BEST GARBLING IS NO GARBLING: PERSUASION IN REAL TIME

CAN URGUN AND MARK WHITMEYER

Abstract. We study continuous-time persuasion where a sender controls both how informative a signal is over time and when to stop providing information to a receiver. Given an exogenous signal process, the sender can both garble the evolving signal path and delay the receiver's decision at a convex, increasing cost of time. We show that, although both instruments are available, any optimal persuasion scheme is *fully transparent*: the sender keeps the signal fully informative and persuades solely by choosing when to stop.

1. Introduction

Many environments–drug approvals, loan underwriting, jury verdicts, hiring and tenure votes, content moderation, even "buy/not-buy" choices–share the same essential features: one agent assembles and releases evidence to sway another agent into taking a favorable binary decision, which takes a cut-off form. For instance, a regulator or internal evaluator approves a drug being tested if and only if her belief that the drug works exceeds a cutoff; a capital-provider funds a loan if and only if default risk falls below a hurdle; and a median juror votes to convict if and only if the guilt probability passes a "reasonable doubt" or "preponderance of evidence" threshold.

The canonical persuasion problem [Kamenica and Gentzkow, 2011] captures this logic in one sentence: a sender, facing a binary state and a receiver with a threshold rule for a binary action, commits to an information structure to maximize the chance the receiver's posterior exceeds the threshold. What the classic model leaves out, and

---

Date: February 2026.
CU: UNC Chapel Hill, curgun@unc.edu & MW: ASU, mark.whitmeyer@gmail.com. We thank Yunus Aybas, Matteo Camboni, Duarte Gonçalves, Fei Li, Udayan Vaidya, Leeat Yariv, and a workshop audience at UNC/Duke/NC State for their comments.

what real cases make unavoidable, is that moving beliefs takes time: evidence must be generated, filtered, and digested, and delay is costly.

In our setting, we keep this threshold-persuasion objective but make the information flow and the stopping time explicit–signals arrive over time, the receiver updates continuously and acts at the first crossing of the threshold, and the sender pays a convex cost of delay–so the design problem becomes: how should a sender shape and pace what is revealed to persuade a receiver *at the lowest (expected) cost*?

More concretely, the sender and the receiver share a common prior about an unknown binary state that drives a Brownian process X in continuous time. This process is not observed directly, but instead through a sender-chosen information structure that garbles X. The sender chooses both the informativeness of the real-time disclosure and a stopping time at which the receiver acts, incurring a convex increasing cost of delay. Upon stopping, the receiver takes a binary action, choosing the sender-preferred action if and only if her belief exceeds a threshold.

Our main result is stark: the "best garbling is no garbling." That is, our main result, Theorem 1 reveals that in any optimal solution to the sender's persuasion problem, she does not garble the exogenous process at all and influences the receiver's behavior through the choice of stopping time alone.[1] The basic intuition for this is that garbling has a net *dampening* effect on the receiver's belief process, which is suboptimal given the convex time cost.

To understand this dampening effect, consider briefly a simple one-shot belief-updating problem in which a normally-distributed state Θ is garbled through independent normal noise ε, parameterized by ε's variance, r. Crucially, when the receiver observes $Y = \Theta + \varepsilon$, the coefficient mapping Y to the posterior mean, $\frac{\text{Var}(\Theta|Y)}{r}$–the responsiveness of the posterior to what the receiver sees–is strictly decreasing in r. That is, the posterior *reacts less to the data.* At the same time, the variance of the distribution of posterior means is also strictly decreasing in r, thus, reducing the

[1]In fact, Theorem 1 reveals that no garbling is the fastest (thus, best) way of embedding any target distribution.

probability that the posterior mean crosses a threshold.[2] Our continuous-time model is the infinitesimal version of this: garbling lowers the posterior's instantaneous gain, slows the belief clock, and thereby raises the expected convex cost of implementing any fixed terminal law.

Returning to our continuous-time setting, we start by establishing an important reduction in Proposition 1. Namely, we show that among all garbling policies that implement some specified distribution over posteriors, it is (weakly) optimal for the principal to choose a garbling policy that depends only on the current posterior. Intuitively, if the designer sometimes runs the clock fast and sometimes slow at the same belief level, she is only adding unnecessary randomness to the stopping time. Averaging this into a single speed at each belief preserves the terminal distribution while trimming the upper tail of stopping times–exactly what any convex delay cost rewards.

Second, once we restrict attention to these belief-based "learning speeds," the key observation is that garbling can only slow belief movements. In the one-shot Gaussian toy model, increasing the noise variance r reduces the responsiveness of the posterior, and the same mechanism operates in our main setting. Although garbling makes the observed path noisier, Bayes' rule discounts that additional noise, so on the whole, the receiver's posterior becomes less volatile: the variance of belief changes per unit time falls, and the belief clock ticks more slowly. As a result, any target distribution of terminal beliefs that can be implemented under garbling can also be implemented under full transparency by an appropriate stopping rule, and it is reached in weakly less calendar time (strictly less if the garbling is nontrivial). With increasing convex delay costs, faster implementation is always better, which delivers our theorem.

Our main finding is perhaps surprising, especially in the context of the other papers in the large persuasion literature. We establish the *robust optimality of transparency:* for any convex and increasing cost, no garbling is optimal. Thus, a company seeking

[2]In particular for any threshold $a, \mathbb{P}\left(\hat{\Theta} > a\right) = 1 - \Phi\left(\frac{a}{s}\sqrt{\frac{r}{\mathrm{Var}(\Theta|Y)}}\right)$, is decreasing in r, where Φ is the normal CDF, and s is the standard deviation of Θ.

to have its drug approved wants to be completely forthright about its research process and only affect approval through its decision on when to stop accumulating evidence. Likewise, a party wishing to have a loan underwritten wants to fully disclose all of the evidence it accumulates, a start-up pursuing funding ought to reveal both the ups *and* downs in its development progress, and a firm conducting a product demonstration for a potential customer should be fully transparent (but stop at the opportune time). Kamenica and Gentzkow [2011]'s prosecutor should fully reveal the evidence he obtains over time.

We also conduct comparative statics. If delay becomes costlier–for two cost functions c_1 and c_2 the difference $c_1 - c_2$ is increasing and convex–then the sender becomes "less persuasive." That is, she chooses to induce a distribution placing a *smaller* mass on the target threshold. Intuitively, experimentation is more costly, so she gives up sooner: the lower quitting threshold increases, making failure more likely. $|\mu_H - \mu_L|/\sigma$ is the signal-to-noise ratio, which benchmarks the speed of persuasion. We show that as this decreases, the sender becomes less persuasive. A slower clock makes every target more expensive, so the optimal policy is quicker to give up.

Later on in the paper, we show that the robust optimality of transparency persists in two extensions. In our main setting, the sender commits *ex ante* both to a garbling process and a stopping time and the receiver's role is merely to take an action upon the sender's stopping of the process. In our first extension, we relax this *ex ante* commitment by the sender and require instead that her instantaneous choice of garbling and her stopping decision are sequentially rational. In Proposition 5, we reveal that this comparative lack of commitment is inconsequential: the sender's *ex ante*-optimal strategy is *interim*-optimal as well.

In our second extension, we allow the receiver to also control the stopping of the process–so that the stopping time is the minimum of his and the sender's–and also allow for the running of the process to impose a constant flow cost on the receiver. Nevertheless, our main insight persists in this setting: in any equilibrium, we show that the sender does not garble (Proposition 6).

1.1. **Roadmap.** We complete our introduction by discussing related work in §1.2. §2 introduces the model, and §3 collects a number of preliminary results. §4 contains our main results. §5 and §6 deliver the aforementioned comparative statics and robustness exercises (respectively). All proofs omitted from the main text as well as some miscellaneous technical details and notation are relegated to Appendix A. Appendix B contains supplementary material, including formulæ for the expected costs of optimal distributions for a broad class of costs.

1.2. **Related Work.** A growing literature studies dynamic persuasion and information design. Relative to this work and the static benchmark [Kamenica and Gentzkow, 2011], our contribution is to make the tempo of information part of the designer's choice. Technically, we bring Brownian methods into persuasion by allowing the sender to choose both how transparent to be about a continuous-time diffusion and when to stop.

Several papers study dynamic persuasion with costly or gradual information, but treat time differently. Che et al. [2023] analyze dynamic persuasion when information takes time to generate and neither player can commit. As persuasion costs vanish, Markov equilibria approximate both the Kamenica and Gentzkow [2011] optimum and full revelation, but the sender has no separate stopping instrument. Escudé and Sinander [2023] study "slow persuasion" under a gradual-accumulation constraint and characterize when the sender benefits from slowing down learning.[3]

Our continuous-time (Brownian) setting is related to Aybas and Callander [2024], who use Brownian motion to represent complexity and study when one-shot cheap talk can be efficient, and to work comparing cheap talk and persuasion, such as Kamenica and Lin [2024]. On the static side, papers that exploit the geometric structure of persuasion problems include Bardhi and Guo [2018] and Guo and Shmaya [2019].

Closest in spirit are models where influence comes from controlling the continuation of an evidence stream. Brocas and Carrillo [2007] study a discrete-time setting

[3]Other recent work combines dynamic disclosure with additional strategic forces like multiple audiences or incentive provision. See, e.g., Li et al. [2025] and Ely et al. [2025].

in which a sender gains influence by deciding when to stop generating public evidence ("influence through ignorance"). Henry and Ottaviani [2019] explore a continuous-time approval environment in which an informer pays a flow cost to run a fixed Brownian experiment and an evaluator decides whether to approve or reject. Starting from Wald's sequential test [Wald, 1945], they analyze how organizational form distorts stopping. Our paper tackles a different design problem: rather than fixing the experiment and varying who controls stopping, we take the continuous-time information technology itself as an object of design, allowing the sender to choose transparency as well as when to stop.[4]

Another strand combines persuasion with search and outside information sources. Bizzotto et al. [2021] study dynamic persuasion when the receiver also observes outside news and can stop listening, while Mekonnen et al. [2025] and Mekonnen and Pakzad-Hurson [2025] analyze information brokers who design signals for a searching agent.

On the information-acquisition side, Zhong [2022] and Chen and Zhong [2025] study single-agent dynamic problems in which a decision maker chooses a signal process subject to constraints or costs and trades off delay against accuracy, while Bloedel and Zhong [2020] characterize which posterior-separable reduced-form costs arise from optimal sequential sampling. Morris and Strack [2019] show that, in the two-state Wald problem, any distribution of posteriors can be implemented by a stopping rule and the induced ex ante cost is posterior-separable. We use a similar embedding perspective, but focus on the sender's dynamic persuasion problem rather than on foundations of information costs.[5]

2. Setup

There are two agents, a sender and a receiver. There is an unknown state, real number $\mu \in \{\mu_h, \mu_l\}$, with $\mu_h > \mu_l$, about which the sender and receiver share a common prior $p_0 \coloneqq \mathbb{P}(\mu_h)$. The unknown state is the drift of a Brownian motion, the

[4] McClellan [2022] pushes the Henry and Ottaviani agenda further.

[5] For related continuous-time sequential-sampling models of response times and learning in rich environments, see, e.g., Fudenberg et al. [2018], Gonçalves [2023, 2024], Barilla [2025], Georgiadis-Harris [2023].

fundamental process X, which evolves according to

$$dX_t = \mu dt + \sigma dB_t,$$

where $\sigma > 0$ is known and $B = (B_t)_{t\geq 0}$ is a standard Brownian motion on a filtered probability space $\left(\Omega, \mathcal{F}, (\mathcal{F}_t^X)_{t\geq 0}, \mathbb{P}\right)$, with $\mathcal{F}_t^X = \sigma\left(X_s \colon s \leq t\right)$.

The sender observes the fundamental process X in real time. The receiver does not observe X directly. Instead, the sender commits at time 0 to an *information policy* consisting of

- an $(\mathcal{F}_t^X)$-adapted continuous signal process $Y = (Y_t)_{t\geq 0}$ constructed from X, and
- a stopping time τ with respect to the filtration generated by Y, $\mathcal{F}_t^Y = \sigma(Y_s \colon s \leq t)$.

We require the observation process Y to be a continuous semimartingale driven by X. Without loss of generality, we can represent it as

$$dY_t = a_t dt + b_t dX_t,$$

for $\mathcal{F}^X$-progressively measurable processes a_t and b_t satisfying $\int_0^T (|a_t| + b_t^2) dt < \infty$ a.s. and $\sup_{t\geq 0} |b_t| < \infty$ a.s. The process Y is the only object the receiver observes: her information at time t is $\mathcal{F}_t^Y$, and $\hat{p}_t$ denotes her posterior belief given this filtration.

At the stopping time τ that marks the end of transmission,[6] the receiver chooses a binary action $a \in \{0, 1\}$. Action $a = 1$ is the sender's desired action ("approve the drug," "grant the loan," etc.). The receiver's preferences over actions and states imply a cutoff rule in terms of her posterior belief about the high state: there exists a threshold $\bar{P} \in (0, 1)$ such that, whenever the receiver is called to act at time t, she chooses action 1 if and only if $\hat{p}_t \geq \bar{P}$, where $\hat{p}_t = \mathbb{P}\left(\mu_h \mid \mathcal{F}_t^Y\right)$ denotes her posterior belief given the information she observes.

At the stopping time τ prescribed by the sender's policy, the interaction ends and the receiver takes the action implied by her cutoff rule applied to $\hat{p}_\tau$. It is convenient for our analysis to view the sender as choosing both the information process Y and the stopping time τ in advance.

[6]The sender can "pause" transmission by setting $a_t = 0$ and $b_t = 0$ on some time intervals, which keeps Y constant without ending the interaction. Accordingly, the stopping time τ is best thought of as the (sender-chosen, possibly random) time at which she permanently ends transmission.

The sender's benefit is 1 if the receiver ultimately takes action 1 and 0 otherwise. In addition, if the sender stops transmitting at stopping time τ, the sender incurs a delay cost $c(\tau)$. We specify that this cost is given by a function $c\colon [0,\infty) \to \mathbb{R}_+$ that is i) continuous, ii) convex and increasing, iii) finite for all $t < \infty$, and iv) normalized so that $c(0) = 0$. We further assume that v) c is Laplace transformable with minimal loss of generality for easier exposition.[7] We restrict attention to policies (Y, τ) for which $\mathbb{E}[c(\tau)] < \infty$.

2.1. **The sender's problem.** Given an information policy (Y, τ), the receiver forms the posterior process $\hat{p}_t = \mathbb{P}\left(\mu_h \mid \mathcal{F}_t^Y\right)$ and, at the stopping time τ, takes the desired action if and only if $\hat{p}_\tau \geq \bar{P}$. The sender solves

$$\sup_{(Y,\tau)} \mathbb{E}\left[\mathbf{1}\{\hat{p}_\tau \geq \bar{P}\} - c(\tau)\right],$$

where the expectation is taken under the common prior and the law of the processes defined above and the supremum is taken over all $(\mathcal{F}_t^X)$-adapted continuous semi-martingales Y of the form above and all $\mathcal{F}^Y$-stopping times τ with $\mathbb{E}[c(\tau)] < \infty$.

In the next section we show that we can simplify the problem by working directly with the posterior belief process and by restricting attention to posterior diffusions whose volatility depends only on the current belief and time.

3. Preliminary Results

In this section we show how the sender's problem can be simplified. Recall that the common prior is p_0 and given the state μ, the fundamental process X evolves according to $dX_t = \mu dt + \sigma dB_t$. The sender observes X and at each time t forms the posterior probability that the state is high, $p_t \coloneqq \mathbb{P}(\mu = \mu_h \mid \mathcal{F}_t^X)$. This posterior process is a continuous martingale taking values in $[0, 1]$ and in Appendix A.1, we

[7] Notably in Appendix B.3 we show that the Laplace-transformable functions are dense in the space of convex increasing and differentiable functions that do not have an asymptote in finite time, see Lemma 10. Laplace transforms allow for an ease of exposition in Proposition 7, and all our results continue to hold for all convex functions when we directly use the stopping-time distribution as opposed to its Laplace transform.

derive the standard result that p_t evolves as

$$dp_t = \frac{\mu_h - \mu_l}{\sigma} p_t(1 - p_t)dW_t,$$

for some $(\mathcal{F}_t^X)$-Brownian motion W.

Under a signal process Y, the receiver's information at time t is $\mathcal{F}_t^Y = \sigma(Y_s \colon s \leq t)$. Rather than work directly with the most general representation $dY_t = a_t dt + b_t dX_t$, it is convenient–and without loss of generality, by an argument we provide in Appendix A.2–to restrict attention to signal processes that are generated by adding noise directly to the posterior process p_t. Specifically, we may assume that the sender chooses a predictable $\mathcal{F}^X$-adapted process $g(t, p_t)$ with $|g(t, p_t)| \leq M$ almost surely for some finite constant M[8] and the receiver observes the process (with abuse of notation) $Y = (Y_t)_{t \geq 0}$ given by

$$dY_t = p_t dt + g(t, p_t) dB_t^Y, \tag{1}$$

where B^Y is a standard Brownian motion that is independent of the Brownian motion B driving the fundamental.

