Constructive Blackwell's Theorem *

Itai Arieli[†] Yakov Babichenko[‡] Fedor Sandomirskiy[§]

Abstract

Blackwell's theorem, connecting majorization to the existence of signals inducing a desired distribution of posterior means, has numerous applications in economics. We give a new proof of this theorem via an explicit construction. Our approach provides a concrete way to generate signals: we demonstrate that any distribution inducible by some signal can also be induced by a "downward-uniform signal," which simply imposes a stochastic lower bound on the realized state. We further study properties of these signals, indicating their suitability in static and dynamic economic environments.

1 Introduction

Blackwell's celebrated theorem Blackwell (1951) is a cornerstone of information economics, providing a fundamental link between the informativeness of signals and the distributions of posterior beliefs they can induce. Specifically, the theorem establishes a necessary and sufficient condition for the existence of a signal that generates a desired distribution of posterior means: a distribution F of posterior means can be induced from a prior distribution G of a state variable if and only if F majorizes G. While Blackwell's theorem characterizes *which* belief distributions are feasible, it offers no guidance on *how* to construct signals that achieve them. The original proof is non-constructive, providing an existence result without a concrete method for signal design.

This paper addresses this gap by providing a novel, simple, and economically meaningful construction of signals that achieve any feasible belief distribution in the context of Blackwell's theorem. We consider a setting where a state, θ, is distributed on an interval Θ according to a distribution G, and an agent seeks to induce a posterior belief distribution F that majorizes G and shares the same mean. Rather than relying on abstract existence arguments, we demonstrate that any such F can be induced by a *downward-uniform signal*.

*We are grateful to Kailin Chen, Théo Durandard, Mira Frick, Ryota Iijima, Andreas Kleiner, Anton Kolotilin, Xiao Lin, Teddy Mekonnen, Jeffrey Mensch, Ulrich Müller, Mehdi Shadmehr, Omer Tamuz, and Alexander Wolitzky for valuable discussions, and to participants at ACM EC 2025, Israeli AGT day, the Toulouse seminar, and the Stony Brook GT Conference for their feedback.

[†]University of Toronto & Technion, iarielitechnion.ac.il. Itai Arieli was supported by the Israel Science Foundation (grant #2030524).

[‡]Technion, yakovbab@technion.ac.il. Yakov Babichenko was supported by BSF grants #2018397 and #2021680.

[§]Princeton, fsandomi@princeton.edu.

A downward-uniform signal is constructed based on a monotone function $h : \Theta \to \mathbb{R}_+$. Given a realized state θ, the agent observes a signal s drawn uniformly from the interval $[0, h(\theta)]$. This implies that the observed signal s provides a lower bound on the true state θ. Importantly, signals of this form are frequently encountered in empirical work where data is only observed for individuals whose underlying type exceeds a variable threshold. For instance, in studies of consumer behavior, a purchase is only observed if a buyer's private value exceeds a random price. This creates a natural censoring mechanism analogous to a downward-uniform signal.

Our result shows that this simple and common signal structure is fully general: for any distribution F and any non-atomic distribution G satisfying the conditions of Blackwell's theorem, there exists a function h such that the resulting downward-uniform signal induces precisely the desired distribution F. Furthermore, we provide an explicit formula for this function h, which is defined through a geometrically intuitive quantity α representing the difference in mass between F and G at pairs of threshold.

This explicit construction offers several advantages beyond demystifying the process of signal design by replacing the abstract existence result of Blackwell's theorem with a concrete recipe. We observe that the family of downward-uniform signals possesses a structural property that makes it particularly convenient for modeling information aggregation across sources and information accumulation over time. Specifically, observing a sequence of conditionally independent downward-uniform signals is equivalent to observing a single downward-uniform signal whose function h equals the product of the individual functions h_i.

This mirrors a well-known invariance property of Gaussian signals under aggregation. However, downward-uniform signals offer greater flexibility: their invariance holds for arbitrary distributions F and G, whereas the Gaussian case requires both to be Gaussian. As a result, in Blackwell's theorem, it is without loss of generality to assume that F is generated by a sequence of i.i.d. signals. That is, for continuous state spaces, signals are effectively infinitely divisible. This is in sharp contrast with the finite-state setting. For instance, in binary-state persuasion, binary signals are known to be optimal, but such signals are generally not equivalent to a combination of informative, conditionally independent signals.

We also examine downward-uniform signals in specific economic environments, highlighting their connection to classical econometric questions and to models of information provision in both static and dynamic settings.

The paper is structured as follows. Section 2, which follows the discussion of related literature, presents our main result: a constructive version of Blackwell's theorem, together with the underlying intuition. Section 3 examines key properties of downward-uniform signals and their applications to specific economic environments. The main proofs are outlined in the body of the paper, while the full formal arguments and technical details are provided in the appendices.

Related Literature Blackwell's seminal work Blackwell (1951, 1953) focused on the comparison of signals by informativeness, characterizing belief distributions that can be induced by garbling a given signal. This result has become one of the central tools in information economics, offering a tractable way to parameterize belief distributions that an informed sender can induce; see, e.g., Gentzkow and Kamenica (2016); Kolotilin (2018); Dworczak and Martini (2019); Ivanov (2021); Candogan and Strack (2023); Bergemann, Heumann, and Morris (2022). While specific classes of belief distributions are known to be inducible by simple signal structures—e.g., bi-pooling signals suffice for optimization of a convex objective Kleiner, Moldovanu, and Strack (2021); Arieli, Babichenko, Smorodinsky, and Yamashita (2023)—the question of *how* general belief distributions can be induced remained unaddressed in the economic literature prior to our work.[1]

While the distribution of posterior means plays a central role in many economic settings, a recent paper by Yang and Zentefis (2024) has shown that in some environments, the relevant object is the distribution of posterior medians or—more generally, posterior quantiles—and characterized such feasible distributions F for a given prior G. As with the Blackwell theorem, their characterization is non-constructive and does not yield a simple recipe for constructing a signal that induces the target distribution. This issue has been recently addressed by Kolotilin and Wolitzky (2024). While conceptually closest to our work, their setting and methods differ substantially since posterior quantiles behave quite differently from means. In particular, the key idea in their construction is to design a signal that generates maximal ambiguity in the quantile, so that different tie-breaking rules yield all feasible distributions. In contrast, posterior means are uniquely defined, and one must construct a separate signal tailored to each target distribution F.

Our results also connect to the growing literature on reduced-form approaches to Bayesian mechanism design Kleiner, Moldovanu, and Strack (2021); Ashlagi, Monachou, and Nikzad (2021); Nikzad (2022). This literature simplifies multi-agent design problems by reformulating them in terms of one-agent marginals of a mechanism—the so-called reduced mechanisms, representing the expected outcome for an agent conditional on her type. Blackwell's theorem characterizes feasible reduced mechanisms when interpreting the agent's type as a signal and the outcome as a state. Hart and Reny (2015) pioneered this connection, studying feasible reduced forms in single-good allocation. Our construction can be seen as a tool for designing mechanisms that achieve desired reduced forms.

In the mathematical literature, Strassen's theorem (Strassen (1965), Theorem 8) generalizes Blackwell's theorem to multiple dimensions and signal sequences. This theorem states, without

[1] Kleiner, Moldovanu, and Strack (2021); Arieli, Babichenko, Smorodinsky, and Yamashita (2023) characterized the extreme points of the set of distributions that majorize a given one, showing that each such extreme distribution can be induced via bi-pooling signals. By the Choquet theorem (Phelps, 2001), any majorizing distribution can be expressed as a weighted average of these extremes, and thus is in principle inducible via a randomization over bi-pooling signals. However, this yields no tractable description of the signal, as the Choquet theory is inherently existential. Finding an explicit decomposition of a majorizing distribution into extreme ones remains an open problem.

economic context, that a distribution F majorizes G if and only if there exists a martingale (X, Y) where $X \sim F$ and $Y \sim G$.[2] Our paper provides an explicit construction of such a joint distribution (X, Y) for given F and G. In general, such a joint distribution is not unique, and several alternative constructions have been proposed in the literature on martingale optimal transport. Notable examples include the *left-curtain coupling* (Beiglböck and Juillet, 2016; Hobson and Norgilas, 2022) and related constructions based on Skorokhod embeddings into Brownian motion (Hobson, 2011; Beiglböck, Henry-Labordère, and Touzi, 2017).

Our construction differs in several important aspects. It is arguably simpler, possesses a clear economic interpretation, and, crucially for information and mechanism design, focuses on generating X given Y (e.g., sampling a signal given the state). Existing mathematical constructions are convenient for sampling Y given X, but the conditional distribution of X given Y lacks a known closed-form expression. Our approach yields equally simple forms for both conditional distributions, $X|Y$ and $Y|X$, though we focus on the former for economic relevance.

Finally, our analysis of optimism in learning connects to a rich literature on dynamic information acquisition and belief updating. Dubins and Gilat (1978) and Hobson (1998) studied maximal martingales, while Koh, Sanguanmoo, and Zhong (2024) analyzed persuasion in optimal stopping problems. Khantadze, Kremer, and Skrzypacz (2025) studied the case of multiple actions. Building on the "conclusive bad news" martingales of Dubins and Gilat (1978), we contribute a novel characterization of the most optimistic martingale under informational constraints.

2 Constructive Blackwell's Theorem

This section presents our main result: a constructive version of Blackwell's theorem using downward-uniform signals. We show that any feasible posterior distribution can be induced by a signal that is uniformly distributed in an interval whose upper bound depends on the realized state. Before describing the construction, we briefly discuss the key concepts of majorization, garbling, and Blackwell's theorem.

Let $\Theta = (\underline{\theta}, \overline{\theta}) \subset \mathbb{R}$ be a possibly unbounded interval, allowing for $\underline{\theta} = -\infty$ or $\overline{\theta} = +\infty$. We consider a state θ drawn from this interval according to a cumulative distribution function (CDF) G. Throughout the paper, we identify distributions with their CDFs. We refer to G as the *prior* distribution and primarily focus on non-atomic priors. The integrated CDF is denoted by $\widehat{G}$

$$\widehat{G}(x) = \int_{\underline{\theta}}^{x} G(y)dy.$$

Definition 1. For CDFs on Θ with finite expectation, we say that F *majorizes* G if

$$\widehat{F}(x) \le \widehat{G}(x) \quad \text{for all} \quad x \in \Theta \qquad \text{and} \qquad \widehat{F}(\overline{\theta}) = \widehat{G}(\overline{\theta}).$$

[2]Recall that (X, Y) is a martingale if $E[Y|X] = X$. Interpreting Y as a state and X as a signal, by the martingale property, the posterior mean $E[Y|X]$ equals the signal itself and thus is distributed according to F.

The majorization is *strict* if the inequality is strict for all $x \in \Theta$.

Majorization formalizes the idea that the distribution F is less dispersed than G. The concept was first introduced by Hardy, Littlewood, and Polya (1929) for vectors in Euclidean space, and our definition follows the formulation in Kleiner, Moldovanu, and Strack (2021).[3]

A *signal* provides information about $\theta \sim G$, and we are interested in the resulting *posterior mean* after observing the signal. Formally, let $\mathcal{S}$ be a measurable set of signals. A signal about θ (or *garbling*) is a mapping $\pi : \Theta \to \Delta(\mathcal{S})$ that provides noisy information in the form of $s \sim \pi(\theta)$ which yields a posterior mean $E[\theta|s]$. The *induced distribution of the posterior means* is the distribution of $E[\theta|s]$ where s is drawn in two steps: first sample $\theta \sim G$ and then draw $s \sim S(\theta)$.

