Bayesian Persuasion with Selective Disclosure*

Yifan Dai† Drew Fudenberg‡ Harry Pei§

August 7, 2026

Abstract: A sender first publicly commits to an experiment and then can privately run additional experiments and selectively disclose their outcomes to a receiver. The sender has private information about the maximal number of additional experiments they can perform (i.e., their *type*). We show that the sender cannot attain their commitment payoff in any equilibrium if (i) the receiver is sufficiently uncertain about their type and (ii) the sender could benefit from selective disclosure after conducting their full-commitment optimal experiment. Otherwise, there can be equilibria where the sender obtains their commitment payoff.

Keywords: Bayesian persuasion, lack of commitment, selective disclosure

JEL Codes: C73, D82, D83.

*We thank Pak Hung Au, Federico Echenique, Jeffrey Ely, Ying Gao, Alexander Jakobsen, Navin Kartik, Xiao Lin, Elliot Lipnowski, Stephen Morris, Paula Onuchic, Alessandro Pavan, Refine.ink, Jean Tirole, Alexander Wolitzky, Kun Zhang, Gabriel Ziegler, and two anonymous referees for helpful comments as well as NSF grants SES-2337566 and SES-2417162 for financial support.

†Department of Economics, Massachusetts Institute of Technology. Email: yfdai@mit.edu

‡Department of Economics, Massachusetts Institute of Technology. Email: drew.fudenberg@gmail.com

§Department of Economics, Northwestern University. Email: harrydp@northwestern.edu

1 Introduction

The credibility of experts depends on their ability to commit to how they acquire and disclose information. For instance, a prosecutor who primarily seeks convictions but lacks commitment power cannot credibly convey useful information to a judge. But when the prosecutor can commit to an information-gathering process and to fully disclosing its results, as in the model of Kamenica and Gentzkow (2011), their statements may influence the judge's decision.

In practice, even when experts can commit to acquiring and truthfully disclosing certain types of information, they may be unable to commit not to conduct additional investigations and not to conceal unfavorable findings. This issue arises in the pharmaceutical industry, where some clinical trials are pre-registered with the FDA and must be reported, but companies can also conduct additional trials after approval (also known as *Phase 4 trials*) in order to monitor the drug's performance in real-world populations. Because these post-approval trials are not pre-registered, companies sometimes conceal or delay the publication of their findings, even though they are legally required to report the results of all trials.[1] This can undermine the expert's credibility, as decision-makers (e.g., doctors, patients) may suspect that unfavorable evidence has been withheld and therefore discount the expert's statements.

Motivated by this concern, we study a model where a sender (the expert) publicly commits to an *initial experiment* and discloses its outcome to a receiver. Unlike in the full commitment benchmark of Kamenica and Gentzkow (2011), the sender in our model cannot commit not to privately conduct additional experiments and conceal unfavorable results. The sender has private information about the maximum number of additional experiments they can run, which is referred to as their *type* or *capacity*.

Our question is when the sender's limited commitment prevents them from attaining their payoff under full commitment (i.e., their *commitment payoff*). Our results show that the answer depends on two factors: (i) the receiver's prior uncertainty about the sender's

[1] DeVito, Bacon, and Goldacre (2020) reported that, among 4,209 trials obligated to submit results to the FDA between 2018 and 2019, only 722 (40.9%) complied within the mandated 1-year time frame. The pain medication Vioxx is a high-profile example of delayed reporting: Its manufacturer, Merck, delayed releasing data from post-approval trials that linked the drug to heart problems. Another example is GSK's antidepressant Paxil—legal filings revealed that the company concealed evidence from clinical trials showing that the drug could induce suicidal thoughts among teenagers. Compared with hiding or delaying disclosure, outright falsification of data is less common and typically subject to harsher penalties.

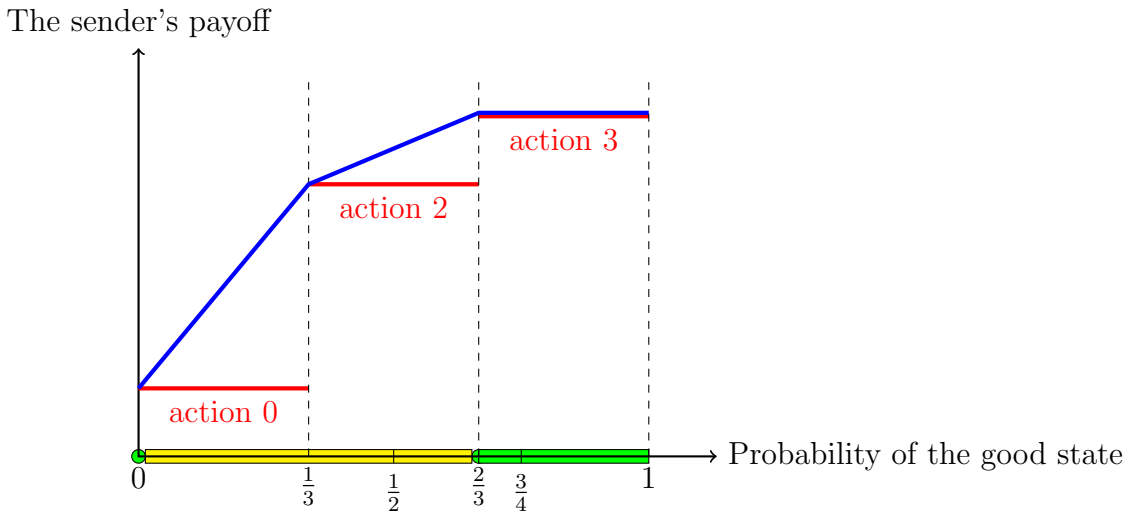

Figure 1: The sender's utility as a function of the receiver's belief (red) and its concave closure (blue), with non-credible beliefs in yellow and credible beliefs in green.

type and (ii) whether the sender benefits from selectively disclosing more information after conducting their full-commitment optimal experiment (hereafter, *the optimal experiment*).

To illustrate our findings, consider the example depicted in Figure 1: There are two states, *good* and *bad*, and three actions 0, 2, and 3. The sender's utility equals the receiver's action. The receiver's optimal action is 0 when the probability of the good state is less than $\frac{1}{3}$, 3 when the probability of the good state is more than $\frac{2}{3}$, and 2 otherwise.

We first show that the sender's limited commitment matters only when the receiver faces uncertainty about the sender's capacity. If the receiver knows that the sender can conduct at most t additional experiments, then the sender can prove that no information is hidden by conducting t uninformative additional experiments and disclosing all the outcomes.

When the receiver does not know the sender's capacity, the sender's ability to attain their commitment payoff hinges on the beliefs induced by the optimal experiment, in particular, whether the sender can benefit from selectively disclosing more information at any of those beliefs. In the example, the sender can benefit from disclosing more information at a belief that assigns probability $\frac{1}{2}$ to the good state: They can conduct an additional experiment that reveals the state and disclose only the outcome which proves that the state is good. This raises the receiver's action from 2 to 3 when the state is good and keeps it at 2 when

the state is bad. In contrast, the sender cannot benefit from disclosing more information at beliefs that assign probability 0 or $\frac{3}{4}$ to the good state since the receiver's action will always be 0 in the first case and will never exceed 3 in the second case. We refer to beliefs where the sender could gain by revealing more information as *non-credible beliefs*. In this example, these are the beliefs that assign probability strictly between 0 and $\frac{2}{3}$ to the good state.

Theorem 1 shows that in *monotone environments* such as our example, the sender can attain their commitment payoff under all type distributions if and only if the optimal experiment does not induce a non-credible belief. If the optimal experiment induces at least one non-credible belief, then there is an open set of type distributions under which all of the sender's equilibrium payoffs are uniformly bounded below their commitment payoff.

Theorem 2 provides a complementary result that does not require monotonicity. It focuses on environments where the sender strictly prefers to fully disclose a certain state θ to any belief induced by the optimal experiment when the state is θ. This is more demanding than the optimal experiment to induce a non-credible belief, which holds in our example when the good state occurs with probability below 1/3. We show that if sufficiently many sender types occur with probability bounded away from 0, the sender's payoff in every equilibrium is uniformly bounded below their commitment payoff. The proof bounds a higher type's payoff from below using the following deviation: Imitate a lower type, then run a fully informative experiment and disclose its outcome exactly when disclosure is profitable. If the optimal experiment induces beliefs that satisfy our conditions, then in every equilibrium where the sender attains their commitment payoff, the higher type's payoff from our deviation exceeds the lower type's equilibrium payoff by a constant. Since payoffs are bounded, a contradiction arises once there are sufficiently many types that occur with positive probability. Theorem 3 then shows that when one sender type occurs with probability close to 1, there is an equilibrium where the sender's payoff is close to their commitment payoff.

Our paper contributes to the persuasion literature by considering a novel form of limited commitment: the sender cannot commit not to conduct additional experiments and selectively disclose their outcomes. Our setting contrasts to ones where the sender either cannot commit to a public experiment (Goltsman et al. (2009), Henry and Ottaviani (2019), Pei (2023), Best and Quigley (2024), Lin and Liu (2024), Mathevet et al. (2024), Titova and Zhang (2025),

Corrao and Dai (2025)) or cannot commit not to truthfully reveal the outcome of the public experiment (Lipnowski et al. (2022)), Kreutzkamp and Lou (2025), Lyu and Suen (2025)).

Arieli and Stewart (2025) studies a model that has economic frictions similar to the ones in ours.[2] Unlike this paper, it allows for state-dependent sender preferences, and shows that the sender obtains their commitment payoff in a weak Perfect Bayesian equilibrium (wPBE), a solution concept that allows for arbitrary off-path beliefs. In contrast, our results imply that the sender may be unable to obtain their commitment payoff when the receiver's beliefs are consistent with the evidence disclosed at *all* histories. In particular we require that the receiver assigns zero probability to any state that is ruled out by a disclosed signal.