Intuitively, the sender first computes her posterior belief p_t and then reports a noisy version of this belief to the receiver. The receiver's time-t posterior belief about the high state is, therefore,

$$\hat{p}_t \coloneqq \mathbb{P}(\mu_h \mid \mathcal{F}_t^Y) = \mathbb{E}\left[p_t \mid \mathcal{F}_t^Y\right].$$

Furthermore, the process $(\hat{p}_t)_{t \geq 0}$ is a continuous $\mathcal{F}^Y$-martingale with $\hat{p}_0 = p_0$. Standard filtering results (see Appendix A.3 for details) imply that there exists an $\mathcal{F}^Y$-Brownian motion B^I such that–letting v_t denote the conditional variance of the sender's posterior given the receiver's information $v_t \coloneqq \mathrm{Var}\left[p_t \mid \mathcal{F}_t^y\right]$–the receiver's posterior satisfies the stochastic differential equation

$$d\hat{p}_t = \frac{v_t}{\sqrt{\mathbb{E}\left[g^2(t, p_t) \mid \mathcal{F}_t^Y\right]}} dB_t^I. \tag{2}$$

[8]The constant M can be taken to be as large as we wish, and we assume it only to ensure the existence of strong solutions with bounded coefficients. Without this restriction, we could conduct the analogous analysis by working with weak solutions. This would not change any of the economic results but would complicate the analysis significantly.

We refer to B^I as the *innovation Brownian motion* associated with the observation process Y.[9]

We can now rewrite the sender's problem entirely in terms of the posterior process $\hat{p}_t$. Under a garbling g and stopping time τ, the posterior process $\hat{p}_t$ solving (2) induces a terminal law ν on $[0,1]$; *viz.*, $\hat{p}_\tau \sim \nu$. Because $\hat{p}_t$ is a martingale with $\hat{p}_0 = p_0$, Bayes plausibility implies that any feasible terminal law ν must satisfy $\int_0^1 p\nu(dp) = p_0$.

Let $m_\nu := \nu\left(\left[\bar{P}, 1\right]\right)$ denote the mass that ν places at or above the target threshold $\bar{P}$.[10] This is the probability, under (g, τ), that the receiver ends up taking the sender's desired action. The sender's expected payoff under (g, τ) is, therefore,

$$\mathbb{E}\left[\mathbf{1}\{\hat{p}_\tau \geq \bar{P}\} - c(\tau)\right] = m_\nu - \mathbb{E}[c(\tau)].$$

Next, for any Bayes-plausible ν, define the minimal expected cost of embedding ν:

$$J(\nu) := \inf\left\{\mathbb{E}\left[c(\tau)\right] \mid \; \exists g, \tau \text{ such that } \hat{p}_0 = p_0, \; d\hat{p}_t = \sigma_g(t, \{\hat{p}_s\})dB_t, \text{ and } \hat{p}_\tau \sim \nu\right\}.$$

The sender's design problem is, thus, equivalent to choosing a Bayes-plausible terminal law ν to maximize the success probability minus the embedding cost:

$$\max_\nu \left\{m_\nu - J(\nu)\right\}, \quad \text{subject to} \quad \nu \text{ a probability measure on } [0,1] \text{ with } \int_0^1 p\nu(dp) = p_0.$$

Equivalently, we can write the problem in three nested steps. Fix a target success probability $m_\nu \in [0,1]$, then choose a (Bayes-plausible) terminal law ν that places a mass m_ν at $\bar{P}$, and finally choose a garbling g and stopping time τ that implement ν:

$$\sup_{m_\nu \in [0,1]} \; \sup_{\substack{\nu :\, \nu(\{\bar{P}\}) = m_\nu, \\ \int p\nu(dp) = p_0}} \; \sup_{g, \tau :\, \hat{p}_\tau \sim \nu} \left\{m_\nu - \mathbb{E}[c(\tau)]\right\}.$$

4. Results

Our first result shows that when there are (weakly) convex costs involved, the best thing the sender can do is to have the posterior be a Markov process.

[9]Formally, we construct B^I from the "innovation process" $dI_t := dY_t - \hat{p}_t dt$–see Appendix A.3.

[10]Note that, in continuous time it is never optimal to place mass strictly above $\bar{P}$ (one could instead stop as soon as the posterior first hits $\bar{P}$, which would keep $\nu\left(\left[\bar{P}, 1\right]\right)$ fixed and weakly reduce the delay cost), so it is without loss of optimality to restrict attention to terminal laws with support contained in $\left[0, \bar{P}\right] \cup \left\{\bar{P}\right\}$ so that $m_\nu = \nu\left(\left\{\bar{P}\right\}\right)$.

Proposition 1. *For any Bayes-plausible terminal law* ν*, any increasing and convex cost* c*, there exists an optimal (cost-minimizing) garbling* g^* *and stopping time* τ^* *implementing* ν *for which the receiver's posterior evolves as a time-homogeneous diffusion* $d\hat{p}_t = \Sigma(\hat{p}_t)dB_t$*, for some Borel* $\Sigma\colon (0,1) \to \mathbb{R}_+$*.*

The formal proof of Proposition 1 is in Appendix A.4 and has two ingredients. First, Lemma 1 concerns the shape of the information process, holding fixed the terminal law. It shows that any time- and history-dependent posterior volatility can be "smoothed" into a single belief-based speed $\Sigma(\hat{p})$ with weakly smaller residual expectations. A useful analogy is driving on a highway to a fixed destination. You can choose your speed at every moment, and you are penalized by some convex function of total travel time. It is natural that your speed depends on *where* you are–slower on tight curves or hills, faster on straight segments–but it should not depend on the time of day (think of a well-lit road): how fast you take the same turn should not change just because you arrived there at 8AM rather than 8PM *ceteris paribus*. Under a complicated policy, at each location y along the highway, you may sometimes crawl and sometimes speed, depending in a messy way on when you arrived there and on the route you took so far. From the point of view of total travel time under a convex penalty, this within-location variation is wasteful: you are mixing very fast and very slow episodes in the same place, even though where you need to end up is fixed.

Second, Lemma 2 (Rost) gives the economic ordering: if one policy has weakly smaller residual expected time

$$R_{\tau^*}(t) := \mathbb{E}\left[(\tau^* - t)_+\right] = \int_t^\infty \mathbb{P}(\tau^* \geq s)ds,$$

at *every* date, then it has weakly lower expected cost for *every* convex increasing delay cost. In this sense, pushing the entire curve $R_\tau(\cdot)$ downward is exactly what convex costs reward.

Putting these pieces together, the smoothing step from Lemma 1 delivers pointwise dominance of residual expectations, and Lemma 2 converts that dominance into cost dominance, yielding Proposition 1.

With our first proposition in hand, our next result pins down the structure of optimal distributions.

Proposition 2. *For any weakly convex cost and any upper atom size $m_\nu > 0$, the optimal embedded distribution is a two atom distribution where the second atom is pinned down by the martingale constraint.*

To put differently, fix $\bar{P}$ and the probability of persuasion m_ν. Then among any distributions over posteriors you can induce, the one that minimizes the residual expected time is the most "polarized" one: all the remaining mass $1 - m_\nu$ is placed at a single lower belief $\underline{P}$ (which satisfies Bayes-plausibility). That is, the optimal terminal law is binary.

The rationale for this result is simple: the objective only requires hitting $\bar{P}$ with likelihood m_ν and is agnostic about the lower failure beliefs. Splitting the "fail" mass across many points forces extra learning just to sort among them, which only takes longer without improving the objective. Collapsing all failure outcomes to one point $\underline{P}$ eliminates that gratuitous delay.

Next, we state our main result. We say that, given a terminal law ν and a weakly convex cost, *optimal garbling is no garbling* if for any garbling process g and stopping time τ that implement ν, there exists a *no-garbling* policy $g = 0$ and stopping time $\tilde{\tau}$ that implement the same ν and satisfy $\mathbb{E}\left[c\left(\tilde{\tau}\right)\right] \leq \mathbb{E}\left[c\left(\tau\right)\right]$, strictly so if the garbling is nontrivial. *Viz.*, any optimal solution to the sender's persuasion problem does not garble the fundamental at all.

Theorem 1. *For any weakly convex, strictly increasing cost, optimal garbling is no garbling.*

Theorem 1 is the continuous-time analog of our normal-plus-noise example in the introduction. Recall that in our one-shot toy model, the state is $\Theta \sim N(0, v)$ and the signal equals the state plus a zero-mean normal noise variable ε that is independent of Θ. The receiver cares about whether her posterior mean $\hat{\Theta} = \mathbb{E}\left[\Theta \middle| Y\right]$ exceeds a cutoff a. When you increase the noise variance r, the posterior mean becomes more diffuse around zero and the probability $\mathbb{P}(\hat{\Theta} > a)$ falls. With n independent observations, the posterior mean $\hat{\Theta}_n$ becomes more precise as n grows, and the probability of crossing the threshold is increasing in n but decreasing in r. If we treat n as a crude clock, then more garbling (larger r) literally slows that clock: for a given n, you are less likely to have pushed the posterior above the cutoff. If you want to maximize the chance of hitting the cutoff given a cost of samples, you never add noise, but instead use the cleanest signal you can, and only choose how many draws n to take.

Theorem 1 says that exactly the same logic is not only present, but demonstrated in a sharper manner (and holds for arbitrary not just binary νs) in the continuous-time persuasion problem. Now the fundamental Brownian motion is the analog of the "clean signal," and the garbling process g_t plays the role of injecting extra noise at each instant. Fix a target terminal law ν, which is the *ex ante* distribution of the receiver's eventual posterior that the sender wants to implement. The lemmas we used to prove Proposition 1 show three things. First, once ν is fixed, the shape of ν pins down a "time budget" at each belief level: any policy that implements ν must, in expectation, spend exactly the same amount of time near each belief y.

Second, given this level-by-level time budget, what matters for convex time costs is how dispersed the stopping time τ is–in particular, how fat its upper tail is. Convex waiting costs are relatively forgiving about small delays but increasingly punitive about very long ones. Third, any extra time- or history-dependence in the volatility of the posterior at a given belief level simply injects unnecessary variation into how fast beliefs move there. From the point of view of convex time costs, these alternations between very slow and very fast episodes at the same belief are harmful, as they make extremely long paths more likely.

Theorem 1 identifies the best way to use the given time budget at each belief level. Among all information policies that implement a given terminal law ν, the optimal one for any convex, increasing delay cost is simply to set $g \equiv 0$, i.e., the sender never deliberately garbles the fundamental. This is the continuous-time counterpart of setting the one-shot noise variance r as low as possible and only choosing how many signals to observe. Under no garbling, the posterior is as responsive as possible to the underlying state at every belief level, so the process uses up its fixed local time budget in the fastest, least-dispersed way. In this sense, full transparency makes the belief "clock" tick as quickly as the model allows. Any additional garbling along the way only slows that clock, leading to more dispersion in stopping times and a higher expected cost under every convex measure of delay.

4.1. **Solving the Persuasion Problem.** Putting Proposition 2 together with Theorem 1, the sender's persuasion problem reduces to a simple one-dimensional choice. Under the optimal design, the sender doesn't garble so that the receiver observes the full-information posterior, i.e., $\hat{p}_t = p_t$, and the sender stops at the first exit time from an interval $\left[\underline{P}, \bar{P}\right]$. The induced terminal law is binary, with success probability

$$m_\nu(\underline{P}) := \mathbb{P}\left(p_{\tau(\underline{P})} = \bar{P}\right) = \frac{p_0 - \underline{P}}{\bar{P} - \underline{P}}.$$

Consequently, the sender solves the one-dimensional program

$$\max_{\underline{P} \in (0, p_0)} \left\{ \frac{p_0 - \underline{P}}{\bar{P} - \underline{P}} - \mathbb{E}\left[c(\tau(\underline{P}))\right] \right\}.$$

When c is Laplace-transformable; *viz.*, $c(t) = \int_0^\infty e^{-st}\mu(ds)$ for some Borel measure μ on $\mathbb{R}_+$, we can write

$$\mathbb{E}\left[c(\tau(\underline{P}))\right] = \int_0^\infty \phi_s(p_0)\mu(ds), \quad \text{where} \quad \phi_s(p_0) := \mathbb{E}\left[e^{-s\tau(\underline{P})}\right],$$

and in the Supplementary Appendix (B), we provide a closed form for ϕ_s (Proposition 7). Consequently, for any given cost, pinning down $\underline{P}$ is a matter of one-dimensional calculus.

Furthermore, even beyond linear costs, $\mathbb{E}[\tau(\underline{P})]$ is a useful/illuminating object: it is the expected time to persuade (or to quit) given the target success probability m_ν. For the linear benchmark $c(t) = t$, Proposition 8 in our Supplementary Appendix (B) computes a closed form. Specifically, writing $\mathfrak{L}(p) := (2p-1)\ln\left(\frac{p}{1-p}\right)$, we obtain

$$\mathbb{E}[\tau(\underline{P})] = \frac{2\sigma^2}{(\mu_h - \mu_l)^2}\left(\frac{\bar{P} - p_0}{\bar{P} - \underline{P}}\mathfrak{L}(\underline{P}) + \frac{p_0 - \underline{P}}{\bar{P} - \underline{P}}\mathfrak{L}(\bar{P}) - \mathfrak{L}(p_0)\right).$$

5. Comparative Statics

We turn our attention to comparative statics. Our analysis so far has allowed us to condense the sender's choice into a simple choice of a probability of persuasion $p \in \left[0, p_0/\bar{P}\right]$. There are two natural comparative statics questions: first, how does the shape of the cost function shape persuasion? That is, how does the optimal probability of persuasion change as we change the cost? Second, how do the primitives of the fundamental affect persuasion?

In Proposition 3, we show that (additively) more convex costs lead to lower persuasiveness, answering the first question. Proposition 4 answers the second: persuasiveness is increasing in the signal-to-noise ratio.

Formally, for a fixed $\bar{P}$ and prior p_0, for each $p \in \left[0, p_0/\bar{P}\right]$, let $\tau(p)$ denote the optimal stopping time that implements $\nu_p = (1-p)\delta_{\underline{P}(p)} + p\delta_{\bar{P}}$ with $\underline{P}(p) = (p_0 - p\bar{P})/(1-p)$. For a gross payoff to persuasion of $V > 0$ and a convex, increasing time cost c with $c(0) = 0$, define

$$J_c(p) \coloneqq \mathbb{E}\left[c\left(\tau(p)\right)\right], \quad \text{and} \quad \Pi_c(p) \coloneqq Vp - J_c(p).$$

For cost functions c_1 and c_2, with $c_1(0) = c_2(0) = 0$, if their difference $c_2 - c_1$ is increasing and convex on $[0, \infty)$ we say that c_2 is *more convex* than c_1.

Definition 1. *Set S_1 dominates set S_2 in the strong set order, $S_1 \geq_{SSO} S_2$, if for any $s_1 \in S_1$ and $s_2 \in S_2$, $\max\{s_1, s_2\} \in S_1$ and $\min\{s_1, s_2\} \in S_2$.*

If $\arg\max_{p \in \left[0, p_0/\bar{P}\right]} \Pi_{c_1}(p) \geq_{SSO} \arg\max_{p \in \left[0, p_0/\bar{P}\right]} \Pi_{c_2}(p)$, we say that the persuader with cost c_1, persuader 1, is *more persuasive* than persuader 2.

Proposition 3. *If c_2 is more convex than c_1, persuader 1 is more persuasive than persuader 2.*

This proposition leans on two technical lemmas. In the first, we show that comparing stopping times in the *increasing convex order* is equivalent to comparing their residual-time curves and their means. Intuitively, if one policy never has less expected remaining time at any date and also has a (weakly) larger mean, then every convex time cost regards it as the "riskier" stopping rule. Our second lemma, compares the exit times of the posterior diffusion from two nested intervals. In particular, we show that lowering the lower boundary increases the expected exit time and leads to pointwise dominance of the residual-time curves.

Putting these pieces together, the remainder of the proof exploits classic comparative statics results. Fix two success probabilities $p' < p$. From the second lemma we know that $\tau(p)$ is larger than $\tau(p')$ in the increasing convex order, so for any increasing convex cost c the incremental cost $\mathbb{E}\left[c(\tau(p)) - c(\tau(p'))\right]$ is nonnegative, and it becomes larger the more convex c is. In other words, as we move from a less convex cost c_1 to a more convex cost c_2 (with $c_2 - c_1$ convex and increasing), the extra cost of raising the success probability from p' to p increases: the map $(p, c) \mapsto c(\tau(p))$ has decreasing differences. Since the benefit term Vp is linear in p and does not depend on c, the net payoff inherits this decreasing-differences property. Standard monotone comparative statics then imply that the $\arg\max$ set of Π_{c_2} is contained in the $\arg\max$ set of Π_{c_1} in the strong set order: as the cost of delay becomes more convex, the sender's optimal success probabilities shift toward more conservative values.

Next, let us change the μs and σs. For $i \in \{1, 2\}$, define

$$\kappa_i \coloneqq \frac{|\mu_H^i - \mu_L^i|}{\sigma_i}, \quad \text{and} \quad \chi \coloneqq \left(\frac{\kappa_2}{\kappa_1}\right)^2.$$

If $\kappa_2 \geq \kappa_1$ ($\chi \geq 1$), we say that the *signal-to-noise ratio is higher for persuader* 2.

Proposition 4. *If the signal-to-noise ratio is higher for persuader* 2*, she is more persuasive.*

The basic intuition behind this result is that changing the signal-to-noise ratio is essentially a time change. Running the diffusion with a higher ratio means that the process moves faster and, therefore, hits the boundaries in less calendar time. Consequently, we have the analogous change in incremental costs as in Proposition 3, hence, decreasing-differences in (p, κ). This produces the result.