If a distribution of posterior means F is induced by some signal s, then a simple application of Jensen's inequality implies that F majorizes G. Blackwell's theorem establishes the other direction.[4]

Theorem 1 (Blackwell (1951))**.** *If a distribution F majorizes a prior distribution G, then there is a signal that induces the distribution of posterior means F.*

To illustrate the applicability as well as the limitations of Blackwell's theorem consider the following example.

Example 1. Let $F = \mathrm{Beta}(2, 2)$ and $G = \mathrm{Uniform}([0, 1])$. Blackwell's theorem makes it elementary to check that F is inducible from G by verifying the majorization condition:

$$\widehat{F}(x) = x^3 - \frac{x^4}{2} \leq \frac{x^2}{2} = \widehat{G}(x) \quad \text{for every} \quad x \in [0, 1].$$

However, the question of *which* garbling of G induces the Beta(2, 2) distribution remains unclear from the theorem and its original non-constructive proof.

As in the example above, we are interested in how, given a prior G and a desired posterior F majorizing G, we can construct a signal that induces this posterior. To formulate the answer, we introduce the concept of a *downward-uniform signal*.

Definition 2. A *downward-uniform signal* is defined by a non-decreasing function $h : \Theta \to \mathbb{R}_+$. Given state θ, the signal s is drawn uniformly from the interval $[0, h(\theta)]$.

That is, a downward-uniform signal can be represented as

$$s = h(\theta) \cdot U, \tag{1}$$

[3]In the economics literature, majorization appears under various names, including second-order stochastic dominance (for distributions with the same mean), mean-preserving contraction, reverse convex order, the Lorenz (1905) order, or having less risk in the sense of Rothschild and Stiglitz (1970).

[4]Blackwell (1951)'s original formulation is a particular case, where the state $\theta \sim G$ itself is the posterior beliefs about some binary $\omega \in \{0, 1\}$ induced by a signal s_1 and the question is what belief distributions F can be induced by garbling s_1. Our formulation is a particular case of (Strassen, 1965, Theorem 8)—a standard reference in math literature—applicable to the multidimensional case and sequential garbling. A related result was obtained by Hardy, Littlewood, and Polya (1929).

where U is uniformly distributed on $[0,1]$ and is independent of θ. Figure 1 illustrates this concept. Such a signal provides a noisy lower bound on the true state θ. The higher the state, the higher the potential range of signals.

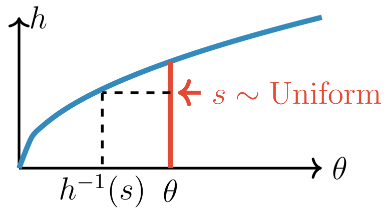


Figure 1: A downward-uniform signal. The myblue curve represents the function $h(\theta)$. For a given state θ, the signal s is drawn uniformly from the interval $[0, h(\theta)]$.

2.1 The Main Theorem

We are now ready to state our main result, the constructive counterpart to Blackwell's theorem.

Theorem 2 (Constructive Blackwell Theorem)**.** *If a distribution F strictly majorizes a non-atomic prior distribution G, then there exists a downward-uniform signal that induces the distribution of posterior means F.*

This theorem refines Blackwell's theorem by guaranteeing that the desired posterior distribution, F, can be garbled from the prior G via a signal that has a specific form—a downward-uniform signal. We stress that while the distribution G is assumed to have no atoms, no continuity assumptions are imposed on F since atomic posterior distributions often originate endogenously as a result of information design. The assumption of strict majorization assumption can be relaxed; see Remark 1 below.

This downward-uniform signal from Theorem 2 can be expressed through the primitives F and G, justifying the constructive nature of the result. The function $h\colon \Theta \to \mathbb{R}_{>0}$, which defines the downward-uniform signal, is pinned down up to a multiplicative factor and is given by

$$h(t) = \alpha(t) \cdot \exp\left(\int_{t_0}^{t} \frac{1}{\alpha(x)}\, dG(x)\right), \tag{2}$$

where $t_0 \in \mathbb{R}$ is an arbitrary interior point of Θ and $\alpha(t)$ is defined geometrically as follows (Figure 2).[5] We trace a tangent line to the graph of the integrated CDF $\widehat{G}(x)$ at a point $x = t$ and find a point $x = T$ to the right of t, where this line intersects the graph of the integrated CDF $\widehat{F}(x)$. The quantity $\alpha(t)$ is then defined as the difference between the slopes of the integrated CDF of F at T and the integrated CDF of G at t. Formally, $\alpha\colon \Theta \to (0,1)$ is given by

$$\alpha(t) = F(T) - G(t), \qquad \text{where} \qquad T \in (t, \underline{\theta}) \qquad \text{solves} \qquad \widehat{G}(t) + (T-t)G(t) = \widehat{F}(T). \tag{3}$$

[5]For every interior point $t \in \Theta$, due to a *strict* majorization, the function $\alpha(x)$ is bounded away from 0 in the interval $[t_0, t]$ (or in the interval $[t, t_0]$, depending on their order). Therefore, for every interior point t the integral in Equation (2) is well defined.

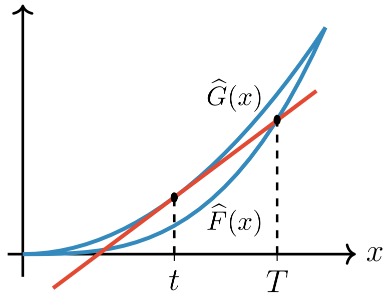


Figure 2: $\alpha(t)$ is the difference between the slope of $\widehat{F}$ (lower curve) at T and the slope of $\widehat{G}$ (upper curve) at t.

Although atoms of F lead to jumps in α and h, both functions remain right-continuous.

Example 2. Let $F = \text{Beta}(2,2)$ and $G = \text{Uniform}([0,1])$ as in Example 1. Theorem 2 suggests the following garbling for this pair of distributions. For every $\theta \sim \text{Uniform}([0,1])$ draw a signal $s \sim \text{Uniform}([0, h(\theta)]) = \text{Uniform}([0,\, \theta^{3/2}])$. A simple calculation using the Bayes formula verifies that s results in the posterior mean distributed according to $\text{Beta}(2,2)$.

The identity $h(\theta) = \theta^{3/2}$ is obtained as follows. First we compute $T(t) = \sqrt{t}$ (see Figure 2 for the functions $\widehat{F}(x) = x^3 - \frac{x^4}{2}$ and $\widehat{G}(x) = \frac{x^2}{2}$), and hence $\alpha(t) = F(T) - G(t) = 2t(1-\sqrt{t})$. By formula (2), we get

$$h(t) = 2t(1-\sqrt{t}) \exp\left(\int_{t_0}^{t} \frac{1}{2x(1-\sqrt{x})} dx\right)$$
$$= 2t(1-\sqrt{t}) \exp\left(\log(\sqrt{t}) - \log(1-\sqrt{t}) + c\right) = 2e^c t^{3/2}$$

The constant factor $2e^c$ plays no role and can be replaced with 1.

The next example provides intuition on how the behavior of h translates into the behavior of F. A sharp increase of h creates an interval in F carrying little mass, while regions where h does not change much lead to condensation of mass. These phenomena are particularly stark in the case of piecewise-constant h that lead to atomic F.

Example 3. Consider atomic $F = \text{Uniform}(\{\frac{1}{3}, \frac{2}{3}\})$ and $G = \text{Uniform}([0,1])$. It is easy to verify that F majorizes G. One can either use formula (2) or infer directly that the step function h equal to $\frac{1}{4}$ for $x \le \frac{1}{3}$ and 1 for $x > \frac{1}{3}$ has the desired properties. The resulting garbling is depicted in Figure 3.

Remark 1 (Non-strict Majorization). The case of non-strict majorization can be reduced to separate strict majorization instances. Consider F majorizing G. The integrated CDFs are continuous, and thus $\{x : \widehat{G}(x) > \widehat{F}(x)\}$ is an open set, which can be represented as the union of countably many open intervals $\{I_n\}_{n\in N}$. Let s_n be a downward-uniform signal for the pair of distributions obtained by conditioning F and G to the interval I_n. To garble G into F, one can send the pair of signals (n, s_n) where $n \in N$ reveals which interval the state belongs. Outside of $\cup_n I_n$, realizations of θ are completely disclosed.

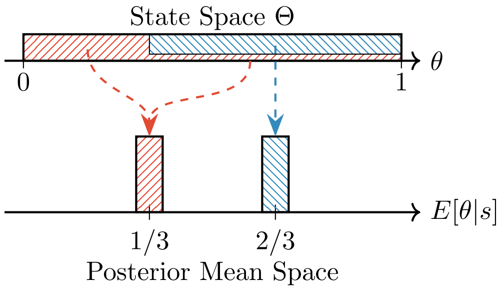


Figure 3: Downward-uniform garbling inducing $F = \text{Uniform}(\{\frac{1}{3}, \frac{2}{3}\})$ from $G = \text{Uniform}([0,1])$.

The economic interpretation of $\alpha(t)$ is explored in Section 3 (see Remark 2). We now discuss an intuition behind the theorem and the origin of the Formula (2). The first step is understanding specific properties of belief distributions induced by downward-uniform signals.

2.2 Beliefs Induced by Downward-Uniform Signals

This section characterizes the distribution of posterior beliefs induced by downward-uniform signals. For clarity of exposition, we assume that the distribution G of the state θ has density g, that is G is absolutely continuous. Conditionally on θ, the downward-uniform signal s is drawn from the uniform distribution on the interval $[0, h(\theta)]$. Thus, the density of s conditional on θ is given by $p(s|\theta) = \mathbb{1}_{\{s \leq h(\theta)\}} \frac{1}{h(\theta)}$. Let $g(\theta|s)$ be the density of the conditional distribution of θ given the signal s. Using Bayes' formula:

$$g(\theta|s) = \frac{p(s|\theta)g(\theta)}{\int_{\mathbb{R}} p(s|\theta)g(\theta)d\theta} = \frac{\mathbb{1}_{\{s \leq h(\theta)\}} \frac{g(\theta)}{h(\theta)}}{\int_{\theta : h(\theta) \geq s} \frac{g(\theta)}{h(\theta)} d\theta}.$$

We see that for any signal realization, $g(\theta|s)$ is an upper tail of the same distribution with density proportional to $\frac{g(\theta)}{h(\theta)}$. In other words, observing different signals only changes our beliefs about the relative probabilities of states within the upper tail, not the shape of the distribution within that tail. We call it the **identical quantiles property**.

The posterior mean is the average of θ with respect to $g(\theta|s)$:

$$E[\theta|s] = \int_{\mathbb{R}} \theta g(\theta|s) d\theta = \frac{\int_{\theta : h(\theta) \geq s} \theta \frac{g(\theta)}{h(\theta)} d\theta}{\int_{\theta : h(\theta) \geq s} \frac{g(\theta)}{h(\theta)} d\theta}. \tag{4}$$

It is easy to see that the identical quantile property implies the following **monotonicity property**: the posterior mean $E[\theta|s]$ is monotone in s; i.e., higher signals lead to higher posterior means.[6]

Corollary 1. *Beliefs induced by downward-uniform signals exhibit identical quantiles and monotonicity properties. In particular, these properties are consistent with inducing any distribution F majorizing G.*

[6] A stronger monotonicity property holds: the belief after observing a signal s—i.e., the conditional distribution of the state given this signal—first-order stochastically dominates the belief after observing a signal $s' < s$. This is immediate as these beliefs are different tails of the same distribution.