Our work also relates to Bayesian persuasion models where the sender has private information (e.g., Perez-Richet, 2014; Koessler and Skreta, 2023; Zapechelnyuk, 2023). Unlike these papers, the sender's private information in our model is payoff-irrelevant for the receiver. Our work also relates to the literature on hard evidence disclosure following Dye (1985), where the receiver is uncertain about the amount of evidence the sender possesses (e.g., Dziuda, 2011; Gao, 2025; Rappoport, 2025; Antic and Pei, 2026). Unlike these models, the evidence in our model is endogenously generated by the sender's choice of private experiments.[3]

2 The Baseline Model

A sender (s) interacts with a receiver (r). The sender's payoff $u^s(a)$ depends only on the receiver's action $a \in A$, so they have *transparent motives* in the sense of Lipnowski and Ravid (2020). The receiver's payoff $u^r(\theta, a)$ depends on both a and the state of the world $\theta \in \Theta$. The state is initially unknown to both players, and is drawn according to a distribution $\pi_0 \in \Delta(\Theta)$ that has full support. We assume that Θ and A are finite sets.

An *experiment* σ is a mapping from Θ to the set of probability distributions over outcomes $s \in S$, where S is a countably infinite set. Let Σ denote the collection of all experiments, and for each experiment $\sigma \in \Sigma$, let $\sigma(s \mid \theta)$ denote the probability of s conditional on θ.

At the beginning of the game, the sender chooses an *initial experiment* σ_0, and then both

[2]Related economic frictions are also considered in Matthews and Postlewaite (1985), Felgenhauer and Schulte (2014), Felgenhauer and Loerke (2017), Ben-Porath et al. (2018), and Lou (2023).

[3]Pei (2025) studies disclosure of endogenously generated signals in a repeated game.

players observe σ_0 and its realized outcome $s_0 \in S$. After observing (σ_0, s_0), the sender may conduct some *additional experiments*.[4] The outcomes of different experiments are assumed to be independent conditional on θ. The maximal number of additional experiments that the sender can conduct is $t \in \{0, 1, ...\} \equiv \mathbb{N}$, which is drawn according to $p \in \Delta(\mathbb{N})$. The sender privately observes t after choosing σ_0. We refer to t as the sender's (interim) *type*.

The sender conducts additional experiments sequentially after observing (σ_0, s_0, t), so their choice of the kth additional experiment can depend on (σ_0, s_0, t) as well as the first $k-1$ additional experiments and their outcomes $\{\sigma_i, s_i\}_{i=1}^{k-1}$. The sender then chooses a subset of outcomes of the additional experiments to disclose before the receiver chooses $a \in A$.

We assume that the receiver observes an additional experiment and its outcome if and only if the sender discloses that outcome. Hence, the receiver cannot directly observe t, the number of additional experiments the sender conducted, and the contents and outcomes of the undisclosed additional experiments.

Let $\mathcal{H}_k$ denote the set of sequences that consist of k pairs of an experiment σ together with one of its outcomes s. Let $\mathcal{H} \equiv \bigcup_{k=1}^{+\infty} \mathcal{H}_k$ with a typical element denoted by $h \in \mathcal{H}$. The receiver's information sets are elements of $\mathcal{H}$ corresponding to the disclosed experiments. The receiver's strategy $\alpha : \mathcal{H} \to \Delta(A)$ maps their information sets to their actions.

The sender's information sets other than the initial node belong to $\mathcal{H} \times \mathbb{N}$. Their pure strategy consists of (i) an initial experiment $\sigma_0 \in \Sigma$, (ii) a mapping from every type $t \geq 1$ and every $h \in \mathcal{H}_k$ with $1 \leq k \leq t$ to $\Sigma \cup \{\emptyset\}$ that captures the sender's choice of the next experiment given their type and the outcomes of the previous experiments, where $\emptyset$ stands for conducting no further experiment, and (iii) a mapping from every type $t \geq 1$ and every $h \in \mathcal{H}_k$ with $1 \leq k \leq t+1$ to subsets of $\{1, 2, ..., k-1\}$, where i belongs to the chosen subset when the ith additional experiment is disclosed. The sender's *mixed strategy* is a distribution of their pure strategies.

The receiver's *naive belief* at $h \in \mathcal{H}$, denoted by $\widehat{\pi}_h \in \Delta(\Theta)$, is defined as their belief about θ if they observe h and believe that the sender has disclosed the outcomes of all

[4]We include an initial experiment to facilitate comparison with the Bayesian persuasion literature following Kamenica and Gentzkow (2011) and to reflect our motivating application, in which pharmaceutical companies conduct pre-registered trials before Phase 4 trials. All of our results except Theorem 3 extend to settings without an initial experiment provided that the sender can conduct at least one additional experiment.

the experiments they conducted. The receiver's naive belief need not coincide with their posterior belief since they may suspect that the sender has concealed some outcomes from the additional experiments.

Our game is neither finite nor a multistage game with observable actions, so neither sequential equilibrium (Kreps and Wilson (1982)) nor Perfect Bayesian Equilibrium (Fudenberg and Tirole (1991)) directly applies. Instead, we will use Fudenberg and Tirole (1991)'s *Perfect Extended Bayesian Equilibrium* (PEBE, or sometimes, equilibrium) as our solution concept. A PEBE for our game consists of a strategy for each player and a mapping from $\mathcal{H}$ to $\Delta(\Theta \times \mathbb{N})$ that captures the receiver's belief about (θ, t), such that (i) at every information set, each player's strategy best replies to their opponent's strategy given their belief, (ii) the receiver's belief is consistent with some conditional probability system induced by the sender's strategy at every information set.[5]

3 Analysis

We examine when the sender's limited commitment prevents them from attaining their *commitment payoff*, i.e., their payoff in the benchmark of Kamenica and Gentzkow (2011). Our results suggest that the answer hinges on two factors: (i) whether the receiver faces enough uncertainty about the sender's type and (ii) whether the sender can benefit from selectively disclosing more information after conducting their optimal experiment.

3.1 Assumptions & Preliminaries

We now introduce Assumptions 1 and 2, which will be maintained throughout. Let $A^*(\pi) \equiv \arg\max_{a \in A} \sum_{\theta \in \Theta} \pi(\theta) u^r(\theta, a)$ denote the set of receiver-optimal actions under $\pi \in \Delta(\Theta)$.

Assumption 1. *For every $a \in A$ and $\pi \in \Delta(\Theta)$ such that $a \in A^*(\pi)$, there exists $\pi' \in \Delta(\Theta)$ with $supp(\pi') \subseteq supp(\pi)$ that satisfies $A^*(\pi') = \{a\}$.*

[5] Appendix A formally defines a conditional probability system and develops an extension of PEBE that is needed to handle the fact that the sender's choice of experiments induces a distribution of verifiable signals. It shows that this extension has the intuitively appealing property that when t is bounded above by t^*, the receiver's belief conditional on observing any $h \in \mathcal{H}_{t^*+1}$ equals their naive belief.

Assumption 1 requires that for every action a that is optimal for the receiver under some belief π about the state, there exists π' with a weakly smaller support under which a is strictly optimal for the receiver. It implies that the receiver has a strict best reply when they know the state. Since Θ and A are finite, Assumption 1 is satisfied for a set of $u^r(\theta, a)$ of full Lebesgue measure in $\mathbb{R}^{|\Theta|\times|A|}$ by Proposition 2 of Lipnowski, Ravid, and Shishkin (2025).

Abusing notation, let $u^s(\theta)$ denote the sender's payoff when the receiver plays their (unique) best reply in state θ. The sender's *full-disclosure payoff* is $V^D(\pi_0) \equiv \sum_{\theta\in\Theta} \pi_0(\theta)u^s(\theta)$.

As in Kamenica and Gentzkow (2011), once the receiver's prior $\pi_0 \in \Delta(\Theta)$ is fixed, each experiment σ is characterized by the distribution it induces over naive beliefs $\pi \in \Delta(\Theta)$, which we will refer to as *beliefs* for short. We use $\sigma[\pi]$ to denote the probability with which belief π occurs in the distribution over beliefs that characterizes σ, and say that an experiment σ *induces* belief π if $\sigma[\pi] > 0$. We use $\Sigma[\pi] \subseteq \Delta(\Delta(\Theta))$ to denote the set of distributions over beliefs where the mean of those distributions equals π. Let $\overline{u}^s(\pi) \equiv \max_{a\in A^*(\pi)} u^s(a)$ denote the sender's highest payoff when the receiver's belief is π. The sender's *commitment payoff* is defined as

$$V^C(\pi_0) \equiv \max_{\sigma\in\Sigma[\pi_0]} \sum_{\pi\in\Delta(\Theta)} \sigma[\pi]\overline{u}^s(\pi). \tag{1}$$

By definition, $V^D(\pi_0) \leq V^C(\pi_0)$. An experiment is *optimal* if it attains the maximum in (1). Assumption 1 guarantees that each optimal experiment is the limit of a sequence of experiments that always induce strict receiver-incentives, as we show in Dai, Fudenberg, and Pei (2026). A standard upper hemi-continuity argument then shows that when the sender can commit not to conduct additional experiments, they can secure payoffs arbitrarily close to $V^C(\pi_0)$, regardless of how the receiver breaks ties, i.e., the *continuity property* in Ali, Kleiner, and Zhang (2024) holds. We focus on (π_0, u^s, u^r) for which the optimal experiment is unique:

Assumption 2. *There is a unique optimal experiment, which is denoted by σ^*.*

When $|\Theta| = 2$ and $V^C(\pi_0) \neq \max_{\pi\in\Delta(\Theta)} \overline{u}^s(\pi)$, Assumption 2 is generically satisfied in the sense that the set of primitives (π_0, u^s, u^r) under which there are multiple optimal experiments has zero Lebesgue measure in $[0, 1] \times \mathbb{R}^{|\Theta|\times|A|^2}$.