6. Extensions

We finish the paper with two extensions. In the first, we relax the global commitment assumption for the sender. Rather than assuming she can commit at time 0 to an entire garbling and stopping policy, we allow her to re-optimize "locally:" at any stopping time ζ of the signal filtration, she may replace the continuation of her policy from ζ onward. This adds a natural sequential-optimality requirement–plans must remain optimal from every future history–but we show it has no bite in our environment: every globally optimal policy is also sequentially optimal, so global and local commitment coincide.

In the second extension, we allow the receiver to stop as well. The sender still commits to a policy, but now the receiver also chooses a stopping time τ_r, so that the realized terminal time is $\tau = \tau_s \wedge \tau_r$. When the receiver has no flow cost of waiting ($\lambda = 0$), she never benefits from preempting the sender, so the model reduces to the baseline sender-only-stopping problem. When $\lambda > 0$, preemption is a genuine concern: because the realized terminal time is $\tau = \tau_s \wedge \tau_r$, the terminal posterior law (the embedded distribution) is now an equilibrium outcome, jointly pinned down by the sender's and receiver's stopping rules. Nevertheless, the main conclusion survives: in any sender-commitment equilibrium, the sender does not garble. Intuitively, the sender can switch to full transparency and embed the same terminal law more quickly, and the receiver is unwilling to punish this deviation by stopping earlier.

6.1. **Local Commitment Equals Global Commitment.**

So far we have treated the sender as a long-lived designer who can commit *ex ante* to an entire policy $\pi = (g, \tau)$: a garbling rule for the flow of information and a stopping rule specifying when to terminate the process and act. In many applications, however, it is more natural to think of the sender as choosing information flows locally, revising them over time as the signal history evolves, and retaining the ability to abandon the interaction. To capture this, we now consider a *local commitment* formulation in which, at any stopping time ζ of the signal filtration $(\mathcal{F}_t^Y)_{t\geq 0}$, the sender may reconsider and replace the continuation of the policy from that point on.

From the sender's perspective, this weaker form of commitment can only make her worse off: it adds the requirement that any policy she chooses must remain optimal when viewed from the perspective of every future history, so that she never wants to abandon her plan midway. Does this additional sequential-rationality constraint actually have bite, or is local commitment just as good as global commitment? Our result is simple: every globally optimal policy is immune to profitable local deviations. Thus, global and local commitment coincide.

Let $\pi = (g^\pi, \tau^\pi)$ be a policy (satisfying the above assumptions), where g^π is an $\mathcal{F}^X$-predictable garbling process and τ^π is an $\mathcal{F}^{Y,\pi}$-stopping time, with Y^π the induced signal process and $\hat{p}^\pi$ the induced posterior process. Let Π denote the set of policies. For each $\pi \in \Pi$, the sender's payoff is $U(\pi) := \mathbb{E}\left[\mathbf{1}_{\{\hat{p}^\pi_{\tau^\pi} \geq \bar{P}\}} - c(\tau^\pi)\right]$. We define the *global commitment value* to be $V^{\mathrm{g}} := \sup_{\pi\in\Pi} U(\pi)$.

A policy $\pi \in \Pi$ is *sequentially optimal* (for the local-control problem) if for every stopping time ζ of the signal filtration $(\mathcal{F}^{Y,\pi}_t)_{t\geq 0}$ induced by π, and every alternative policy $\tilde{\pi} \in \Pi$ that coincides with π up to time ζ,[11] we have

$$\mathbb{E}\left[\mathbf{1}_{\{\hat{p}^\pi_{\tau^\pi} \geq \bar{P}\}} - c(\tau^\pi)\middle|\mathcal{F}^{Y,\pi}_\zeta\right] \geq \mathbb{E}\left[\mathbf{1}_{\{\hat{p}^{\tilde{\pi}}_{\tau^{\tilde{\pi}}} \geq \bar{P}\}} - c(\tau^{\tilde{\pi}})\middle|\mathcal{F}^{Y,\pi}_\zeta\right] \quad \text{a.s. on } \{\zeta < \tau^\pi\}. \tag{3}$$

We define the *local commitment value* to be $V^{\mathrm{l}} := \sup\{U(\pi) \colon \pi \in \Pi \text{ is sequentially optimal}\}$.

Proposition 5. *The global commitment value V^{g} equals the local commitment value V^{l}.*

We defer the full proof to Appendix A.9, but here is a sketch. Clearly, $V^{\mathrm{g}} \geq V^{\mathrm{l}}$, since any sequentially optimal policy is a feasible policy for the global problem. To prove the reverse inequality, suppose that a globally optimal policy π^* were not sequentially optimal. Then there exists a stopping time ζ of the signal filtration and an alternative policy $\tilde{\pi} \in \Pi$ such that $\tilde{\pi}$ coincides with π^* up to ζ, but on some event $A \in \mathcal{F}^Y_\zeta$ with $\mathbb{P}(A) > 0$ the conditional continuation payoff under $\tilde{\pi}$ (starting from the history at ζ) is strictly higher than under π^*.

[11]That is, $Y^{\tilde{\pi}}_t = Y^\pi_t$ and $\tau^{\tilde{\pi}} \wedge \zeta = \tau^\pi \wedge \zeta$ almost surely for all $t \leq \zeta$.

Define a new policy π' by "splicing" $\tilde{\pi}$ into π^*: π' behaves like π^* up to time ζ, and from ζ onward follows $\tilde{\pi}$ on A and π^* on A^c. By construction, π' is feasible and delivers a strictly higher *ex ante* expected payoff than π^*, contradicting the global optimality of π^*. Consequently, every globally optimal policy must be sequentially optimal, so $V^{\mathrm{g}} \leq V^{\mathrm{l}}$. Together with the first inequality this implies $V^{\mathrm{g}} = V^{\mathrm{l}}$.

6.2. **Bilateral Stopping.** We now extend the baseline model by allowing *both* players to stop. We continue to work in the continuous time persuasion environment of §2 with an unknown binary state $\mu \in \{\mu_\ell, \mu_h\}$ and a binary receiver action $a \in \{0, 1\}$.

Now, in addition to the sender's chosen stopping time τ_s, the receiver also chooses a stopping time τ_r (both of which are with respect to the filtration $\mathcal{F}^Y_t$), and the realized terminal time is $\tau \coloneqq \tau_s \wedge \tau_r$. Upon stopping, the receiver also chooses binary action $a_\tau \in \{0, 1\}$. The receiver's objective is to maximize $\mathbb{E}\left[u(a_\tau, \mu) - \lambda\tau\right]$ $(\lambda \geq 0)$, whereas the sender's payoff is as before: $U_S(Y; \tau, a) \coloneqq \mathbb{E}\left[u_S(a_\tau, \mu)\right] - \mathbb{E}\left[c(\tau)\right]$, where $c\colon [0, \infty) \to \mathbb{R}_+$ is convex, strictly increasing, and continuous, with $c(0) = 0$; and u_S depends on the state and the receiver's action.

We specify that the sender is a Stackelberg leader: she commits to (Y, τ_s) at time 0. The receiver observes (Y, τ_s) then chooses (τ_r, a), which must be sequentially rational with respect to $\mathcal{F}^Y$. Furthermore, as the receiver always optimizes the terminal action given her belief, we define the receiver's static value to be

$$V(p) \coloneqq \max_{a \in \mathcal{A}} \ \mathbb{E}\left[u_R(a, \mu) \mid \hat{p} = p\right],$$

so the receiver's stopping problem given (Y, τ_s) is

$$\sup_{\tau_r} \ \mathbb{E}\left[V(\hat{p}_{\tau_s \wedge \tau_r}) - \lambda(\tau_s \wedge \tau_r)\right].$$

We adapt our earlier definition, saying that *optimal garbling is no garbling* if all (subgame-perfect) equilibria of the bilateral-stopping game feature full-revelation on the continuation path.

Proposition 6. *For any strictly increasing, weakly convex sender cost, in the bilateral-stopping game, optimal garbling is no garbling.*

With bilateral stopping, the receiver's only credible punishment of a sender deviation is to preempt–i.e., stop *earlier* than the sender—because the realized terminal time is $\tau = \tau_s \wedge \tau_r$ (she cannot punish by waiting longer). The sender's deviation to no garbling effectively "tightens the clock," delivering each unit of information in weakly less calendar time. Since $\lambda > 0$, this makes waiting to the same "information time" weakly cheaper while leaving the benefit from additional information unchanged. Therefore, if the receiver did not preempt under the original (slower) clock, she will still not preempt under the deviation, so the deviation goes unpunished.

Appendix A. Omitted Proofs and Derivations

Throughout the appendix we use the following notation consistently:

- g_t denotes noise added to the *posterior* p_t .
- p_t is the "true" posterior belief (the sender's belief about the high state).
- $\hat{p}_t$ is the receiver's posterior belief given the garbled observations.
- $v_t := \mathrm{Var}[p_t \mid \mathcal{F}_t^Y]$ is the posterior variance.
- Σ denotes a (Markovian) diffusion coefficient in posterior space obtained from Gyöngy's mimicking theorem.[12]
- $R_\tau(t)$ is the residual of a stopping time τ at time t (see Definition 2).[13]

Appendices A.1-A.3 derive the posterior process and the filtering equations, Appendix A.4 proves the state-only diffusion result (Proposition 1), Appendix A.5 proves the binary-terminal-law result (Proposition 2), and the remaining appendices collect the moment and comparative-statics calculations used in the main text.

A.1. Sender's Beliefs.

The sender observes

$$dX_t = \mu\, dt + \sigma\, dB_t, \quad \text{with} \quad \mu \in \{\mu_h, \mu_l\}, \quad \text{and} \quad \mathbb{P}(\mu = \mu_h) = p_0.$$

[12]When time-homogeneous we write $d\hat{p}_t = \Sigma(\hat{p}_t)dB_t$, and when we want to emphasize the dependence on a specific garbling, we write Σ_g.

[13]To highlight dependence to various objects we write, for example, $R_{\tau|\nu}(t)$ for "R_τ when τ embeds ν," $R_{\tau|\sigma}(t)$ or $R_{\tau|\Sigma}(t)$ for "under diffusion coefficient σ or Σ," and $R_{\tau|g}(t)$ for "under garbling g."

The likelihood ratio is

$$L_t = \exp\left(\frac{\mu_h - \mu_l}{\sigma^2} X_t - \frac{\mu_h^2 - \mu_l^2}{2\sigma^2} t\right),$$

so that the posterior $p_t := \mathbb{P}\left(\mu = \mu_h \mid \mathcal{F}_t^X\right)$ satisfies

$$\frac{p_t}{1-p_t} = \frac{p_0}{1-p_0} L_t \quad \Longrightarrow \quad p_t = \frac{p_0 L_t}{p_0 L_t + (1-p_0)}.$$

Defining the innovation Brownian motion

$$W_t := \frac{1}{\sigma}\left(X_t - \int_0^t \mathbb{E}[\mu \mid \mathcal{F}_s^X] ds\right),$$

we can rewrite the observation as

$$dX_t = \mathbb{E}[\mu \mid \mathcal{F}_t^X]\, dt + \sigma\, dW_t,$$

and the posterior satisfies the innovation form

$$dp_t = \frac{\mu_h - \mu_l}{\sigma} p_t(1-p_t)\, dW_t.$$

A.2. **Reporting Signals is Reporting Beliefs.** This subsection shows that any signal process Y of the form

$$dY_t = a_t dt + b_t dX_t$$

is, from the receiver's point of view, equivalent to a process in which the sender reports her own posterior belief plus suitable Gaussian noise. Formally, for any such (a_t, b_t) the induced receiver posterior $(r_t)_{t\geq 0}$ can also be generated by an alternative signal Z satisfying

$$dZ_t = dp_t + g_t dB'_t,$$

where $(p_t)_{t\geq 0}$ is the sender's posterior process given X, B' is a Brownian motion, and g_t is a predictable "posterior noise" coefficient. That is, information policies can be represented without loss of generality directly in posterior space.

Let $\mu \in \{\mu_l, \mu_h\}$ with prior $p_0 := \mathbb{P}(\mu = \mu_h)$, and let the fundamental process X evolve according to

$$dX_t = \mu dt + \sigma dB_t, \quad \text{where} \quad \sigma > 0,$$

on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t^X)_{t\geq 0}, \mathbb{P})$, where B is a standard Brownian motion and $\mathcal{F}_t^X = \sigma(X_s \colon s \leq t)$ is the completed natural filtration of X.

The sender observes X and commits at time 0 to an $(\mathcal{F}_t^X)$-adapted signal process Y of the form

$$dY_t = a_t\,dt + b_t\,dX_t,$$

where (a_t, b_t) is $\mathcal{F}^X$–progressively measurable and satisfies the integrability and boundedness conditions in the main text so that Y is a continuous semimartingale. The receiver observes only Y so that her information at time t is $\mathcal{F}_t^Y \coloneqq \sigma(Y_s \colon s \leq t)$.

Since (a_t, b_t) is $\mathcal{F}^X$-adapted and there is no additional primitive noise in Y, we have

$$\mathcal{F}_t^Y \subseteq \mathcal{F}_t^X, \quad \text{and} \quad \sigma(X_s, Y_s \colon s \leq t) = \mathcal{F}_t^X, \quad \text{for all } t \geq 0:$$

the joint filtration generated by (X, Y) coincides with the sender's filtration $\mathcal{F}^X$.

Define the sender's posterior $p_t \coloneqq \mathbb{P}(\mu = \mu_h \mid \mathcal{F}_t^X)$ and the receiver's posterior under the policy Y, $r_t \coloneqq \mathbb{P}(\mu = \mu_h \mid \mathcal{F}_t^Y)$. It is standard in this binary-drift Gaussian setting that $(p_t)_{t\geq 0}$ is a bounded continuous $(\mathcal{F}_t^X)$-martingale, and in fact solves

$$dp_t = \frac{\mu_h - \mu_l}{\sigma}\, p_t(1-p_t)\,dW_t,$$

for the innovation Brownian motion W of X in $\mathcal{F}^X$ (see e.g. Liptser and Shiryaev [2013], Ch. 7). In what follows we only use that p is a bounded continuous martingale.

Let $H \coloneqq 1_{\{\mu=\mu_h\}}$, and note that by definition, $p_t = \mathbb{E}[H \mid \mathcal{F}_t^X]$ and $r_t = \mathbb{E}[H \mid \mathcal{F}_t^Y]$. Because Y is constructed from X and there is no additional primitive randomness, we have $\mathcal{F}_t^Y \subseteq \mathcal{F}_t^X$ for all t. By the tower property,

$$r_t = \mathbb{E}[H \mid \mathcal{F}_t^Y] = \mathbb{E}\left[\mathbb{E}[H \mid \mathcal{F}_t^X] \,\middle|\, \mathcal{F}_t^Y\right] = \mathbb{E}\left[p_t \mid \mathcal{F}_t^y\right]. \tag{4}$$

Thus, $(r_t)_{t\geq 0}$ is the optional projection of $(p_t)_{t\geq 0}$ onto the smaller filtration $(\mathcal{F}_t^Y)$, and is itself a bounded continuous $(\mathcal{F}_t^Y)$-martingale with $r_0 = p_0$.

Let $\mathbb{F}^X = (\mathcal{F}_t^X)_{t\geq 0}$ denote the sender's filtration. Consider the sender's posterior martingale

$$M_t \coloneqq p_t - p_0, \quad \text{for all} \quad t \geq 0,$$

which is a continuous $(\mathbb{F}^X)$-martingale. Let $\mathbb{G} = (\mathcal{F}_t^Y)_{t\geq 0}$ be the receiver's filtration, regarded as a subfiltration of $\mathbb{F}^X$.

By the Kunita-Watanabe decomposition with respect to $\mathbb{G}$ (see, e.g., Revuz and Yor [2013], Ch. III, Thm. 3.9), there exists a unique decomposition

$$M_t = N_t + O_t, \quad \text{for all} \quad t \geq 0, \tag{5}$$

where

- N is a continuous $(\mathbb{F}^X)$-martingale which is also a $(\mathbb{G})$-martingale,
- O is a continuous $(\mathbb{F}^X)$-martingale, and
- O is strongly orthogonal to every $(\mathbb{G})$-martingale (in particular, $\langle L, N\rangle \equiv 0$).

Using (4), we identify N as the receiver's martingale:

$$N_t = \mathbb{E}[M_t \mid \mathcal{F}_t^Y] = \mathbb{E}[p_t - p_0 \mid \mathcal{F}_t^Y] = r_t - p_0.$$

Thus, $O_t = M_t - N_t = (p_t - p_0) - (r_t - p_0) = p_t - r_t$, and (5) becomes

$$p_t - p_0 = (r_t - p_0) + (p_t - r_t). \tag{6}$$

Since $\mathbb{F}^X$ is the Brownian filtration generated by B, every continuous $(\mathbb{F}^X)$-martingale admits a stochastic integral representation with respect to B. In particular, there exists an $\mathbb{F}^X$-predictable process h_t such that:[14]

$$O_t = p_t - r_t = \int_0^t h_s \, dB_s. \tag{7}$$

Substituting (7) into (6) yields

$$p_t - p_0 = (r_t - p_0) + \int_0^t h_s dB_s \quad \Rightarrow \quad r_t = p_t + \int_0^t g_s dB_s, \quad \text{where} \quad g_t \coloneqq -h_t. \tag{8}$$

Consequently, the receiver's posterior process r is the sender's posterior process p plus an $\mathbb{F}^X$-martingale given by an Itô integral with respect to the same Brownian motion B. In particular, the process g_t is adapted to $\mathbb{F}^X$.

[14] We do not need an explicit expression for h_t; only its existence and predictability matter.

