no matter how specific or complex—was destined to occur under state G. In other words, she can effectively "curve-fit" a unique logic to every individual observation. For example, consider the realized path of a stock market index over N days, with irregular increases, sharp drops, and rebounds. The sender can impose a specific interpretation on this exact sequence, arguing that "this is exactly what happens in state G."

3.2. The scope of persuasion and entropy of signal histories

The IID constraint places an upper bound on the likelihood of a history conditional on the favored state. I characterize this bound by separating the likelihood into two distinct components. The first component is entropy, which depends only on the empirical distribution of the observed history and captures its degree of heterogeneity. The second component is the divergence between the proposed model and the history, which captures how closely the model aligns with the observations. This decomposition provides the key tool for characterizing the cost of consistency in the next section.

For any IID model π, let $p_\pi \in (\Delta S^N)^{|\Omega|}$ denote the induced model over histories, given by $p_\pi(h \mid \omega) = \prod_{s \in S} \pi(s \mid \omega)^{h(s)}$ for each $\omega \in \Omega$. Define the *empirical distribution of signals in history* h by

$$\hat{\pi}_h = (\hat{\pi}_h(s))_{s \in S} = \left(\frac{h(s)}{N} \right)_{s \in S}.$$

The *entropy of the empirical distribution* $\hat{\pi}_h$ is $\mathcal{H}(\hat{\pi}_h) = -\sum_{s \in S} \hat{\pi}_h(s) \log \hat{\pi}_h(s)$.[3] For each $\omega \in \Omega$, the *Kullback-Leibler (KL) divergence* from $\hat{\pi}_h$ to $\pi(\cdot \mid \omega) \in \Delta S$ is

$$\mathcal{D}_{\mathrm{KL}}(\hat{\pi}_h \| \pi(\cdot \mid \omega)) = \sum_{s \in S} \hat{\pi}_h(s) \log \left(\frac{\hat{\pi}_h(s)}{\pi(s \mid \omega)} \right).$$

Lemma 1. *For each $h \in S^N$ and $\omega \in \Omega$, a distribution $\pi(\cdot \mid \omega) \in \Delta S$ induces*

$$p_\pi(h \mid \omega) = \exp\left(-N[\mathcal{H}(\hat{\pi}_h) + \mathcal{D}_{\mathrm{KL}}(\hat{\pi}_h \| \pi(\cdot \mid \omega))]\right).$$

Proof. Begin with the likelihood written as a product $p_\pi(h \mid \omega) = \prod_{s \in S} \pi(s \mid \omega)^{h(s)}$. Insert

[3] The Shannon entropy $\mathcal{H}(q)$ defined for a probability distribution q over signals measures the expected "surprise" in a random draw from q (see Shannon, 1948). In this context, $\mathcal{H}(\hat{\pi}_h)$ captures the degree of heterogeneity in the observed history: the value is higher when the empirical distribution is more dispersed across signals and lower when it is more concentrated.

the empirical distribution $\hat{\pi}_h(s)$ by multiplying and dividing inside the product, which factors the expression into

$$p_\pi(h \mid \omega) = \prod_{s \in S} \hat{\pi}_h(s)^{h(s)} \left(\frac{\pi(s \mid \omega)}{\hat{\pi}_h(s)} \right)^{h(s)}.$$

Using the identity $x = \exp(\log x)$ and substituting $h(s) = N\hat{\pi}_h(s)$ then yield

$$p_\pi(h \mid \omega) = \exp\left(N \sum_{s \in S} \hat{\pi}_h(s) \log \hat{\pi}_h(s) \right) \exp\left(N \sum_{s \in S} \hat{\pi}_h(s) \log \left(\frac{\pi(s \mid \omega)}{\hat{\pi}_h(s)} \right) \right).$$

Rearranging the second term introduces a minus sign and produces

$$\begin{aligned} p_\pi(h \mid \omega) &= \exp\left(N \sum_{s \in S} \hat{\pi}_h(s) \log \hat{\pi}_h(s) \right) \exp\left(-N \sum_{s \in S} \hat{\pi}_h(s) \log \left(\frac{\hat{\pi}_h(s)}{\pi(s \mid \omega)} \right) \right) \\ &= \exp\left(-N\mathcal{H}(\hat{\pi}_h) \right) \exp\left(-N\mathcal{D}_{\mathrm{KL}}(\hat{\pi}_h \| \pi(\cdot \mid \omega)) \right) \end{aligned}$$

which is the desired result. □

The lemma decomposes the likelihood of h under state ω into two multiplicative terms. The first, $\exp(-N \cdot \mathcal{H}(\hat{\pi}_h))$, depends only on the empirical distribution of the history and is entirely outside the sender's control. The second, $\exp(-N \cdot \mathcal{D}_{\mathrm{KL}}(\hat{\pi}_h \| \pi(\cdot|\omega)))$, depends on the proposed model and is maximized when $\pi(\cdot \mid \omega) = \hat{\pi}_h$, since KL divergence is minimized when the two distributions coincide. It follows that, under the IID constraint, the maximum likelihood the sender can assign to the history in state G is $\exp(-N \cdot \mathcal{H}(\hat{\pi}_h))$, by proposing $\pi(\cdot|G) = \hat{\pi}_h$. The entropy of the history places a hard ceiling on the sender's ability to make the history look likely under state G, a ceiling that does not exist in the benchmark where the sender can freely set $p(h \mid G) = 1$ given any history.

3.3. Persuading with consistency requirement

I now characterize the cost of imposing the IID constraint on top of the FD constraint. The following result shows that the ratio of the maximum posterior achievable under the FD constraint alone to that achievable under both constraints is determined entirely by the entropy of the empirical distribution of the history: the more heterogeneous the observed signals, the weaker the sender's persuasive power. The result assumes $\mu_0(G) < \Pr(h \mid p_d)$, the case where full persuasion is not achievable under the FD constraint alone.

Proposition 2 (Cost of Consistency)**.** *Given h, p_d, μ_0 where $p_d \in \mathbb{M}_{iid}$ and $\mu_0(G) < \Pr(h \mid p_d)$, let $\mu^{\mathrm{FD\&IID}}(G)$ denote the maximum posterior achievable under the FD and IID con-*

straints. Then

$$\frac{\mu^{\mathrm{FD}}(G)}{\mu^{\mathrm{FD\&IID}}(G)} = \exp\Big(N \cdot \mathcal{H}(\hat{\pi}_h)\Big)$$

where equality holds if and only if $\hat{\pi}_h$ is degenerate.

Proof. Let $p_\pi = (p_\pi(\cdot \mid G), p_\pi(\cdot \mid B))$ be a model induced by an IID model π. Analogous to the baseline case, the sender's objective is to maximize $p_\pi(h \mid G)$ and minimize $p_\pi(h \mid B)$ as much as possible. Let p^*_π be the optimal model and π^* the corresponding IID model. Then the induced posterior belief about state G is

$$\mu_h(G \mid p^*_\pi) = \frac{\mu_0(G)p^*_\pi(h \mid G)}{\mu_0(G)p^*_\pi(h \mid G) + \mu_0(B)p^*_\pi(h \mid B)} = \frac{\mu_0(G)p^*_\pi(h \mid G)}{\Pr(h \mid p_d)}$$

where the second equality comes from the binding FD constraint: $\mu_0(G)p^*_\pi(h \mid G)+\mu_0(B)p^*_\pi(h \mid B) = \Pr(h \mid p_d)$. By Lemma 1, it follows that

$$p^*_\pi(h \mid G) = \max_{\pi(\cdot|G)\in\Delta S} \exp\left(-N[\mathcal{H}(\hat{\pi}_h) + \mathcal{D}_{\mathrm{KL}}(\hat{\pi}_h \| \pi(\cdot \mid G)]\right).$$

This implies that the optimal signal distribution $\pi^*(\cdot \mid G)$ in state G is the KL divergence minimizer: the sender sets $\pi^*(\cdot \mid G) = \hat{\pi}_h$, eliminating the KL divergence term entirely. It follows that $p^*_\pi(h \mid G) = \exp\left(-N\mathcal{H}(\hat{\pi}_h)\right)$.

Note that the FD constraint pins down $p^*_\pi(h \mid B) = \frac{\Pr(h|p_d)-\mu_0(G)p^*_\pi(h|G)}{\mu_0(B)}$, which is feasible because the default model p_d is also induced by an IID model. Lemma 1 also implies that the default model's h-fit is also bounded above: $\Pr(h \mid p_d) \leq \exp(-N \cdot \mathcal{H}(\hat{\pi}_h))$. Hence, given $p^*_\pi(h \mid G) = \exp\left(-N\mathcal{H}(\hat{\pi}_h)\right)$, we can always find $\pi^*(\cdot \mid B) \in \Delta S$ inducing $p^*_\pi(h \mid B)$ that binds the FD constraint. This completes the proof. □

The IID constraint introduces an exponential cost captured by $\exp(N \cdot \mathcal{H}(\hat{\pi}_h)) \geq 1$, which is exactly the ratio between the two maximum posteriors. By standard properties, entropy is always nonnegative, with $\mathcal{H}(\hat{\pi}_h) = 0$ if and only if the data are degenerate, meaning all signals in the history are identical: in this case the posterior ratio is equal to 1, which means the IID constraint is costless. For any non-degenerate empirical distribution, entropy is strictly positive and the posterior ratio strictly exceeds one, implying that the IID constraint always reduces persuasive power. The cost is largest when the empirical distribution is uniform, $\hat{\pi}_h = \left(\frac{1}{|S|}, \ldots, \frac{1}{|S|}\right)$, in which case $\mathcal{H}(\hat{\pi}_h) = \log|S|$ and the posterior ratio reaches $|S|^N$. Thus, when signals are evenly distributed across all possible values, the cost of consistency is greatest.