Proposition 1 shows that the sender attains their commitment payoff in all equilibria unless (i) their commitment payoff exceeds their full-disclosure payoff, and (ii) the receiver

faces nontrivial uncertainty about the sender's capacity to conduct additional experiments.

Proposition 1.

1. *For any $p \in \Delta(\mathbb{N})$, the sender's payoff belongs to $[V^D(\pi_0), V^C(\pi_0)]$ in every PEBE.*

2. *For any $p \in \Delta(\mathbb{N})$ that is degenerate, the sender's payoff is $V^C(\pi_0)$ in every PEBE.*

The proof is in Appendix B. Intuitively, the sender can secure $V^D(\pi_0)$ by fully disclosing the state in the initial experiment. If the sender's capacity t is common knowledge, then they can verify to the receiver that no additional outcomes are concealed by conducting t uninformative experiments and revealing their outcomes. This makes the sender's inability to commit irrelevant, and they attain their commitment payoff in every equilibrium.

3.2 Main Results

Our main results examine when the sender's limited commitment prevents them from attaining their commitment payoff in *all* equilibria. These results use the idea of a *credible belief*:

Definition 1. *A belief $\pi \in \Delta(\Theta)$ about the state is* credible *if for every $\theta \in supp(\pi)$, we have $\overline{u}^s(\pi) \geq u^s(\theta)$. Otherwise, this belief is* non-credible.

Intuitively, a belief is credible if the sender cannot benefit from revealing any state in the support of that belief. By definition, degenerate beliefs and beliefs where the sender attains their maximum utility are credible. Our notion of credible beliefs is related to *incentive-compatible outcomes* in Titova and Zhang (2025) and *revelation-proofness* in Shishkin, Titova, and Zhang (2025), both of which require that the sender's payoff in any state be weakly greater than their payoff from revealing that state. It is also related to Koessler and Skreta (2023)'s characterization of *interim optimal mechanisms*. They show that when the sender's payoff is a quasi-convex function of the receiver's belief, interim optimality requires the sender's interim payoff to be greater than their payoff from revealing any state.

Theorem 1 shows that in *monotone* payoff environments, which have been a primary focus of the persuasion literature, the sender can attain their commitment payoff under all type distributions *if and only if* the optimal experiment σ^* does not induce a non-credible belief.

Definition 2. (u^s, u^r) *is* monotone *if there exist complete orders on* Θ *and* A*, respectively, under which* $u^s(a)$ *is strictly increasing in* a *and* $u^r(\theta, a)$ *has strictly increasing differences.*

Theorem 1. *Suppose* (u^s, u^r) *is monotone.*

1. *If all beliefs induced by* σ^* *are credible, then for every* $p \in \Delta(\mathbb{N})$*, there exists a PEBE in which the sender's payoff is* $V^C(\pi_0)$*.*

2. *If* σ^* *induces at least one non-credible belief, then there exist* $\eta > 0$ *and an open set of type distributions (in the product topology in* $[0,1]^\infty$*) such that, under each type distribution* p *in this set, the sender's payoff in every PEBE is no more than* $V^C(\pi_0) - \eta$*.*

In the prosecutor–judge example of Kamenica and Gentzkow (2011), all beliefs induced by the optimal experiment are credible. As a result, the sender can attain the commitment payoff in some equilibrium, irrespective of the type distribution. In contrast, the optimal experiment induces a non-credible belief in the example of the introduction when the prior probability of the good state lies in $(0, 2/3)$. Consequently, by Theorem 1, the sender's payoff is strictly below the commitment payoff for an open set of type distributions.

Proof of Theorem 1 Part 1: Suppose the sender chooses σ^* as the initial experiment and conducts no additional experiments on the equilibrium path; the receiver breaks ties in favor of the sender; the receiver's posterior belief equals their naive belief if σ^* is the initial experiment *or* the number of additional outcomes disclosed equals $\bar{t} \equiv \max \operatorname{supp}(p)$;[6] at any other information set h, their posterior belief assigns probability 1 to state $\arg\min_{\theta \in \operatorname{supp}(\widehat{\pi}_h)} u^s(\theta)$, i.e., the worst state for the sender in the support of the receiver's naive belief.

First, we show that when the initial experiment is σ^*, the sender has no incentive to disclose any additional experiment. This is because in monotone environments, π being credible implies that $\overline{u}^s(\pi) \geq \overline{u}^s(\pi')$ for every $\operatorname{supp}(\pi') \subseteq \operatorname{supp}(\pi)$.[7] If all beliefs induced

[6] If the support of p is unbounded, then $\bar{t} = \infty$, in which case the number of additional experiments disclosed is always strictly below $\bar{t}$.

[7] Using the same constructive proof, one can show that in non-monotone environments, there exists a PEBE in which the sender's payoff is $V^C(\pi_0)$ for every $p \in \Delta(\mathbb{N})$ if every $\pi \in \Delta(\Theta)$ induced by the optimal experiment σ^* satisfies a stronger notion of credibility: $\overline{u}^s(\pi) \geq \underline{u}^s(\pi')$ for every π' that satisfies $\operatorname{supp}(\pi') \subseteq \operatorname{supp}(\pi)$ where $\underline{u}^s(\pi) \equiv \min_{a \in A^*(\pi)} u^s(a)$. This stronger notion reduces to credibility in monotone environments.

by σ^* are credible, then the sender cannot strictly benefit from disclosing any additional information and shifting the receiver's posterior to any π' with $\text{supp}(\pi') \subseteq \text{supp}(\pi)$.

Next, we verify that the sender cannot obtain payoff more than $V^C(\pi_0)$ by choosing a different initial experiment. This is because any type $t < \bar{t}$ cannot disclose $\bar{t}$ additional outcomes, so their expected payoff is no more than the full disclosure payoff $V^D(\pi_0)$, which is no more than $V^C(\pi_0)$. Type $\bar{t}$'s payoff is no more than $V^C(\pi_0)$, since $V^C(\pi_0)$ is an upper bound on their payoff when they conduct and disclose $\bar{t}$ additional outcomes, and given the receiver's off-path belief, concealing outcomes can only weakly lower the sender's payoff. □

The proof of Part 2 uses a lemma that applies beyond monotone environments:

Lemma 1. *If σ^* induces a non-credible belief, then there exist $\eta > 0$, $n \in \mathbb{N}$, and an open set of type distributions that contains the uniform distribution on $\{0, 1, ..., n\}$ such that for every p in this open set, the sender's payoff in any PEBE is at most $V^C(\pi_0) - \eta$.*

The proof is in Appendix C. The key step of our proof bounds type $t+1$ sender's payoff from below by computing their payoff from (i) initially using the equilibrium strategy of type t and (ii) at every non-credible posterior belief π induced by type t, conducting an additional experiment that fully reveals the state, and (iii) disclosing the resulting outcome if and only if the corresponding state θ satisfies $u^s(\theta) > \overline{u}^s(\pi)$. This leads to a lower bound on the difference between type $t+1$'s equilibrium payoff and type t's, which is a linear function of the probability with which type t's equilibrium strategy induces non-credible beliefs.

Suppose by way of contradiction that there are many types and there exists an equilibrium where the sender's ex ante expected payoff is close to $V^C(\pi_0)$. Then given the uniqueness of the optimal experiment, there will be many types inducing non-credible posterior beliefs with probability strictly above zero. Since the set of feasible payoffs is bounded, this will lead to a contradiction once the number of types that occur with positive probability is large enough.

Arieli and Stewart (2025) shows in a model with more general sender preferences that the sender can attain their commitment payoff in at least one wPBE. In contrast to wPBE that allows arbitrary off-path beliefs, PEBE requires that the support of the receiver's posterior belief be contained within the support of the corresponding naive belief. The following example illustrates how this additional belief restriction leads to different conclusions.