Define a new signal process Z by $Z_t := r_t$ for all $t \geq 0$, and let $\mathcal{F}_t^Z := \sigma(Z_s \colon s \leq t)$ be its natural filtration. By (8), Z satisfies

$$Z_0 = p_0, \quad \text{and} \quad dZ_t = dp_t + g_t \, dB_t,$$

for the Brownian motion B and $\mathbb{F}^X$-adapted process g_t constructed above. Thus, Z is exactly of the form "sender's posterior plus Gaussian noise".

We now verify that the policy Z induces the same posterior process as the original policy Y. Since $Z_t = r_t$ and r_t is $\mathcal{F}_t^Z$-measurable,

$$\mathbb{P}(\mu = \mu_h \mid \mathcal{F}_t^Z) = \mathbb{E}[H \mid \mathcal{F}_t^Z] = \mathbb{E}[\,\mathbb{E}[H \mid \mathcal{F}_t^Y] \mid \mathcal{F}_t^Z] = \mathbb{E}[r_t \mid \mathcal{F}_t^Z] = r_t.$$

In other words, observing Z yields exactly the same posterior process $(r_t)_{t\geq 0}$ as observing Y.

If desired, we may also rescale the noise term so that $g_t \geq 0$ a.s.: define $\tilde{B}_t := \int_0^t \operatorname{sgn}(g_s)\, dB_s$, which is again a Brownian motion in $\mathbb{F}^X$, and set $\tilde{g}_t := |g_t|$. Then $dZ_t = dp_t + \tilde{g}_t \, d\tilde{B}_t$.

For any admissible linear signal policy $dY_t = a_t dt + b_t dX_t$, the induced receiver posterior process (r_t) coincides with the posterior induced by an alternative policy in which the sender reports a belief-based signal

$$dZ_t = dp_t + g_t \, dB_t$$

for some $\mathbb{F}^X$-adapted process g_t. Thus, without loss of generality, we may model information policies directly in *posterior space* as choices of a Gaussian noise process g_t added to the sender's posterior p_t, with g_t adapted to the sender's information about X. Since the meaning is clear, With abuse of notation we will refer to this Z_t directly as Y_t.

A.3. **How to Garble: Adding Mean-Preserving Noise.** We now fix the posterior process p_t from Appendix A.1 and specify how the sender garbles it before passing information to the receiver.

The posterior evolves as a one-dimensional martingale diffusion:

$$dp_t = \frac{\mu_h - \mu_l}{\sigma}\, p_t(1-p_t)dW_t, \quad \text{where} \quad p_0 \in (0,1),$$

W_t is a standard Brownian motion, and $\mathcal{F}_t^p := \sigma(p_s \colon s \le t)$ is the natural filtration of p_t.

The sender communicates to the receiver a "noise-added" observation process Y_t of the form

$$dY_t = p_t\, dt + g(t,p_t)\, dB_t^Y,$$

where B_t^Y is a Brownian motion independent of W_t, and $|g(t,p_t)| \le M$ a.s. for all $t \ge 0$ for some (as large as we want) constant $M > 0$. This boundedness assumption ensures that the stochastic integral $\int_0^t g(s,p_s)\, dB_s^Y$ is well defined for all finite t, the SDE admits a unique strong solution adapted to the filtration generated by (p, B^Y), and Y is a continuous semimartingale with square-integrable martingale part.[15]

We define the observation filtration to be $\mathcal{F}_t^Y := \sigma(Y_s \colon s \le t)$, and denote the receiver's posterior mean and variance by $\hat{p}_t := \mathbb{E}[p_t \mid \mathcal{F}_t^Y]$ and $v_t := \mathrm{Var}[p_t \mid \mathcal{F}_t^Y]$. The receiver is interested in $\hat{p}_t = \mathbb{E}[p_t \mid \mathcal{F}_t^Y]$. We can calculate this using the Kushner–Stratonovich (K–S) equation from nonlinear filtering.

For any sufficiently smooth test function $\phi(p)$, the K–S equation gives

(K-S)
$$\begin{aligned} d\mathbb{E}[\phi(p_t) \mid \mathcal{F}_t^Y] &= \mathbb{E}[\mathcal{A}\phi(p_t) \mid \mathcal{F}_t^Y]dt \\ &\quad + \frac{\mathbb{E}[\phi(p_t)h(p_t) \mid \mathcal{F}_t^Y] - \mathbb{E}[\phi(p_t) \mid \mathcal{F}_t^Y]\mathbb{E}[h(p_t) \mid \mathcal{F}_t^Y]}{\mathbb{E}[g^2(t,p_t) \mid \mathcal{F}_t^Y]}\,(dY_t - \mathbb{E}[h(p_t) \mid \mathcal{F}_t^Y]dt), \end{aligned}$$

where $\mathcal{A}\phi(p) := \frac{1}{2}\left(\frac{\mu_h-\mu_l}{\sigma}\right)^2 p^2(1-p)^2\phi''(p)$ is the generator of p_t, and $h(p) = p$ is the linear observation function.

We define the *innovation process* as

$$dI_t := dY_t - \mathbb{E}[p_t \mid \mathcal{F}_t^Y]\, dt = dY_t - \hat{p}_t\, dt,$$

[15]The assumption is merely for technical convenience, an analogous but much more cumbersome analysis can be done with weak solutions.

which is a continuous $\mathcal{F}^Y$-local martingale. Setting $\phi(p) = p$ in (K-S), the generator term vanishes, $\mathcal{A}\phi = 0$, and the observation term becomes the posterior variance v_t. Hence, the posterior mean satisfies the *innovation-form SDE*:

$$d\hat{p}_t = \frac{v_t}{\mathbb{E}[g^2(t,p_t) \mid \mathcal{F}_t^Y]} dI_t.$$

By standard filtering results (equivalently, the martingale representation theorem applied to the bounded continuous martingale $\hat{p}$), there exists an $\mathcal{F}^Y$-Brownian motion B^I such that the same posterior process admits the representation given in (2).

A.4. **Proof of Proposition 1.** Our goal is to show that, for any fixed target law over posteriors, among all garblings that implement it, one can find an optimal design in which the receiver's posterior $\hat{p}_t$ is a time-homogeneous diffusion with a volatility that depends only on the current belief level. Together with Lemma 2, this implies Proposition 1.

Definition 2. *Let τ be a stopping time. Its residual expected time is*[16]

$$R_\tau(t) := \mathbb{E}[(\tau - t)_+] = \int_t^\infty \mathbb{P}(\tau \geq s) ds, \quad \text{for all} \quad t \geq 0.$$

Lemma 1. *Let X be a one-dimensional continuous local martingale taking values in $(0,1)$, with dynamics $dX_s = \sigma(s, X_{[0,s]})dB_s$, where σ is an adapted (possibly path-dependent) diffusion coefficient satisfying $0 < \sigma_{\min} \leq \sigma(s, X_{[0,s]}) \leq \sigma_{\max} < \infty$. Let τ be a stopping time. Then there exists a* ***level-dependent*** *diffusion coefficient $\bar{\Sigma}(y)$ such that $R_{\tau|\sigma}(t) \geq R_{\tau|\bar{\Sigma}}(t)$ for all $t \geq 0$.*

Proof. By Gyöngy's Markovian projection theorem (see Gyöngy [1986]) there exists a measurable function

$$\Sigma(s,y) := \sqrt{\mathbb{E}\big[\sigma^2(s, X_{[0,s]}) \,\big|\, X_s = y\big]}$$

[16]Recall that when we want to highlight dependence, we use a conditioning bar instead of superscripts:

$$R_{\tau|\sigma}(t), \quad R_{\tau|\Sigma}(t), \quad R_{\tau|\nu}(t), \quad \text{and} \quad R_{\tau|g}(t).$$

These are purely book-keeping tags indicating, respectively, the diffusion coefficient, embedded law, or garbling relative to which R_τ is being computed.

such that the Markov process $\tilde{X}$ solving $d\tilde{X}_s = \Sigma(s, \tilde{X}_s)dB_s$ satisfies $X_s \overset{d}{=} \tilde{X}_s$ for all s. Thus, the processes X and $\tilde{X}$ have identical one-dimensional marginals, and, consequently, $R_{\tau|\sigma}(t) = R_{\tau|\Sigma}(t)$ for all t.

Applying the occupation-time formula to the stopped process $\tilde{X}_{s\wedge\tau}$, we have

$$(\tau - t)_+ = \int_t^\tau ds = \int_0^1 \int_t^\tau \frac{1}{\Sigma^2(s,y)}\, dL_s^y(\tilde{X})\, dy,$$

where $L_s^y(\tilde{X})$ denotes the local time of $\tilde{X}_{s\wedge\tau}$ at level y.

Taking expectations,

$$R_\tau(t) = \mathbb{E}[(\tau - t)_+] = \int_0^1 \mathbb{E}\left[\int_t^\tau \frac{1}{\Sigma^2(s,y)}\, dL_s^y\right] dy.$$

For each fixed $y \in (0,1)$, define the probability measure

$$\Gamma_y(ds) := \frac{\mathbb{E}[\, dL_s^y\,]}{\mathbb{E}[L_\tau^y - L_t^y]}, \quad \text{for all } s \ge t,$$

where $L_\tau^y := L_\tau^y(\tilde{X})$ is the total local time at level y. Let

$$\bar{\Sigma}^2(y) := \int_t^\tau \Sigma^2(s,y)\, d\Gamma_y(s).$$

By Jensen's inequality for the convex function $x \mapsto 1/x$,

$$\mathbb{E}\left[\int_t^\tau \frac{1}{\Sigma^2(s,y)}\, dL_s^y\right] = \mathbb{E}[L_\tau^y - L_t^y] \int_t^\tau \frac{1}{\Sigma^2(s,y)}\, d\Gamma_y(s) \ge \frac{\mathbb{E}[L_\tau^y - L_t^y]}{\bar{\Sigma}^2(y)}.$$

Integrating over $y \in (0,1)$ gives

$$R_\tau(t) \ge \int_0^1 \frac{\mathbb{E}[L_\tau^y - L_t^y]}{\bar{\Sigma}^2(y)}\, dy = \int_0^1 \mathbb{E}\left[\int_t^\tau \frac{1}{\bar{\Sigma}^2(y)}\, dL_s^y\right] dy = R_{\tau|\bar{\Sigma}}(t).$$

Thus, introducing explicit time or path dependence in the diffusion coefficient $\sigma(s, X_{[0,s]})$ cannot reduce the residual expectation. □

We now apply Lemma 1 to the posterior-mean process $\hat{p}_t$ obtained in Section A.3. Recall,

$$d\hat{p}_t = \frac{v_t}{\mathbb{E}[g^2(t,p_t) \mid \mathcal{F}_t^Y]} dI_t,$$

where the innovation process I_t is a continuous $\mathcal{F}^Y$-local martingale and $v_t = \operatorname{Var}\left[p_t \mid \mathcal{F}_t^Y\right]$. Under the boundedness assumption $0 \le |g(t,p_t)| \le M$ a.s. for all t and $0 \le v_t \le \frac{1}{4}$, the quadratic variation $\langle \hat{p} \rangle_t$ is absolutely continuous and uniformly bounded. In particular, there exists a $\mathcal{F}^Y$-Brownian motion B_t^Y and a bounded progressively measurable coefficient $\sigma(t,\omega)$ such that $d\hat{p}_t = \sigma(t,\omega)dB_t^Y$.

Thus, all standard conditions for Gyöngy's mimicking theorem are satisfied: there exists a (time-inhomogeneous) Markov diffusion $\tilde{p}_t$ and a measurable function $\Sigma(t,x)$ such that $d\tilde{p}_t = \Sigma(t,\tilde{p}_t)dB_t^Y$, where B_t^Y is a Brownian motion adapted to the observation filtration $\mathcal{F}_t^y$, and $\tilde{p}_t \overset{d}{=} \hat{p}_t$ for all $t \geq 0$.

Applying Lemma 1 to $\tilde{p}_t$ we obtain a *time-homogeneous*, state-dependent diffusion coefficient $\bar{\Sigma}(\cdot)$ such that the residual expectation curve is weakly smaller at every t when we run $d\hat{p}_t = \bar{\Sigma}(\hat{p}_t)dB_t^Y$ instead of the original path-dependent specification; *viz.*, $R_{\tau|\bar{\Sigma}}(t) \ \leq \ R_{\tau|\Sigma}(t)$ for all $t \geq 0$. For notational simplicity we henceforth write $\Sigma(\cdot)$ for this level-dependent speed and, with a slight abuse, again denote the resulting Markov process by $\hat{p}_t$.

Corollary 1. *For any admissible garbling process g that implements a given terminal law ν and any convex, increasing cost of delay, there exists another garbling $\tilde{g}$ such that the induced posterior evolves as $d\hat{p}_t = \Sigma_g(\hat{p}_t)dB_t$ for some Borel $\Sigma_g \colon (0,1) \to \mathbb{R}_+$, and such that $R_{\tau|\tilde{g}}(t) \ \leq \ R_{\tau|g}(t)$ for all $t \geq 0$. Consequently, it is without loss of generality to restrict attention to time-homogeneous, state-dependent posterior diffusions when minimizing any convex time cost (cf. Lemma 2 below).*

The following result, due to Rost [2006],[17] expresses all convex waiting costs in terms of the residual expectation.

Lemma 2 (Rost)**.** *Let τ and $\tilde{\tau}$ be stopping times. Suppose that for all $t \geq 0$, $R_\tau(t) \leq R_{\tilde{\tau}}(t)$. Then for any convex and increasing function $c \colon [0,\infty) \to \mathbb{R}$, $\mathbb{E}[c(\tau)] \leq \mathbb{E}[c(\tilde{\tau})]$.*

Proof of Lemma 2. Since c is convex and increasing, it is locally absolutely continuous and its right derivative $c'_+(t)$ exists for every $t \geq 0$, is nondecreasing, and satisfies

[17]Specifically, see parts (a) and (b) in the introduction of Rost [2006].

$c(t) = c(0) + \int_0^t c'_+(s)ds$ for all $t \geq 0$. Thus,

$$\mathbb{E}[c(\tau)] = c(0)+\mathbb{E}\left[\int_0^\tau c'_+(s)ds\right] = c(0)+\int_0^\infty c'_+(t)\mathbb{P}(\tau > t)dt = c(0)-\int_0^\infty c'_+(t)d\mathbb{E}[(\tau-t)_+].$$

Integration by parts for Stieltjes integrals implies that for every $T > 0$, the integral

$$-\int_0^T c'_+(t)d\mathbb{E}[(\tau - t)_+] = -c'_+(T)\mathbb{E}[(\tau - T)_+] + c'_+(0)\mathbb{E}[\tau] + \int_0^T \mathbb{E}[(\tau - t)_+]dc'_+(t).$$

By convexity, for every $T \geq 0$ and every $x \geq T$ we have $c(x) \geq c(T) + c'_+(T)(x - T)$. Applying this with $x = \tilde{\tau}$ yields $c'_+(T)(\tilde{\tau} - T)_+ \leq c(\tilde{\tau})\mathbf{1}_{\{\tilde{\tau}>T\}}$, so $c'_+(T)\mathbb{E}[(\tilde{\tau} - T)_+] \leq \mathbb{E}[c(\tilde{\tau})\mathbf{1}_{\{\tilde{\tau}>T\}}] \to 0$ as $T \to \infty$. Since $\mathbb{E}[(\tau - T)_+] \leq \mathbb{E}[(\tilde{\tau} - T)_+]$ for all T, we also have $c'_+(T)\mathbb{E}[(\tau - T)_+] \to 0$. Taking $T \to \infty$ produces

$$\mathbb{E}[c(\tau)] = c(0) + c'_+(0)\,\mathbb{E}[\tau] + \int_0^\infty \mathbb{E}[(\tau - t)_+]\,dc'_+(t),$$

which also holds with $\tilde{\tau}$ in place of τ. Using $\mathbb{E}[\tau] = \mathbb{E}[(\tau - 0)_+] \leq \mathbb{E}[(\tilde{\tau} - 0)_+] = \mathbb{E}[\tilde{\tau}]$, the pointwise dominance $\mathbb{E}[(\tau - t)_+] \leq \mathbb{E}[(\tilde{\tau} - t)_+]$ for all $t \geq 0$, and the fact that $c'_+(0) \geq 0$ and $dc'_+(t) \geq 0$, we conclude $\mathbb{E}[c(\tau)] \leq \mathbb{E}[c(\tilde{\tau})]$. □

Combining Corollary 1 with Lemma 2 yields Proposition 1.

A.5. **Proof of Proposition 2.** We first restate the proposition formally in the diffusion setting of the previous subsection.

Let $\hat{p}_t$ be a continuous local martingale in natural scale with diffusion coefficient $\Sigma(\hat{p}_t)$, started at $\hat{p}_0 \in (0, 1)$: $d\hat{p}_t = \Sigma(\hat{p}_t)dB_t$. Let $\mathcal{M}^p_{\bar{P}}$ denote the set of probability measures ν supported on $[0, \bar{P}]$ such that ν has an atom of size p at $\bar{P} \in (0, 1)$ and satisfies the martingale constraint $\int_0^{\bar{P}} \hat{p}\nu(d\hat{p}) = \hat{p}_0$.