3.4. When is persuasion impossible?

A natural question is when this cost is so severe that persuasion becomes entirely impossible, i.e., the sender cannot move the receiver's posterior beyond his prior. The following corollary characterizes this case: persuasion is impossible if and only if the receiver's default model treats the signals as uninformative about the state, and that interpretation is consistent with the empirical distribution of the history.

Corollary 1. *Given h, p_d, μ_0 where $\mu_0(G) < \Pr(h \mid p_d)$, suppose p_d is induced by an IID model π_d. Then $\mu^{\text{FD\&IID}}(G) = \mu_0(G)$ if and only if $\pi_d(\cdot \mid \omega) = \hat{\pi}_h$ for every $\omega \in \Omega$.*

Proof. By the previous result, $\mu_0(G) < \Pr(h \mid p_d)$ implies that under the IID constraint the sender's maximal attainable posterior belief about state G is

$$\mu^{\text{FD\&IID}}(G) = \exp(-NH(\hat{\pi}_h))\mu^{\text{FD}}(G) = \exp(-NH(\hat{\pi}_h)) \left(\frac{\mu_0(G)}{\Pr(h \mid p_d, \mu_0)} \right).$$

Since $\mu_0(G) > 0$, we have $\mu^{\text{FD\&IID}}(G) = \mu_0(G)$ if and only if $\Pr(h \mid p_d, \mu_0) = \exp(-NH(\hat{\pi}_h))$. Because p_d is induced by an IID model π_d, Lemma 1 implies

$$\Pr(h \mid p_d, \mu_0) = \sum_{\omega \in \Omega} \mu_0(\omega) \prod_{s \in S} \pi_d(s \mid \omega)^{h(s)} \leq \sum_{\omega \in \Omega} \mu_0(\omega) \exp(-NH(\hat{\pi}_h)) = \exp(-NH(\hat{\pi}_h))$$

where equality holds if and only if $\pi_d(\cdot \mid \omega) = \hat{\pi}_h$ for all ω. □

The condition $\pi_d(\cdot \mid \omega) = \hat{\pi}_h$ for every ω means that the receiver's default model treats signals as uninformative about the state, and the observed history is fully consistent with this interpretation. When this holds, the default model achieves the maximum h-fit, leaving no room for the sender to propose a model that fits the history better. The sender is therefore unable to improve upon the receiver's default, and persuasion collapses entirely.

This stands in sharp contrast to the benchmark case where the sender faces only the FD constraint, where persuasion is impossible only when $\Pr(h \mid p_d) = 1$, meaning the default model perfectly fits the history. Moreover, under the FD constraint alone, whether or not the default model treats signals as uninformative about the state does not influence the scope of persuasion—only the overall fit $\Pr(h \mid p_d)$ matters. The IID constraint substantially lowers this bar: even when the default model does not perfectly fit the history, if it treats signals as uninformative and matches the empirical distribution, the sender is powerless.

The corollary also shows that persuasion is possible even when the receiver's default model is the true model. To see this, suppose $S = \{s, s'\}$ and the true model is π_d, with $\pi_d(s \mid G) \neq \pi_d(s \mid B)$. Suppose further that the true state is B and the history is long

enough, so that the empirical distribution $\hat{\pi}_h$ is close to $\pi_d(\cdot \mid B)$ but not to $\pi_d(\cdot \mid G)$. Hence the default model does not match the empirical distribution conditional on every state, and therefore fails to attain the upper bound on h-fit, $\exp(-NH(\hat{\pi}_h))$. This leaves room for the sender to improve upon the default model and achieve persuasion even with a long history.

Example 1. *Consider a receiver who is deciding whether or not to invest in an asset, with prior belief $\mu_0(G) = 20\%$, where G represents a good time to invest in an asset. Suppose he observes two coin flips that result in the history $h = (Head, Tail)$ where the signal space is $S = \{Head, Tail\}$. He does not believe these coin flips are informative about the state, which is reflected in his default model: $p_d(h \mid G) = 25\%$ and $p_d(h \mid B) = 25\%$. Without any intervention, his posterior belief would remain the same as his prior belief: $\mu_h(G \mid p_d) = 20\%$.*

Suppose a fund manager seeks to persuade the receiver that the state is G based on the two coin flips. Suppose the manager only faces the FD constraint. Then the manager would propose that the history $h = (Head, Tail)$ is exactly what happens in state G. Formally, he proposes $\bar{p}(h \mid G) = 100\%$ and $\bar{p}(h \mid B) = 6.25\%$, which binds the FD constraint: $20\% \cdot \bar{p}(h \mid G) + 80\% \cdot \bar{p}(h \mid B) = 20\% \cdot p_d(h \mid G) + 80\% \cdot p_d(h \mid B) = 25\%$. Without the IID constraint, the manager convinces the receiver that the state is G with probability $\mu^{\mathrm{FD}}(G) = \frac{20\% \cdot 100\%}{25\%} = 80\%$.

Now suppose the manager also faces the IID constraint: the receiver demands a consistent explanation, requiring a fundamental law that governs the informativeness of a single coin flip about the state. The manager proposes the IID model π that matches the empirical distribution: $\pi^*(Head \mid G) = \frac{1}{2}$ and $\pi^*(Head \mid B) = \frac{1}{2}$. That is, the best the manager can do is to merely admit that the history is simply a result of two coin flips, providing no information about the state. Any other likelihood $\pi(Head \mid B)$ that is smaller than the empirical distribution would violate the FD constraint. Under this model, the likelihood of the history h conditional on state G becomes $p^*_\pi(h \mid G) = \left(\frac{1}{2}\right) \cdot \left(\frac{1}{2}\right) = 25\%$, which is exactly the term $\exp(-N\mathcal{H}(\hat{\pi}_h))$. Hence, the manager cannot change the receiver's posterior belief. It remains the same as his prior belief: $\mu^{\mathrm{FD\&IID}}(G) = 25\% \cdot \mu^{\mathrm{FD}}(G) = 20\%$.

3.5. When does persuasion leave the sender worse off?

I also consider the case where the receiver's default model is not induced by an IID model, even though he demands consistency. Is the sender willing to bear the cost of consistency? To address this, suppose the sender can choose not to propose any model, in which case the receiver relies on the default model. Proposing a model is optimal only if it induces a posterior belief that exceeds the one generated by the default. This is when $\mu^{\mathrm{FD\&IID}}(G) \geq \mu_h(G \mid p_d)$, which always holds whenever the default model is also induced by an IID model; otherwise,

opting out may be optimal for the sender.

Corollary 2. *Given h, p_d, μ_0, suppose $\Pr(h \mid p_d) \leq \exp(-N\mathcal{H}(\hat{\pi}_h))$. Then the sender is strictly better off not proposing a model if and only if*

$$\exp(-N\mathcal{H}(\hat{\pi}_h)) < p_d(h \mid G). \tag{1}$$

Without a proposed model, the receiver's default posterior belief about state G is

$$\mu_h(G \mid p_d) = \frac{\mu_0(G)p_d(h \mid G)}{\Pr(h \mid p_d)} = \mu^{\text{FD}}(G)p_d(h \mid G). \tag{2}$$

By Proposition 2, condition (2) implies Corollary 2. Since the sender's optimal IID model induces $p^*_\pi(h \mid G) = \exp(-N\mathcal{H}(\hat{\pi}_h))$, condition (1) implies that the receiver's default likelihood of the history given state G exceeds the likelihood under the sender's optimal IID model. In other words, the cost of consistency outweighs the value of persuasion. This occurs when the receiver's initial understanding is not induced by any IID model, yet he demands an IID model. That is, he holds an inconsistent view of the history but insists on consistency from the sender. In such cases, the sender is better off allowing the receiver to maintain his default interpretation. The assumption $\Pr(h \mid p_d) \leq \exp(-N\mathcal{H}(\hat{\pi}_h))$ ensures that the FD constraint is nonempty: if $\Pr(h \mid p_d) > \exp(-N\mathcal{H}(\hat{\pi}_h))$, then for any model p_π induced by an IID model π, the h-fit is strictly worse than that of the default.