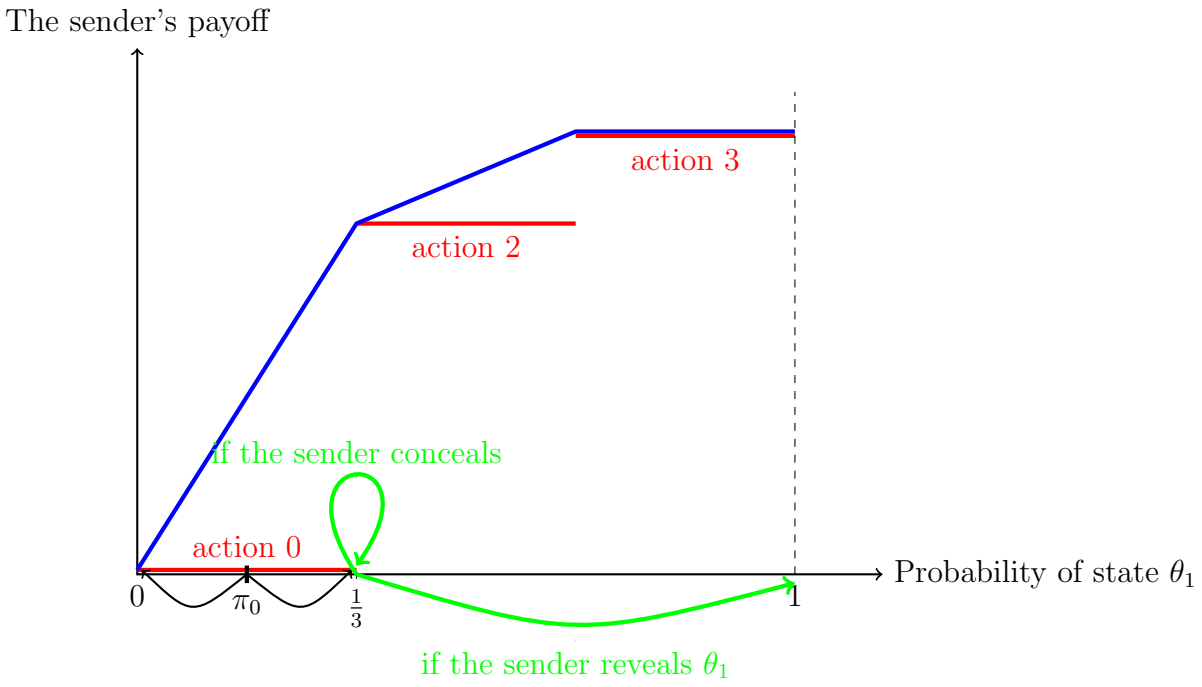

Figure 2: The sender's payoff function $\overline{u}^s(\pi)$ (red) and its concave closure (blue). The green arrows represent type 1 sender's profitable deviation after observing the outcome of the initial experiment that leads to naive belief $\frac{1}{3}$.

Example 1: Suppose $A = \{0, 2, 3\}$, $\Theta = \{\theta_0, \theta_1\}$, $t \in \{0, 1\}$. The receiver's prior assigns probability strictly between 0 and $\frac{1}{3}$ to state θ_1 and $\frac{1}{2}$ to type $t = 1$. The sender's payoff equals a. Let $\pi \in [0, 1]$ denote the probability the receiver's posterior belief assigns to state θ_1. The receiver's payoff is such that it is optimal for them to choose action 0 when $\pi \in [0, \frac{1}{3}]$, action 2 when $\pi \in [\frac{1}{3}, \frac{2}{3}]$, and action 3 when $\pi \in [\frac{2}{3}, 1]$ (See Figure 2).

For there to be a PEBE where the sender's payoff is $V^C(\pi_0)$, π would need to be $\frac{1}{3}$ when the state is θ_1. Since type 0 occurs with positive probability, there exists $h \in \mathcal{H}_1$ (an information set that is reached after disclosing the initial experiment and its outcome) at which the receiver's posterior belief is $\frac{1}{3}$. But then type 1 has a strict incentive to conduct a fully informative experiment at h and reveal the outcome if and only if $\theta = \theta_1$, since doing so increases the receiver's action in state θ_1 while leaving it unchanged in state θ_0. This contradicts the receiver's posterior being $\frac{1}{3}$ conditional on state θ_1. By contrast, in wPBE, after the sender has revealed state θ_1, the receiver's posterior may not assign probability 1 to θ_1 if that happens off-path. This allows Arieli and Stewart (2025) to construct wPBEs where

the sender obtains their commitment payoff.

The second part of Theorem 1 hinges on a *dispersion property* of the receiver's belief p about the sender's type, that it assigns probability bounded above 0 to sufficiently many types. In fact, in many natural settings, any sufficiently dispersed belief about the sender's type is enough to prevent the sender from attaining their commitment payoff. For each $\theta \in \Theta$, let $\Pi^*_\theta \subseteq \Delta(\Theta)$ denote the set of beliefs induced by σ^* that have θ in their support.

Theorem 2. *Suppose there exists a state $\theta \in \Theta$ such that $u^s(\theta) > \max_{\pi \in \Pi^*_\theta} \overline{u}^s(\pi)$. Then there exist $\lambda > 0$, $n \in \mathbb{N}$, and $\overline{\eta} > 0$ such that for every $\eta \in (0, \overline{\eta})$ and every $p \in \Delta(\mathbb{N})$ that assigns probability more than $\lambda\eta$ to at least n distinct types, the sender's payoff in every PEBE is no more than $V^C(\pi_0) - \eta$.*

Theorem 2 requires the existence of a state θ such that whenever the receiver's belief belongs to Π^*_θ, the sender strictly prefers to fully reveal θ, i.e., $\overline{u}^s(\pi) < u^s(\theta)$. In the example of the introduction, this additional requirement is satisfied when the good state occurs with probability between 0 and $\frac{1}{3}$. This requirement is more demanding than the assumption in Lemma 1 that σ^* induces at least one non-credible belief. We provide an example showing that this stronger requirement is not redundant in our working paper (Dai et al. (2026)).

The proof of Theorem 2, in Appendix D, parallels that of Lemma 1: We again examine a higher-type sender's payoff from imitating the equilibrium strategy of a lower type and then conducting an additional experiment that fully reveals the state, disclosing the result if and only if it yields a strictly higher payoff. When $u^s(\theta) > \max_{\pi \in \Pi^*_\theta} \overline{u}^s(\pi)$, the higher type can guarantee a strictly higher payoff than any lower type's equilibrium payoff, regardless of the probability that the lower type's strategy induces non-credible beliefs. The contrapositive—namely, there exists an equilibrium yielding a sender payoff arbitrarily close to the commitment payoff—must be false once sufficiently many types occur with positive probability, which completes the argument.

Next, we explain why the receiver having a dispersed belief about the sender's type is not redundant. We first present an example where the sender can attain their commitment payoff in some equilibrium even when (i) the optimal experiment induces a non-credible belief and (ii) the receiver faces non-trivial uncertainty about the sender's type. We then establish

Theorem 3, which shows that when p is concentrated on a single type, there always exists an equilibrium in which the sender's payoff is close to $V^C(\pi_0)$.

Example 2: Suppose $A \equiv \{0, 2, 3, \frac{7}{2}\}$, $\Theta \equiv \{\theta_0, \theta_1\}$, $t \in \{0, 1\}$, $\pi_0 = \frac{3}{8}$ and $t = 0$ with probability $\frac{1}{3}$. We will use the probability of θ_1 to represent the receiver's belief about the state. The sender's payoff equals a. The receiver's payoff is such that $a = 0$ is optimal when $\pi \in [0, \frac{1}{4}]$, $a = 2$ is optimal when $\pi \in [\frac{1}{4}, \frac{1}{2}]$, $a = 3$ is optimal when $\pi \in [\frac{1}{2}, \frac{3}{4}]$, and $a = \frac{7}{2}$ is optimal when $\pi \in [\frac{3}{4}, 1]$. We construct an equilibrium in which the sender's expected payoff is $V^C(\pi_0) = \frac{5}{2}$, which is attained by inducing posterior beliefs $\frac{1}{4}$ and $\frac{1}{2}$ with equal probability.[8]

In this equilibrium, the sender chooses an uninformative initial experiment. Type 1 conducts an additional experiment with outcome $\overline{s}$ occurring with probability 1 when $\theta = \theta_1$ and probability $\frac{3}{5}$ when $\theta = \theta_0$, and discloses only $\overline{s}$. The receiver's posterior belief equals $\frac{1}{2}$ after observing $\overline{s}$ and $\frac{1}{4}$ after observing no additional signal. At all off-path information sets, their posterior belief equals their naive belief. Type 1 has no incentive to fully reveal state θ_1 via additional experiments. Instead, they prefer to disclose a signal that partially reveals the state and induces posterior belief $\frac{1}{2}$. When the receiver only assigns positive probability to types 0 and 1, this leads to an equilibrium where the sender obtains their commitment payoff. However, no such equilibrium exists when the receiver's belief p is sufficiently dispersed.

Next, we show that if the type distribution p assigns probability sufficiently close to 1 to one type, there is an equilibrium where the sender's payoff is close to $V^C(\pi_0)$.

Theorem 3. *For every $\delta > 0$, there exists $\varepsilon > 0$ such that for every $p \in \Delta(\mathbb{N})$ where $p(n) > 1 - \varepsilon$ for some $n \in \mathbb{N}$, there exists a PEBE in which the sender's payoff is more than $V^C(\pi_0) - \delta$.*

Our three theorems together imply that, when payoffs are monotone, whenever the optimal experiment induces some non-credible beliefs, whether the sender can attain their commitment payoff in their best equilibrium is governed by the dispersion of the receiver's belief about their type. If this belief is sufficiently concentrated on a single type (so that the receiver faces little uncertainty about the sender's capacity), then there exist equilibria in which the

[8] More generally, for every $\pi_0 \in (\frac{1}{4}, \frac{1}{2})$, there is a non-degenerate distribution on $\{0, 1\}$ and a corresponding equilibrium in which the sender attains $V^C(\pi_0)$.