For a given $\nu \in \mathcal{M}^p_{\bar{P}}$ and an embedding τ of ν by $\hat{p}_t$, the residual expectation can be written (see, e.g., Rost [2006]) as

$$R_{\tau|\nu}(0) = \int_0^{\bar{P}} \frac{U^\nu(\hat{p}) - |\hat{p} - \hat{p}_0|}{\Sigma(\hat{p})}\,d\hat{p}, \quad \text{where} \quad U^\nu(\hat{p}) := \int_0^{\bar{P}} |\hat{p} - z|\,\nu(dz).$$

Proposition 2 states that $R_{\tau|\nu}(0)$ is minimized over $\nu \in \mathcal{M}^p_{\bar{P}}$ by a binary measure with exactly two atoms: one at $\bar{P}$ and one at $\underline{P} = \frac{\hat{p}_0 - p\bar{P}}{1-p}$.

Proof of Proposition 2. Write $\nu = p\delta_{\bar{P}} + \eta$, where η is a measure with total mass $1-p$ supported on $[0, \bar{P})$ and satisfying the martingale constraint $\int z\eta(dz) = \hat{p}_0 - p\bar{P}$.

Then the residual expectation can be rewritten as

$$R_{\tau|\nu}(0) = \int_0^{\bar{P}} \frac{p|\hat{p} - \bar{P}| + \int_0^{\bar{P}} |\hat{p} - z|\,\eta(dz) - |\hat{p} - \hat{p}_0|}{\Sigma^2(\hat{p})}\, d\hat{p}.$$

Define the function

$$F(\eta) := \int_0^{\bar{P}} \frac{\int_0^{\bar{P}} |\hat{p} - z|\,\eta(dz)}{\Sigma^2(\hat{p})} d\hat{p},$$

Write $\eta = (1-p)\mu$ with μ a probability measure on $\left[0, \bar{P}\right)$ and $\int z\mu(dz) = \underline{P}$ so that

$$F(\eta) = (1-p)\int_0^{\bar{P}} \frac{\int_0^{\bar{P}} |\hat{p} - z|\mu(dz)}{\Sigma^2(\hat{p})} d\hat{p}.$$

For each $\hat{p}$, the function $z \mapsto |\hat{p} - z|$ is convex. Thus, by Jensen's inequality,

$$\int_0^{\bar{P}} |\hat{p} - z|\mu(dz) \geq \left|\hat{p} - \underline{P}\right|,$$

and so substituting into the expression for F, we have

$$F(\eta) \geq (1-p)\int_0^{\bar{P}} \frac{|\hat{p} - \underline{P}|}{\Sigma^2(\hat{p})} d\hat{p} = F((1-p)\delta_{\underline{P}}),$$

with equality if and only if $\mu = \delta_{\underline{P}}$. □

A.6. Proof of Theorem 1.

Proof of Theorem 1. Fix a convex, increasing cost c with $c(0) = 0$, and fix an arbitrary Bayes-plausible terminal law ν on $(0, 1)$ with mean p_0.

Let τ and g be arbitrary designs that implement ν, so that $\hat{p}_\tau \sim \nu$, where by Corollary 1, we may assume that under g the receiver's posterior evolves as a time-homogeneous martingale diffusion in natural scale: $d\hat{p}_t = \Sigma_g(\hat{p}_t)dB_t$ (and $\hat{p}_0 = p_0$), for some Borel $\Sigma_g \colon (0, 1) \to (0, \infty)$.

Next, under no garbling, the receiver observes the sender's information and her posterior coincides with the canonical full-information posterior $(p_t)_{t\geq 0}$ that solves

$$dp_t = \frac{\mu_h - \mu_l}{\sigma} p_t(1-p_t)dW_t, \quad \text{where } p_0 \in (0, 1),$$

which is a bounded continuous martingale in natural scale with diffusion coefficient $\Sigma_0(p)$.

Recall that under garbling g, $\mathcal{F}_t^Y$ is the receiver's observation filtration, with $\hat{p}_t \coloneqq \mathbb{E}[p_t \mid \mathcal{F}_t^y]$. Moreover, the same Kunita-Watanabe decomposition referenced in Appendix A.2 yields a continuous martingale $(N_t)_{t\geq 0}$ such that

$$p_t = \hat{p}_t + N_t, \quad \text{and} \quad \langle p \rangle_t = \langle \hat{p} \rangle_t + \langle N \rangle_t, \quad \text{for all } t \geq 0,$$

which implies $\langle \hat{p} \rangle_t \leq \langle p \rangle_t$ for all $t \geq 0$.[18]

Now we establish volatility dominance, i.e., $\Sigma_g(x) \leq \Sigma_0(x)$ for all Lebesgue-a.e. $x \in (0,1)$. Take an arbitrary $x \in (0,1)$ and consider the full-information posterior p_t started at $p_0 = x$ and the garbled posterior $\hat{p}_t$ started at $\hat{p}_0 = x$. Applying $\langle \hat{p} \rangle_t \leq \langle p \rangle_t$ to these restarted processes, dividing both sides by $t > 0$, and taking expectations of both sides with respect to the law of the process started from initial condition x gives

$$\frac{1}{t}\mathbb{E}_x\left[\int_0^t \Sigma_g^2(\hat{p}_s)ds\right] \leq \frac{1}{t}\mathbb{E}_x\left[\int_0^t \Sigma_0^2(p_s)ds\right], \quad \text{for all } t > 0.$$

Taking the limit as $t \downarrow 0$ yields $\Sigma_g^2(x) \leq \Sigma_0^2(x)$ for Lebesgue-a.e. $x \in (0,1)$, as desired.[19]

Next, we apply the Dambis-Dubins-Schwarz theorem to the continuous local martingale $(\hat{p}_t - p_0)_{t\geq 0}$. Consequently, there exists a Brownian motion B such that, as $\hat{p}_\tau \sim \nu$,

$$\hat{p}_t = p_0 + B_{\langle \hat{p} \rangle_t}, \quad \text{for all } t \geq 0 \quad \Rightarrow \quad \hat{p}_\tau = p_0 + B_{\langle \hat{p} \rangle_\tau} \quad \Rightarrow \quad p_0 + B_{\langle \hat{p} \rangle_\tau} \sim \nu.$$

[18] $\langle p \rangle_t$ denotes the quadratic variation of p up to time t (and likewise for $\langle \hat{p} \rangle_t$). Explicitly, $\langle \hat{p} \rangle_t = \int_0^t \Sigma_g^2(\hat{p}_s)ds$ and $\langle p \rangle_t = \int_0^t \Sigma_0^2(p_s)ds$.

[19] Under the uniform bounds in Lemma 3, any diffusion X solving $dX_t = \Sigma_g(X_t)dB_t$ with $X_0 = x$ admits a transition density $q(t,x,y)$ with Gaussian upper and lower bounds–see Bass [1998], Theorems 5.5 and 6.8. Hence, for any bounded Borel f,

$$\frac{1}{t}\mathbb{E}_x \int_0^t f(X_s)ds = \int f(y)\left(\frac{1}{t}\int_0^t q(s,x,y)ds\right)dy,$$

and the averaged kernel is an approximate identity as $t \downarrow 0$ (in particular $\int q(s,x,y)dy = 1$ and the bounds imply $\int_{|y-x|>\varepsilon} \frac{1}{t}\int_0^t q(s,x,y)dsdy \to 0$ for all $\varepsilon > 0$). Therefore, by the Lebesgue differentiation theorem for approximate identities, the left-hand side converges to $f(x)$ at Lebesgue points of f, thus, for Lebesgue-a.e. x. Applying this with $f = \Sigma_g^2$ yields the stated a.e. limit.

Using $\langle \hat{p} \rangle_t = \int_0^t \Sigma_g^2(\hat{p}_s)ds$ and $\hat{p}_s = p_0 + B_{\langle \hat{p} \rangle_s}$, a change of variables ($u = \langle \hat{p} \rangle_s$) delivers

$$\tau = \int_0^{\langle \hat{p} \rangle_\tau} \frac{1}{\Sigma_g^2(p_0 + B_u)} du.$$

By the same time-change representation for martingale diffusions in natural scale, we may realize the no-garbling posterior diffusion with coefficient Σ_0 on the same Brownian motion B, so that for each t,

$$t = \int_0^{\langle p \rangle_t} \frac{1}{\Sigma_0^2(p_0 + B_u)} du, \quad \text{and} \quad p_t = p_0 + B_{\langle p \rangle_t}.$$

Defining the no-garbling stopping time τ_0 by

$$\tau_0 := \int_0^{\langle \hat{p} \rangle_\tau} \frac{1}{\Sigma_0^2(p_0 + B_u)} du,$$

we have, by construction, $\langle p \rangle_{\tau_0} = \langle \hat{p} \rangle_\tau$, which implies

$$p_{\tau_0} = p_0 + B_{\langle p \rangle_{\tau_0}} = p_0 + B_{\langle \hat{p} \rangle_\tau} = \hat{p}_\tau \sim \nu,$$

i.e., τ_0 embeds the same terminal law ν under no garbling.

Finally, as $\Sigma_g(x) \leq \Sigma_0(x)$ for Lebesgue-a.e. $x \in (0,1)$, we have, for a.e. u,

$$\frac{1}{\Sigma_g^2(p_0 + B_u)} \geq \frac{1}{\Sigma_0^2(p_0 + B_u)},$$

which yields $\tau \geq \tau_0$ a.s., which implies that $\mathbb{E}[c(\tau_0)] \leq \mathbb{E}[c(\tau)]$, by the monotonicity of c. □

A.7. **Proof of Proposition 3.** We start with two lemmas. For a nonnegative random variable τ, and $t \geq 0$, recall our definition of the residual expected time: $R_\tau(t) := \mathbb{E}\left[(\tau - t)_+\right]$. For two nonnegative random variables X, Y, we say that X dominates Y in the increasing convex order, $X \succeq_{icx} Y$, if $\mathbb{E}[\varphi(X)] \geq \mathbb{E}[\varphi(Y)]$ for every increasing convex φ (for which the expectations exist).

Lemma 3. *For two nonnegative random variables X and Y, $X \succeq_{icx} Y$ if and only if $R_X(t) \geq R_Y(t)$ for all $t \geq 0$ and $\mathbb{E}[X] \geq \mathbb{E}[Y]$.*

Proof. ($\Rightarrow$) Trivial: for each fixed t, the function $x \mapsto (x-t)_+$ is increasing and convex. Hence,

$$R_X(t) = \mathbb{E}\left[(x-t)_+\right] \geq \mathbb{E}\left[(y-t)_+\right] = R_Y(t).$$

($\Leftarrow$) Let φ be any increasing convex function on $[0,\infty)$ with finite expectations under X and Y. By the absolute continuity of φ, for every $x \geq 0$, we may write

$$\varphi(x) = \varphi(0) + \int_0^x \varphi'_+(s)ds,$$

where φ'_+ is the right derivative of φ. Since φ'_+ is weakly increasing on $[0,\infty)$, it induces a (non-negative) Borel measure via $\mu_\varphi\left((a,b]\right) = \varphi'_+(b) - \varphi'_+(a)$. Thus, we have

$$\int_0^x \left[\varphi'_+(s)ds - \varphi'_+(0)ds\right] = \int_0^x \int_0^s d\mu_\varphi(t)ds = \int_0^x (x-t)d\mu_\varphi(t) = \int_0^\infty (x-t)_+ d\mu_\varphi(t),$$

and so we have the representation

$$\varphi(x) = \varphi(0) + \varphi'_+(0)x + \int_0^\infty (x-t)_+ d\mu_\varphi(t).$$

Taking expectations and subtracting yields

$$\mathbb{E}\left[\varphi(X)\right] - \mathbb{E}\left[\varphi(Y)\right] = \varphi'_+(0)\left(\mathbb{E}\left[X\right] - \mathbb{E}\left[Y\right]\right) + \int_0^\infty \left(R_X(t) - R_Y(t)\right) d\mu_\varphi(t).$$

By assumption $\mathbb{E}\left[X\right] - \mathbb{E}\left[Y\right] \geq 0$ and $R_X(t) - R_Y(t) \geq 0$ for every t, while $\varphi'_+(0) \geq 0$ and μ_φ is a nonnegative measure. Consequently, the right-hand side is ≥ 0, so $\mathbb{E}\left[\varphi(X)\right] \geq \mathbb{E}\left[\varphi(Y)\right]$ for all increasing convex φ. We conclude that $X \succeq_{\text{icx}} Y$. □

Now let $0 \leq p' < p < \bar{P} \leq 1$ and fix a starting point of the posterior diffusion $x \in \left[p, \bar{P}\right]$. For any lower boundary $a \in \{p', p\}$, define the *first exit time* $\tau(a) \coloneqq \inf\left\{t \geq 0\colon\ p_t \notin (a, \bar{P})\right\}$ and set $u_a(x) \coloneqq \mathbb{E}_x[\tau(a)]$ to be the mean exit time when the process starts at x. Our next lemma shows that this statistic is monotone in an intuitive way:

Lemma 4. *For every $x \in [p, \bar{P}]$ we have $u_{p'}(x) \geq u_p(x)$. Consequently, for every $t \geq 0$ the conditional residual expectations satisfy almost surely*

$$R_{\tau(p')}(t) \coloneqq \mathbb{E}\left[(\tau(p') - t) \mid \mathcal{F}_t\right] \mathbf{1}_{\{t<\tau(p')\}} \geq R_{\tau(p)}(t) \coloneqq \mathbb{E}\left[(\tau(p) - t) \mid \mathcal{F}_t\right] \mathbf{1}_{\{t<\tau(p)\}},$$

and, in particular, $\mathbb{E}[\tau(p')] \geq \mathbb{E}[\tau(p)]$.

Proof. The diffusion coefficient is $\sigma_p(x) = \frac{\mu_h - \mu_l}{\sigma} x(1-x)$, which is strictly positive on $(0,1)$. Standard results for one-dimensional diffusions (via Dynkin's formula) imply that for each fixed lower boundary $a \in (0, \bar{P})$, the function u_a is C^2 on $(a, \bar{P})$, continuous on $[a, \bar{P}]$, and satisfies

$$\frac{1}{2}\sigma_p(x)^2 u_a''(x) = -1, \quad x \in (a, \bar{P}), \quad u_a(a) = 0, \quad \text{and} \quad u_a(\bar{P}) = 0. \tag{9}$$

Fix $p' < p$ and denote $u_{p'}$ and u_p as the corresponding mean exit time functions. Both satisfy (9) on $(p, \bar{P})$. Letting $v \coloneqq u_{p'} - u_p$, we have $\frac{1}{2}\sigma_p(x)^2 v''(x) = 0$ for all $x \in (p, \bar{P})$. Since $\sigma_p^2(x) > 0$, we have $v''(x) = 0$, hence, v is affine on $(p, \bar{P})$.

The boundary values are

$$v(\bar{P}) = u_{p'}(\bar{P}) - u_p(\bar{P}) = 0, \quad \text{and} \quad v(p) = u_{p'}(p) - u_p(p) = u_{p'}(p) \geq 0.$$

A linear function that is nonnegative at p and zero at $\bar{P}$ remains nonnegative on $[p, \bar{P}]$. Consequently, $v(x) \geq 0$ for all $x \in [p, \bar{P}]$, i.e., $u_{p'}(x) \geq u_p(x)$ for all $x \in [p, \bar{P}]$.

By the strong Markov property,

$$\mathbb{E}\left[(\tau(a) - t) \mid \mathcal{F}_t\right] \mathbf{1}_{\{t<\tau(a)\}} = \mathbb{E}_{p_t}[\tau(a)]\, \mathbf{1}_{\{t<\tau(a)\}} = u_a(p_t)\, \mathbf{1}_{\{t<\tau(a)\}}.$$

Fix $t \geq 0$. On $\{t < \tau(p) \cap \tau(p')\}$ we have $p_t \in [p, \bar{P}]$, and the above argument implies $u_{p'}(p_t) \geq u_p(p_t)$. Accordingly,

$$R_{\tau(p')}(t) = u_{p'}(p_t)\mathbf{1}_{\{t<\tau(p')\}} \geq u_p(p_t)\mathbf{1}_{\{t<\tau(p)\}} = R_{\tau(p)}(t),$$

and taking expectations yields $\mathbb{E}[\tau(p')] \geq \mathbb{E}[\tau(p)]$. □

Proof of Proposition 3. Fix any p', p with $0 \leq p' < p \leq p_0/\bar{P}$. For each p in this range, let $\tau(p)$ denote the optimal stopping time that implements

$$\nu_p = (1-p)\,\delta_{\underline{P}(p)} + p\,\delta_{\bar{P}}, \quad \text{where} \quad \underline{P}(p) = \frac{p_0 - p\bar{P}}{1-p}.$$

As p increases, the corresponding lower belief $\underline{P}(p)$ decreases, so the continuation interval $(\underline{P}(p), \bar{P})$ expands. By Lemma 4 applied to the lower boundaries $\underline{P}(p')$ and

$\underline{P}(p)$ (and the diffusion p_t), we have $R_{\tau(p)}(t) \geq R_{\tau(p')}(t)$ for all $t \geq 0$ and $\mathbb{E}[\tau(p)] \geq \mathbb{E}[\tau(p')]$. Lemma 3 then implies $\tau(p') \preceq_{\text{icx}} \tau(p)$.

Let c_1 and c_2 be two convex, increasing cost functions with $c_1(0) = c_2(0) = 0$, and suppose c_2 is more convex than c_1 in the sense that $\Delta(t) \coloneqq c_2(t) - c_1(t)$ is increasing and convex on $[0, \infty)$. For each $i \in \{1, 2\}$ and p define $J_{c_i}(p) \coloneqq \mathbb{E}\left[c_i(\tau(p))\right]$ and $\Pi_{c_i}(p) \coloneqq Vp - J_{c_i}(p)$.