Example 2. *Consider a fund manager (sender) and a client (receiver) with prior belief $\mu_0(G) = 10\%$, where G represents a good time to invest in an asset. Let the signal space be $S = \{Up, Down\}$, where Up and Down denote periods of positive and negative asset returns, respectively. Suppose the observed history is*

$$h = (Down, Up, Up)$$

representing an initial decline followed by two consecutive increases.

A common finding in behavioral finance is that non-expert investors exhibit memory-based recency bias, over-weighting recent trends (see Malmendier and Nagel, 2011; Jiang et al., 2025). Suppose this bias is reflected in the client's default model p_d: he overemphasizes the two recent consecutive upsides and interprets them as a good sign for the asset. Specifically, $p_d(h \mid G) = 55\%$ and $p_d(h \mid B) = 10\%$, giving an overall fit of $\Pr(h \mid p_d) = 14.5\%$ with prior $\mu_0(G) = 10\%$. Despite the favorable recent trend, the client's default posterior remains low: $\mu_h(G \mid p_d) = \frac{10\% \times 55\%}{14.5\%} \approx 38\%$.

The manager seeks to persuade the client that the state is G. Under the FD constraint

alone, the manager proposes that the two consecutive positive returns are almost conclusive evidence for a high-quality asset. Formally, she proposes $\bar{p}(h \mid G) = 100\%$ *and* $\bar{p}(h \mid B) = 5\%$*, which binds the FD constraint:* $10\% \cdot \bar{p}(h \mid G) + 90\% \cdot \bar{p}(h \mid B) = 14.5\% = \Pr(h \mid p_d)$*. The manager raises the posterior to* $\mu^{\text{FD}}(G) = \frac{10\% \times 100\%}{14.5\%} \approx 69\%$*.*

Now suppose the manager also faces the IID constraint: the client demands a consistent principle governing how each positive or negative return relates to the quality of the asset. The manager proposes $\pi^*(Up \mid G) = 2/3$ *and* $\pi^*(Down \mid G) = 1/3$*, giving* $p^*_\pi(h \mid G) = (1/3)^1(2/3)^2 = 4/27 \approx 14.8\% = \exp(-N\mathcal{H}(\hat{\pi}_h))$*. Since* $14.8\% < 55\% = p_d(h \mid G)$*, proposing this model would reduce the client's posterior to* $\mu^{\text{FD\&IID}}(G) = \frac{10\% \times 14.8\%}{14.5\%} \approx 10\%$*, far below the default posterior of* 38%*. The manager therefore chooses not to propose any model, leaving the client to rely on his default interpretation.*

4. Consistency and Strategic Disclosure

The previous section characterized the scope of persuasion when the history of signals is public. This section asks what happens when the sender privately observes the history and can strategically choose which signals to disclose. Figure 2 summarizes the timeline.

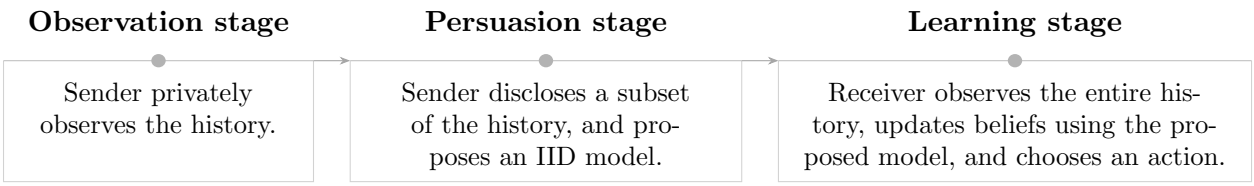


Figure 2: Timeline of disclosure and persuasion

In the observation stage, the sender privately observes the full history $h = (s_1, \ldots, s_N)$. In the persuasion stage, she discloses a *sub-history* $h_I = (s_i)_{i \in I}$, where $I \subseteq \{1, \ldots, N\}$, and proposes a model. Finally, in the learning stage, the receiver observes the full history h, updates his beliefs using the adopted model, and takes an action. The sender's problem is therefore to select a subset of observations that makes the proposed model compelling, while anticipating that the same model will be used to interpret the full history.

The IID constraint is unchanged: the sender must propose an IID model π alongside h_I. The FD constraint is revised as follows: the receiver adopts π if it fits the disclosed sub-history (rather than the full history) at least as well as his default model. As before, I assume the default model is induced by an IID model π_d. Formally, for each ω, the likelihood of the sub-history h_I conditional on state ω is $p_\pi(h_I \mid \omega) := \prod_{s \in S} \pi(s \mid \omega)^{h_I(s)}$ where $h_I(s)$ denotes the number of times signal s appears in the sub-history h_I. Define the h_I-fit of IID

model π by

$$\Pr(h_I \mid \pi) \coloneqq \sum_{\omega \in \Omega} \mu_0(\omega) p_\pi(h_I \mid \omega).$$

Then the FD constraint is $\Pr(h_I \mid \pi) \geq \Pr(h_I \mid \pi_d)$. After observing the entire history, the receiver's posterior belief induced by the adopted model π is $\mu_h(\omega \mid \pi) = \frac{\mu_0(\omega) p_\pi(h|\omega)}{\Pr(h|\pi)}$ for each ω.

Given history h and default model π_d, suppose the sender discloses a sub-history h_I. Then the sender solves

$$\pi^*(h_I) \in \arg \max_{\pi \in (\Delta S)^{|\Omega|}} \mu_h(G \mid \pi) \text{ subject to } \Pr(h_I \mid \pi) \geq \Pr(h_I \mid \pi_d).$$

Let $\mu_h(G \mid \pi^*(h_I))$ denote the posterior belief induced by the model $\pi^*(h_I)$. The sender's optimal disclosure is then

$$I^*(h, \pi_d) \in \arg \max_{I \subseteq \{1,\ldots,N\}} \mu_h(G \mid \pi^*(h_I)).$$

Three features of this environment warrant discussion. First, the key departure from the earlier section is that while the sender's objective is to maximize the posterior given the entire history, persuasion is achieved by explaining only the sub-history h_I. This separates the moment of persuasion from the moment of the receiver's learning. Second, I impose intertemporal consistency: the sender proposes a model only once and cannot revise it after the full history is observed. In practice, experts who repeatedly revise their stated view of the world, or insist that "this time is different" whenever new data arrives, often lose credibility. Third, at the learning stage the receiver does not revisit his default model and does not compare the adopted model against his original interpretation. Once he accepts the sender's proposed model, he applies it to the full history without questioning its validity.

4.1. Optimal disclosure given heterogeneous history

I say the history $h = (s_1, \ldots, s_N)$ is heterogeneous if it contains two or more different signals: i.e., $s_i \neq s_j$ for some $i, j \in \{1, \ldots, N\}$. Whenever the history is heterogeneous, the sender can always achieve full persuasion, regardless of the receiver's default model. This result follows from a simple but powerful disclosure strategy: the sender can withhold at least one signal type from her disclosure and propose a model under which that withheld signal type is conclusive evidence for state G.

Proposition 3. *If history h is heterogeneous, then $\mu_h(G \mid \pi^*(h_I^*)) = 100\%$, i.e., the sender achieves maximal persuasion.*

Proof. See Appendix A. □

It is sufficient to illustrate the point by assuming that the history contains two signal types $s, s' \in S$: that is, $h = (s, \ldots, s, s', \ldots, s')$. Suppose the sender discloses only the observations of signal s and proposes a model as follows:

$$\pi(s \mid G) = 99\%, \qquad \pi(s' \mid G) = 1\%, \qquad \pi(s \mid B) = 100\%, \qquad \pi(s' \mid B) = 0\%.$$

In words, the sender argues that "s always happens regardless of states, but there is a small chance that s' could occur in the good state." Under this model, s is almost entirely uninformative about the state, while s' is conclusive evidence for G. The receiver finds the model compelling because it fits the disclosed sub-history—consisting entirely of signal s—nearly perfectly. Once the receiver adopts this interpretation, he observes the full history, which also contains s'. Since the adopted model assigns zero probability to s' in state B, the full history is impossible under state B, and the receiver becomes fully convinced that the state is G.

As long as there are two or more types of signals in the history, the sender can achieve maximal persuasion similarly as above: she would disclose any sub-history that excludes at least one signal type. By doing so, she gains full flexibility in interpreting the undisclosed signals when they are eventually revealed. Furthermore, even when the default model is not induced by an IID model, maximal persuasion is always feasible as long as the default h-fit is not 100% (see Appendix B).

This extreme persuasive power ultimately stems from the receiver's naivete: once he adopts the proposed model, he does not revisit its validity after observing the full history. This naivete, combined with the sender's private information about the history, renders the FD and IID constraints largely irrelevant. The sender can therefore achieve full persuasion against any default model the receiver holds, suggesting that the combination of private observation of the history and selective disclosure can substantially weaken the disciplining role of consistency, and may overstate the extent of persuasion achievable in practice.