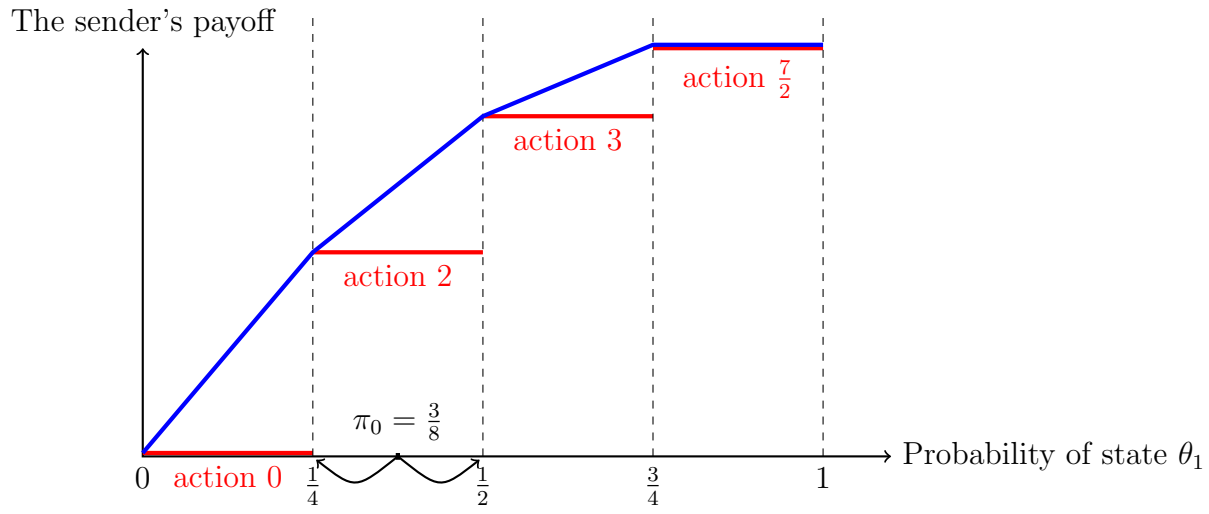

Figure 3: The sender's indirect utility as a function of the receiver's belief $\overline{u}^s(\pi)$ (red) and its concave closure (blue) in the example.

sender's payoff is approximately $V^C(\pi_0)$. By contrast, if the receiver's belief assigns positive probability to many sender types and each such type occurs with probability bounded above 0, then the sender's equilibrium payoff is bounded strictly below their commitment payoff.

Theorem 2 and Theorem 3 are complementary, as their proofs lead to lower and upper bounds on the gap between the sender's commitment payoff and their highest equilibrium payoff. The proof of Theorem 2 implies that the gap is bounded below by a linear function of the probability of the nth most likely type, where n is a function of the sender's benefit from selectively disclosing additional information at non-credible beliefs, the prior belief about the state, and the probability with which the optimal experiment induces those non-credible beliefs. Theorem 3 provides an upper bound on that gap, which is a linear function of the probability of types other than the most likely one.

The proof of Theorem 3, which is in Appendix E, constructs equilibria that approximately attain the sender's commitment payoff. For some intuition, consider the simple case where type 0 occurs with probability zero. In our construction, the sender chooses an uninformative initial experiment and (i) when their realized type is between 1 and $n-1$, they conduct one additional experiment that reveals the state and fully disclose the outcome, (ii) when their realized type is at least n, they conduct an additional experiment that is close to σ^*, fully

disclose its outcome, as well as the outcome of $n-1$ uninformative experiments, (iii) when their realized type is above n, they may conduct more than n additional experiments and selectively disclose their outcomes, and (iv) at every off-path information set, the receiver's posterior belief assigns probability 1 to the worst state in the support of their naive belief, i.e., the state θ that minimizes $u^s(\theta)$. The sender's ex ante expected payoff is close to $V^C(\pi_0)$ given that type n occurs with probability close to 1.

This construction does not work when type 0 occurs with positive probability, because the receiver will then assign probability 1 to type 0 when no additional outcome is disclosed, which can induce types 1 to $n-1$ not to disclose the outcome of the fully informative additional experiment. Our proof adjusts the strategies of types 1 to n in order to accommodate this. In equilibrium, the sender chooses an experiment close to σ^* as their initial experiment. The additional experiments conducted by each type as well as the outcomes disclosed depend on the relative probabilities of type 0 and types 1 to $n-1$. When type 0 is relatively likely compared to types 1 to $n-1$, type n will imitate these lower types with positive probability, which lowers the lower-types' equilibrium payoffs. When type 0 is relatively unlikely compared to types 1 to $n-1$, types 1 to $n-1$ conduct an additional experiment that fully reveals the state and disclose if and only if their indirect utility under that state is high enough.

4 Gap Between Commitment and Equilibrium Payoffs

Section 3 provides conditions under which the sender's equilibrium payoffs fall short of their commitment payoff. This section computes this gap in an example, and compares it with the difference between the sender's commitment payoff $V^C(\pi_0)$ and full-disclosure payoff $V^D(\pi_0)$.

Suppose $A \equiv \{0, 1, 1+\eta\}$, $\Theta \equiv \{\theta_0, \theta_1\}$, and $t \in \{0, 1\}$. Let $p \in (0, 1)$ denote the probability that $t = 0$ and let $\pi_0 \in (0, \pi^*)$ denote the prior probability that $\theta = \theta_1$. The sender's payoff equals a. For the receiver, $a = 0$ is optimal when $\pi \in [0, \pi^*]$, $a = 1$ is optimal when $\pi \in [\pi^*, \pi^{**}]$, and $a = 1+\eta$ is optimal when $\pi \in [\pi^{**}, 1]$, with $0 < \pi^* < \pi^{**} < 1$. We assume that $\frac{1+\eta}{\pi^{**}} < \frac{1}{\pi^*}$, which implies that $V^C(\pi_0) \equiv \pi_0/\pi^*$ and $V^D(\pi_0) \equiv \pi_0(1+\eta)$. Let $V_{\max}(\pi_0)$ denote the sender's highest equilibrium payoff and let $R \equiv \frac{V^C(\pi_0) - V_{\max}(\pi_0)}{V^C(\pi_0) - V^D(\pi_0)}$.

Proposition 2. *If* $\pi_0 \in (0, \pi^*)$ *and* $\frac{1+\eta}{\pi^{**}} < \frac{1}{\pi^*}$*, then*

$$V_{\max}(\pi_0) = \pi_0 \max\left\{ \frac{1}{\pi^*} - \frac{p\eta}{\pi^{**} - \pi^*}, \ \frac{p}{\pi^*} + \frac{(1-p)(1+\eta)}{\pi^{**}} \right\},$$

$$R = \frac{1}{1 - \pi^*(1+\eta)} \min\left\{ \frac{p\,\eta\,\pi^*}{\pi^{**} - \pi^*}, \ (1-p)\left(1 - \frac{\pi^*(1+\eta)}{\pi^{**}}\right) \right\}.$$

The proof is in Appendix F. Proposition 2 characterizes the sender's highest equilibrium payoff $V_{\max}(\pi_0)$ as well as the ratio R that quantifies the gap between the sender's highest equilibrium payoff and their commitment payoff. This gap can be arbitrarily close to the gap between the commitment payoff and the full disclosure payoff, as R is arbitrarily close to 1 in the limit $\pi^* \to 0$, $\pi^{**} \to 1$, $\eta \to \infty$ with $\pi^*(1+\eta) \to 1$, and $p \to 0$.

5 Worst Equilibrium Payoff

Our theorems examine whether the sender can obtain their commitment payoff in their best equilibrium. Our next result examines the sender's payoff in their worst equilibrium.

Proposition 3. *If* $p \in \Delta(\mathbb{N})$ *assigns positive probability to infinitely many types, then there is a PEBE in which the sender fully reveals* θ *and conducts no additional experiments on the equilibrium path.*

Proof. Let $\Theta \equiv \{\theta_1, \theta_2, ..., \theta_n\}$. Without loss of generality, we assume that $u^s(\theta_n) \geq ... \geq u^s(\theta_1)$. Consider a strategy profile and belief in which the sender chooses an initial experiment that fully reveals the state and conducts no additional experiments on the equilibrium path. For every $k \geq 1$ and every off-path history $\tilde{h}_k \in \mathcal{H}_k$, let $i(\tilde{h}_k)$ denote the lowest i such that $\theta_i \in \text{supp}(\widehat{\pi}_{\tilde{h}_k})$. Sender types $t \geq k$ will conduct one fully informative additional experiment at $\tilde{h}_k$ and disclose the outcome unless the state is $\theta_{i(\tilde{h}_k)}$. At every off-path information set $\tilde{h}_k \in \mathcal{H}_k$, the receiver's posterior belief assigns probability 1 to state $\theta_{i(\tilde{h}_k)}$, namely, the worst state for the sender in the support of the receiver's naive belief at that information set. One can verify that this is a PEBE in which the state is fully revealed to the receiver on the equilibrium path and the sender's expected payoff is $V^D(\pi_0)$. □

In settings where the receiver's belief assigns probability close to 1 to one sender type but also assigns strictly positive but small probability to infinitely many types, Theorem 3 and Proposition 3 together imply that there are equilibria in which the sender approximately attains their commitment payoff although there are also equilibria in which the sender obtains their full disclosure payoff, their payoff lower bound established in Proposition 1.

Because our model has many equilibria, one might wonder whether there are refinements that can rule some of them out. The refinements in the existing literature on disclosure games, such as the truth-leaning refinement in Hart, Kremer, and Perry (2017) and Gao (2025) and the monotonicity or threshold-strategy refinements in Guttman, Kremer, and Skrzypacz (2014) and Antic and Pei (2026) do not have much bite in our model, because an important source of multiplicity is the receiver's uncertainty about the additional experiments conducted by the sender, which is itself endogenous. Moreover, the existence of equilibria where the sender attains their commitment payoff as well as ones where the sender attains their full-disclosure payoff are robust to a number of perturbations. For example, they are robust when the sender faces a small additive cost to conduct each additional experiment, and when the sender faces a small additive cost to disclose the outcome of an additional experiment. Whether there are other well-motivated refinements that can select among the set of PEBEs in our model is a question left for future research.