We will see that these properties are important in future sections.

2.3 Intuition Behind the Main Theorem

This section provides a heuristic derivation of the formula for h, the function defining the downward-uniform signal that induces the posterior distribution F from the prior G. While Appendix A establishes the formula's validity for general F and G, here we focus on the intuition behind its functional form without paying attention to technical assumptions we make along the way. For further simplicity, we focus on the case where both G and F admit densities g and f, which are positive on $[0,1]$ and zero elsewhere.

We begin by considering the cumulative distribution function of the signal s, denoted by $K(s)$. Recall that the conditional density is given by $p(s|\theta) = \mathbb{1}_{\{s\le h(\theta)\}}\frac{1}{h(\theta)}$. Thus,

$$K(s) = \int_0^1 \left(\int_0^s p(s'|\theta)ds'\right) g(\theta)d\theta = \int_0^1 \left(\int_0^s \frac{\mathbb{1}_{\{s'\le h(\theta)\}}}{h(\theta)}ds'\right) g(\theta)d\theta.$$

Since only the indicator function depends on s' in the inner integral, we can rewrite this as

$$\begin{aligned} K(s) &= \int_0^1 \frac{g(\theta)}{h(\theta)} \left(\int_0^s \mathbb{1}_{\{s'\le h(\theta)\}}ds'\right) d\theta \\ &= \int_0^1 \frac{g(\theta)}{h(\theta)} \min\{h(\theta), s\} d\theta \\ &= \int_{\theta:h(\theta)\le s} \frac{g(\theta)}{h(\theta)} h(\theta)d\theta + \int_{\theta:h(\theta)\ge s} \frac{g(\theta)}{h(\theta)} s\, d\theta \\ &= \int_{\theta:h(\theta)\le s} g(\theta)d\theta + s\int_{\theta:h(\theta)\ge s} \frac{g(\theta)}{h(\theta)}d\theta. \end{aligned}$$

Now, consider a value $t \in [0,1]$ and define $T = E[\theta|s]$ for $s = h(t)$. By the monotonicity property of downward-uniform signals, $E[\theta|s] \le T$ if and only if $s \le h(t)$. By the choice of T and t, the following identity holds:

$$F(T) = K(h(t)).$$

Substituting the formula for $K(s)$ and simplifying $h(\theta) \le h(t)$ to $\theta \le t$ (which is valid because h is monotone), we obtain

$$F(T) = G(t) + h(t)\int_t^1 \frac{g(\theta)}{h(\theta)}d\theta. \tag{5}$$

Since T can be viewed as a function of t, we define $\alpha(t) = F(T) - G(t)$. Thus,

$$\alpha(t) = h(t)\int_t^1 \frac{g(\theta)}{h(\theta)}d\theta. \tag{6}$$

We now show that h can be expressed in terms of α via formula (2). At this stage, α is not yet expressed in terms of the primitives F and G as it depends on T. However, we will later demonstrate that T can be expressed in terms of the primitives via the identity (3).

Taking the logarithm of both sides of (6) and then differentiating with respect to t, we get

$$(\log\alpha)' = (\log h)' - \frac{g}{h\cdot\int_t^1 \frac{g(\theta)}{h(\theta)}d\theta}.$$

Expressing the integral in the denominator using (6) and rearranging, we arrive at

$$(\log h)' = (\log \alpha)' + \frac{g}{\alpha}.$$

Integrating both sides yields

$$\log h = \log \alpha + \int \frac{g}{\alpha},$$

which is equivalent to (2).

Finally, we derive the identity (3) for $T = T(t)$. Using the formula (4) for $E[\theta|s]$ and the definition of T, we have

$$T = \frac{\int_t^1 \theta \frac{g(\theta)}{h(\theta)} d\theta}{\int_t^1 \frac{g(\theta)}{h(\theta)} d\theta}.$$

Multiplying by the denominator and differentiating both sides with respect to t gives

$$-T(t) \cdot \frac{g(t)}{h(t)} + T'(t) \int_t^1 \frac{g(\theta)}{h(\theta)} d\theta = -t \cdot \frac{g(t)}{h(t)}.$$

Multiplying both sides by $h(t)$ and expressing the integral using (5) results in

$$-T(t) \cdot g(t) + T'(t) \cdot (F(T(t)) - G(t)) = -t \cdot g(t).$$

Since $T'G(t) + Tg(t)$ is the derivative of $TG(t)$ and $T'F(T)$ is the derivative of $\widehat{F}(T)$, we can integrate both sides and obtain

$$\widehat{F}(T) - TG(t) = C - \int_0^t xg(x)dx,$$

where C is some constant. Integrating by parts on the right-hand side and reshuffling the terms, we get

$$\widehat{G}(t) + G(t) \cdot (T - t) + C = \widehat{F}(T)$$

Setting $t = 1$ we get $T = 1$ and find that $C = 0$, thus establishing (3).

3 Properties of Downward-Uniform Signals, Examples, and Applications

For a given prior distribution G of a state θ and a target distribution F of induced beliefs, downward-uniform signals are just one among many possible ways to contract the former to the latter. In this section, we study the specific properties of the downward-uniform garbling that single it out from other information structures capable of inducing the same distribution of posterior means.

To see that many different garblings can induce the same F from a given G, consider an example where the prior is $G = \text{Uniform}([0, 1])$ and the target distribution of posteriors is $F = \text{Uniform}(\{\frac{1}{3}, \frac{2}{3}\})$. The downward-uniform signal provides one way to achieve this by contracting a fraction of $\frac{3}{4}$ of the mass of G in the interval $[\frac{1}{3}, 1]$ into an atom on $\frac{2}{3}$ while the remaining mass is contracted into an atom on $\frac{1}{3}$ as depicted in Figure 3. There are however infinitely many other garblings that result in the same F but differ in the distribution of posterior means induced *conditionally on realized state* θ. For instance, Figure 4 illustrates a structurally different garbling that contracts the mass of the interval $[\frac{1}{12}, \frac{7}{12}]$ into an atom on $\frac{1}{3}$ and the remaining mass into an atom on $\frac{2}{3}$.

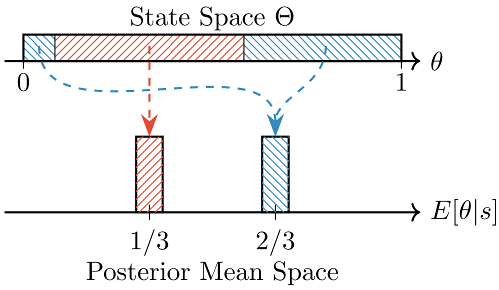


Figure 4: A non-monotone garbling inducing $F = \text{Uniform}(\{\frac{1}{3}, \frac{2}{3}\})$ from $G = \text{Uniform}([0, 1])$ maps intermediate states to the low posterior and extreme states to the high posterior.

A social planner whose objective depends only on the induced distribution of posterior means F, as in Dworczak and Martini (2019), would be indifferent between all such signals. For such a planner, downward-uniform signals simply offer one convenient choice among many. However, in many settings the planner's cares about the posterior means induced for each state realization, which makes the choice of garbling no longer a matter of convenience but decision that affects outcomes.[7]

Below, we discuss various environments where downward-uniform signals arise naturally. We use the context provided by these environments to highlight specific properties of downward-uniform signals and joint distributions of states and beliefs they induce.

3.1 The German tank problem and independent downward-uniform signals

The German tank problem is a classic statistical problem of estimating the size of a population from which a random sample is drawn.[8] A version of this problem is tightly related to downward-uniform signals.

[7]For instance, the joint distribution of the state and induced beliefs is critical, for instance, in persuasion to handle state-dependent objectives (Dworczak and Kolotilin, 2024; Kolotilin, Corrao, and Wolitzky, 2025) and welfare effects in persuasion Doval and Smolin (2024), recommendation system and public recognition schemes design (Saeedi and Shourideh, 2020; Vaeth, 2024), and martingale optimal transport (Beiglböck and Juillet, 2016).

[8]The name originates from its application by Allied forces during World War II to estimate the monthly production rate θ of German tanks from the serial numbers on captured vehicles. The same statistical technique has since been used to estimate the scope of production in various contexts and also in numerous other applications, ranging from software bugs to ecology; see a survey by Simon (2024).

Let the state θ be the total number of items produced, which we model as a continuous variable for simplicity. An analyst holds a prior belief G about θ. The items are enumerated from 0 to θ, so a randomly drawn item s has a serial number uniformly distributed on $[0, \theta]$. The observation of a single serial number s is therefore a downward-uniform signal with the function $h(\theta) = \theta$.[9]

A more realistic scenario involves an analyst observing a sample of k items with serial numbers $s_1, \ldots, s_k$, drawn independently from the uniform distribution on $[0, h(\theta)]$. This is equivalent to observing a collection of downward-uniform signals that are i.i.d. conditional on the state. Sequences of conditionally i.i.d. signals also arise in models of learning, where agents accumulate information over time or from multiple sources.[10]

In general, combining such signals does not preserve the original signal structure. For example, the joint observation of two conditionally i.i.d. binary signals $s_i \in 0, 1$, in general, cannot be reduced to a single binary signal, as the realizations $\{(0, 0)\}$, $\{(1, 0), (0, 1)\}$, and $\{(1, 1)\}$ yield different posteriors. Partially for this reason, the literature has emphasized asymptotic learning, where the number of signals grows large and detail-free conclusions become possible.

A notable feature of downward-uniform signals is that the family is closed under combination: any finite collection of conditionally independent downward-uniform signals—possibly with different distributions—is equivalent to a single downward-uniform signal. This holds even when the individual signals differ, as long as each is downward-uniform relative to its own bounding function.

Two signals s and s' are said to be *equivalent* if they induce the same joint distribution over the state and the posterior mean.

Proposition 1. *Let $s_1, \ldots, s_k$ be conditionally independent downward-uniform signals with corresponding bounding functions $h_1(\theta), \ldots, h_k(\theta)$. Then the collection $(s_1, \ldots, s_k)$ is equivalent to a single downward-uniform signal with bounding function*

$$h(\theta) = \prod_{i=1}^{k} h_i(\theta).$$

This property of downward-uniform signals suggests that they form a tractable class for modeling sequential learning and information aggregation from multiple sources. It is instructive to compare them with the widely used Gaussian framework. In that setting, the prior G is Gaussian, and each signal takes the form $s_i = \theta + \eta_i$, where the shocks η_i are Gaussian. Aggregating such signals yields another signal of the same form. Downward-uniform signals exhibit a similar invariance under aggregation, but offer greater flexibility: they accommodate arbitrary prior distributions G and allow each signal to induce an arbitrary posterior mean distribution F_i.

[9] Downward-uniform signals with general h, can capture the situation where $h(\theta)$ is the number of potentially available produced items and θ is a parameter of affecting the scope of production. For instance, θ may represent the number of produced items including those not yet distributed, a factor of production such as capital or labor, or just a statistic of the total production that the analyst is interested in.

[10] See, for instance, Moscarini and Smith (2002); Azrieli (2014); Cripps, Ely, Mailath, and Samuelson (2008); Mu, Pomatto, Strack, and Tamuz (2021); Frick, Iijima, and Ishii (2023, 2024).