Example 3. *Consider again the vaccine example. Suppose the vaccine requires two doses, and after receiving the first dose, the receiver decides whether to complete vaccination by taking the second. Let s denote feeling well and s' denote symptoms such as fatigue or fever. Suppose the firm knows that the complete history after the first dose will be $h = (s, s', s)$: the receiver initially feels well, subsequently develops symptoms, and then finally recovers. Suppose, however, that under the receiver's default interpretation, s' is bad news: it suggests that the vaccine may be harmful, and thus after observing s', the receiver would be inclined to refuse the second dose. Before he experiences s', when his observed sub-history $h_I = (s)$,*

the firm can propose the following IID model:

$$\pi(s \mid G) = 99\%, \qquad \pi(s' \mid G) = 1\%, \qquad \pi(s \mid B) = 100\%, \qquad \pi(s' \mid B) = 0\%.$$

In words, feeling well is essentially uninformative, since it occurs in both states, while symptoms are decisive evidence for G. Once the receiver adopts this model, the later realization of s' leads him to interpret the symptoms as evidence that the vaccine is working, and he returns for the second dose.[4]

4.2. Optimal disclosure given homogeneous history

I say the history $h = (s_1, \ldots, s_N)$ is homogeneous if it is not heterogeneous: i.e., there is a signal $s \in S$ such that $s_i = s$ for all $i \in \{1, \ldots, N\}$. Given this history of length $N \geq 2$, the sender discloses $k \in \{1, \ldots, N\}$ signals, and proposes a model p_π induced by an IID model π in the persuasion stage.

4.2.1. Setup

I let $p_\pi(k \mid \omega)$ denote the likelihood of k signals conditional on state ω under IID model π: i.e., $p_\pi(k \mid \omega) \coloneqq \pi(s \mid \omega)^k$ and define the k-fit of model π by $\Pr(k \mid \pi) \coloneqq \sum_{\omega \in \Omega} \mu_0(\omega)\pi(s \mid \omega)^k$ for each $k \in \{1, \ldots, N\}$. The FD constraint conditional on disclosing k signals is therefore $\Pr(k \mid \pi) \geq \Pr(k \mid \pi_d)$. If the receiver adopts π, his posterior belief about state G given the entire history h is

$$\mu_h(G \mid p_\pi) = \frac{\mu_0(G)p_\pi(N \mid G)}{\mu_0(G)p_\pi(N \mid G) + \mu_0(B)p_\pi(N \mid B)}.$$

Hence, the sender's persuasion value conditional on disclosing k signals is

$$U(k) \coloneqq \max_{\pi \in (\Delta S)^{|\Omega|}} \mu_h(G \mid p_\pi) \quad \text{subject to} \quad \Pr(k \mid \pi) \geq \Pr(k \mid \pi_d). \tag{3}$$

Then the sender's optimal disclosure is

$$k^*(h, \pi_d) \in \arg\max_{k \in \{1, \ldots, N\}} U(k).$$

[4] This example points to the role of competition. An anti-vaccine sender could propose the opposite interpretation: feeling well is uninformative, but symptoms are decisive evidence for B. This competing model can fit the same disclosed history equally well and also satisfy the IID constraint. Thus, when opposed senders can offer equally coherent interpretations of the same data, persuasion may be neutralized. Schwartzstein and Sunderam (2021) show that when two senders with opposing objectives compete, the equilibrium belief collapses to the prior: competition neutralizes the data entirely.

I assume N is small enough so that $\Pr(N \mid \pi_d) > \mu_0(G)$. Otherwise, the sender is always better off disclosing the entire history. This is because the default model's N-fit converges to zero as N increases, which implies that the FD constraint effectively disappears at $k = N$ for large N. This allows the sender to disclose the entire history and argue that s is conclusive evidence for state G. However, when the size of the history is small enough—i.e., the default N-fit is larger than the prior $\mu_0(G)$—the sender can no longer present such an argument. Since the k-fit of the default model can only decrease in k, the assumption also implies $\Pr(k \mid \pi_d) > \mu_0(G)$ for all $k \in \{1, \ldots, N\}$, and thus finding the optimal disclosure $k^*(h, \pi_d)$ is no longer straightforward.[5]

4.2.2. Trade-off between disclosing more vs. less signals

Given a homogeneous history, the sender faces a trade-off. Disclosing more observations makes it easier to outperform the receiver's default model: a longer sequence of identical signals is increasingly unlikely under any given model initially held by the receiver, and thus the FD constraint becomes less demanding. This counts as the marginal benefit of disclosing more signals. On the other hand, the IID constraint becomes more demanding: explaining a longer sequence with an IID model requires assigning higher per-signal likelihood in unfavorable states. This is the marginal cost. I show that this trade-off can be characterized tractably, and that the optimal disclosure is always at a corner: she either discloses a single observation or the entire history.

I first reduce problem (3) to a simpler form. Unlike the case of the heterogeneous history, the entropy of the empirical distribution of any homogeneous history is zero, and thus there is no cost of consistency characterized in Proposition 2: i.e., $\exp(-N \cdot \mathcal{H}(\hat{\pi}_h)) = 1$ for any homogeneous h. This means that given any disclosure k, the sender can argue that s happens in state G with probability 1. Therefore, the optimal disclosure $k^*(h, \pi_d)$ must be the one that allows the sender to propose the least possible likelihood of s in state B.

Lemma 2. *Suppose history $h = (s, \ldots, s)$ of $N \geq 2$ signals is homogeneous and $\Pr(h \mid \pi_d) > \mu_0(G)$. Let π_k be the optimal IID model conditional on disclosing k signals. Then the optimal disclosure $k^*(h, \pi_d)$ minimizes $\pi_k(s \mid B)$.*

Proof. Suppose the sender discloses k occurrences of signal s. Analogous to the proofs of Propositions 1-2, it is straightforward to show that she proposes an IID model π_k such that

[5] Notice that assuming $\Pr(N \mid \pi_d) > \mu_0(G)$ also rules out $\pi_d(s \mid B) = 0$. If $\pi_d(s \mid B) = 0$, then the receiver initially believes that s can never happen in the bad state. Once the receiver sees the signal s, regardless of how many of them are disclosed, his posterior belief about state G becomes 100%, without the sender's intervention.

the likelihoods of k occurrences for each state are $p_{\pi_k}(k \mid G) = 1$ and

$$p_{\pi_k}(k \mid B) = \frac{\Pr(k \mid \pi_d) - \mu_0(G)}{1 - \mu_0(G)}. \tag{4}$$

Note that since I assumed $\Pr(k \mid \pi_d) > \mu_0(G)$, the sender cannot argue that $p_{\pi_k}(k \mid B) = 0$. By proposing this model, the IID constraint implies that the per-signal likelihoods for each state must satisfy $\pi_k(s \mid G) = 1$ and

$$p_{\pi_k}(k \mid B) \equiv \pi_k(s \mid B)^k. \tag{5}$$

It follows that the likelihood of N occurrences in state B is $p_{\pi_k}(k \mid B)^{\frac{N}{k}}$. Then the sender achieves the posterior belief given N signals as follows

$$U(k) = \frac{\mu_0(G)}{\mu_0(G) + (1 - \mu_0(G))p_{\pi_k}(k \mid B)^{\frac{N}{k}}} = \frac{\mu_0(G)}{\mu_0(G) + (1 - \mu_0(G))\pi_k(s \mid B)^N}.$$

Therefore, the optimal $k^*(h, \pi_d)$ is the one that minimizes $\pi_k(s \mid B)$. More precisely,

$$k^*(h, \pi_d) \text{ uniquely solves } \min_{k \in \{1,\dots,N\}} \left(\frac{\Pr(k \mid \pi_d) - \mu_0(G)}{1 - \mu_0(G)} \right)^{\frac{1}{k}}.$$

This proves Lemma 2. □

4.2.3. How many signals does the sender disclose?

Proposition 4. *Suppose history $h = (s, \dots, s)$ of $N \geq 2$ signals is homogeneous and $\Pr(h \mid \pi_d) > \mu_0(G)$. Then $k^*(h, \pi_d) \in \{1, N\}$, i.e., the optimal disclosure is either the entire history or a single observation of s.*

Proof. See Appendix C. □

I sketch the proof here. The key trade-off is captured by identity (5), which is the combination of the FD and the IID constraints. The marginal benefit (MB) of increasing the number k of disclosed signals comes from the FD constraint. Since the receiver's default model is also induced by an IID model, the default model's k-fit can only worsen with more observations: i.e., $\Pr(k \mid \pi_d) > \Pr(k + 1 \mid \pi_d)$ for any $k < N$. It follows from (4) that the sender can propose a lower $p_{\pi_k}(k \mid B)$—the L.H.S. of identity (5). On the other hand, the marginal cost (MC) stems from the IID constraint, which requires that the same per-signal likelihood explain k observations. Therefore, increasing k forces the sender to spread the total likelihood over more observations, which pushes up $\pi_k(s \mid B)$ on the R.H.S. of identity (5).