Appendices

A PEBE in Games with Random Successors

This Appendix extends the definition of PEBE to games where the termination and successor nodes are random and are affected by players' actions. Consider an extensive-form game with nodes $x \in X$, information sets $h \in H$, actions $a \in A(h)$ available at information set h, and signals $s \in S$. We use $h(x)$ to denote the information set that contains x. For every $x \in X$ and $a \in A(h(x))$, signal $s \in S$ is realized with probability $q(s \mid x, a)$, leading to successor node $(x, a, s) \in X$. Let Z denote the set of terminal nodes. Given $Y \subseteq X$, we use $Z(Y)$ to

denote the subset of terminal nodes that are preceded by some element in Y.

A *conditional probability system* (CPS) on Z is a function $\nu(\cdot \mid \cdot) : 2^Z \times 2^Z \setminus \{\emptyset\} \to [0,1]$ such that (i) for all non-empty $\Omega_0 \subseteq Z$, $\nu(\cdot \mid \Omega_0)$ is a probability distribution on Ω_0 and (ii) for all $\Omega_3 \subseteq \Omega_2 \subseteq \Omega_1 \subseteq \Omega$ with $\Omega_2 \neq \emptyset$, we have $\nu(\Omega_3 \mid \Omega_2)\,\nu(\Omega_2 \mid \Omega_1) = \nu(\Omega_3 \mid \Omega_1)$. A CPS μ on Z induces conditional beliefs $\mu(\cdot \mid \cdot)$: $\mu(A \mid B) \coloneqq \nu(Z(A) \mid Z(B))$ for every $A, B \subseteq X$ such that $Z(A) \subseteq Z(B)$.

Given a profile of strategies π, let $\pi(a \mid h)$ denote the induced probability of taking action $a \in A(h)$ at information set h. We say a profile of strategies π and a CPS ν on terminal nodes is an *extended assessment*.

Definition 3. *Let μ denote the conditional beliefs associated with ν. The extended assessment (ν, π) is a PEBE if*

1. *π is a best response to conditional beliefs μ.*

2. *For every information set $h \in H$, action $a \in A(h)$, signal $s \in S$, and node $x \in h$,*

$$q(s \mid a, x)\,\pi(a \mid h) = \mu((x, a, s) \mid x);$$

3. *For every information set $h \in H$, action $a \in A(h)$, signal $s \in S$, and nodes $x, y \in h$,*

$$\frac{\mu((x,a,s) \mid (x,a,s),(y,a,s))}{\mu((y,a,s) \mid (x,a,s),(y,a,s))} = \frac{q(s \mid a, x)}{q(s \mid a, y)} \frac{\mu(x \mid x, y)}{\mu(y \mid x, y)}.$$

Lemma 2. *Suppose (ν, π) is a PEBE and μ is the associated conditional beliefs. For any receiver information set h,*

1. *If $h \in \mathcal{H}_k$, then $\mu(\{t \geq k-1\} \mid h) = 1$.*

2. *$\mu(\{\theta \in \mathit{supp}(\widehat{\pi}_h)\} \mid h) = 1$.*

3. *If $n = \max\{\mathit{supp}(p)\}$ and $h \in \mathcal{H}_{n+1}$, then $\mu(\theta \mid h) = \widehat{\pi}_h(\theta)$ for every $\theta \in \Theta$.*

Proof. (i) If $h \in \mathcal{H}_k$, then for every $x \in h$, the corresponding sender type $t(x) \geq k-1$, so statement 1 holds. (ii) If $\theta \notin \operatorname{supp}(\widehat{\pi}_h)$. For every sender information set h' that is

a predecessor of information set h, the corresponding sender posterior belief also assigns probability 0 to θ. Condition 3 in the definition of PEBE then implies that the receiver's posterior at h must assign probability 0 to θ as well. (iii) If $n = \max\{\text{supp}(p)\}$ and $h \in \mathcal{H}_{n+1}$, then for every $x \in h$, the sender type is $t(x) = n$. This implies the sender revealed all additional experiments, so the receiver's posterior is $\mu(t = n \mid h) = 1$. Consequently, by condition 3 of the PEBE definition, the receiver's posterior matches their naive belief $\widehat{\pi}_h$. □

B Proof of Proposition 1

In any equilibrium, the sender's strategy induces a distribution of the receiver's posterior beliefs about θ whose expected value equals their prior π_0. Hence, the sender's equilibrium payoff cannot exceed their payoff in an auxiliary game where they choose any posterior distribution subject to this mean-preserving constraint.

By definition, this upper bound is $V^C(\pi_0)$. The sender also has the option to deviate by choosing an initial experiment that fully reveals θ and conducting no additional experiments regardless of the realized θ. Under such a deviation, for every $\theta \in \Theta$, the sender obtains payoff $u^s(\theta)$ when the realized state is θ. Hence, the sender obtains an expected payoff $V^D(\pi_0)$ from this deviation, which implies that their equilibrium payoff must be at least $V^D(\pi_0)$.

In order to show that when p is degenerate the sender's payoff is $V^C(\pi_0)$ in all equilibria, let $\underline{u}^s(\pi) \equiv \min_{a \in A^*(\pi)} u^s(a)$ for every $\pi \in \Delta(\Theta)$ and let

$$\overline{U}[\sigma] \equiv \sum_{\pi \in \Delta(\Theta)} \sigma[\pi] \overline{u}^s(\pi) \quad \text{and} \quad \underline{U}[\sigma] \equiv \sum_{\pi \in \Delta(\Theta)} \sigma[\pi] \underline{u}^s(\pi).$$

Since Θ and A are finite, the set $\arg\max_{\sigma \in \Sigma} \overline{U}[\sigma]$ is non-empty. Recall that σ^* is the optimal experiment. By definition, $V^C(\pi_0) = \overline{U}[\sigma^*]$, and for every $\varepsilon > 0$, there exists an experiment σ^ε such that $\underline{U}[\sigma^\varepsilon] > \overline{U}[\sigma^*] - \varepsilon$.

Suppose p assigns probability 1 to t^*. Fix any PEBE and $\varepsilon > 0$. Consider a deviation where the sender conducts an initial experiment σ^ε satisfying $\underline{U}[\sigma^\varepsilon] > V^C(\pi_0) - \varepsilon$, followed by t^* uninformative experiments, and reveals all t^* outcomes. After observing t^* additional outcomes, the receiver is certain that every outcome was disclosed, so their posterior belief

about θ must coincide with their naive belief. Thus the sender's payoff from the deviation is at least $V^C(\pi_0) - \varepsilon$, so the sender's payoff in every equilibrium is at least $V^C(\pi_0)$.

C Proof of Lemma 1

Since every degenerate belief is credible and σ^* is assumed to be the unique optimal experiment, the hypothesis that σ^* induces a non-credible belief implies that $V^C(\pi_0) > V^D(\pi_0)$. Let $\underline{\pi} \in \Delta(\Theta)$ denote a non-credible belief induced by σ^*. By definition, there exists $\theta \in \text{supp}(\underline{\pi})$ such that $u^s(\theta) > \overline{u}^s(\underline{\pi})$. Recall that $B(\pi, d)$ denotes the open ball with radius $d > 0$ centered at π. For every $\sigma \in \Sigma[\pi_0]$, denote the conditional distribution over beliefs given state θ by σ^θ, where $\sigma^\theta[B] \equiv \int_B \frac{\pi(\theta)}{\pi_0(\theta)} d\sigma[\pi]$ for every Borel measurable set $B \subseteq \Delta(\Theta)$.

Lemma 3. *Suppose r is such that $\pi(\theta) > \underline{\pi}(\theta)/2$ for every $\pi \in B(\underline{\pi}, r)$. There exist $\eta, \varepsilon > 0$ such that for every $\sigma \in \Sigma[\pi_0]$, if $\mathbb{E}_{\pi\sim\sigma}\overline{u}^s(\pi) > V^C(\pi_0) - \eta$, then $\sigma^\theta[B(\underline{\pi}, r)] \geq \varepsilon$.*

Proof of Lemma 3. Let $m \equiv \sigma^*[B(\underline{\pi}, r)] > 0$ and let $\mathcal{C} := \{\sigma \in \Sigma[\pi_0] : \sigma[B(\underline{\pi}, r)] \leq m/2\}$. Since $B(\underline{\pi}, r)$ is an open set, the function that maps σ to $\sigma[B(\underline{\pi}, r)]$ is lower semi-continuous by the Portmanteau Theorem, so the lower sublevel set $\{\sigma \in \Delta(\Delta(\Theta)) : \sigma[B(\underline{\pi}, r)] \leq m/2\}$ is closed. Since Θ is finite, $\Sigma[\pi_0]$ is compact, and hence $\mathcal{C}$ is a closed subset of a compact set, which is also compact.