The proposition is proved in Appendix B. To build intuition for the product formula, consider the likelihood of observing a single downward-uniform signal s' with bounding function h', and compare it to the likelihood of observing a collection of signals $s = (s_1, \ldots, s_k)$:

$$L(s'|\theta) = \frac{1}{h'(\theta)} \mathbb{1}_{\{s' \le h'(\theta)\}}, \qquad L(s|\theta) = \prod_{i=1}^{k} \frac{1}{h_i(\theta)} \mathbb{1}_{\{s_i \le h_i(\theta)\}}.$$

The second expression simplifies by noting that the product of indicators equals a single indicator $\mathbb{1}_{\{m(s) \le \theta\}}$, where $m(s) = \max_i h_i^{-1}(s_i)$. Setting $h(\theta) = \prod_{i=1}^{k} h_i(\theta)$, the likelihood becomes

$$L(s|\theta) = \frac{1}{h(\theta)} \mathbb{1}_{\{m(s) \le \theta\}}.$$

Comparing the two expressions suggests that, to match the posterior beliefs induced by the collection of signals s using a single signal s', one should take $h' = h$. This observation motivates the product formula. In Appendix B, we formally establish the equivalence by showing that a suitable monotone transformation of the sufficient statistic $m(s)$ yields a valid downward-uniform signal, and address technicalities arising from possible discontinuities and flat regions in the functions h_i.[11]

Proposition 1 implies a general structural property of garblings: their infinite divisibility under a non-atomic prior.

Corollary 2 (IID Blackwell Theorem)**.** *If a distribution F strictly majorizes a non-atomic prior G, then for any integer $k \ge 1$, there exist signals $s_1, \ldots, s_k$, i.i.d. conditional on the state, such that F is induced by the joint observation of $s_1, \ldots, s_k$.*

Indeed, by Theorem 1, the distribution F can be induced by a downward-uniform signal with bounding function h given by (2). By Proposition 1, a collection of k conditionally i.i.d. downward-uniform signals with bounding functions $h_i = h^{1/k}$ also induces F.

The IID Blackwell Theorem shows that information about a continuous state can always be decomposed into a sequence of conditionally i.i.d. signals. This implies that acquiring information over time or from multiple informative independent sources imposes no constraint on the eventual belief distribution of the receiver. Consequently, in persuasion problems with a continuous state space, requiring the receiver to obtain information through several i.i.d. signals does not reduce the sender's optimal value.

This stands in sharp contrast to the binary-state setting of Kamenica and Gentzkow (2011), where optimal signals are binary. Due to the atomic nature of the state space, a binary signal cannot be replicated by any collection of informative, conditionally independent signals unless one of the original signal's realizations fully reveals the state. As a result, imposing a multi-signal structure in that context results in a loss for the sender.

[11] In fact, we prove a stronger result: the single-signal and multi-signal models induce not only the same joint distribution over $(\theta, E[\theta|s])$, but also the same joint distribution over (θ, μ), where $\mu \in \Delta(\Theta)$ denotes the posterior belief.

3.2 Optimism in learning

Consider a Bayesian agent whose posterior mean evolves over time due to information arrival. We refer to this evolving posterior mean as a *learning process*. Suppose an analyst observes snapshots of the agent's belief distribution at two points: at $t = 0$, it's given by F, and at $t = 1$, by G, where G is majorized by F. From these snapshots, the analyst wants to make inferences about the agent's beliefs during the period $t \in [0, 1]$. A natural question arises: how optimistic can the agent's beliefs become during this interval? This notion is captured by τ-optimism (Definition 3). A related question is: which learning processes maximize optimism?

Formally, the posterior mean of the Bayesian agent is captured by a continuous-time martingale $(X_t)_{t\in[0,1]}$ with initial distribution $X_0 = F$ and terminal distribution $X_1 = G$. By a splitting lemma argument, exposure to information results in a martingale of posterior means and vice versa; every martingale of posterior means can be a result of some exposure to information.[12] Henceforth, we identify learning processes and continuous-time martingales.

Definition 3. The *τ-optimism* of an agent in the learning process $X = (X_t)_{t\in[0,1]}$ is defined by

$$\mathrm{Opt}_\tau(X) = P\big[\exists t \in [0, 1] : X_t \geq \tau\big] = P\left[\max_{t\in[0,1]} X_t \geq \tau\right]$$

Consider a scenario in which the analyst observes the correlation between F and G. Namely, the analyst can trace which types of agents ended up having which posterior mean as a function of their initial beliefs. *Given a correlation between F and G can one characterize the most optimistic learning process?*

For clarity of exposition, we assume that the state space is $\Theta = [0, 1]$.[13] We consider the following revelation strategy.

Definition 4. Given a continuous terminal distribution G, the *Gradual Exposure to Bad News (GEBN)* learning process at time $t \in [0, 1]$ exposes the agent to the realization of G in case it lies in the interval $[0, t]$. Otherwise, no further information is provided.

The most optimistic learning process turns out to be the above GEBN.

Proposition 2 (Essentially, Dubins and Gilat (1978))**.** *For every $\tau \in [0, 1]$, the GEBN learning process maximizes τ-optimism across all learning processes with the same correlation of initial and terminal distributions F and G.*

The proof of Proposition 2 is relegated to Appendix B. Dubins and Gilat (1978) proved this result for a point mass F, and the idea behind the proof in the appendix is to apply their result

[12] See Strassen (1965) for discrete time and Kellerer (1961) for continuous time.

[13] For a general Θ, the Definition 4 is modified so that, at time $t \in [0, 1]$, all the realizations that belong to the bottom t-quantile of G are revealed rather than realizations whose value is below t. This analog of the GEBN process is the relevant one also in the more general case in which G admits atoms. All the results in this section hold in these more general cases for the quantile variant of the GEBN process.

pointwise, conditional on each realization of X_0. For two-point G and one-point F, a version of this result has recently appeared in an application to dynamic implementation (Koh, Sanguanmoo, and Uzui, 2023).

This simple extension of Dubins and Gilat (1978) shows, given a correlation between F and G, the best way to reveal information to maximize $\tau-$optimism is by gradually revealing bad news. A natural question is whether there exists an "optimal correlation" between F and G that globally maximizes $\tau-$optimism. We provide a positive answer to this question and show that the correlation induced by the downward-uniform garbling is the optimal one.

We start by providing a general upper bound on optimism that applies to all correlations. Recall that $(t, \widehat{G}(t))$ for $t = T^{-1}(\tau)$ is the tangency point of the tangent line to $\widehat{G}$ that passes through $(\tau, F(\tau))$ (see Figure 2), and $G(T^{-1}(\tau))$ is the slope of this tangent.

Proposition 3. *For every learning process with initial distribution F and terminal distribution G, the agent is τ-optimistic with a probability of at most $1 - G(T^{-1}(\tau))$.*

The proof intuition of Proposition 3 is relegated to the end of this section, while the formal proof appears in Appendix B.

Proposition 3 is a generalization of the Hardy-Littlewood maximal inequality, which corresponds to F being a point mass. Hardy-Littlewood inequality provides a bound on τ-optimism as a function of the terminal distribution G only. The probability for τ-optimism is bounded from above by the mass of the top quantile whose conditional expectation is τ.[14] In our notation, it claims that the probability of τ-optimism is bounded by $1 - s$ where s is the slope of the tangent to $\widehat{G}$ that passes through the point $(\tau, E[G] + \tau - 1)$. Proposition 3 argues that this bound can be improved by considering the initial distribution too; The point through which the tangent passes increases from $(\tau, E[G] + \tau - 1)$ to $(\tau, F(\tau))$ which increases this slope; see Figure 5.

The following example demonstrates that the upper bound on the optimism from Proposition 3 may be (and typically is) strictly higher than the level of optimism achieved by the GEBN process of Proposition 2 for given correlation of F and G.

Example 4. Let $F = \text{Uniform}(\{\frac{1}{3}, \frac{2}{3}\})$ and let $G = \text{Uniform}([0, 1])$. The correlation of F and G is given by $G|(F = \frac{1}{3}) = \text{Uniform}([\frac{1}{12}, \frac{7}{12}])$ and $G|(F = \frac{2}{3}) = \text{Uniform}([0, \frac{1}{12}) \cup (\frac{7}{12}, 1])$.

Figure 6 depicts the dynamic posterior belief conditional on the event that the agent has not been exposed to the realization of G yet in the GEBN policy.

[14] The Hardy-Littlewood inequality immediately implies Doob's maximal inequality. For a distribution G with zero mean $E[G] = 0$ and $P[G > 0] = p$, we denote by G_q the distribution that equals 0 in the top q-quantile of G and 0 otherwise, and we denote by $q(\tau)$ the mass of the quantile whose expectation is τ. Now the Hardy-Littlewood inequality implies

$$\text{Opt}_\tau \leq q(\tau) = \frac{q(\tau) \cdot \tau}{\tau} = \frac{E[G_{q(\tau)}]}{\tau} \leq \frac{E[G_p]}{\tau} = \frac{E[\max(G, 0)]}{\tau}.$$

which is precisely Doob's maximal inequality.

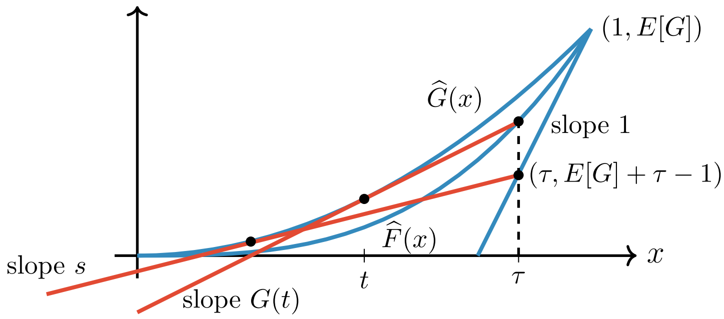
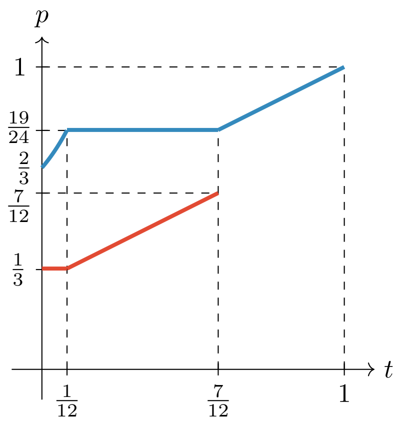


Figure 5: $\alpha(t)$ is the difference between the slope of $\widehat{F}$ (lower curve) at T and the slope of $\widehat{G}$ (upper curve) at t.

Figure 6: The upper function is the posterior of the agent with initial belief $\frac{2}{3}$ conditional on the state not being revealed till time t. The lower function is the posterior of the agent with initial belief $\frac{1}{3}$ conditional on the state not being revealed by time t. If the agent has initial belief $\frac{1}{3}$, the state is revealed with probability 1 by time $t = \frac{7}{12}$, therefore the function is defined only for $t \leq \frac{7}{12}$.