To analyze the MB and MC, consider the log-likelihood of s

$$\log \pi_k(s \mid B) = \frac{1}{k} \cdot \log p_{\pi_k}(k \mid B)$$

where the equality comes from (5). Assume that the function $k \mapsto \log \pi_k(s \mid B)$ is defined on the positive real line. Then taking the derivative yields

$$\frac{d}{dk} \log \pi_k(s \mid B) = -\frac{1}{k} \left(\underbrace{\left| \frac{d}{dk} \log p_{\pi_k}(k \mid B) \right|}_{\text{MB of increasing } k} - \underbrace{\left| \frac{1}{k} \cdot \log p_{\pi_k}(k \mid B) \right|}_{\text{MC of increasing } k} \right).$$

The term $\left|\frac{d}{dk} \log p_{\pi_k}(k \mid B)\right|$ captures the MB of disclosing more signals, while $\left|\frac{1}{k} \cdot \log p_{\pi_k}(k \mid B)\right|$ captures the MC. Note that both terms $\frac{d}{dk} \log p_{\pi_k}(k \mid B)$ and $\frac{1}{k} \cdot \log p_{\pi_k}(k \mid B)$ are strictly negative since $\Pr(k \mid \pi_d)$ is decreasing and $p_{\pi_k}(k \mid B) \in (0, 1)$. Then the sender's objective is to maximize the MC over $k \in \{1, \ldots, N\}$ since the MC is equal to $-\log \pi_k(s \mid B)$ where $\log \pi_k(s \mid B)$ is a logarithmic transformation of the minimization objective function.[6]

The optimal disclosure problem is governed by a standard marginal–average relationship between the MB and the MC. The MB is the marginal rate at which the log-likelihood of k occurrences declines, while the MC is the average log-likelihood per observation; indeed,

$$MC(k) = \frac{1}{k} \int_0^k MB(x)\, dx.$$

The marginal pulls the average: the MC decreases where the MB is below the MC and increases where the MB is above the MC. Lemma 4 in the appendix shows that the MB and the MC coincide at $k = 0$ and both diverge to infinity as $k \to k_0$, where k_0 is the disclosure level at which $p_{\pi_k}(k \mid B) = 0$, i.e. the point at which the sender can argue that signal s is conclusive of state G; these frame the comparison on $(0, k_0)$. Assuming $\Pr(h \mid \pi_d) > \mu_0(G)$ ensures that $k_0 > N$ so that the MB and MC do not diverge in our domain $[1, N]$.

The optimum then depends on the monotonicity of the MB. Lemma 3 in the appendix shows that when $\mu_0(G) \geq 0.5$ or $\pi_d(s \mid G) \geq \pi_d(s \mid B)$, the MB is strictly increasing; since the MB and MC coincide at $k = 0$, the MC stays below the rising MB and therefore increases throughout, and thus disclosing the entire history is optimal, i.e., $k^*(h, \pi_d) = N$.

When $\mu_0(G) < 0.5$ and $\pi_d(s \mid G) < \pi_d(s \mid B)$, I show that the MB has at most one turning point, and because it diverges at k_0, that turning point can only be a minimum; the MB is thus either increasing throughout or declining before rising. In either shape, Lemma 5

[6] Hence, the standard equality MB = MC is not the optimality condition here.

in the appendix confirms that the MB crosses the MC at most once, so the MC changes direction at most once. Combined with the divergence of the MC at k_0, this implies that the MC is either increasing throughout or declining before rising. In both cases the MC's maximum over $\{1, \ldots, N\}$ falls at $k = 1$ or $k = N$, giving $k^*(h, \pi_d) \in \{1, N\}$.

4.2.4. When is disclosing one signal optimal?

The next question is: When is disclosing one signal better than disclosing the entire history? The answer is twofold. First, the receiver's prior must be unfavorable to the sender, placing more weight on the bad state. Second, the signal s must be strong evidence for state B under the default model.

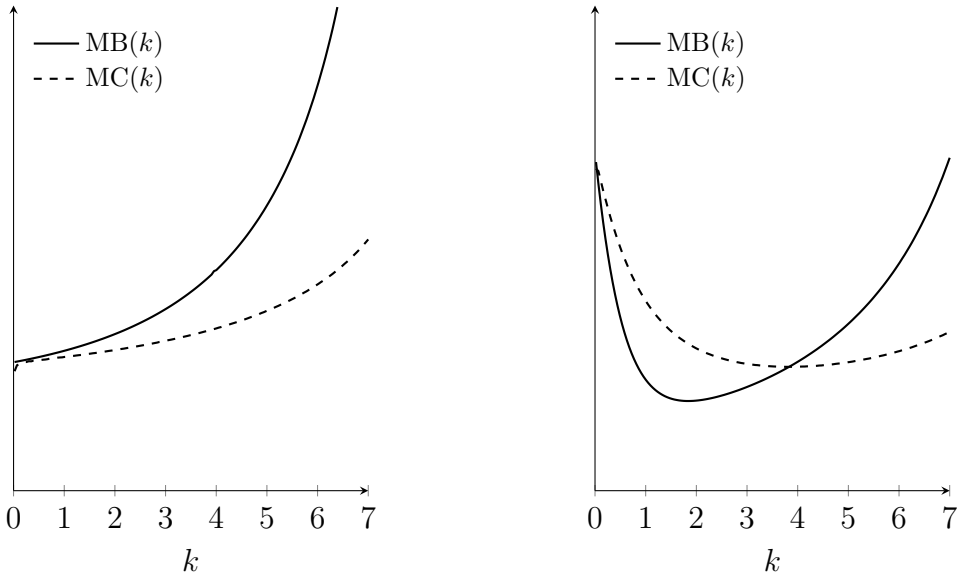
Proposition 5. *Suppose history $h = (s, \ldots, s)$ of $N \geq 2$ signals is homogeneous and $\Pr(h \mid \pi_d) > \mu_0(G)$. If $\mu_0(G) \geq 50\%$, then the optimal disclosure is $k^*(h, \pi_d) = N$. If $\mu_0(G) < 50\%$, there exists a threshold $\delta^* > 0$ such that $k^*(h, \pi_d) = 1$ whenever $\pi_d(s \mid B) - \pi_d(s \mid G) \geq \delta^*$.*

Proof. See Appendix D. □

When $\mu_0(G) \geq 50\%$, the prior belief is already favorable to the sender. This corresponds to the case where the MB and MC never cross. Intuitively, the favorable prior relaxes the FD constraint. As a result, the sender can assign an extremely small likelihood of k signals in state B, and changes in k does not significantly affect that magnitude. In particular, $p_{\pi_k}(k \mid B) = \frac{\Pr(k|\pi_d) - \mu_0(G)}{1 - \mu_0(G)}$ approaches zero as $\mu_0(G) \to \Pr(k \mid \pi_d)$. Therefore, when $\mu_0(G)$ is above 50%, the MB becomes arbitrarily large for any disclosure k. This implies that the MC is strictly increasing in k. Hence the MB dominates throughout and disclosing the entire history is optimal.

When $\mu_0(G) < 50\%$, the MB may initially decrease depending on how pessimistically the receiver initially interprets signal s. Specifically, I say the receiver is initially more pessimistic about s if s is sufficiently more indicative of the bad state than of the good state under the default model. Formally, this is characterized by the distance $\pi_d(s \mid B) - \pi_d(s \mid G)$. Once this gap exceeds a threshold, the optimal disclosure switches from the full history to a single observation. Intuitively, if the receiver is initially pessimistic about s, then observing more realizations of s pushes her posterior further toward state B. This strengthens her initial view and makes it harder for the sender to persuade her toward state G. For this reason, when the sender wants to persuade the receiver to adopt the alternative view that s is evidence of state G, she prefers to disclose as little data as possible.

Figure 3 plots the MB and MC, fixing $\mu_0(G) = 10\% < 50\%$ across two panels and varying only the default model π_d. Panel (a) sets $\pi_d(s \mid G) = 85\% \geq \pi_d(s \mid B) = 75\%$: the receiver's



(a) $\pi_d(s \mid G) = 85\%,\ \pi_d(s \mid B) = 75\%$ (b) $\pi_d(s \mid G) = 10\%,\ \pi_d(s \mid B) = 80\%$

Figure 3: Marginal benefit (MB) and marginal cost (MC) of disclosing k signals, with prior belief $\mu_0(G) = 0.1$ and history length $N = 7$. Panels (a) and (b) differ in the receiver's default model π_d.

default model treats s as a good sign, since it is more likely in state G than in state B. The MB is increasing throughout and dominates the MC at every $k \in \{1, \ldots, N\}$; the sender optimally discloses the entire history, yielding $U(1) \approx 49\%$ versus $U(7) \approx 66\%$. Panel (b) sets $\pi_d(s \mid G) = 10\% < \pi_d(s \mid B) = 80\%$: the receiver's default model treats s as a bad sign, since it is far more likely in state B than in state G, and observing more of it would only reinforce this view. The MB initially declines before rising and crosses the MC exactly once; among $k \in \{1, \ldots, N\}$, the MC attains its maximum at $k = 1$, and the sender optimally discloses a single signal, yielding $U(1) \approx 57\%$ versus $U(7) \approx 53\%$.