By definition, $\sigma^* \notin \mathcal{C}$. Since σ^* is the unique optimal experiment (Assumption 2), $\sup_{\sigma\in\mathcal{C}} \mathbb{E}_{\pi\sim\sigma}(\overline{u}^s(\pi)) < V^C(\pi_0)$. Let $\eta \equiv \min\{V^C(\pi_0) - \sup_{\sigma\in\mathcal{C}} \mathbb{E}_\sigma(\overline{u}^s), m/2\} > 0$ and $\varepsilon \equiv \frac{\underline{\pi}(\theta)}{2\pi_0(\theta)}\eta > 0$. It follows that for any $\sigma \in \Sigma[\pi_0]$ with $\mathbb{E}_{\pi\sim\sigma}(\overline{u}^s(\pi)) > V^C(\pi_0) - \eta$, $\sigma \notin \mathcal{C}$ and hence $\sigma[B(\underline{\pi}, r)] > m/2 \geq \eta$. By definition,

$$\sigma^\theta[B(\underline{\pi}, r)] = \int_{B(\underline{\pi},r)} \frac{\pi(\theta)}{\pi_0(\theta)} d\sigma[\pi] \geq \frac{\underline{\pi}(\theta)}{2\pi_0(\theta)} \sigma[B(\underline{\pi}, r)] \geq \frac{\underline{\pi}(\theta)}{2\pi_0(\theta)}\eta = \varepsilon,$$

where the first inequality follows from the assumption that $\pi(\theta) > \underline{\pi}(\theta)/2$ for every $\pi \in B(\underline{\pi}, r)$. □

Let $\overline{v}$ and $\underline{v}$ denote the highest and lowest feasible payoff for the sender, respectively. Since the action set is finite and the receiver's best-reply correspondence is upper hemi-continuous,

there exists $r > 0$ such that for every $\pi \in B(\underline{\pi}, r)$, we have $\overline{u}^s(\pi) \leq \overline{u}^s(\underline{\pi}) < u^s(\theta)$ and $\pi(\theta) > \underline{\pi}(\theta)/2 > 0$. By Lemma 3, there exist $\eta, \varepsilon > 0$ such that for every $\sigma \in \Sigma[\pi_0]$, if $\mathbb{E}_{\pi\sim\sigma}\overline{u}^s(\pi) > V^C(\pi_0) - \eta$, then $\sigma^\theta[B(\underline{\pi}, r)] \geq \varepsilon$. Fix this selection of r, η, ε and choose $N \in \mathbb{N}$ with

$$N > \frac{3(\overline{v} - \underline{v})}{2\varepsilon\pi_0(\theta)(u^s(\theta) - \overline{u}^s(\underline{\pi}))} + \frac{2}{\varepsilon}.$$

Endow $\Delta(\mathbb{N})$ with the product topology. For any $p \in \Delta(\mathbb{N})$ that belongs to

$$U \equiv \left\{ p \in \Delta(\mathbb{N}) : |p(t) - \tfrac{1}{N}| < \tfrac{1}{2N} \; \forall t = 0, \ldots, N-1, \text{ and } p(\{t \geq N\}) < \tfrac{1}{2N} \right\},$$

we show that the sender's expected payoff in any equilibrium is no more than $V^C(\pi_0) - \eta$.

Suppose by way of contradiction that there exists an equilibrium in which the sender's payoff is strictly greater than $V^C(\pi_0) - \eta$. For any $t \in \{0, ..., N-2\}$, we construct a deviation for type $t+1$ under which their payoff is strictly bounded above type t's equilibrium payoff.

For any pure strategy γ that is feasible for type t, let q^γ denote the probability with which type t plays γ in the above equilibrium. Let $\mathcal{H}^\gamma$ denote the collection of information sets h reached with positive probability under strategy γ such that type t conducts no further experiments after h, and the receiver's posterior belief at the disclosed history following h belongs to $B(\underline{\pi}, r)$. Consider type $t+1$'s payoff from the following deviation: for every γ with $q^\gamma > 0$, the deviation assigns probability q^γ to a pure strategy modified from γ: Use strategy γ until after reaching histories that belong to $\mathcal{H}^\gamma$. At every $h \in \mathcal{H}^\gamma$, conduct one additional experiment that fully reveals the state, and discloses the outcome of that experiment if and only if the state is θ. Let V_t denote type t's equilibrium payoff. Lemma 2 implies that after type t_i discloses the outcome of a fully informative experiment that identifies state θ, the receiver's posterior assigns probability one to θ: the naive belief at that information set has support θ, and PEBE requires the receiver's posterior support to be contained in the support of the naive belief. Hence the receiver chooses a best reply to state θ, and the sender obtains payoff $u^s(\theta)$.

Type $t+1$'s payoff from this deviation is at least $V_t + \widetilde{\sigma}_t^\theta[B(\underline{\pi}, r)] \cdot \pi_0(\theta) \cdot (u^s(\theta) - \overline{u}^s(\underline{\pi}))$, where $\widetilde{\sigma}_t^\theta[\cdot]$ denotes the induced distribution of the receiver's posterior beliefs under type t's equilibrium strategy conditional on state θ. This is because under the deviation we described,

type $t+1$ will induce the same receiver-action as type t except when the state is θ and type t reaches information sets that induce beliefs close to the non-credible one $\underline{\pi}$, in which case type $t+1$'s deviation will induce the receiver to take their optimal action for state θ.

Type $t+1$'s equilibrium payoff is weakly greater than their payoff from any deviation, so $V_{t+1} - V_t \geq \widetilde{\sigma}_t^\theta[B(\underline{\pi}, r)]\pi_0(\theta)\Big(u^s(\theta) - \overline{u}^s(\underline{\pi})\Big)$, and hence

$$V_{N-1} - V_0 \geq \left(\sum_{t=0}^{N-2} \widetilde{\sigma}_t^\theta[B(\underline{\pi}, r)]\right) \pi_0(\theta)\Big(u^s(\theta) - \overline{u}^s(\underline{\pi})\Big).$$

Since $|p(t) - \frac{1}{N}| \leq \frac{1}{2N}$ for every $t = 0, \ldots, N-1$, we have

$$(\tfrac{1}{N} + \tfrac{1}{2N}) \sum_{t=0}^{N-2} \tilde{\sigma}_t^\theta[B(\underline{\pi}, r)] \geq \sum_{t=0}^{N-2} p(t)\widetilde{\sigma}_t^\theta[B(\underline{\pi}, r)] = \widetilde{\sigma}^\theta[B(\underline{\pi}, r)] - \sum_{t=N-1}^{\infty} p(t)\widetilde{\sigma}_t^\theta[B(\underline{\pi}, r)].$$

By Lemma 3, we have

$$\widetilde{\sigma}^\theta[B(\underline{\pi}, r)] - \sum_{t=N-1}^{\infty} p(t)\widetilde{\sigma}_t^\theta[B(\underline{\pi}, r)] \geq \varepsilon - p(N-1) - p(\{t \geq N\}) \geq \varepsilon - \tfrac{2}{N},$$

where the last inequality follows from $|p(N-1) - \frac{1}{N}| \leq \frac{1}{2N}$ and $p(\{t \geq N\}) \leq \frac{1}{2N}$. Therefore,

$$V_{N-1} - V_0 \geq (\tfrac{2}{3}N\varepsilon - \tfrac{4}{3})\pi_0(\theta)(u^s(\theta) - \overline{u}^s(\underline{\pi})).$$

This together with the construction of N leads to a contradiction.

D Proof of Theorem 2

For every $\Pi \subseteq \Delta(\Theta)$ and $\pi \in \Delta(\Theta)$, let $d(\pi, \Pi) \equiv \inf_{\pi' \in \Pi} \|\pi - \pi'\|_1$ and $B(\Pi, \varepsilon) \equiv \{\pi \in \Delta(\Theta) : d(\pi, \Pi) < \varepsilon\}$ for every $\varepsilon > 0$. Let $\Pi^* \subseteq \Delta(\Theta)$ denote the set of posteriors induced by σ^*.

Lemma 4. *Fix $\theta \in \Theta$ and $\varepsilon > 0$. There exists $\lambda_\theta > 0$ such that for every $\eta > 0$ and every $\sigma \in \Sigma[\pi_0]$ satisfying $\int \overline{u}^s(\pi)\, d\sigma(\pi) \geq V^C(\pi_0) - \eta$, we have $\sigma^\theta\big(B(\Pi^*_\theta, \varepsilon)\big) \geq 1 - \lambda_\theta \eta$.*

Proof of Lemma 4. We use the assumption that σ^* is the unique optimal experiment. Apply-

ing Hoffman's error-bound theorem, there exists $C > 0$ such that, for every $\sigma \in \Sigma[\pi_0]$,

$$\int_{\Delta(\Theta)} d(\pi, \Pi^*)\, d\sigma(\pi) \leq C \left(V^C(\pi_0) - \int_{\Delta(\Theta)} \overline{u}^s(\pi)\, d\sigma(\pi) \right).$$

We first show that, for every $\pi \in \Delta(\Theta) \setminus B(\Pi^*_\theta, \varepsilon)$, $\pi(\theta) \leq \max\{1, \varepsilon^{-1}\}\, d(\pi, \Pi^*)$. Since Π^* is finite, there exists $\pi^* \in \Pi^*$ that is closest to π. If $\pi^* \notin \Pi^*_\theta$, then $\pi^*(\theta) = 0$, and hence $\pi(\theta) = |\pi(\theta) - \pi^*(\theta)| \leq \|\pi - \pi^*\|_1 = d(\pi, \Pi^*)$. If instead $\pi^* \in \Pi^*_\theta$, then $\pi \notin B(\Pi^*_\theta, \varepsilon)$ implies that $d(\pi, \Pi^*) = \|\pi - \pi^*\|_1 \geq \varepsilon$. Since $\pi(\theta) \leq 1$, it follows that $\pi(\theta) \leq d(\pi, \Pi^*)/\varepsilon$.