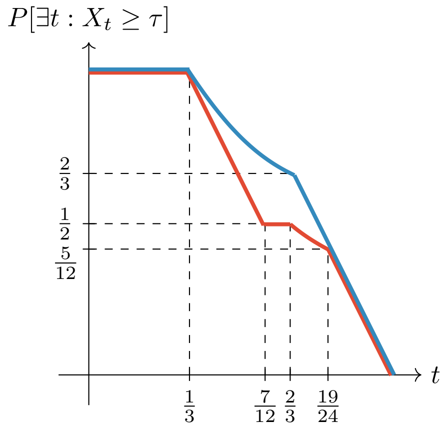
From these calculations, we can deduce the probability of τ-optimism for every value τ, and it is depicted in Figure 7. For the case of $F = \text{Uniform}(\{\frac{1}{3}, \frac{2}{3}\})$ and $G = \text{Uniform}([0, 1])$, one can calculate the bound of Proposition 3 on τ-optimism (for arbitrary correlations of F and G); its graph is also depicted in Figure 7.

Even though the learning process is the most optimistic one (for the exogenously given correlation), one can see that the GEBN learning process fails to reach the upper bound for $\frac{1}{3} < \tau < \frac{19}{24}$.

This example raises the question: *Are there correlations that induce higher optimism than others under the most optimistic GEBN learning process?* The answer to this question is positive. There exists a correlation that simultaneously maximizes τ-optimism for all values of τ. This correlation is the one induced by the downward-uniform garbling from Theorem 2.

Theorem 3. *If the correlation between F and G is given by the downward-uniform signals, then the GEBN learning process induces τ-optimism with the best-possible probability $1 - G(T^{-1}(\tau))$ for*

Figure 7: The lower function captures the probability of τ-optimism for the given correlation in the example (i.e., the upper-bound follows from Proposition 2). The upper function captures the upper bound on optimism given by Proposition 3.



every $\tau \in [0, 1]$.

In particular, we see that the bound from Proposition 3 is not only tight but also admits a correlation that is simultaneously optimal for all thresholds. The proof intuition of Theorem 3 is relegated to the end of this section, while the formal proof appears in Appendix B.

Intuitions via Persuasion Optimism has a natural interpretation in the context of dynamic persuasion, where the learning process is not exogenously given but rather is designed by a strategic sender. This interpretation is also convenient for building intuition about the results on $\tau-$optimism presented above.

Consider a dynamic setting in which a partially informed sender reveals information over continuous time to persuade a partially informed receiver (i.e., the agent) to make an irreversible adoption decision. Let $\{L, H\}$ be a binary state reflecting a product's low or high quality. The distribution $G \in \Delta([0, 1])$ captures the partial information (i.e., distribution over beliefs) of the sender about the quality being high. We assume that in the dynamic setting, the agent adopts at the moment once her posterior mean (i.e., her posterior about the high quality) exceeds some threshold τ.[15] Notice that τ-optimism is equivalent to adoption in this dynamic persuasion setting, which the sender aims to maximize.

More formally, F is the receiver's initial information. The sender is more informed than the receiver and holds the private information G that is majorized by F. The sender is allowed to reveal information over time and hence can design the martingale $(X_t)_{t \in [0,1]}$. Since the adoption

[15]Immediate adoption can be rationalized if the interaction occurs over time $[0, \infty)$, the receiver is impatient (namely, has a discount factor $\delta < 1$), and the sender is patient (namely, has an average-limit utility). In such a setting, immediate adoption is without loss of generality since the sender can arbitrarily slow down any information revelation policy, thus enforcing immediate adoption.

action is irreversible, we can assume without loss of generality that $X_1 = G$ because revealing further information can never harm the sender.

The correlation between F and G can be either exogenously given or designed by the sender. We first describe an interpretation of an exogenously given correlation. Before the interaction, both the sender and the receiver observe a public signal S_{pub} and the sender additionally observes a private signal S_{priv}. This creates a setting in which the correlation is exogenously given. Proposition 2 can be formulated in the context of dynamic persuasion as follows.

Corollary 3 (Reformulation of Proposition 2)**.** *In a setting with an exogenously given correlation between F and G, the GEBN policy is optimal for the sender.*

Another variant of the persuasion problem considers a sender who designs the correlation between F and G. The sender is partially informed about the state with the distribution of posteriors G. Before the dynamic interaction starts, the sender is required to send partial information to the receiver that will be at least as informative as F.[16] In this setting, the policy of the sender consists of two parts: the garbling of G to F and a continuous time information revelation policy. Proposition 3 and Theorem 3 specify the optimal policy in this setting.

Corollary 4 (Reformulation of Proposition 3)**.** *The sender cannot induce adoption with a probability higher than $1 - G(T^{-1}(\tau))$.*

Corollary 5 (Reformulation of Theorem 3)**.** *In a setting with a designed correlation between F and G, the downward-uniform signaling followed by the GEBN policy is optimal for the sender and induces adoption with probability $1 - G(T^{-1}(\tau))$ for every $\tau \in [0, 1]$.*

Notice that the policy in Corollary 5 is independent of τ. Therefore, in more general settings in which the threshold τ is unknown to the sender, or alternatively, the sender faces multiple receivers with different thresholds τ the same policy is the one to maximize the expected number of adopters.

Remark 2. We recall that the function $\alpha = \alpha(t) = F(T(t)) - G(t)$ played a central role in the explicit construction of the function $h(t)$—the generator of the downward-uniform signals; see Section 2.1. An economic interpretation for the function α can be deduced from Corollary 5. Given a threshold τ the persuasion value for the sender is $1 - G(T^{-1}(\tau))$. If, instead, the sender doesn't try to manipulate the receiver and simply sends no information the probability of adoption would be $1 - F(\tau)$. The *gain from persuasion* is the difference between these two probabilities and it equals $F(\tau) - G(T^{-1}(\tau)) = F(T(t)) - G(t) = \alpha(t)$ for $t = T^{-1}(\tau)$.

The persuasion perspective not only allows us to deduce insights about an intriguing dynamic persuasion model but also can serve as a mathematical tool for proving these results. Below, we

[16] Such a requirement might arise from regulations that require the producer to provide some information about the product before entering the market.

sketch the ideas of Corollaries 4 and 5, which are just equivalent formulations of Proposition 3 and Theorem 3.

Proof sketch of Proposition 3 and Corollary 4 The key idea is to bound the value of the dynamic persuasion problem by the analogous static persuasion problem. In the static problem, the sender reveals information in a single round; this information must be less informative than G (because this is what she knows) but more informative than F (because this is what the receiver initially knows).

Notice that any persuasive dynamic policy can be translated into a static one with the same probability of adoption: the sender commits to perform a realization of the martingale and reports whether this realization visits the interval $[\tau, 1]$. If it does, by the martingale property, the receiver's posterior lies in $[\tau, 1]$ and she takes the adoption action. This implies that indeed the value of the dynamic problem is bounded by the value of the static one.

To compute the value of the static problem, we formulate the optimization problem in the integrated CFDs space. We optimize over the set of concave functions $\widehat{H}$ sandwiched in between $\widehat{G} \leq \widehat{H} \leq \widehat{F}$ (i.e., the informativeness restrictions). The left derivative at the point τ is the probability of non-adoption and hence we wish to minimize it. This objective is simple enough to be solved explicitly. The minimizing concave function $\widehat{H} = \widehat{M}$ is the one depicted in Figure 8.

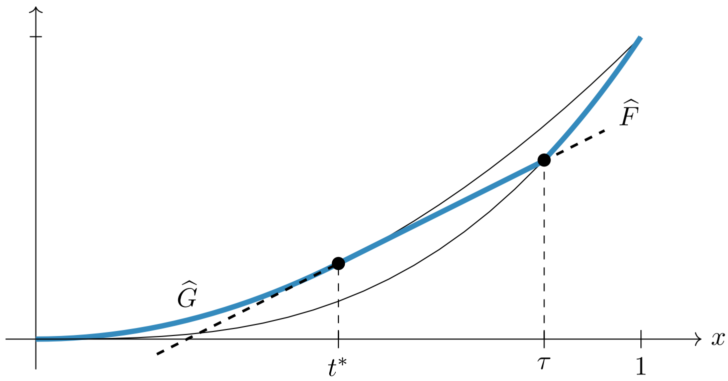


Figure 8: The optimal function $\widehat{M_\tau}$ is displayed in blue.

Proof sketch of Theorem 3 and Corollary 5 We show that the distribution of posterior means along the GEBN process for $t \in [0, 1]$ passes through all the distributions $\{M_\tau\}_{\tau \in [0,1]}$ where M_τ is the distribution whose integrated CDF is $\widehat{M_\tau}$ in Figure 8. This proves the theorem because Proposition 3 shows that the same distribution serves as an upper bound on τ-optimism.

To see that the downward-uniform garbling passes through all the M_τ distributions,[17] we trace the distribution of posteriors at every time t. We show that agent types who initially receive a signal $s \leq h(t)$ have the same posterior mean as the agent who received the signal $s = h(t)$.

[17]In fact, the downward-uniform garbling is the *unique* garbling that enjoys this property.

Additionally, the posterior mean of agent types who initially receive a signal $s > h(t)$ remains unchanged. See Figure 9. These two properties pin down the distribution to be exactly of the form M_τ.

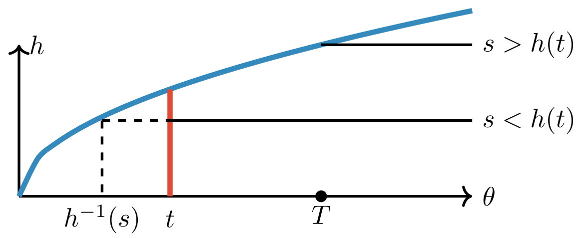


Figure 9: The top horizontal line is the belief's support of an agent whose initial posterior mean is above T. The bottom horizontal line is the belief's support of an agent whose initial posterior mean is below T, and hence at time t it is exactly T.

References

L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford university press, 2000.

I. Arieli, Y. Babichenko, R. Smorodinsky, and T. Yamashita. Optimal persuasion via bi-pooling. *Theoretical Economics*, 18(1):15–36, 2023.

I. Ashlagi, F. Monachou, and A. Nikzad. Optimal dynamic allocation: Simplicity through information design. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 101–102, 2021.

Y. Azrieli. Comment on "the law of large demand for information". *Econometrica*, 82(1):415–423, 2014.

M. Beiglböck and N. Juillet. On a problem of optimal transport under marginal martingale constraints. *Annals of Probability*, 44(1):42–106, 2016.

M. Beiglböck, P. Henry-Labordère, and N. Touzi. Monotone martingale transport plans and skorokhod embedding. *Stochastic Processes and their Applications*, 127(9):3005–3013, 2017.

D. Bergemann, T. Heumann, and S. Morris. Screening with persuasion. *arXiv preprint arXiv:2212.03360*, 2022.

D. Blackwell. Comparison of experiments. In *Proceedings of the second Berkeley symposium on mathematical statistics and probability*, volume 1, page 26, 1951.

D. Blackwell. Equivalent comparisons of experiments. *The annals of mathematical statistics*, pages 265–272, 1953.

O. Candogan and P. Strack. Optimal disclosure of information to privately informed agents. *Theoretical Economics*, 18(3):1225–1269, 2023.

M. W. Cripps, J. C. Ely, G. J. Mailath, and L. Samuelson. Common learning. *Econometrica*, 76 (4):909–933, 2008.

L. Doval and A. Smolin. Persuasion and welfare. *Journal of Political Economy*, 132(7):2451–2487, 2024.

L. E. Dubins and D. Gilat. On the distribution of maxima of martingales. *Proceedings of the American Mathematical Society*, 68(3):337–338, 1978.