An interesting implication of Proposition 5 is that, in the disclosure problem, the internal structure of the default model matters, not just its fit. This differs from the benchmark framework (Schwartzstein and Sunderam, 2021), where the sender's persuasion problem is driven only by how well the default model fits the realized history. In the disclosure problem given a homogeneous history, the receiver's detailed initial interpretation of the signal also becomes important. Given a common prior belief leaning toward state B, suppose two receivers' default model have the same h-fit, but they differ in how they interpret the signal: one interprets it more pessimistically than the other. Proposition 5 implies that the sender's optimal disclosure may differ across them. For the pessimistic receiver, the sender optimally discloses a single signal; for the other, she optimally discloses the entire history.

Corollary 3. *Suppose $\mu_0(G) < 50\%$ and the history is homogeneous with two observations, $h = (s, s)$. Then for each $\pi_d(s \mid G) \in [0, 1)$, the threshold δ^* is*

$$\delta^* = \frac{1 - \pi_d(s \mid G)}{2\mu_0(B)}. \tag{6}$$

Proof. See Appendix E. □

Corollary 3 shows that when the homogeneous history contains two observations, the threshold δ^* in Proposition 5 is in closed form. Equivalently, the sender optimally discloses a single signal if and only if

$$\mu_{(s)}(G \mid \pi_d) < \mu_0(G) \left(\frac{2 \Pr(1 \mid \pi_d) - 1}{\Pr(1 \mid \pi_d)} \right), \tag{7}$$

where $\mu_{(s)}(G \mid \pi_d) = \frac{\mu_0(G)\pi_d(s|G)}{\Pr(1|\pi_d)}$ denotes the posterior belief about state G under the default model after observing a single signal s.[7] The right-hand side depends only on the prior and the default model's fit of a single signal. Holding this single-signal fit fixed, the left-hand side is smaller when s is more indicative of state B under the default model, that is, when the receiver is more pessimistic about s. Two receivers with the same single-signal fit but different pessimism can therefore call for different disclosure: the sender optimally discloses a single signal to the more pessimistic one and the entire history to the other. The corollary therefore illustrates in a simple setting why a pessimistic receiver is better persuaded with a single observation: disclosing the full homogeneous history would push her posterior further toward state B, making persuasion harder.

Example 4. *Consider academic presentations. Suppose a researcher wrote a paper arguing that the presence of a certain local amenity lowers nearby housing prices. She applies the same empirical strategy across N different settings, such as case studies from different regions, and in each case obtains the same qualitative finding: the estimated effect is negative and supports her hypothesis. In terms of the framework here, she privately observes a homogeneous history $h = (s, \ldots, s)$, where each realization s represents the same kind of finding, namely a result supportive of her claim.*

The timing matches the framework closely. First, before the audience sees the full set of findings, the researcher chooses what to disclose in the presentation. She may present only one of the supportive cases or all N cases. Second, based on this initial disclosure, the audience forms an interpretation of what the finding implies about the underlying hypothesis.

[7] Inequality (7) rearranges $\pi_1(s \mid B) < \pi_2(s \mid B)$. By Lemma 2, the sender optimally discloses whichever minimizes $\pi_k(s \mid B)$.

Third, the remaining findings are eventually revealed, for example when the audience reads the full paper, and are interpreted through the lens established by the initial presentation. Finally, the audience makes its decision, such as whether a hiring committee views the paper and the candidate favorably.

Proposition 5 suggests that when the audience is initially skeptical, presenting many similar supportive findings at the outset can be counterproductive: additional examples can invite doubt and weaken the force of a clean argument. In that case, the researcher is better off presenting only one case initially, so as to establish a more favorable interpretation before the rest of the evidence is seen. When the audience is not sufficiently skeptical, it is optimal to present as many cases as possible initially.

Appendix

A. Proof of Proposition 3

If the history is impossible under the default model (i.e., $\Pr(h \mid \pi_d) = 0$), then the sender achieves maximal persuasion by disclosing the entire history. Now, consider $\Pr(h \mid \pi_d) > 0$. Let $S_h \subseteq S$ denote the set of distinct signals appearing in h. The sender can achieve full persuasion by disclosing any sub-history h_I that excludes at least one signal type $s^* \in S_h$ (i.e., $h_I(s^*) = 0$ while $h(s^*) > 0$). Note that $\Pr(h \mid \pi_d) > 0$ implies that every such sub-history satisfies

$$\Pr(h_I \mid \pi_d) < \exp\big(-|h_I| \cdot \mathcal{H}(\hat{\pi}_{h_I})\big) \tag{8}$$

where $|h_I|$ denotes the total number of signals in h_I and $\hat{\pi}_{h_I}$ is the empirical distribution of signals in h_I (i.e., for each s in h_I, $\hat{\pi}_{h_I}(s) = \frac{h_I(s)}{|h_I|}$). By Lemma 1, $\exp(-|h_I| \cdot \mathcal{H}(\hat{\pi}_{h_I}))$ is the maximum h_I-fit achievable by any IID model. However, the default model's h_I-fit can never be at its maximum. To show this, note that by Corollary 1, the equality $\Pr(h_I \mid \pi_d) = \exp(-|h_I| \cdot \mathcal{H}(\hat{\pi}_{h_I}))$ holds if and only if $\pi_d(\cdot \mid \omega) = \hat{\pi}_{h_I}$ for all $\omega \in \Omega$. Since h_I excludes s^*, the equality would require $\pi_d(s^* \mid \omega) = 0$ for all ω. But since h contains s^* and $\Pr(h \mid \pi_d) > 0$, the default model assigns strictly positive probability to s^* in at least one state ω, and thus $\pi_d(\cdot \mid \omega) \neq \hat{\pi}_{h_I}$ for at least one ω. Hence, condition (8) holds for any sub-history excluding s^*.

Given condition (8), construct for small $\varepsilon > 0$ the IID model π_ε as follows:

$$\begin{aligned} &\pi_\varepsilon(s^* \mid G) = \varepsilon, \qquad \pi_\varepsilon(s \mid G) = (1-\varepsilon)\hat{\pi}_{h_I}(s) \quad \text{for all } s \in S_{h_I}, \\ &\pi_\varepsilon(s^* \mid B) = 0, \qquad \pi_\varepsilon(s \mid B) = \hat{\pi}_{h_I}(s) \quad \text{for } s \in S_{h_I}. \end{aligned}$$

where $S_{h_I} \subseteq S$ denotes the set of distinct signals appearing in h_I. By construction, this model almost attains the maximal h_I-fit: $\Pr(h_I \mid \pi_\varepsilon)$ approaches $\exp(-|h_I| \cdot \mathcal{H}(\hat{\pi}_{h_I}))$ as $\varepsilon \to 0$. Condition (8) ensures that the sender can choose ε small enough so that the FD constraint is satisfied. Since $h(s^*) > 0$ and $\pi_\varepsilon(s^* \mid B) = 0$, the likelihood of h under model p_π in state B becomes zero: $p_{\pi_\varepsilon}(h \mid B) = 0$. On the other hand, h happens with positive probability in state G: $p_{\pi_\varepsilon}(h \mid G) > 0$. Consequently, the receiver's posterior belief about state G using the model p_{π_ε} becomes

$$\mu_h(G \mid p_{\pi_\varepsilon}) = \frac{\mu_0(G)p_{\pi_\varepsilon}(h \mid G)}{\Pr(h \mid \pi_\varepsilon)} = \frac{\mu_0(G)p_{\pi_\varepsilon}(h \mid G)}{\mu_0(G)p_{\pi_\varepsilon}(h \mid G) + (1-\mu_0(G))p_{\pi_\varepsilon}(h \mid B)} = 100\%.$$

This completes the proof.