Since Δ is increasing and convex and $\tau(p') \preceq_{\text{icx}} \tau(p)$, we have

$$\mathbb{E}[\Delta(\tau(p))] - \mathbb{E}[\Delta(\tau(p'))] \geq 0,$$

which is equivalent to

$$J_{c_2}(p) - J_{c_2}(p') - (J_{c_1}(p) - J_{c_1}(p')) \geq 0 \quad \Leftrightarrow \quad J_{c_2}(p) - J_{c_2}(p') \geq J_{c_1}(p) - J_{c_1}(p').$$

Accordingly,

$$\begin{aligned} \Pi_{c_2}(p) - \Pi_{c_2}(p') &= V(p - p') - (J_{c_2}(p) - J_{c_2}(p')) \\ &\leq V(p - p') - (J_{c_1}(p) - J_{c_1}(p')) = \Pi_{c_1}(p) - \Pi_{c_1}(p'), \end{aligned} \tag{10}$$

so the family $\{\Pi_c(\cdot)\}$ has *decreasing differences* in (p, c) in the (additively) more convex partial order.

Now let $p_1 \in \arg\max_{p \in [0, p_0/\bar{P}]} \Pi_{c_1}(p)$ and $p_2 \in \arg\max_{p \in [0, p_0/\bar{P}]} \Pi_{c_2}(p)$. We need to show that $\min\{p_1, p_2\} \in \arg\max \Pi_{c_1}$ and $\max\{p_1, p_2\} \in \arg\max \Pi_{c_2}$.

Suppose, without loss of generality, that $p_2 > p_1$. Set $p' = p_1$ and $p = p_2$ in (10). Since p_1 maximizes Π_{c_1}, we have $\Pi_{c_1}(p_2) - \Pi_{c_1}(p_1) \leq 0$. Then (10) implies

$$\Pi_{c_2}(p_2) - \Pi_{c_2}(p_1) \leq \Pi_{c_1}(p_2) - \Pi_{c_1}(p_1) \leq 0.$$

But p_2 maximizes Π_{c_2}, so $\Pi_{c_2}(p_2) - \Pi_{c_2}(p_1) \geq 0$. Combining the inequalities yields

$$\Pi_{c_2}(p_2) - \Pi_{c_2}(p_1) = 0, \quad \text{and} \quad \Pi_{c_1}(p_2) - \Pi_{c_1}(p_1) = 0.$$

Thus, both p_1 and p_2 are maximizers of both objectives, and the argmax set for c_1 dominates that for c_2 in the strong set order. $\square$

A.8. **Proof of Proposition 4.**

Proof. For $i \in \{1,2\}$, let $(\mu_H^i, \mu_L^i, \sigma_i)$ be the drift and volatility parameters of persuader i and define the signal-to-noise ratio

$$\kappa_i \coloneqq \frac{|\mu_H^i - \mu_L^i|}{\sigma_i}, \quad \text{and} \quad \chi \coloneqq \left(\frac{\kappa_2}{\kappa_1}\right)^2.$$

If $\kappa_2 \geq \kappa_1$, then $\chi \geq 1$, and we say that persuader 2 has a higher signal-to-noise ratio.

For each p and each volatility σ_i, let $\tau_p^{\sigma_i}$ denote the optimal stopping time that embeds ν_p under volatility σ_i. As is standard for one-dimensional diffusions, there is a time-change relationship between the two models: $\tau_p^{\sigma_1} \overset{d}{=} \chi \tau_p^{\sigma_2}$, so, for each p,

$$\Pi_c^{\sigma_1}(p) - \Pi_c^{\sigma_2}(p) = -\left(\mathbb{E}\left[c(\chi \tau_p^{\sigma_2})\right] - \mathbb{E}\left[c(\tau_p^{\sigma_2})\right]\right) =: -\mathbb{E}\left[\Delta_\chi(\tau_p^{\sigma_2})\right], \tag{11}$$

where $\Pi_c^{\sigma_i}(p)$ is the value under cost c and volatility σ_i, and $\Delta_\chi(t) \coloneqq c(\chi t) - c(t)$.

Next, define the hitting time

$$T_a^\sigma \coloneqq \inf\{t \geq 0 \colon p_t^\sigma = a\}, \quad \text{for } a \in [0,1],$$

and note that for all $\sigma > 0$ and $p' < p''$ we have $\underline{P}(p'') < \underline{P}(p')$, whence we have

$$\tau_{p'}^\sigma = \min\{T_{\underline{P}(p')}, T_{\bar{P}}\} \leq \min\{T_{\underline{P}(p'')}, T_{\bar{P}}\} = \tau_{p''}^\sigma.$$

Hence, $\tau_{p'}^\sigma \leq \tau_{p''}^\sigma$ almost surely, so $\tau_{p''}^\sigma$ first-order stochastically dominates (FOSD) $\tau_{p'}^\sigma$.

Now fix $\chi > 1$. For any $t' > t$ the convexity of c implies

$$\frac{c(\chi t') - c(t')}{t'(\chi - 1)} \geq \frac{c(\chi t) - c(t)}{t(\chi - 1)} \quad \Rightarrow \quad c(\chi t') - c(t') \geq c(\chi t) - c(t),$$

so $\Delta_\chi(t) = c(\chi t) - c(t)$ is increasing in t. Combining this with (11) and the FOSD relation between $\tau_{p'}^{\sigma_2}$ and $\tau_{p''}^{\sigma_2}$ yields that $\Pi_c^{\sigma_1}(p) - \Pi_c^{\sigma_2}(p)$ is decreasing in p: there are decreasing differences in (p, σ).

By the same monotone-comparative-statics argument as in Proposition 3, the argmax set of persuader 2 (with higher signal-to-noise ratio) dominates that of persuader 1 in the strong set order. □

A.9. **Proof of Proposition 5.**

Proof. Obviously, $V^{\mathrm{l}} \leq V^{\mathrm{g}}$ so it remains to show $V^{\mathrm{g}} \leq V^{\mathrm{l}}$. Let π^* be such that $U(\pi^*) = V^{\mathrm{g}}$. We claim that π^* satisfies (3), i.e., is sequentially optimal.

Suppose for the sake of contradiction not. Then there exist i) a stopping time ζ of $(\mathcal{F}_t^{Y,\pi^*})_{t\geq 0}$, ii) an alternative policy $\tilde{\pi} \in \Pi$ that coincides with π^* up to ζ, and iii) an event $A \in \mathcal{F}_\zeta^{Y,\pi^*}$ with $\mathbb{P}(A) > 0$, such that the continuation value under $\tilde{\pi}$ is strictly better than under π^* on $A \cap \{\zeta < \tau^{\pi^*}\}$. That is, there is a set

$$B \coloneqq A \cap \{\zeta < \tau^{\pi^*}\} \in \mathcal{F}_\zeta^{Y,\pi^*}, \quad \text{with } \mathbb{P}(B) > 0,$$

such that

$$\mathbb{E}\left[\mathbf{1}_{\{\hat{p}^{\tilde{\pi}}_{\tau^{\tilde{\pi}}} \geq \bar{P}\}} - c(\tau^{\tilde{\pi}}) \middle| \mathcal{F}_\zeta^{Y,\pi^*}\right] > \mathbb{E}\left[\mathbf{1}_{\{\hat{p}^{\pi^*}_{\tau^{\pi^*}} \geq \bar{P}\}} - c(\tau^{\pi^*}) \middle| \mathcal{F}_\zeta^{Y,\pi^*}\right] \quad \text{on } B. \tag{12}$$

Because $\tilde{\pi}$ and π^* coincide up to ζ, they induce the same signal process and filtration up to ζ, so the conditional expectations in (12) are well-defined with respect to the common sigma-algebra $\mathcal{F}_\zeta^{Y,\pi^*}$.

We now construct a new policy $\pi' \in \Pi$ by "splicing" $\tilde{\pi}$ into π^* after ζ on B. Formally, we define the garbling $g^{\pi'}$ by

$$g_t^{\pi'} \coloneqq \begin{cases} g_t^{\pi^*}, & \text{if} \quad t \leq \zeta, \\ \mathbf{1}_B g_t^{\tilde{\pi}} + \mathbf{1}_{B^c} g_t^{\pi^*}, & \text{if} \quad t > \zeta, \end{cases}$$

and define the stopping time $\tau^{\pi'} \coloneqq \mathbf{1}_B \tau^{\tilde{\pi}} + \mathbf{1}_{B^c} \tau^{\pi^*}$. Because $B \in \mathcal{F}_\zeta^{Y,\pi^*} \subseteq \mathcal{F}_\zeta^X \subseteq \mathcal{F}^{X_t}$ for all $t \geq \zeta$ and both g^{π^*} and $g^{\tilde{\pi}}$ are $\mathcal{F}^X$-predictable, $g^{\pi'}$ is $\mathcal{F}^X$-predictable and $\tau^{\pi'}$ is an $\mathcal{F}^{Y,\pi'}$-stopping time.

Furthermore, i) up to time ζ, π' coincides with π^* and induces the same filtration $\mathcal{F}_t^{Y,\pi'} = \mathcal{F}_t^{Y,\pi^*}$ for $t \leq \zeta$; ii) on B, after ζ, $(Y^{\pi'}, \tau^{\pi'})$ coincides with $(Y^{\tilde{\pi}}, \tau^{\tilde{\pi}})$; and iii) on B^c, $(Y^{\pi'}, \tau^{\pi'})$ coincides with $(Y^{\pi^*}, \tau^{\pi^*})$. Now define

$$D \coloneqq \left[\mathbf{1}_{\{\hat{p}^{\pi'}_{\tau^{\pi'}} \geq \bar{P}\}} - c(\tau^{\pi'})\right] - \left[\mathbf{1}_{\{\hat{p}^{\pi^*}_{\tau^{\pi^*}} \geq \bar{P}\}} - c(\tau^{\pi^*})\right].$$

By construction,

$$D = \left[\mathbf{1}_{\{\hat{p}^{\tilde{\pi}}_{\tau^{\tilde{\pi}}} \geq \bar{P}\}} - c(\tau^{\tilde{\pi}})\right] - \left[\mathbf{1}_{\{\hat{p}^{\pi^*}_{\tau^{\pi^*}} \geq \bar{P}\}} - c(\tau^{\pi^*})\right],$$

on B and $D = 0$ on B^c. Using the law of iterated expectations and conditioning on $\mathcal{F}_{\zeta}^{Y,\pi^*}$, we obtain

$$U(\pi') - U(\pi^*) = \mathbb{E}[D] = \mathbb{E}\left[\mathbb{E}\left[D \middle| \mathcal{F}_{\zeta}^{Y,\pi^*}\right]\right].$$

On B,

$$\mathbb{E}\left[D \mid \mathcal{F}_{\zeta}^{Y,\pi^*}\right] = \mathbb{E}\left[Z^{\tilde{\pi}} \mid \mathcal{F}_{\zeta}^{Y,\pi^*}\right] - \mathbb{E}\left[Z^{\pi^*} \mid \mathcal{F}_{\zeta}^{Y,\pi^*}\right] > 0,$$

by (12), whereas on B^c it is zero. Hence, $\mathbb{E}[D \mid \mathcal{F}_{\zeta}^{Y,\pi^*}] > 0$ on a set of positive probability, so $U(\pi') - U(\pi^*) = \mathbb{E}\left[\mathbb{E}\left[D \mid \mathcal{F}_{\zeta}^{Y,\pi^*}\right]\right] > 0$, which contradicts the *ex ante* optimality of π^*. $\square$

A.10. **Proof of Proposition 6.** We begin by noting a sequence of observations and lemmas, which we will then assemble.

The first easy fact is that if τ_s and τ_r are stopping times with respect to the filtration $\mathcal{F}_t^Y$, then so is $\tau = \tau_s \wedge \tau_r$, which is just a consequence of the fact that for each $t \geq 0$, $\{\tau \leq t\} = \{\tau_s \leq t\} \cup \{\tau_r \leq t\} \in \mathcal{F}_t^Y$. This sanity-check ensures that all optional-sampling and stopped-process manipulations we have used so far continue to apply after replacing τ by $\tau_s \wedge \tau_r$.

Second, we divide the remainder of the analysis into two cases: i) when the receiver incurs no cost from the process running ($\lambda = 0$), and ii) when the receiver does incur a flow cost ($\lambda > 0$).

The first case is easy: if $\lambda = 0$, the receiver has no direct cost of delay. In that case, stopping strictly before the sender's committed τ_s cannot help the receiver.

Lemma 5. *For any sender strategy (Y, τ_s) and any receiver stopping time τ_r, $\mathbb{E}\left[V(\hat{p}_{\tau_s})\right] \geq \mathbb{E}\left[V(\hat{p}_{\tau_s \wedge \tau_r})\right]$. Thus, the receiver has a best response $\tau_r \geq \tau_s$, so that $\tau = \tau_s$ (on-path).*

Proof. The process $V(\hat{p}_t)$ is a submartingale: for $0 \leq t \leq s$, for any $a \in \mathcal{A}$,

$$\mathbb{E}[u_R(a,\mu) \mid \mathcal{F}_t^Y] = \mathbb{E}\left[\mathbb{E}[u_R(a,\mu) \mid \mathcal{F}_s^Y] \mid \mathcal{F}_t^Y\right] \leq \mathbb{E}\left[\max_{a'} \mathbb{E}[u_R(a',\mu) \mid \mathcal{F}_s^Y] \mid \mathcal{F}_t^Y\right] = \mathbb{E}[V(\hat{p}_s) \mid \mathcal{F}_t^Y].$$

Taking the maximum over a yields $V(\hat{p}_t) \leq \mathbb{E}[V(\hat{p}_s) \mid \mathcal{F}_t^Y]$. Optional sampling gives the stated inequality. $\square$

Corollary 2. *The realized stopping time equals τ_s, so the bilateral-stopping extension reduces to the baseline sender-stopping model.*

If $\lambda > 0$, the receiver may strictly prefer to stop before τ_s for some sender policies, and the equilibrium embedded law $\eta := \mathcal{L}(\hat{p}_{\tau_s \wedge \tau_r})$ may differ from the sender-optimal embedded law in the baseline model (and may vary across equilibria). Our key insight is that even though η is endogenous, the sender's incentive to garble (rather, to not garble) is unchanged: garbling only slows the information clock and is, therefore, (weakly) suboptimal.

Fix any equilibrium (Y, τ_s, τ_r, a) with $\lambda > 0$ and let $\tau = \tau_s \wedge \tau_r$. Consider the *stopped posterior* $X_t := \hat{p}_{t\wedge\tau}$ for all $t \geq 0$. X is a bounded continuous $(\mathcal{F}_t^Y)$-martingale, so there exists a predictable process $\sigma_t \geq 0$ and an $(\mathcal{F}_t^Y)$-Brownian motion B such that

$$dX_t = \sigma_t dB_t, \quad \text{with} \quad \sigma_t = 0 \text{ for } t \geq \tau.$$

Let $\Sigma^2(t,x) := \mathbb{E}[\sigma_t^2 \mid X_t = x]$. By Gyöngy's Markovian projection theorem, there exists a time-inhomogeneous Markov diffusion $\tilde{X}$ solving

$$d\tilde{X}_t = \Sigma(t, \tilde{X}_t)\, d\tilde{B}_t, \quad \text{where} \quad \tilde{X}_0 = X_0,$$

such that $\tilde{X}_t \overset{d}{=} X_t$ for every $t \geq 0$. Since X is bounded and $X_t \to X_\infty = X_\tau$ a.s. as $t \to \infty$, we also have $\tilde{X}_t \Rightarrow \tilde{X}_\infty$ and, therefore, $\mathcal{L}(\tilde{X}_\infty) = \mathcal{L}(X_\infty) = \mathcal{L}(X_\tau) = \eta$. Thus, Gyöngy provides a convenient way to interpret the equilibrium *outcome* using a time-inhomogeneous Markov state $\tilde{X}$ without changing any payoff-relevant one-time marginals or the embedded law η.[20]

Let $A_t := \langle X\rangle_t = \int_0^t \sigma_s^2 ds$ and let $T(u) := \inf\{t \geq 0\colon A_t \geq u\}$ be its right-continuous inverse. By Dambis-Dubins-Schwarz (DDS), there exists a Brownian motion W such that $X_t = X_0 + W_{A_t}$ for all $t \geq 0$. We interpret $A_t = \langle X\rangle_t$ as accumulated information: by DDS the posterior can be written as a Brownian motion run at the random clock A_t, so one unit of "information time" equals one unit of posterior quadratic variation. We work with the information-time filtration $\mathcal{G}_u := \mathcal{F}^Y_{T(u)}$ and call $U := \langle X\rangle_\tau$

[20]This is the sense in which Gyöngy is used here: it is a reduction of the state/strategy space for analyzing equilibrium *outcomes*, not a one-to-one mapping of the underlying stopping strategies.

the terminal information time. Then $X_\tau = X_0 + W_U$ and, since $T(A_t) \le t$ for all t, $T(U) = T(A_\tau) \le \tau$, with equality whenever A is strictly increasing on $[0, \tau]$ (as is the case on path when $\lambda > 0$, since any interval with $\sigma \equiv 0$ yields no learning while strictly increasing the flow cost λt).

Let p denote the full-information posterior (*viz.*, no garbling). By the Kunita-Watanabe decomposition again, we have $\langle \hat{p} \rangle_t \le \langle p \rangle_t$ for all $t \ge 0$; hence, for the stopped process $X_t = \hat{p}_{t\wedge\tau}$, $A_t = \langle X \rangle_t \le \langle p \rangle_t$ for all $t \ge 0$. Let $T_0(u) \coloneqq \inf\{t \ge 0 \colon \langle p \rangle_t \ge u\}$ be the inverse clock under no garbling. Then, $T_0(u) \le T(u)$ for all $u \ge 0$.