P. Dworczak and A. Kolotilin. The persuasion duality. *Theoretical Economics*, 19(4):1701–1755, 2024.

P. Dworczak and G. Martini. The simple economics of optimal persuasion. *Journal of Political Economy*, 127(5):1993–2048, 2019.

M. Frick, R. Iijima, and Y. Ishii. Learning efficiency of multiagent information structures. *Journal of Political Economy*, 131(12):3377–3414, 2023.

M. Frick, R. Iijima, and Y. Ishii. Multidimensional screening with rich consumer data. *arXiv preprint arXiv:2411.06312*, 2024.

M. Gentzkow and E. Kamenica. A Rothschild-Stiglitz approach to bayesian persuasion. *American Economic Review*, 106(5):597–601, 2016.

E. M. Goggin. Convergence in distribution of conditional expectations. *The Annals of Probability*, pages 1097–1114, 1994.

G. H. Hardy, J. E. Littlewood, and G. Polya. Some simple inequalities satisfied by convex functions. *Messenger Math.*, 58:145–152, 1929.

S. Hart and P. J. Reny. Implementation of reduced form mechanisms: a simple approach and a new characterization. *Economic Theory Bulletin*, 3(1):1–8, 2015.

D. Hobson. The Skorokhod embedding problem and model-independent bounds for option prices. *Paris-Princeton Lectures on Mathematical Finance 2010*, pages 267–318, 2011.

D. Hobson and D. Norgilas. A construction of the left-curtain coupling. *Electronic Journal of Probability*, 27:1–46, 2022.

D. G. Hobson. The maximum maximum of a martingale. *Séminaire de probabilités de Strasbourg*, 32:250–263, 1998.

M. Ivanov. Optimal monotone signals in bayesian persuasion mechanisms. *Economic Theory*, 72 (3):955–1000, 2021.

E. Kamenica and M. Gentzkow. Bayesian persuasion. *American Economic Review*, 101(6):2590–2615, 2011.

H. G. Kellerer. Funktionen auf produkträumen mit vorgegebenen marginal-funktionen. *Mathematische Annalen*, 144(4):323–344, 1961.

D. Khantadze, I. Kremer, and A. Skrzypacz. Persuasion with multiple actions. *Journal of Political Economy*, 133(5):1497–1526, 2025.

A. Kleiner, B. Moldovanu, and P. Strack. Extreme points and majorization: Economic applications. *Econometrica*, 89(4):1557–1593, 2021.

A. Koh, S. Sanguanmoo, and K. Uzui. Full implementation with dynamic information. *Available at SSRN 4669928*, 2023.

A. Koh, S. Sanguanmoo, and W. Zhong. Persuasion and optimal stopping. *arXiv preprint arXiv:2406.12278*, 2024.

A. Kolotilin. Optimal information disclosure: A linear programming approach. *Theoretical Economics*, 13(2):607–635, 2018.

A. Kolotilin and A. Wolitzky. Distributions of posterior quantiles via matching. *Theoretical Economics*, 19(4):1399–1413, 2024.

A. Kolotilin, R. Corrao, and A. Wolitzky. Persuasion and matching: Optimal productive transport. *Journal of Political Economy*, 133(4):1334–1381, 2025.

M. O. Lorenz. Methods of measuring the concentration of wealth. *Publications of the American statistical association*, 9(70):209–219, 1905.

G. Moscarini and L. Smith. The law of large demand for information. *Econometrica*, 70(6): 2351–2366, 2002.

X. Mu, L. Pomatto, P. Strack, and O. Tamuz. From blackwell dominance in large samples to rényi divergences and back again. *Econometrica*, 89(1):475–506, 2021.

A. Nikzad. Constrained majorization: Applications in mechanism design. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 330–331, 2022.

R. R. Phelps. *Lectures on Choquet's theorem*. Springer, 2001.

M. Rothschild and J. E. Stiglitz. Increasing risk: I. a definition. *Journal of Economic theory*, 2(3): 225–243, 1970.

M. Saeedi and A. Shourideh. Optimal rating design under moral hazard. *arXiv preprint arXiv:2008.09529*, 2020.

C. M. Simon. A bayesian treatment of the german tank problem. *The Mathematical Intelligencer*, 46(2):117–127, 2024.

V. Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36(2):423–439, 1965.

M. Vaeth. The optimal design of public recognition schemes. 2024.

K. H. Yang and A. K. Zentefis. Monotone function intervals: Theory and applications. *American Economic Review*, 114(8):2239–2270, 2024.

A Proof of Theorem 2

The proof has two parts. First, we verify that the function h given in (2) and functions T and α are well-defined and satisfies the properties needed to define a downward-uniform signal, e.g., h is positive and monotone. Then, we effectively reverse the steps of the heuristic derivation from Section 2.3, starting with the formula and confirming that the corresponding downward-uniform induces the desired distribution of beliefs.

Recall the setting. A state θ is distributed on an open interval $\Theta = (\underline{\theta}, \overline{\theta}) \subset \mathbb{R}$ with $\underline{\theta}, \overline{\theta} \in \mathbb{R} \cup \{-\infty, +\infty\}$ according to a non-atomic distribution G. We do not assume that G has a density or that, if it has a density, that this density satisfies any regularity assumptions. Without loss of generality, $\underline{\theta}$ and $\overline{\theta}$ are the leftmost and the rightmost points of the support of G, respectively. Let F be another distribution on Θ that has a finite first moment and strictly majorizes G, i.e., $\widehat{F}(x) < \widehat{G}(x)$ for all $x \in \Theta$ and $\widehat{F}(\overline{\theta}) - \widehat{G}(\overline{\theta}) = 0$, where $\widehat{F}$ and $\widehat{G}$ denote the integrated CDFs.[18] No additional assumptions are imposed on F, in particular, it is allowed to have atoms.

Given G and F, we consider a downward-uniform signal s with a function h given by

$$h(t) = \alpha(t) \cdot \exp\left(\int_{t_0}^{t} \frac{1}{\alpha(x)}\, dG(x)\right), \tag{7}$$

where t_0 is some fixed point in Θ,

$$\alpha(t) = F(T(t)) - G(t),$$

and $T = T(t)$ is the solution to

$$\widehat{G}(t) + (T - t)G(t) = \widehat{F}(T) \tag{8}$$

with the property $T \in (t, \overline{\theta})$.

Our goal is to show that the distribution of posterior means $E[\theta|s]$ induced by the signal s equals F. The first step is showing that the signal s is well-defined, which boils down to checking that the function h is well-defined and non-decreasing.

[18] Here and below, the values of an expression at $\underline{\theta}, \overline{\theta}$ in the case of $\underline{\theta} = -\infty$ or $\underline{\theta} = +\infty$ are to be understood as the corresponding limits.

Part of the difficulty in our proof is that F and G may fail to admit densities. In that case, the functions T, α, and h, which are differentiable when F and G have densities, need not possess classical pointwise derivatives. We address this issue directly by working with distributional derivatives.[19]

A reader who prefers to assume that F and G have densities may interpret all derivatives in the classical sense and will find the proof to be an exercise in elementary analysis.

Checking that T is well-defined, continuous, and monotone We begin with verifying that, for any $t \in \Theta$, a solution T to equation (8) with the property $T \in (t, \overline{\theta})$ exists and is unique. Fix $t \in \Theta$ and consider a function $\varphi_t(T) = \widehat{F}(T) - \widehat{G}(t) - (T - t)G(t)$. Hence, $\varphi_t(T) = 0$ is equivalent to (8). Since, φ_t is continuous, we invoke the intermediate value theorem to deduce the existence of a zero. For $T = t$, we have $\varphi_t(t) < 0$ by the assumption of strict majorization. On the other hand, $\varphi_t(T) > 0$ for T in the vicinity of $\overline{\theta}$. To see that, we rewrite φ_t as follows:

$$\varphi_t(T) = \Big(\widehat{F}(T) - \widehat{G}(T)\Big) + \Big(\widehat{G}(T) - (\widehat{G}(t) + (T - t)G(t))\Big).$$

The first term converges to zero as $T \to \overline{\theta}$, by the definition of majorization. For the second term, note that $\widehat{G}$ is convex, so its graph lies above the tangent line at t, i.e., $\widehat{G}(T) \geq \widehat{G}(t) + (T - t)G(t)$ for all T. Moreover, the difference between the left-hand side and the right-hand side is non-decreasing in T for $T \geq t$. Hence, the second term is nonnegative and non-decreasing in T. If it were identically zero for all $T \geq t$, then $\widehat{G}$ would coincide with its tangent line and thus be linear on $[t, \overline{\theta})$, implying that G places no mass on this interval. This contradicts the assumption that $\overline{\theta}$ is the upper endpoint of the support of G. We conclude that $\varphi_t(T) > 0$ for T in the vicinity of $\overline{\theta}$. Since φ_t is continuous, $\varphi_t(t) < 0$, and $\varphi_t(T)$ is positive for T in the vicinity of $\overline{\theta}$, there exists $T \in (t, \overline{\theta})$ such that $\varphi_t(T) = 0$. Such T is unique since φ_t is a convex function taking both positive and negative values. Thus, a function $t \to T(t)$ is well-defined.

We now check that $T(t)$ is continuous and non-decreasing. Consider $t, t' \in \Theta$ such that $t < t'$. Denote $T = T(t)$ and $T' = T(t')$ and show $T \leq T'$. By definition, $\varphi_t(T) = 0$ and $\varphi_{t'}(T') = 0$. By the convexity of $\widehat{G}$, we get $\widehat{G}(t') \geq \widehat{G}(t) + (t' - t)G(t)$. Plugging this into the equation satisfied by T', we obtain $\varphi_t(T') \geq 0$. Since φ_t is convex and T is its only zero, we conclude that $T' \geq T$ and thus T is a non-decreasing function of[20] t. We verify continuity of $T(t)$ for all $t \in \Theta$. By monotonicity, T admits left and right limits at t which we denote by T_- and T_+, respectively. Approaching t from the left and from the right in (8) and taking into account the continuity of $\widehat{F}, \widehat{G}$ and G, we obtain that both T_- and T_+ must be solutions at the point t. Since the solution

[19]An alternative approach would be to first prove the result under the assumption that F and G have densities and all relevant maps are classically differentiable, and then extend the conclusion to arbitrary distributions by approximation, for example by applying the convergence theorem of (Goggin, 1994).

[20] In fact, the inequality $T(t') \geq T(t)$ is strict if and only if the interval (t, t') carries positive G mass, i.e., $G(t') > G(t)$. Indeed, for continuous G, this is exactly the case when $\widehat{G}(t') \geq \widehat{G}(t) + (t' - t)G(t)$ holds as a strict inequality.

is unique, $T_- = T_+ = T$, and thus T is a continuous function of[21] t.

Checking that h is well-defined, positive, and monotone Consider the function $\alpha(t) = F(T(t)) - G(t)$ from the definition (7) of the function h. First, we show positivity of α which yields positivity of h and ensures the convergence of the integral in (7).