B. Disclosure given heterogeneous history against arbitrary default model

Proposition 6. *Consider the sender's disclosure problem given a heterogeneous history h. Against any arbitrary default model p_d with $\Pr(h \mid p_d) < 1$, the sender achieves maximal persuasion, i.e., $\mu_h(G \mid \pi^*(h_I^*)) = 100\%$.*

Proof. Let $S_h \subseteq S$ be the set of signals that the history h contains. Define the set $I(s) \subseteq \{1, \dots, N\}$ by $s_i = s$ for all $i \in I(s)$. That is, $I(s)$ collects all realizations in h that are $s \in S$. Since the h-fit of the default model is not 100%, there are two or more possible histories under the default model. Then the sender can choose a sub-history $h_{I(s)}$ such that $\Pr(h_{I(s)} \mid p_d) < 1$. Consider an IID model π satisfying

$$\pi(s \mid G) = 1 - \varepsilon \qquad \pi(s' \mid G) = \frac{\varepsilon}{|S_h \setminus \{s\}|} \quad \text{for all } s' \in S_h \setminus \{s\}$$
$$\pi(s \mid B) = 1$$

for some small $\varepsilon > 0$. Let p_π be the model induced by this π. Then p_π satisfies the FD constraint since it almost perfectly fits h_I: i.e., $\Pr(h_I \mid p_\pi) \approx 1$. Since h is heterogeneous, $S_h \setminus \{s\}$ is nonempty. Hence, the likelihood of h under model p_π in state B becomes zero: i.e., $p_\pi(h \mid B) = \pi(s \mid B)^{h(s)} \left(\prod_{s' \in S_h \setminus \{s\}} \pi(s' \mid B)^{h(s')} \right) = 0$. On the other hand, for state G, h happens with positive probability: $p_\pi(h \mid G) = \pi(s \mid G)^{h(s)} \left(\prod_{s' \in S_h \setminus \{s\}} \pi(s' \mid G)^{h(s)} \right) > 0$. Consequently, the receiver's posterior belief about state G using the model p_π becomes

$$\mu_h(G \mid p_\pi) = \frac{\mu_0(G)p_\pi(h \mid G)}{\Pr(h \mid p_\pi)} = \frac{\mu_0(G)p_\pi(h \mid G)}{\mu_0(G)p_\pi(h \mid G) + (1-\mu_0(G))p_\pi(h \mid B)} = 1.$$

which completes the proof. □

C. Proof of Proposition 4

Let $MB(k) := \left|\frac{d}{dk}\log p_{\pi_k}(k \mid B)\right|$ and $MC(k) := \left|\frac{1}{k}\cdot \log p_{\pi_k}(k \mid B)\right|$. Let $k_0 \in (N, \infty)$ be such that $p_{\pi_{k_0}}(k_0 \mid B) = 0$. Note that $\log p_{\pi_k}(k \mid B) < 0$ for $k \in (0, k_0)$ and zero at $k = 0$ since $p_{\pi_0}(0 \mid B) = 1$. Moreover, $\log p_{\pi_k}(k \mid B)$ is decreasing in k, with $\log p_{\pi_k}(k \mid B) \to -\infty$ as $k \to k_0$, since $p_{\pi_{k_0}}(k_0 \mid B) = 0$. It follows that $MB(k) = -\frac{d}{dk}\log p_{\pi_k}(k \mid B)$ and $MC(k) = -\frac{1}{k}\log p_{\pi_k}(k \mid B)$ on $(0, k_0)$. For brevity, write $d_\omega = \pi_d(s \mid \omega)$ and $\lambda = \frac{\mu_0(G)}{1-\mu_0(G)}$ so that $p_{\pi_k}(k \mid B) = d_B^k + \lambda(d_G^k - 1)$.

I consider two cases: (i) either $\mu_0(G) \geq 0.5$ or $\pi_d(s \mid G) \geq \pi_d(s \mid B)$ and (ii) $\mu_0(G) < 0.5$ and $\pi_d(s \mid G) < \pi_d(s \mid B)$. For case (i), I use the following lemma to prove that $k^*(h, \pi_d) = N$.

Lemma 3. *$MB(k)$ is strictly increasing on $(0, k_0)$ if either $\mu_0(G) \geq 0.5$ or $\pi_d(s \mid G) \geq \pi_d(s \mid B)$.*

Proof. Since $MB(k) = -\frac{d}{dk}\log p_{\pi_k}(k \mid B)$ on $(0, k_0)$, it suffices to show $\frac{d^2}{dk^2}\log p_{\pi_k}(k \mid B) < 0$ under either condition. Directly computing the derivative of MB gives

$$MB'(k) = -\frac{d^2}{dk^2}\log p_{\pi_k}(k \mid B) = \frac{\lambda\left(-d_B^k d_G^k\left[\log\frac{d_B}{d_G}\right]^2 + d_B^k(\log d_B)^2 + \lambda d_G^k(\log d_G)^2\right)}{p_{\pi_k}(k \mid B)^2}. \quad (9)$$

It suffices to show the numerator is strictly positive. First suppose $d_G \geq d_B$. Then $d_B \leq \frac{d_B}{d_G} \leq 1$, hence $[\log\frac{d_B}{d_G}]^2 \leq (\log d_B)^2$, and since $d_G^k < 1$ we obtain $d_B^k d_G^k[\log\frac{d_B}{d_G}]^2 < d_B^k(\log d_B)^2$; as the remaining term $\lambda d_G^k(\log d_G)^2$ is positive, the numerator is strictly positive. Next suppose $\mu_0(G) \geq \frac{1}{2}$, i.e. $\lambda \geq 1$, and consider the case $d_B > d_G$ (the case $d_G \geq d_B$ has been covered above). Since $\log d_B < 0$ and $0 < \log d_B - \log d_G = \log\frac{d_B}{d_G} < -\log d_G$, we have $d_B^k d_G^k\left[\log\frac{d_B}{d_G}\right]^2 < \lambda d_G^k[\log d_G]^2$, which implies that the numerator is strictly positive. Hence, in either case, we have $MB'(k) > 0$ for $k \in (0, k_0)$. □

Since $MB(k) = -\frac{d}{dk}\log p_{\pi_k}(k \mid B)$ on $(0, k_0)$ and $\log p_{\pi_0}(0 \mid B) = 0$, we have

$$MC(k) = \frac{1}{k}\int_0^k MB(x)\,dx$$

for every $k \in (0, k_0)$; that is, $MC(k)$ is the average of MB over $[0, k]$. When $\mu_0(G) \geq 0.5$ or $\pi_d(s \mid G) \geq \pi_d(s \mid B)$, MB is strictly increasing on $(0, k_0)$ by Lemma 3, and a strictly increasing function exceeds its average over $[0, k]$, which gives $MB(k) > MC(k)$ for every $k \in (0, k_0)$. The identity $MC'(k) = \frac{1}{k}\big(MB(k) - MC(k)\big)$ then yields $MC'(k) > 0$ for every

$k \in (0, k_0)$, hence MC is strictly increasing on $(0, k_0)$, and its maximum over $\{1, \ldots, N\}$ is attained at $k^* = N$.

I now consider case (ii).

Lemma 4. *(a)* $MB(0) = \lim_{k\to 0^+} MC(k)$*, and (b)* $\lim_{k\to k_0} MB(k) = \lim_{k\to k_0} MC(k) = \infty$.

Proof. By the definition of the derivative at $k = 0$, we have

$$\lim_{k\to 0^+} MC(k) = \lim_{k\to 0^+} \frac{-\log p_{\pi_k}(k \mid B)}{k} = -\frac{d}{dk} \log p_{\pi_k}(k \mid B)\big|_{k=0} = MB(0)$$

which is finite and strictly positive since $p_{\pi_k}(k \mid B)$ is smooth and strictly positive on a neighborhood of 0. This proves (a). To prove (b), consider the two limits separately. First, we have $\lim_{k\to k_0} MC(k) = -\lim_{k\to k_0} \left(\frac{1}{k} \log p_{\pi_k}(k \mid B)\right) = \infty$ since $\log p_{\pi_k}(k \mid B) \to -\infty$ and $k_0 \in (N, \infty)$ is finite and positive. Next, for MB, write $MB(k) = -\frac{\frac{d}{dk} p_{\pi_k}(k|B)}{p_{\pi_k}(k|B)}$, whose numerator converges to $-\left(d_B^{k_0} \log d_B + \lambda d_G^{k_0} \log d_G\right) > 0$ while its denominator satisfies $p_{\pi_k}(k \mid B) \to 0^+$, hence the ratio diverges to ∞. □

Lemma 5. *If* $\mu_0(G) < 0.5$ *and* $\pi_d(s \mid B) > \pi_d(s \mid G)$*, then* $MB(k)$ *and* $MC(k)$ *intersect at most once on* $(0, k_0)$.

Proof. I first show that $MB(k)$ has at most one turning point on $(0, k_0)$. A turning point k of MB is such that $MB'(k) = 0$. Using equation (9), we have

$$MB'(k) \left(\frac{p_{\pi_k}(k \mid B)^2}{\lambda d_G^k}\right) = -d_B^k \left[\log \tfrac{d_B}{d_G}\right]^2 + \frac{d_B^k}{d_G^k} (\log d_B)^2 + \lambda (\log d_G)^2.$$

Since $\frac{p_{\pi_k}(k|B)^2}{\lambda d_G^k} > 0$ for all $k \in (0, k_0)$, the sign of $MB'(k)$ equals that of the R.H.S. Then

$$\frac{d}{dk}\left[MB'(k) \left(\frac{p_{\pi_k}(k \mid B)^2}{\lambda d_G^k}\right)\right] = -d_B^k (\log d_B) \left[\log \tfrac{d_B}{d_G}\right]^2 + \frac{d_B^k}{d_G^k} \left[\log \tfrac{d_B}{d_G}\right] (\log d_B)^2 > 0. \quad (10)$$

Since $d_B > d_G$ gives $\log \frac{d_B}{d_G} > 0$, both terms of (10) are positive. It follows that the function $k \mapsto MB'(k) \left(\frac{p_{\pi_k}(k|B)^2}{\lambda d_G^k}\right)$ is strictly increasing, which implies that $MB'(k)$ has at most one zero. The single turning point implies that MB either increases throughout or declines before rising. Then by Lemma 4, in the former case MB exceeds MC everywhere on $(0, k_0)$, and they never cross; in the latter, MB intersects MC at most once at MC's minimum because MC is the average of MB. □

Through $MC'(k) = \frac{1}{k}(MB(k) - MC(k))$, Lemma 5 implies that MC changes direction at most once. Lemma 4 leaves only two possibilities: MC increases throughout, or declines

before rising. In either case, MC attains its maximum over $\{1, \dots, N\}$ at either $k = 1$ or $k = N$. This completes the proof of Proposition 4.