Lemma 6. *Let W be a Brownian motion with filtration $(\mathcal{G}_u) \coloneqq \mathcal{F}^Y_{T(u)}$ and let $T, \tilde{T}$ be nondecreasing maps. Take an information-time stopping time U and consider the restricted class of stopping times $U' \le U$. If*

$$U' \le U \quad \Longrightarrow \quad \mathbb{E}\left[\tilde{T}(U) - \tilde{T}(U')\right] \;\le\; \mathbb{E}\left[T(U) - T(U')\right], \tag{13}$$

then

$$U \in \operatorname*{arg\,max}_{U' \le U} \left\{\mathbb{E}\left[V(X_0 + W_{U'}) - \lambda T(U')\right]\right\} \quad \Rightarrow \quad U \in \operatorname*{arg\,max}_{U' \le U} \left\{\mathbb{E}\left[V(X_0 + W_{U'}) - \lambda \tilde{T}(U')\right]\right\}.$$

Proof. For any $U' \le U$, the payoff gain of waiting until U rather than stopping at U' under clock T is

$$\Delta_T(U, U') = \mathbb{E}\left[V(X_0 + W_U) - V(X_0 + W_{U'})\right] - \lambda \mathbb{E}\left[T(U) - T(U')\right].$$

The first term does not depend on the clock. By (13) and $\lambda > 0$,

$$\Delta_{\tilde{T}}(U, U') = \mathbb{E}\left[V(X_0 + W_U) - V(X_0 + W_{U'})\right] - \lambda \mathbb{E}\left[\tilde{T}(U) - \tilde{T}(U')\right] \ge \Delta_T(U, U').$$

If U is optimal under T, then $\Delta_T(U, U') \ge 0$ for all $U' \le U$, which implies $\Delta_{\tilde{T}}(U, U') \ge 0$ for all $U' \le U$, proving optimality under $\tilde{T}$. □

Next, we recall our observation (made in the proof of Theorem 1) that garbling slows the posterior's quadratic variation. Recall our convention that $p_t \coloneqq \mathbb{P}(\mu = \mu_h \mid \mathcal{F}^X_t)$ denotes the full-information posterior and let $\hat{p}_t \coloneqq \mathbb{P}(\mu = \mu_h \mid \mathcal{F}^Y_t)$ be the receiver's posterior under an arbitrary (possibly time- and path-dependent) garbling Y, where $\mathcal{F}^Y_t \subseteq \mathcal{F}^X_t$ for all t and $\hat{p}_t = \mathbb{E}[p_t \mid \mathcal{F}^Y_t]$. We also maintain our assumption that

p is a continuous square-integrable martingale with absolutely continuous quadratic variation $d\langle p\rangle_t = \Sigma_0^2(p_t)dt$ for Borel $\Sigma_0\colon (0,1) \to (0,\infty)$.

Lemma 7. *There exists an* $(\mathcal{F}_t^Y)$*-predictable process* $\phi_t \in [0,1]$ *such that* $d\langle \hat{p}\rangle_t = \phi_t d\langle p\rangle_t = \phi_t \Sigma_0^2(p_t)dt$ *for a.e.* t.

Proof. By construction, $\hat{p}$ is the $(\mathcal{F}_t^Y)$-optional projection of p. Since p is a continuous square-integrable martingale, the Kunita-Watanabe decomposition applies: there exists a continuous $(\mathcal{F}_t^Y)$-martingale $\hat{p}$ and a continuous martingale N orthogonal to every $(\mathcal{F}_t^Y)$-martingale (in particular, to $\hat{p}$) such that $p - p_0 = (\hat{p} - p_0) + N$. Taking quadratic variations yields $\langle p\rangle = \langle \hat{p}\rangle + \langle N\rangle$; hence, $d\langle \hat{p}\rangle_t \le d\langle p\rangle_t$. As $d\langle p\rangle_t$ is absolutely continuous with respect to dt, the same is true for $d\langle \hat{p}\rangle_t$, and the Radon-Nikodym derivative $\phi_t := \frac{d\langle \hat{p}\rangle_t}{d\langle p\rangle_t}$ is well-defined dt-a.e. and satisfies $0 \le \phi_t \le 1$. □

A consequence of this lemma is that conditional on the receiver's current information (conditional on $\hat{p}_t$), the expected marginal calendar time per unit posterior quadratic variation is larger under any garbling than under no garbling. To elaborate, we define the (instantaneous) quadratic-variation accumulation rates

$$\beta_t^2 := \frac{d\langle \hat{p}\rangle_t}{dt}, \quad \text{and} \quad \beta_{0,t}^2 := \Sigma_0^2(\hat{p}_t),$$

and the corresponding slowness processes (calendar time per unit of q.v.):

$$s_t := \frac{dt}{d\langle \hat{p}\rangle_t} = \frac{1}{\beta_t^2}, \quad \text{and} \quad s_t^0 := \frac{1}{\Sigma_0^2(\hat{p}_t)}.$$

Corollary 3. *We have* $\frac{1}{\phi_t \Sigma_0^2(p_t)} \ge \frac{1}{\Sigma_0^2(p_t)}$ dt*-a.e. and* $\mathbb{E}\left[s_t \mid \mathcal{F}_t^Y\right] \ge s_t^0$.

Proof. Lemma 7 gives us $d\langle \hat{p}\rangle_t = \phi_t \Sigma_0^2(p_t)dt$, so that $\beta_t^2 = \phi_t \Sigma_0^2(p_t)$; and, thus, $s_t = 1/(\phi_t \Sigma_0^2(p_t)) \ge 1/\Sigma_0^2(p_t)$, since $\phi_t \le 1$. If $\Sigma_0(p) = Kp(1-p)$, then $p \mapsto 1/\Sigma_0^2(p) = K^{-2}p^{-2}(1-p)^{-2}$ is convex on $(0,1)$, and Jensen's inequality together with $\hat{p}_t = \mathbb{E}[p_t \mid \mathcal{F}_t^Y]$ yields the conditional expectation inequality. □

We now prove the proposition by showing that any equilibrium with nontrivial garbling admits a profitable deviation by the sender that will go unpunished.

Proof. If $\lambda = 0$, the result is implied by Corollary 2, so let $\lambda > 0$. Fix a purported equilibrium (Y, τ_s, τ_r, a) and let $\tau = \tau_s \wedge \tau_r$, $X_t = \hat{p}_{t\wedge\tau}$, and $U = \langle X \rangle_\tau$ as above.

Consider the sender deviation in which she switches to no garbling and chooses stopping time $\tau_s' := T_0(U) = \inf\{t \geq 0 \colon \langle p \rangle_t \geq U\}$. Under no garbling, the receiver observes p and may stop at any stopping time τ_r'. The realized terminal time under the deviation is $\tau' = \tau_s' \wedge \tau_r'$.

By DDS, there exists a Brownian motion W such that $X_\tau = X_0 + W_U$. Under no garbling, the full-information posterior satisfies $p_t = p_0 + \widetilde{W}_{\langle p \rangle_t}$ for some Brownian motion $\widetilde{W}$. Coupling $\widetilde{W}$ and W to be the same Brownian motion (on an enlarged probability space if necessary) yields

$$p_{\tau_s'} = p_0 + W_{\langle p \rangle_{\tau_s'}} = p_0 + W_U \overset{d}{=} X_0 + W_U = X_\tau;$$

viz., the deviation embeds the same posterior law $\eta = \mathcal{L}(X_\tau)$ at the sender stopping time τ_s'.

As $T_0(u) \leq T(u)$ for all $u \geq 0$, and by the definition of $\tau_s' = T_0(U)$, we have $\tau_s' \leq T(U) \leq \tau$ a.s., whence $\mathbb{E}[c(\tau_s')] \leq \mathbb{E}[c(\tau)]$ for every increasing c, with strict inequality if c is strictly increasing and the equilibrium features nontrivial garbling ($\langle X \rangle_t < \langle p \rangle_t$ on a set reached with positive probability before τ).

In the original equilibrium, the receiver did not stop strictly before τ on-path. In information time, this means that among all information-time stopping rules $U' \leq U$, stopping at U is optimal under the original clock $T(\cdot)$, because any deviation that induces an earlier realized stopping time corresponds to some $U' \leq U$. Under the deviation, the relevant clock is $T_0(\cdot)$. To apply Lemma 6 (with $\tilde{T} = T_0$), it suffices to verify the incremental dominance condition (13), namely that for every $U' \leq U$,

$$\mathbb{E}\left[T_0(U) - T_0(U')\right] \;\leq\; \mathbb{E}\left[T(U) - T(U')\right].$$

Corollary 3 tells us that dt-a.e., $\mathbb{E}\left[s_t \mid \mathcal{F}_t^Y\right] \geq s_t^0$. Moreover, for $u \leq U$, we have $T(u) \leq \tau$ and, hence, $X_{T(u)} = \hat{p}_{T(u)} = X_0 + W_u$. As $\mathcal{G}_u$ is the information-time filtration (so $\mathcal{G}_u = \mathcal{F}^Y_{T(u)}$), this implies

$$\mathbb{E}\left[s_{T(u)} \mid \mathcal{G}_u\right] = \mathbb{E}\left[s_{T(u)} \mid \mathcal{F}^Y_{T(u)}\right] \;\geq\; \frac{1}{\Sigma_0^2(\hat{p}_{T(u)})} = \frac{1}{\Sigma_0^2(X_0 + W_u)} \quad \text{for a.e. } u \leq U.$$

Now fix any stopping time $U' \leq U$. Changing variables via $dt = s_t\, d\langle \hat{p}\rangle_t$,

$$T(U) - T(U') = \int_{U'}^{U} s_{T(u)}\, du = \int_0^\infty \mathbf{1}_{\{U' < u \leq U\}}\, s_{T(u)}\, du.$$

Taking expectations and using that $\mathbf{1}_{\{U'<u\leq U\}}$ is $\mathcal{G}_u$-measurable, Fubini and the tower property yield

$$\begin{aligned}\mathbb{E}\left[T(U) - T(U')\right] &= \int_0^\infty \mathbb{E}\left[\mathbf{1}_{\{U'<u\leq U\}}\, \mathbb{E}[s_{T(u)} \mid \mathcal{G}_u]\right] du \\ &\geq \int_0^\infty \mathbb{E}\left[\mathbf{1}_{\{U'<u\leq U\}} \frac{1}{\Sigma_0^2(X_0 + W_u)}\right] du = \mathbb{E}\left[\int_{U'}^{U} \frac{1}{\Sigma_0^2(X_0 + W_u)}\, du\right].\end{aligned}$$

Under no garbling, $d\langle p\rangle_t = \Sigma_0^2(p_t)\, dt$, so $dt = (1/\Sigma_0^2(p_t))\, d\langle p\rangle_t$ and, therefore,

$$T_0(U) - T_0(U') = \int_{U'}^{U} \frac{1}{\Sigma_0^2(p_{T_0(u)})}\, du.$$

By the same DDS coupling used above, $p_{T_0(u)} = p_0 + W_u = X_0 + W_u$ for $u \leq U$, thus,

$$\mathbb{E}\left[T_0(U) - T_0(U')\right] = \mathbb{E}\left[\int_{U'}^{U} \frac{1}{\Sigma_0^2(X_0 + W_u)}\, du\right] \;\leq\; \mathbb{E}\left[T(U) - T(U')\right],$$

which is exactly (13). Lemma 6 then implies that U remains optimal for the receiver under the tighter clock T_0. Therefore, the receiver has a best response with $\tau'_r \geq \tau'_s$ a.s., so the realized stopping time under the deviation is $\tau' = \tau'_s$ and the embedded law is indeed η.

Since the deviation preserves the embedded posterior law (and so preserves the distribution of the receiver's terminal action under optimal play) while weakly reducing $\mathbb{E}[c(\tau)]$ (strictly if garbling is nontrivial), the sender's payoff weakly increases (strictly under nontrivial garbling). This contradicts equilibrium. We conclude that no equilibrium can feature nontrivial garbling on the continuation path. $\square$

References

Yunus C. Aybas and Steven Callander. Cheap talk in complex environments. Mimeo, 2024.

Arjada Bardhi and Yingni Guo. Modes of persuasion toward unanimous consent. *Theoretical Economics*, 13(3):1111–1149, 2018.

César Barilla. When and what to learn in a changing world. Mimeo, 2025.

Richard F Bass. *Diffusions and elliptic operators*. Springer, 1998.

Jacopo Bizzotto, Jesper Rüdiger, and Adrien Vigier. Dynamic persuasion with outside information. *American Economic Journal: Microeconomics*, 13(1):179–194, 2021.

Alexander W. Bloedel and Weijie Zhong. The cost of optimally acquired information. Mimeo, 2020.

Isabelle Brocas and Juan D. Carrillo. Influence through ignorance. *RAND Journal of Economics*, 38(4):931–947, 2007.

Yeon-Koo Che, Kyungmin Kim, and Konrad Mierendorff. Keeping the listener engaged: A dynamic model of Bayesian persuasion. *Journal of Political Economy*, 131(7):1797–1844, 2023.

Daniel Chen and Weijie Zhong. Information acquisition and time-risk preference. *American Economic Review: Insights*, 7(2):213–230, 2025.

Jeffrey C. Ely, George Georgiadis, and Luis Rayo. Feedback design in dynamic moral hazard. *Econometrica*, 93(2):597–621, March 2025.

Matteo Escudé and Ludvig Sinander. Slow persuasion. *Theoretical Economics*, 18(1): 129–162, 2023.

Drew Fudenberg, Philipp Strack, and Tomasz Strzalecki. Speed, accuracy, and the optimal timing of choices. *American Economic Review*, 108(12):3651–3684, December 2018.

Alkis Georgiadis-Harris. Preparing to act: Information acquisition and the timing of actions. Mimeo, 2023.

Duarte Gonçalves. Sequential sampling equilibrium. Mimeo, 2023.

Duarte Gonçalves. Speed, accuracy, and complexity. Mimeo, 2024.

Yingni Guo and Eran Shmaya. The interval structure of optimal disclosure. *Econometrica*, 87(2):653–675, 2019.

István Gyöngy. Mimicking the one-dimensional marginal distributions of processes having an itô differential. *Probability theory and related fields*, 71(4):501–516, 1986.

Emeric Henry and Marco Ottaviani. Research and the approval process: The organization of persuasion. *American Economic Review*, 109(3):911–955, 2019.

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

Emir Kamenica and Xiao Lin. Commitment and randomization in communication. Mimeo, 2024.

Xuelin Li, Martin Szydlowski, and Fangyuan Yu. Dynamic information design in an entry game. *Journal of Economic Theory*, 226:106018, 2025.

Robert S Liptser and Albert N Shiryaev. *Statistics of random processes: I. General theory*, volume 5. Springer Science & Business Media, 2013.

Andrew McClellan. Experimentation and approval mechanisms. *Econometrica*, 90 (5):2215–2247, 2022.

Teddy Mekonnen and Bobak Pakzad-Hurson. Competition, persuasion, and search. Mimeo, 2025.

Teddy Mekonnen, Zeky Murra-Anton, and Bobak Pakzad-Hurson. Persuaded search. *Journal of Political Economy*, 133(10):3167–3207, 2025.

Stephen Morris and Philipp Strack. The wald problem and the relation of sequential sampling and ex-ante information costs. Mimeo, 2019.

Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion*, volume 293. Springer Science & Business Media, 2013.

Hermann Rost. Skorokhod stopping times of minimal variance. In *Séminaire de Probabilités X Université de Strasbourg*, pages 194–208. Springer, 2006.

A Wald. Sequential tests of statistical hypotheses. *The Annals of Mathematical Statistics*, 16(2):117–186, 1945.

Weijie Zhong. Optimal dynamic information acquisition. *Econometrica*, 90(4):1537–1582, 2022.

Appendix B. Supplemental (Online) Material

B.1. **Local-Time Identity.** We record the following local-time identity.

Lemma 8. *Let X be a continuous local martingale in natural scale started at $X_0 = x \in (0,1)$, and let τ be a stopping time such that $X_\tau \sim \nu$, where ν is a probability measure supported on $[0,1]$. For each fixed $y \in [0,1]$ define the Borel measure on $[0,\infty)$*

$$\Lambda_y(A) := \mathbb{E}\left[\int_A dL_s^y\right], \quad \textit{for Borel } A \subset [0,\infty),$$

where L_s^y is the local time of X at level y. Then Λ_y is a finite Borel measure and

$$\Lambda_y([0,\infty)) = \mathbb{E}[L_\tau^y] = U^\nu(y) - |x-y| < \infty,$$

where $U^\nu(y) := \int |z-y|\,\nu(dz)$ is the potential of the measure.