By the convexity of $\widehat{F}$, we have $\widehat{F}(t) \geq \widehat{F}(T) + (t - T)F(T)$. Expressing $\widehat{F}(T)$ from this inequality and plugging it into (8), we obtain $F(T) - G(t) \geq \frac{\widehat{G}(t) - \widehat{F}(t)}{T - t}$. Equivalently,

$$\alpha(t) \geq \frac{\widehat{G}(t) - \widehat{F}(t)}{T(t) - t} > 0. \tag{9}$$

By strict majorization, the right-hand side is positive. Since $\widehat{G}, \widehat{F}$, and T are continuous, we conclude that α is bounded from below by a strictly positive function that is continuous on Θ. Consequently, $1/\alpha$ is also non-negative and bounded from above by a continuous function ensuring the convergence of $\int_{t_0}^{t} \frac{1}{\alpha(x)}\, dG(x)$ for $t \in \Theta$. Thus h is well-defined. Since α is positive, so is h.

We now verify that h is non-decreasing. Recall that a function is of bounded variation if it can be expressed as the difference of two non-decreasing functions. For instance, α is of bounded variation. Any such function β admits a generalized derivative—the Stieltjes derivative $d\beta$—a signed measure whose cumulative distribution function (CDF) is β. Classical derivatives are a special case: differentiable β yields $d\beta = \beta'(x)\, dx$, while jumps in β correspond to atoms of $d\beta$. Crucially, Stieltjes derivatives obey the classical rules of calculus (Ambrosio, Fusco, and Pallara, 2000): If α and β are functions of bounded variation with no common discontinuities, then the product $\alpha \cdot \beta$ is also of bounded variation, and $d(\alpha \cdot \beta) = \beta\, d\alpha + \alpha\, d\beta$. If γ is of bounded variation and f is continuously differentiable, then $\beta = f(\gamma)$ is of bounded variation and $d(f(\gamma)) = f'(\gamma)\, d\gamma$. Finally, an indefinite Stieltjes integral $\gamma(t) = \int_{t_0}^{t} q(x)\, d\lambda(x)$, for measurable q such that the integral converges absolutely, defines a function of bounded variation with $d\gamma = q\, d\lambda$.

Define $\gamma(t) = \int_{t_0}^{t} \frac{1}{\alpha(x)}\, dG(x)$. Then γ is of bounded variation as an indefinite Stieltjes integral, and so is $h = \alpha \cdot \exp(\gamma)$, since bounded variation is preserved under multiplication and smooth composition. Because G is non-atomic, γ is continuous, and thus $\exp(\gamma)$ shares no discontinuities

[21] If G admits a continuous density g, then $\varphi_t(T) = 0$ satisfies the conditions of the implicit function theorem and thus T in not only continuous but also continuously differentiable with

$$T'(t) = \frac{\big(T(t) - t\big) g(t)}{F(T(t)) - G(t)}.$$

In particular, we see that the derivative is strictly positive unless $g(t) = 0$ in agreement with Footnote 20. Note that the denominator is never zero: it equals $\alpha(t)$ and we verify that $\alpha > 0$ below.

with α. We can therefore apply the product rule and obtain

$$\begin{aligned} dh &= \exp(\gamma)\, d\alpha + \alpha\, d\exp(\gamma) \\ &= \exp(\gamma)\, d\alpha + \alpha \exp(\gamma)\, d\gamma \\ &= \exp(\gamma)\, d\alpha + \alpha \exp(\gamma) \cdot \frac{1}{\alpha}\, dG \\ &= \exp(\gamma)(d\alpha + dG) \\ &= \exp(\gamma)\, dF(T), \end{aligned}$$

where we used $d\alpha + dG = dF(T)$ by definition. Since F and T are non-decreasing, $dF(T)$ is a non-negative measure. Hence dh is non-negative, and its CDF h is non-decreasing as claimed.

Expressing α through h We now derive an expression for α in terms of h. Rewriting (7), we obtain

$$\frac{\alpha(t)}{h(t)} = \exp\left(-\int_{t_0}^{t} \frac{1}{\alpha(x)}\, dG(x)\right).$$

The right-hand side is of bounded variation, as is its logarithm, so both $\frac{\alpha}{h}$ and $\ln \frac{\alpha}{h}$ inherit this property. Taking logarithms and computing the Stieltjes derivative yields

$$d\left(\ln \frac{\alpha}{h}\right) = -\frac{1}{\alpha}\, dG.$$

By the chain rule for Stieltjes derivatives,

$$d\left(\ln \frac{\alpha}{h}\right) = \frac{1}{\alpha/h}\, d\left(\frac{\alpha}{h}\right),$$

and hence, multiplying both sides by $\frac{\alpha}{h}$,

$$d\left(\frac{\alpha}{h}\right) = -\frac{1}{h}\, dG.$$

We integrate from t_0 and get

$$\frac{\alpha(t)}{h(t)} - \frac{\alpha(t_0)}{h(t_0)} = -\int_{t_0}^{t} \frac{1}{h(x)}\, dG(x).$$

Using the identity $h(t_0) = \alpha(t_0)$, we simplify

$$\alpha(t) = h(t)\left(1 - \int_{t_0}^{t} \frac{1}{h(x)}\, dG(x)\right). \tag{10}$$

This representation admits an alternative form

$$\alpha(t) = h(t) \cdot \int_{t}^{\overline{\theta}} \frac{1}{h(x)}\, dG(x), \tag{11}$$

since that $\alpha(t) \to 0$ as $t \to \overline{\theta}$. Indeed, $\overline{\theta}$ is the highest point of the support of G. Since G is continuous, we have $G(t) \to 1$ as $t \to \overline{\theta}$. Moreover, because F majorizes G, the distribution F cannot have an atom at $\overline{\theta}$, and thus $F(T(t)) \to 1$ as $T(t)$ is squeezed between t and $\overline{\theta}$. Consequently,

$\alpha(t) = F(T(t)) - G(t) \to 0$. Given that h is strictly positive and increasing, this asymptotic behavior is only compatible with (10) if

$$\int_{t_0}^{\bar{t}} \frac{1}{h(x)}\, dG(x) = 1,$$

which[22] leads to (11).

Computing the induced distribution of posterior means We now show that the downward-uniform signal defined by the function h induces the distribution of posterior means F. The argument proceeds in two steps. First, we establish that the posterior mean corresponding to a signal realization s is equal to $T(h^{-1}(s))$. Second, we use this identity to prove that the unconditional distribution of posterior means is indeed F.

To compute the posterior mean induced by a realization s, observe that the conditional density of the signal given the state θ is $\frac{1}{h(\theta)}\mathbb{1}_{\{h(\theta)\geq s\}}$, and θ is distributed according to G. Therefore, the pair (θ, s) is distributed according to $\frac{1}{h(\theta)}\mathbb{1}_{\{h(\theta)\geq s\}}\, dG(\theta)\, ds$. The posterior distribution of θ given s is thus proportional to $\frac{1}{h(\theta)}\mathbb{1}_{\{h(\theta)\geq s\}}\, dG(\theta)$, so the posterior mean is

$$E[\theta|s] = \frac{\int_{\{\theta:\, h(\theta)\geq s\}} \frac{\theta}{h(\theta)}\, dG(\theta)}{\int_{\{\theta:\, h(\theta)\geq s\}} \frac{1}{h(\theta)}\, dG(\theta)}. \tag{12}$$

We now establish the identity relating the posterior mean and the function T

$$E[\theta|s] = T(h^{-1}(s)), \tag{13}$$

where $h^{-1}(s)$ is the generalized inverse defined as $\min\{\theta\colon h(\theta) \geq s\}$. The minimum is attained by the right-continuity of h. Comparing (12) and (13), we see that it is enough to prove the identity

$$T(t) = \frac{\int_t^{\overline{\theta}} \frac{\theta}{h(\theta)}\, dG(\theta)}{\int_t^{\overline{\theta}} \frac{1}{h(\theta)}\, dG(\theta)}, \tag{14}$$

which expresses T, defined in (8), in terms of G and h. To derive this, we differentiate both sides of (8). Since T is continuous and non-decreasing, both sides admit Stieltjes derivatives

$$G(t)\, dt + (T(t) - t)\, dG(t) + G(T)\, dT(t) - G(t)\, dt = F(T)\, dT(t).$$

Canceling $G(t)\, dt$ and using $\alpha(t) = F(T(t)) - G(t)$, we get

$$\alpha(t)\, dT(t) - T(t)\, dG(t) = -t\, dG(t).$$

Expressing α using (6) and dividing both sides by $h(t)$ gives

$$\left(\int_t^{\overline{\theta}} \frac{1}{h(\theta)}\, dG(\theta)\right) dT(t) + T(t)\left(\frac{1}{h(t)}\, dG(t)\right) = -\frac{t}{h(t)}\, dG(t).$$

[22] This identity may be of independent interest, as it constrains the growth of h.

The left-hand side is the Stieltjes derivative of the product

$$d\left(T(t) \cdot \int_t^{\overline{\theta}} \frac{1}{h(\theta)}\, dG(\theta)\right) = -\frac{t}{h(t)}\, dG(t).$$

Integrating both sides and rearranging terms yields

$$T(t) = \frac{\int_t^{\overline{\theta}} \frac{\theta}{h(\theta)}\, dG(\theta) + C}{\int_t^{\overline{\theta}} \frac{1}{h(\theta)}\, dG(\theta)},$$

where C is a constant independent of t. Letting $t \to \overline{\theta}$, we conclude that $C = 0$ thus confirming identity (13).

This identity lets us express the distribution of posterior means through the distribution of signals. Let $K(s)$ be the CDF of signals

$$K(s) = \int_\Theta P(s' \le s \mid \theta)\, dG(\theta) = \int_\Theta \frac{\min(s, h(\theta))}{h(\theta)}\, dG(\theta).$$

Setting $s = h(t)$ and using the fact that h is non-decreasing, we obtain

$$K(h(t)) = \int_{\{\theta<t\}} dG(\theta) + h(t) \int_{\{\theta \ge t\}} \frac{dG(\theta)}{h(\theta)} = G(t) + h(t) \int_t^{\overline{\theta}} \frac{dG(\theta)}{h(\theta)}.$$

From identity (11), the second term equals $\alpha(t)$. By the definition $\alpha(t) = F(T(t)) - G(t)$, we conclude

$$K(h(t)) = F(T(t)). \tag{15}$$

Finally, let $M(s) = E[\theta \mid s]$ be the posterior mean induced by the signal s, and let F_M denote its CDF. We now show that $F_M = F$. We fix $z \in \Theta$ and demonstrate that $F_M(z) = F(z)$. Choose the maximal t in the closure of Θ such that $T(t) = z$; such a t exists because T is continuous and maps Θ onto itself. We compute:

$$\begin{aligned}
F_M(z) &= F_M(T(t)) \\
&= P\big(M(s) \le T(t)\big) \\
&= P\big(T(h^{-1}(s)) \le T(t)\big) \quad \text{by (13)} \\
&= P\big(h^{-1}(s) \le t\big) \quad \text{by maximality of } t \\
&= P\big(s \le h(t)\big) \quad \text{since } h^{-1} \text{ is the minimal inverse} \\
&= K\big(h(t)\big) \\
&= F\big(T(t)\big) = F(z) \quad \text{by (15)}.
\end{aligned}$$

Therefore the distribution of posterior means induced by the downward-uniform signal s with bounding function h given in (7) coincides with the target distribution F, completing the proof.

B Omitted Proofs from Section 3

Proof of Proposition 1. We establish the proof by identifying a one-dimensional sufficient statistic for the collection of signals and showing it is distributionally equivalent to the sufficient statistic of the asserted single signal.