D. Proof of Proposition 5

Consider a default model p_d induced by an IID model π_d such that $\pi_d(s \mid B) = \pi_d(s \mid G) + \delta$ for some $\delta \in (0, 1 - \pi_d(s \mid G)]$. Denote the k-fit of this model π_d by

$$\hat{D}(k; \delta) \coloneqq \Pr(k \mid \pi_d) = \mu_0(G)\pi_d(s \mid G)^k + (1 - \mu_0(G))(\pi_d(s \mid G) + \delta)^k.$$

By Proposition 4, we know that the optimal disclosure is either $k^*(h, \pi_d) = 1$ or $k^*(h, \pi_d) = N$. Then Lemma 2 implies that given δ, we have $k^*(h, \pi_d) = 1$ only if the sender's choice of models for $k = 1$ and $k = N$ satisfy $\pi_1(s \mid B)^N < \pi_N(s \mid B)$. In other words,

$$\left(\frac{\hat{D}(1; \delta) - \mu_0(G)}{1 - \mu_0(G)}\right)^N < \frac{\hat{D}(N; \delta) - \mu_0(G)}{1 - \mu_0(G)}.$$

For each $\delta \in [0, 1 - \pi_d(s \mid G)]$, define a function T by

$$T(\delta) \coloneqq \frac{\hat{D}(N; \delta) - \mu_0(G)}{1 - \mu_0(G)} - \left(\frac{\hat{D}(1; \delta) - \mu_0(G)}{1 - \mu_0(G)}\right)^N.$$

It is then sufficient to show that given any $\pi_d(s \mid G) < 1$, there exists $\delta^* \in (0, 1 - \pi_d(s \mid G))$ such that $T(\delta) > 0$ for all $\delta > \delta^*$.

Lemma 6. *(a) The function T is strictly increasing on $[0, 1 - \pi_d(s \mid G)]$, (b) $T(0) < 0$, and (c) $T(1 - \pi_d(s \mid G)) > 0$.*

Proof. I first rewrite the function T as

$$\begin{aligned} T(\delta) &= (\pi_d(s \mid G) + \delta)^N + \lambda\,\pi_d(s \mid G)^N - \lambda - (\pi_d(s \mid G) + \delta + \lambda\,\pi_d(s \mid G) - \lambda)^N \\ &= (\pi_d(s \mid G) + \delta)^N - \lambda\big(1 - \pi_d(s \mid G)^N\big) - \big(\pi_d(s \mid G) + \delta - \lambda\big(1 - \pi_d(s \mid G)\big)\big)^N \end{aligned}$$

where $\lambda = \frac{\mu_0(G)}{1 - \mu_0(G)}$.

For (a), we can show that for $\delta \in [0, 1 - \pi_d(s \mid G)]$,

$$T'(\delta) = N(\pi_d(s \mid G) + \delta)^{N-1} - N\Big((\pi_d(s \mid G) + \delta) - \lambda(1 - \pi_d(s \mid G))\Big)^{N-1}.$$

Since $\lambda(1 - \pi_d(s \mid G)) > 0$, we have $\pi_d(s \mid G) + \delta > (\pi_d(s \mid G) + \delta) - \lambda(1 - \pi_d(s \mid G))$. Because $N - 1 \geq 1$, the function $t \mapsto t^{N-1}$ is strictly increasing on $\mathbb{R}_+$. Hence $(\pi_d(s \mid G) + \delta)^{N-1} >$

$((\pi_d(s \mid G) + \delta) - \lambda(1 - \pi_d(s \mid G)))^{N-1}$ and thus $T'(\delta) > 0$ for all $\delta \in [0, 1 - \pi_d(s \mid G)]$. This proves (a).

For (b), note that

$$\begin{aligned}
T(0) &= \pi_d(s \mid G)^N - \lambda(1 - \pi_d(s \mid G)^N) - (\pi_d(s \mid G) - \lambda(1 - \pi_d(s \mid G)))^N \\
&= (1+\lambda)\pi_d(s \mid G)^N - \lambda - ((1+\lambda)\pi_d(s \mid G) - \lambda)^N \\
&= (1+\lambda)\pi_d(s \mid G)^N - \left[\lambda + ((1+\lambda)\pi_d(s \mid G) - \lambda)^N\right] \\
&= (1+\lambda)\left(\pi_d(s \mid G)^N - \left[\left(\frac{\lambda}{1+\lambda}\right) \cdot 1^N + \left(\frac{1}{1+\lambda}\right)((1+\lambda)\pi_d(s \mid G) - \lambda)^N\right]\right).
\end{aligned}$$

Since $\pi_d(s \mid G) < 1$, we have $(1+\lambda)\pi_d(s \mid G) - \lambda \neq 1$. Since $N \geq 2$ and the function $t \mapsto t^N$ is strictly convex, we have

$$\pi_d(s \mid G)^N < \frac{\lambda}{1+\lambda} + \left(\frac{1}{1+\lambda}\right)((1+\lambda)\pi_d(s \mid G) - \lambda)^N$$

by Jensen's inequality. Hence, $T(0) = (1+\lambda)\pi_d(s \mid G)^N - \lambda - y^N < 0$. This proves (b).

For (c), note that

$$\begin{aligned}
T(1 - \pi_d(s \mid G)) &= 1 - \lambda(1 - \pi_d(s \mid G)^N) - (1 - \lambda(1 - \pi_d(s \mid G)))^N \\
&= (1-\lambda) \cdot 1^N + \lambda\pi_d(s \mid G)^N - [(1-\lambda) \cdot 1 + \lambda\pi_d(s \mid G)]^N.
\end{aligned}$$

Since $\pi_d(s \mid G) < 1$, we have $(1-\lambda) \cdot 1 + \lambda\pi_d(s \mid G) \neq 1$. Notice that since $\mu_0(G) < 0.5$, we have $\lambda < 1$ and thus $(1-\lambda) \cdot 1^N + \lambda\pi_d(s \mid G)^N$ is a convex combination between 1^N and $\pi_d(s \mid G)^N$. Again, since $t \mapsto t^N$ is strictly convex, we have $[(1-\lambda) \cdot 1 + \lambda\pi_d(s \mid G)]^N < (1-\lambda) \cdot 1^N + \lambda\pi_d(s \mid G)^N$ by Jensen's inequality. Hence $T(1 - \pi_d(s \mid G)) > 0$. This completes the proof of Lemma 6. □

Since T is continuous, Lemma 6 and the intermediate value theorem imply that there exists a unique $\delta^* \in (0, 1 - \pi_d(s \mid G))$ such that $T(\delta^*) = 0$ and $T(\delta) > 0$ for all $\delta > \delta^*$. This completes the proof of Proposition 5.

E. Proof of Corollary 3

In the proof of Proposition 5, the sender discloses a single signal if and only if $T(\delta) \geq 0$, where $\delta = \pi_d(s \mid B) - \pi_d(s \mid G)$, and T is increasing in δ with a unique root δ^*. Setting $N = 2$ and solving $T(\delta) = 0$ gives $\delta^* = \frac{1 - \pi_d(s|G)}{2\,\mu_0(B)}$, which completes the proof.

References

Aina, C. (2025). Tailored stories. Working paper.

Ba, C., Dmitriev, D., Hang, Z., and Papazyan, F. (2026). Strategically Controlling Worldviews. Working paper.

Burkovskaya, A. and Starkov, E. (2026). Causal Persuasion. Working paper.

Grossman, S. J. and Hart, O. D. (1980). Disclosure Laws and Takeover Bids. *The Journal of Finance*, 35 (2): 323–334.

Jain, A. (2023). Informing agents amidst biased narrative. Working paper.

Jiang, Z., Liu, H., Peng, C., and Yan, H. (2025). Investor Memory and Biased Beliefs: Evidence from the Field. *The Quarterly Journal of Economics*, 140 (4): 2749–2804.

Kamenica, E. and Gentzkow, M. (2011). Bayesian persuasion. *American Economic Review*, 101 (6).

Malmendier, U. and Nagel, S. (2011). Depression Babies: Do Macroeconomic Experiences Affect Risk Taking? *The Quarterly Journal of Economics*, 126 (1): 373–416.

Milgrom, P. R. (1981). Good News and Bad News: Representation Theorems and Applications. *The Bell Journal of Economics*, 12 (2): 380–391.

Schwartzstein, J. and Sunderam, A. (2021). Using models to persuade. *American Economic Review*, 111 (1): 276–323.

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27 (3).