Proof of Lemma 8. Apply Itô-Tanaka to the convex function $z \mapsto |z-y|$ and take expectations at the stopping time τ. Since $X_\tau \sim \nu$ and the local martingale term has zero expectation, we obtain

$$\mathbb{E}[|X_\tau - y|] = |x-y| + \mathbb{E}[L_\tau^y].$$

Thus, $\mathbb{E}[L_\tau^y] = \int |z-y|\,\nu(dz) - |x-y| = U^\nu(y) - |x-y|$. Because ν is a probability on $[0,1]$, $U^\nu(y) < \infty$ for every $y \in [0,1]$, so $\mathbb{E}[L_\tau^y] < \infty$. Finally, define $\Lambda_y(A) := \mathbb{E}[\int_A d_s L_s^y]$. By monotone convergence, the total mass of this finite measure equals $\mathbb{E}[L_\tau^y]$. $\square$

B.2. **Closed Forms.** If we assume that the cost c is Laplace-transformable, we can derive the closed-form expected cost of an arbitrary (Bayes-plausible) optimally-implemented binary distribution with support $\{\underline{P}, \bar{P}\}$. Specifically, we know that no garbling is optimal, which means that we merely need to compute the Laplace transform of the expected hitting time of the posterior diffusion leaving $[\underline{P}, \bar{P}]$, which allows us to derive the formula for the cost.[21]

[21]To elaborate, we compute the Laplace transform of the exit time τ of the posterior process from $[\underline{P}, \bar{P}]$ and then represent $\mathbb{E}[c(\tau)]$ as an integral over this transform. A direct representation using the exit density of a diffusion from two fixed levels can also be done, but the density, although very

Proposition 7. *Suppose c is Laplace-transformable: $c(t) = \int_0^\infty e^{-st}\mu(ds)$ for some Borel measure μ on $\mathbb{R}_+$. Then, $\mathbb{E}[c(\tau)] = \int_0^\infty \phi_s(p_0)\mu(ds)$, where*

$$\phi_s(p) = \frac{\sqrt{p(1-p)}}{e^{\frac{\gamma(s)}{2}\ln\frac{\bar{x}}{\underline{x}}} - e^{-\frac{\gamma(s)}{2}\ln\frac{\bar{x}}{\underline{x}}}}\left[\frac{e^{\frac{\gamma(s)}{2}\ln\frac{\bar{x}}{x}} - e^{-\frac{\gamma(s)}{2}\ln\frac{\bar{x}}{x}}}{\sqrt{\underline{P}(1-\underline{P})}} + \frac{e^{\frac{\gamma(s)}{2}\ln\frac{x}{\underline{x}}} - e^{-\frac{\gamma(s)}{2}\ln\frac{x}{\underline{x}}}}{\sqrt{\bar{P}(1-\bar{P})}}\right],$$

where $\gamma(s) = \sqrt{1 + \frac{8s\sigma^2}{(\mu_h-\mu_l)^2}}$.

Proof. Let $\tau \coloneqq \inf\{t \geq 0 \colon p_t \notin [\underline{P}, \bar{P}]\}$ be the first exit time from the interval. We are interested in $\mathbb{E}[c(\tau)]$ where $c(t)$ is Laplace-transformable, i.e., $c(t) = \int_0^\infty e^{-st}\mu(ds)$ for some Borel measure μ on $\mathbb{R}_+$.

By Fubini's theorem,

$$\mathbb{E}[c(\tau)] = \int_0^\infty \phi_s(p_0)\,\mu(ds), \quad \text{for} \quad \phi_s(p_0) \coloneqq \mathbb{E}_{p_0}[e^{-s\tau}].$$

By the Feynman–Kac formula, for each fixed $s > 0$, the function $\phi_s(p) \coloneqq \mathbb{E}_p[e^{-s\tau}]$ satisfies

$$\frac{1}{2}\left(\frac{\mu_h-\mu_l}{\sigma}\right)^2 p^2(1-p)^2\,\phi_s''(p) - s\phi_s(p) = 0, \quad \text{where} \quad \phi_s(\underline{P}) = \phi_s(\bar{P}) = 1.$$

Introduce the odds transform $x \coloneqq p/(1-p)$, which implies $p = x/(1+x)$. The derivatives transform as $d/dp = (1+x)^2 d/dx$ and $d^2/dp^2 = (1+x)^4 d^2/dx^2 + 2(1+x)^3 d/dx$. Defining the auxiliary function $g_s(x) \coloneqq (1+x)\phi_s(p(x))$, the ODE simplifies to the Cauchy-Euler equation:

$$x^2 g_s''(x) - \frac{2s\sigma^2}{(\mu_h-\mu_l)^2} g_s(x) = 0.$$

The general solution is $g_s(x) = Ax^{m_+} + Bx^{m_-}$. The characteristic exponents $m_\pm$ are determined by the roots of the indicial equation $m(m-1) = \frac{2s\sigma^2}{(\mu_h-\mu_l)^2}$. Explicitly, $m_\pm = \frac{1\pm\gamma(s)}{2}$, where the discriminant is:

$$\gamma(s) \coloneqq \sqrt{1 + \frac{8s\sigma^2}{(\mu_h-\mu_l)^2}}.$$

well known, is not particularly amenable to analysis since it is a sine series. We opt for the Laplace transformation route for the relative ease of representation.

The boundary conditions $\phi_s(\underline{P}) = \phi_s(\bar{P}) = 1$ transform to the linear system:

$$\begin{cases} A\underline{x}^{m_+} + B\underline{x}^{m_-} = 1 + \underline{x} \\ A\bar{x}^{m_+} + B\bar{x}^{m_-} = 1 + \bar{x}, \end{cases}$$

where $\underline{x}$ and $\bar{x}$ represent the odds at the boundaries. Solving for A and B via Cramer's rule yields a determinant of $D = -2\sqrt{\underline{x}\bar{x}}\sinh\left(\frac{\gamma(s)}{2}\ln\frac{\bar{x}}{\underline{x}}\right)$.

Algebraic simplification of the numerator groups the terms by boundary coefficients, leading to the explicit solution for the Laplace transform:

$$\phi_s(p) = \frac{\sqrt{p(1-p)}}{\sinh\left(\frac{\gamma(s)}{2}\ln\frac{\bar{x}}{\underline{x}}\right)}\left[\frac{\sinh\left(\frac{\gamma(s)}{2}\ln\frac{\bar{x}}{x}\right)}{\sqrt{\underline{P}(1-\underline{P})}} + \frac{\sinh\left(\frac{\gamma(s)}{2}\ln\frac{x}{\underline{x}}\right)}{\sqrt{\bar{P}(1-\bar{P})}}\right].$$

Expanding the hyperbolic sine terms provides the equivalent exponential representation:

$$\phi_s(p) = \frac{\sqrt{p(1-p)}}{e^{\frac{\gamma(s)}{2}\ln\frac{\bar{x}}{\underline{x}}} - e^{-\frac{\gamma(s)}{2}\ln\frac{\bar{x}}{\underline{x}}}}\left[\frac{e^{\frac{\gamma(s)}{2}\ln\frac{\bar{x}}{x}} - e^{-\frac{\gamma(s)}{2}\ln\frac{\bar{x}}{x}}}{\sqrt{\underline{P}(1-\underline{P})}} + \frac{e^{\frac{\gamma(s)}{2}\ln\frac{x}{\underline{x}}} - e^{-\frac{\gamma(s)}{2}\ln\frac{x}{\underline{x}}}}{\sqrt{\bar{P}(1-\bar{P})}}\right].$$

Finally, for any Laplace-transformable cost c, the expected cost is recovered via $\mathbb{E}\left[c(\tau)\right] = \int_0^\infty \phi_s(p_0)\mu(ds)$. □

If we specialize to the case where the cost is linear in time, $c(t) = t$, Proposition 7's formula simplifies further. Let $\mathcal{A}$ be the infinitesimal generator of the posterior diffusion under no garbling, so that for smooth f, $(\mathcal{A}f)(p) = \frac{1}{2}k^2p^2(1-p)^2f''(p)$, where $k \coloneqq (\mu_h - \mu_l)/\sigma$. A linear time cost corresponds to a unit running cost per unit time, so we look for a twice continuously-differentiable function $\mathfrak{T}$ on $(0,1)$ that solves the Poisson equation $\mathcal{A}\mathfrak{T}(p) = 1$ for $p \in \left(\underline{P}, \bar{P}\right)$. We define $\psi \coloneqq \mathfrak{T}'$ to be the derivative of this solution. Then,

Proposition 8. *We have*

$$\psi(p) = \int_{p_0}^{p} \frac{2}{\left(\frac{\mu_h-\mu_l}{\sigma}\right)^2 r^2(1-r)^2}\,dr.$$

Moreover,

$$\mathbb{E}[\tau] = (1-m_\nu)\int_{p_0}^{\underline{P}}\psi(s)ds + m_\nu\int_{p_0}^{\bar{P}}\psi(s)ds.$$

Integrating fully and letting $\mathfrak{L}(p) := (2p-1)\ln\left(\frac{p}{1-p}\right)$, *we have*

$$\mathbb{E}[\tau] = \frac{2\sigma^2}{(\mu_h-\mu_l)^2}\left(\frac{\bar{P}-p_0}{\bar{P}-\underline{P}}\,\mathfrak{L}(\underline{P}) + \frac{p_0-\underline{P}}{\bar{P}-\underline{P}}\,\mathfrak{L}(\bar{P}) - \mathfrak{L}(p_0)\right).$$

Toward proving the proposition, we first record a general identity for the expected embedding time of the posterior diffusion by solving the associated generator equation.

Lemma 9. *Let* p_t *solve*

$$dp_t = k\,p_t(1-p_t)\,dW_t, \quad \textit{with} \quad p_0 \in (0,1), \quad \textit{and} \quad k := \frac{\mu_h-\mu_l}{\sigma}.$$

Let $\mathfrak{T}(p)$ *be a solution to the differential equation* $\frac{1}{2}k^2p^2(1-p)^2\mathfrak{T}''(p) = 1$. *Then, for any stopping time* τ *where* p_τ *is bounded away from* 0 *and* 1, *we have*

$$\mathbb{E}[\tau] = \mathbb{E}[\mathfrak{T}(p_\tau)] - \mathfrak{T}(p_0).$$

Proof. The infinitesimal generator of p_t is given by $\mathcal{A}f(p) = \frac{1}{2}k^2p^2(1-p)^2f''(p)$. By hypothesis, $\mathcal{L}\mathfrak{T}(p) = 1$. Applying Itô's formula to $\mathfrak{T}(p_t)$, we obtain

$$d\mathfrak{T}(p_t) = \mathfrak{T}'(p_t)\,dp_t + \mathcal{L}\mathfrak{T}(p_t)\,dt = \mathfrak{T}'(p_t)\,dp_t + dt.$$

Integrating from 0 to τ yields $\mathfrak{T}(p_\tau) - \mathfrak{T}(p_0) = \int_0^\tau \mathfrak{T}'(p_t)\,dp_t + \tau$. Taking expectations and applying the optional stopping theorem to the martingale term $\int \mathfrak{T}'(p_t)\,dp_t$, we find $\mathbb{E}[\mathfrak{T}(p_\tau) - \mathfrak{T}(p_0)] = \mathbb{E}[\tau]$. □

Proof of Proposition 8. We explicitly construct $\mathfrak{T}(p)$ by integrating the equation $\mathfrak{T}''(p) = \frac{2}{k^2p^2(1-p)^2}$. Using the partial fraction decomposition $\frac{1}{p^2(1-p)^2} = \frac{1}{p^2} + \frac{2}{p} + \frac{2}{1-p} + \frac{1}{(1-p)^2}$, the first integration yields

$$\begin{aligned}\mathfrak{T}'(p) &= \frac{2}{k^2}\int\left(\frac{1}{p^2} + \frac{2}{p} + \frac{2}{1-p} + \frac{1}{(1-p)^2}\right)dp \\ &= \frac{2}{k^2}\left[\frac{2p-1}{p(1-p)} + 2\ln\left(\frac{p}{1-p}\right)\right] + C_1.\end{aligned}$$

We integrate a second time to determine $\mathfrak{T}(p)$. For the rational term, we observe that $\int \frac{2p-1}{p(1-p)}\,dp = -\ln(p(1-p))$. For the logarithmic term, integration by parts yields $\int 2\ln\left(\frac{p}{1-p}\right)\,dp = 2p\ln p + 2(1-p)\ln(1-p) - 2$. Combining these terms and absorbing constants into the linear coefficients, we recover the particular solution

$$\begin{aligned}\mathfrak{L}(p) &= -\ln p - \ln(1-p) + 2p\ln p + 2(1-p)\ln(1-p)\\ &= (2p-1)\ln\left(\frac{p}{1-p}\right).\end{aligned}$$

The general solution is, therefore, $\mathfrak{T}(p) = \frac{2}{k^2}\mathfrak{L}(p) + Ap + B$.

We now evaluate $\mathbb{E}[\tau] = \mathbb{E}[\mathfrak{T}(p_\tau) - \mathfrak{T}(p_0)]$. Since p_t is a martingale, $\mathbb{E}[p_\tau] = p_0$, which implies that the linear component vanishes from the expectation:

$$\mathbb{E}[Ap_\tau + B] - (Ap_0 + B) = A(\mathbb{E}[p_\tau] - p_0) = 0.$$

Consequently, $\mathbb{E}[\tau]$ depends only on the nonlinear term $\mathfrak{L}(p)$. Substituting $k = (\mu_h - \mu_l)/\sigma$, we obtain

$$\mathbb{E}[\tau] = \frac{2\sigma^2}{(\mu_h - \mu_l)^2}\left[\mathbb{E}[\mathfrak{L}(p_\tau)] - \mathfrak{L}(p_0)\right].$$

Plugging in the two-atom value of p_τ, we have

$$\mathbb{E}[\tau] = \frac{2\sigma^2}{(\mu_h - \mu_l)^2}\left(\frac{\bar{P} - p_0}{\bar{P} - \underline{P}}\,\mathfrak{L}(\underline{P}) + \frac{p_0 - \underline{P}}{\bar{P} - \underline{P}}\,\mathfrak{L}(\bar{P}) - \mathfrak{L}(p_0)\right),$$

as desired. □

These propositions highlight how useful Theorem 1 and Proposition 2 are practically. With these propositions in hand, the persuasion problem is now a simple one-dimensional problem, parametrized only by $\underline{P}$, and which can, thus, be solved directly for any (Laplace-transformable) c. Concretely, the sender's problem simplifies to

$$\max_{\underline{P}\in[0,p_0)}\left\{\frac{p_0 - \underline{P}}{\bar{P} - \underline{P}} - \int_0^\infty \phi_s(p_0)\mu(ds)\right\},$$

with Proposition 7 detailing the closed form of ϕ_s.

B.3. **Denseness of Laplace Transformable Costs.** Define

$$\mathcal{C}_\infty := \left\{c\colon\ [0,\infty) \to [0,\infty)\ \middle|\ c \text{ is increasing, convex, differentiable, and finite on } [0,\infty)\right\},$$

and

$$\mathbf{L} := \left\{c \in \mathcal{C}_\infty\ \middle|\ \exists\,\delta > 0 : \int_0^\infty |c(t)|e^{-\delta t}dt < \infty\right\}.$$

Lemma 10. *Let τ be an a.s. finite stopping time. Then, for any $c \in \mathcal{C}_\infty$, there exists a family of functions $\tilde{c}^{(T,\delta,\eta)} \in \mathbf{L}$ with $T, \delta, \eta > 0$ such that*

$$\lim_{\eta\to 0}\lim_{\delta\to 0}\lim_{T\to\infty}\sup_{t\in[0,T]}|c(t)-\tilde{c}^{(T,\delta,\eta)}(t)| = 0, \quad \text{and} \quad \lim_{\eta\to 0}\lim_{\delta\to 0}\lim_{T\to\infty}\mathbb{E}\left[\left|c(\tau)-\tilde{c}^{(T,\delta,\eta)}(\tau)\right|\right] = 0.$$

Proof. For $T > 0$, define the truncation $c^{(T)}(t) := c(t \wedge T), \quad t \geq 0$. Since $\tau < \infty$ a.s., the tail beyond T satisfies

$$\mathbb{E}[|c(\tau) - c^{(T)}(\tau)|] \to 0 \quad \text{as } T \to \infty.$$

Let $\rho \in C_c^\infty(\mathbb{R})$ be a standard mollifier and define $\rho_\delta(t) := \delta^{-1}\rho(t/\delta)$. Extend $c^{(T)}$ outside $[0,T]$ by setting $c^{(T)}(t) = c(T)$ for $t > T$. Define the mollified function

$$c_\delta^{(T)}(t) := (c^{(T)} * \rho_\delta)(t), \quad \text{for all } t \geq 0.$$

Then $c_\delta^{(T)} \in C^\infty([0,\infty))$, and $c_\delta^{(T)} \to c^{(T)}$ uniformly on $[0,T]$ as $\delta \to 0$, preserving convexity and monotonicity.

Define finally

$$\tilde{c}^{(T,\delta,\eta)}(t) := c_\delta^{(T)}(t)e^{-\eta t}, \quad \text{for all } t \geq 0.$$

For any $\eta > 0$, $\tilde{c}^{(T,\delta,\eta)} \in \mathbf{L}$, because

$$\int_0^\infty |\tilde{c}^{(T,\delta,\eta)}(t)|e^{-st}dt < \infty \quad \text{for some } s > 0.$$

For any stopping time $\tau < \infty$ a.s.:

$$\mathbb{E}[|c(\tau)-\tilde{c}^{(T,\delta,\eta)}(\tau)|] \leq \mathbb{E}[|c(\tau)-c^{(T)}(\tau)|]+\mathbb{E}[|c^{(T)}(\tau)-c_\delta^{(T)}(\tau)|]+\mathbb{E}[|c_\delta^{(T)}(\tau)-\tilde{c}^{(T,\delta,\eta)}(\tau)|].$$

The first term goes to 0 as $T \to \infty$. The second term goes to 0 as $\delta \to 0$ (uniform convergence on $[0,T]$). The third term goes to 0 as $\eta \to 0$ (exponential damping).

Taking limits in the order $T \to \infty$, $\delta \to 0$, $\eta \to 0$ delivers the result. $\square$