For a non-decreasing right-continuous function $h : \Theta \to \mathbb{R}_+$ we define its generalized inverse $h^{-1} : \mathbb{R}_+ \to \Theta$ as the quantile function $h^{-1}(y) = \min\{\theta \in [\underline{\theta}, \overline{\theta}] \colon h(\theta) \geq y\}$. This construction ensures that the inverse is well-defined and non-decreasing. Crucially, for any signal realization y, the two events $\{y \leq h(\theta)\}$ and $h^{-1}(y) \leq \theta$ coincide.

We now consider a collection of conditionally independent downward-uniform signals $s = (s_1, \ldots, s_k)$ with bounding functions $h_1, \ldots, h_k$. Without loss of generality, we can assume that these functions are right-continuous. Indeed, replacing a non-decreasing function h with its right-continuous version $h_+(t) = \lim_{\varepsilon \to 0^+} h(t)$ requires changing it at at most countable number of points. Since the distribution of the state is non-atomic, such a change corresponds to a zero measure of states and thus results in an equivalent signal.

The likelihood of observing s given θ is:

$$L(s|\theta) = \prod_{i=1}^{k} \frac{1}{h_i(\theta)} \mathbb{1}_{\{s_i \leq h_i(\theta)\}}.$$

Rewriting each indicator using the generalized inverse and denoting $h(\theta) = \prod_i h_i(\theta)$, we obtain

$$L(s|\theta) = \frac{1}{h(\theta)} \prod_{i=1}^{k} \mathbb{1}_{\{h_i^{-1}(s_i) \leq \theta\}} = \frac{1}{h(\theta)} \mathbb{1}_{\{\max_i h_i^{-1}(s_i) \leq \theta\}}.$$

The likelihood depends on the signal vector s only through the statistic $m(s) = \max_i h_i^{-1}(s_i)$, which is therefore sufficient for θ.

Next, we show that the state-conditional distribution of $m(s)$ is identical to that of a sufficient statistic from a single downward-uniform signal with bounding function $h(\theta) = \prod_i h_i(\theta)$. We compute the CDF of $m(s)$ conditional on θ. For any $t \in \Theta$:

$$\begin{aligned} F_{m(s)|\theta}(t) = P(m(s) \leq t|\theta) &= P\left(\max_i h_i^{-1}(s_i) \leq t \middle| \theta\right) \\ &= \prod_{i=1}^{k} P(h_i^{-1}(s_i) \leq t|\theta), \end{aligned}$$

where the last identity holds by conditional independence. The event $h_i^{-1}(s_i) \leq t$ is identical to $s_i \leq h_i(t)$. Since s_i is uniformly distributed on $[0, h_i(\theta)]$ conditional on θ, the probability of this event is $\frac{h_i(t)}{h_i(\theta)}$. Thus, the CDF of the sufficient statistic is:

$$F_{m(s)|\theta}(t) = \prod_{i=1}^{k} \frac{h_i(t)}{h_i(\theta)} = \frac{h(t)}{h(\theta)}.$$

Now, consider a single downward-uniform signal s' with bounding function $h(\theta)$. Its sufficient statistic is $m'(s') = h^{-1}(s)$, where h^{-1} is the generalized inverse of h. The conditional CDF of this statistic is:

$$F_{m'(s)|\theta}(t) = P(h^{-1}(s') \leq t|\theta) = P(s' \leq h(t)|\theta) = \frac{h(t)}{h(\theta)}.$$

Since the sufficient statistics $m(s)$ and $m'(s')$ have identical conditional distributions for any θ, the two information structures induce the same distribution of posterior beliefs and are therefore equivalent. □

Proof of Proposition 2. Dubins and Gilat (1978) formulate and prove a particular case of Proposition 2 for a Dirac measure $F = \delta_x$. We now show how to derive the general case from their result.

Let X be the GEBN learning process and let $M_x^\tau = P[\max_{t\in[0,1]} X_t \geq \tau | X_0 = x]$ for every x and τ. Let $Y = (Y_t)_{t\in[0,1]}$ be any other learning process with law Q such that $Y_0 \sim F$ and $Y_1 \sim G$ and the conditional distributions $Q(Y_1 \in B|Y_0 = x) = P(X_1 \in B|X_0 = x)$ for any Borel subset $B \subseteq \mathbb{R}$ and almost every x with respect to F. By Dubins and Gilat (1978),

$$Q[\max_{t\in[0,1]} Y_t \geq \tau | Y_0 = x] \leq M_x^\tau$$

for every x and for every τ. Therefore, for every τ,

$$\mathrm{Opt}_\tau(Y) = \int Q[\max_{t\in[0,1]} Y_t \geq \tau | Y_0 = x] dF(x) \leq \int M_x^\tau dF(x) = \mathrm{Opt}_\tau(X),$$

as desired. □

Proof of Proposition 3. We denote by $\mathcal{D}_\tau$ the dynamic persuasion problem in which the sender reveals information to a receiver who immediately adopts once her posterior exceeds τ. The martingale, which captures receiver's posterior over time is restricted to have initial distribution F and terminal distribution G. We denote by $\mathcal{S}_\tau$ the static persuasion problem in which the sender reveals information to a receiver and is restricted to reveal information that is less informative than G but more informative than F; Namely she can reveal any H such that $F \leq_m H \leq_m G$. Proposition 3 essentially states that $\mathrm{val}(\mathcal{D}_\tau) = 1 - G(T^{-1}(\tau))$.

Notice that $\mathrm{val}(\mathcal{D}_\tau) \leq \mathrm{val}(\mathcal{S}_\tau)$ because a possible signaling policy in the static interaction is the one in which the sender draws a realization of the martingale and reports whether at some time t it exceeded τ. Such a policy is persuasive because the martingale condition ensures that once τ has been reached, the expectation at time $t = 1$ will be τ. Namely, every policy in the dynamic setting has a corresponding signaling policy in the static setting with the same value for the sender.

For a CDF H we denote $H_-(x) = \lim_{\epsilon\to 0, \epsilon>0} H(x - \epsilon)$ the CDF that excludes the atom on x (if such exists). From the left-differentiable function $\widehat{H}(x) = \int_0^x H(y)dy$ we denote its left derivative by

$$\widehat{H}'_-(x) = \lim \epsilon \to 0, \epsilon > 0 \frac{\widehat{H}(x) - \widehat{H}(x-\epsilon)}{\epsilon}$$

and we notice that $\widehat{H}'_-(x) = H_-(x)$. We denote by $\mathcal{H}$ the set of all convex functions $\widehat{H}(x)$ with left derivative bounded by $0 \leq \widehat{H}'_-(x) \leq 1$ that are sandwiched in between $\widehat{F}(x) \leq \widehat{H}(x) \leq \widehat{G}(x)$

for all $x \in [0,1]$. Now $\text{val}(\mathcal{S}_\tau)$ can be elegantly written in the domain of the integrated CDFs.

$$\text{val}(\mathcal{S}_\tau) = \max_{H: F \leq_m H \leq_m G} (1 - H_-(\tau)) = \max_{\widehat{H} \in \mathcal{H}} (1 - \widehat{H}'_-(\tau)) = 1 - \min_{\widehat{H} \in \mathcal{H}} \widehat{H}'_-(\tau) \tag{16}$$

The $\widehat{M_\tau} \in \mathcal{H}$ that minimizes the left derivative at the point τ is the one depicted in Figure 8. Namely, we draw a tangent to the function $\widehat{G}$ from the point $(\tau, \widehat{F}(\tau))$ and denote the tangent point by t^*. Now $\widehat{M_\tau}$ equals $\widehat{G}$ in the interval $[0, t^*]$, it equals the tangent in the interval $[t^*, \tau]$, and it equals $\widehat{F}$ in the interval $[\tau, 1]$.

The argument for $\widehat{M_\tau}$ being the minimizer of the left derivative follows from two arguments. First, for every $y \in [\widehat{F}(\tau), \widehat{G}(\tau)]$ the solution for the minimization problem

$$\min_{\widehat{H} \in \mathcal{H}, \widehat{H}(\tau) = y} \widehat{H}'_-(\tau)$$

is obtained by drawing a tangent of the function $\widehat{G}$ from the point (τ, y) with the tangent point t and letting $\widehat{H}$ being equal to the tangent in the interval $[t, \tau]$. Indeed any function $\widehat{H} \leq \widehat{G}$ with a lower left derivative will violate convexity. Second, the slope of the tangent to $\widehat{G}$ from (τ, y) (i.e., the left derivative at τ) is monotonically increasing in y. Therefore, the minimizing choice is $y = \widehat{F}(\tau)$ which exactly yields the minimizer $\widehat{M_\tau}$.

Notice that the minimal value for the minimizer $\widehat{M_\tau}$ is $\widehat{M}'_{\tau -}(\tau) = \widehat{G}'(t^*) = G(t^*) = G(T^{-1}(\tau))$ and hence we conclude by Equation (16) that

$$\text{val}(\mathcal{D}_\tau) \leq \text{val}(\mathcal{S}_\tau) = 1 - \widehat{M}'_-(\tau) = 1 - G(T^{-1}(\tau)).$$

To see that $1 - G(T^{-1}(\tau))$ is achievable, we observe that $\widehat{F} \leq \widehat{M_\tau} \leq \widehat{G}$. So, by Blackwell's theorem, there exists a two-step martingale (say times $t = 0, \frac{1}{2}, 1$) that consists of spreading F to M_τ at time $t = \frac{1}{2}$ and spreading M_τ to G at time $t = 1$. The maximal optimism will be achieved at time $t = \frac{1}{2}$. □

Proof of Theorem 3. Denote by $Y \in \Delta([0,1])$ the belief of a receiver who gets the signal $s = 0$ about the state $\theta \in [0,1]$ in the downward-uniform garbling. For a threshold τ, let $t \in [0,1]$ be the unique value for which $E[Y|Y \geq t] = \tau = T$. We argue that at time t of the GEBN process the distribution of posteriors means of the receiver is precisely M_τ from the proof of Proposition 3. This will conclude the proof.

If the state has not been revealed yet and the receiver's signal is $s \leq h(t)$ then her posterior mean is $E[Y|Y \geq t] = T$. Saying it differently, if the state has not been revealed yet and the receiver's posterior mean at time $t = 0$ was (weakly) below T, then her current posterior mean is T. See Figure 9.

If the state has not been revealed yet and the receiver's signal is $s > h(t)$ then her posterior remains unchanged because she initially knew that $\theta > t$ with probability 1. Saying it differently, if the state has not been revealed yet and the receiver's posterior mean at time $t = 0$ exceeded T, it remains unchanged. See Figure 9.

Finally, if the state has been revealed the receiver adopts this state as her posterior mean.

Therefore at every point of time t the CDF of the distribution of posterior means denoted by H_t has the following properties:

- $H_t(x) = G(x)$ for every $x \in [0, t]$; this corresponds to the event of revealing the state.
- $H_t(x) = F(x)$ for every $x \in [T, 1]$; this corresponds to the population of receivers whose signal is $s > h(t)$ in the event of not revealing the state.
- All the remaining mass of H_t is concentrated on an atom on T; this corresponds to the population of receivers whose signal is $s \leq h(t)$ in the event of not revealing the state.

Expressing these three properties in the integrated CDF space implies $\widehat{H}_t(x) = \widehat{G}(x)$ for every $x \in [0, t]$, $\widehat{H}_t(x) = \widehat{F}(x)$ for every $x \in [T, 1]$, and $\widehat{H}_t(x)$ is linear in (t, T). Notice that these three properties pin down uniquely the function H_t and this function is exactly M_τ for $T = \tau$; see Figure 8. □