Now suppose that $\int \overline{u}^s(\pi)\, d\sigma(\pi) \geq V^C(\pi_0) - \eta$. Then

$$\begin{aligned} 1 - \sigma^\theta(B(\Pi^*_\theta, \varepsilon)) =& \frac{1}{\pi_0(\theta)} \int_{\Delta(\Theta)\setminus B(\Pi^*_\theta, \varepsilon)} \pi(\theta)\, d\sigma(\pi) \\ \leq& \frac{\max\{1, \varepsilon^{-1}\}}{\pi_0(\theta)} \int_{\Delta(\Theta)} d(\pi, \Pi^*)\, d\sigma(\pi) \leq \frac{C \max\{1, \varepsilon^{-1}\}}{\pi_0(\theta)} \eta, \end{aligned}$$

where the equality follows from the definition of σ^θ, the first inequality follows from the pointwise bound on $\pi(\theta)$, and the second inequality follows from Hoffman's bound above. Therefore, the lemma holds with $\lambda_\theta = C \max\{1, \varepsilon^{-1}\}/\pi_0(\theta)$. □

Now we prove the theorem. Suppose there exists $\theta \in \Theta$ such that $u^s(\theta) > \max_{\pi \in \Pi^*_\theta} \overline{u}^s(\pi)$. Since the receiver's best reply correspondence is upper hemi-continuous, there exists $\varepsilon > 0$ such that $\max_{\pi \in B(\Pi^*_\theta, \varepsilon)} \overline{u}^s(\pi) = \max_{\pi \in \Pi^*_\theta} \overline{u}^s(\pi)$. Fix such θ and ε in the rest of the proof. Let $\overline{v}$ and $\underline{v}$ denote the sender's highest and lowest feasible payoffs. Let $\lambda \equiv 2\lambda_\theta$ and

$$n \equiv \left\lceil \frac{2(\overline{v} - \underline{v})}{\pi_0(\theta)\left\{ u^s(\theta) - \max\limits_{\pi \in \Pi^*_\theta} \overline{u}^s(\pi) \right\}} \right\rceil + 1.$$

Pick any $\eta > 0$ such that $\eta < \bar{\eta} \equiv \frac{1}{\lambda n}$, and suppose by way of contradiction that p assigns probability more than $\lambda \cdot \eta$ to at least n types $\{t_1, ..., t_n\}$ with $t_1 < t_2 < ... < t_n$ yet there exists an equilibrium where the sender's payoff is more than $V^C(\pi_0) - \eta$. Lemma 4 shows that conditional on state θ, the receiver's posterior belief belongs to $B(\Pi^*_\theta, \varepsilon)$ with probability more than $1 - \lambda\eta/2$.

Let $\Pr(\cdot \mid \theta, t)$ denote the induced probability measure given the equilibrium strat-

egy of type t when the state is θ. We use D to denote the Hausdorff distance between the induced receiver posterior and Π^*_θ, which is a random variable. Taking expectation across types, we have $\mathbb{E}_{t\sim p}\left[\Pr(D \geq \varepsilon \mid \theta, t)\right] < \lambda\eta/2$. Markov's inequality then implies that $p(\{t \in \mathbb{N} : \Pr(D \geq \varepsilon \mid \theta, t) \geq 1/2\}) < \lambda\eta$. Because each selected type t_i has probability at least $\lambda\eta$, none of them can be in that set, so for every t_i, we have $\Pr(D < \varepsilon \mid \theta, t_i) > 1/2$.

For every $i \in \{1, \ldots, n\}$, choose a pure strategy γ_i in the support of type t_i's equilibrium mixed strategy such that, conditional on state θ, the probability that γ_i induces a receiver posterior in $B(\Pi^*_\theta, \varepsilon)$ is greater than $1/2$. Such a pure strategy exists by the preceding inequality and an averaging argument. Since γ_i is played with positive probability in equilibrium, it yields type t_i's equilibrium payoff.

Let $\mathcal{H}^*_i$ denote the set of sender histories reached with positive probability under γ_i, conditional on state θ, such that after each $h \in \mathcal{H}^*_i$ the sender conducts no further experiments under γ_i, and the receiver's posterior belief at h belongs to $B(\Pi^*_\theta, \varepsilon)$.

For every $i \geq 2$, consider type t_i's payoff when they deviate to the following strategy: Play γ_{i-1} except for histories that belong to $\mathcal{H}^*_{i-1}$. At every $h \in \mathcal{H}^*_{i-1}$, conduct one additional experiment that fully reveals the state, and discloses the outcome of that experiment if and only if the state is θ. Let V_i denote the equilibrium payoff of type t_i. Type t_i's payoff from the above deviation is at least

$$V_{i-1} + \frac{1}{2}\pi_0(\theta)\Big\{u^s(\theta) - \max_{\pi \in \Pi^*_\theta} \overline{u}^s(\pi)\Big\},$$

since they will induce the same receiver action as one of the pure strategy types t_{i-1} uses with positive probability in equilibrium, except that when the state is θ and type t_{i-1} reaches history $\mathcal{H}^*_{i-1}$, they will induce receiver action $a^*(\theta)$. Hence, for every $i \geq 2$, we have

$$V_i - V_{i-1} \geq \frac{1}{2}\pi_0(\theta)\Big\{u^s(\theta) - \max_{\pi \in \Pi^*_\theta} \overline{u}^s(\pi)\Big\}.$$

This together with the construction of n leads to a contradiction.

E Proof of Theorem 3

If σ^* fully reveals the state, then $V^C(\pi_0) = V^D(\pi_0)$, so the claim follows from Proposition 1. Next, suppose that σ^* is not fully revealing. By Lemma 3 in Lipnowski, Ravid, and Shishkin (2025), for every $\delta > 0$, there exists $\varepsilon > 0$ and an experiment σ^*_ε such that for every $\pi \in \Pi^*$, there exists $\pi_\varepsilon(\pi) \in B(\pi, \varepsilon)$ that is induced by σ^*_ε such that $a^*(\pi)$ is strictly optimal for the receiver when their belief about the state is $\pi_\varepsilon(\pi)$, and σ^*_ε assigns probability more than $1-\varepsilon$ to beliefs in $\Pi^*_\varepsilon = \{\pi_\varepsilon(\pi)\}_{\pi\in\Pi^*}$. The sender's expected payoff from committing to any such σ^*_ε is thus at least $V^C(\pi_0) - \delta$.

Fix a type distribution $p \in \Delta(\mathbb{N})$ that assigns probability $p(n) > 1-\varepsilon$ to some $n \in \mathbb{N}$. We show that there exists a PEBE where the sender's payoff is more than $V^C(\pi_0) - \delta$. If $\text{supp}(p)$ is bounded above, let $N \equiv \max \text{supp}(p)$ denote the highest possible sender type. We focus on the case where $n < N$ or $\text{supp}(p)$ is not bounded above (there's no such N), because if $n = N$, using a similar argument as Proposition 1, the type n sender can secure a payoff that is close to their commitment payoff. It follows that the sender's ex ante payoff is also close to the commitment payoff since $p(n) \approx 1$.

If $n < N$ or $\text{supp}(p)$ is not bounded above, we an equilibrium after each interim belief induced by σ^*_ε. At credible interim beliefs, all types except possibly the maximal type conduct no additional experiments. At non-credible interim beliefs, type n obtains approximately $\overline{u}^s(\pi)$ by disclosing n uninformative outcomes, lower types use the disclosure rules constructed below, and higher types disclose n uninformative outcomes before using any extra capacity. Off path, the receiver assigns probability one to the sender-worst state in the support of the naive belief, except at histories where all feasible additional outcomes have been disclosed.

In this equilibrium, following any interim belief $\pi \in \Pi^*_\varepsilon$, type n's on-path payoff is close to $\overline{u}^s(\pi)$ and hence the sender's ex ante expected payoff is close to $V^C(\pi_0)$ as $p(n) \approx 1$. Given the receiver's off-path posterior, it is without loss of generality to focus on subgames where the sender does not deviate to other initial experiments. In the paragraph below, we analyze subgames following a credible interim belief in Π^*_ε. In subsection E.1, we consider subgames following non-credible interim beliefs in Π^*_ε and describe the details of the strategy profiles.

Following any credible interim belief $\pi \in \Pi^*_\varepsilon$, all types of the sender except the highest

type N (if it exists) conduct no additional experiments in equilibrium. Doing so is optimal for all types of the sender given the receiver's off-path belief and the facts that (i) π is credible and (ii) type N has negligible probability, so the receiver's posterior when observing no additional experiments is close to π and also credible.

E.1 Non-credible Interim Beliefs

Consider the subgame following a non-credible interim belief $\pi \in \Pi^*_\varepsilon$. Let $\text{supp}(\pi) \equiv \{\theta_1, ..., \theta_k\}$, $\pi_j \equiv \pi(\theta_j)$, and $u_j \equiv u^s(\theta_j)$ for $j \in \{1, \ldots, k\}$. Without loss of generality, we assume that $u_1 \geq u_2 \geq ... \geq u_k$. Let $u^* \equiv u^s(a^*(\pi))$. Since $\pi \in \Pi^*_\varepsilon$ is non-credible and full disclosure is suboptimal at π, we know that $u_1 > u^* > u_k$.

Let u_i^* be defined via the following equation:

$$u^* = \sum_{j=1}^{i} \pi_j u_j + \Big(1 - \sum_{j=1}^{i} \pi_j\Big) u_i^*. \tag{2}$$

Intuitively, u_i^* is the sender's payoff from no disclosure so that they are indifferent between receiving u^* and disclosing the state if and only if it belongs to $\{\theta_1, ..., \theta_i\}$. By definition, $u_k^* = +\infty$. The sender's incentive to reveal states with index below i and to conceal states with index above i requires that $u_i \geq u_i^* \geq u_{i+1}$.

Lemma 5 shows that this incentive constraint is equivalent to u_i^* reaching a local minimum at i and that such a local minimum exists.

Lemma 5. *$u_i \geq u_i^* \geq u_{i+1}$ if and only if $\min\{u_{i+1}^*, u_{i-1}^*\} \geq u_i^*$. Furthermore, there exists $i \in \{1, 2, ..., k\}$ such that $u_i \geq u_i^* \geq u_{i+1}$.*

Proof. Applying equation (2) to i and $i+1$, we have:

$$\sum_{j=1}^{i} \pi_j u_j + \Big(1 - \sum_{j=1}^{i} \pi_j\Big) u_i^* = \sum_{j=1}^{i+1} \pi_j u_j + \Big(1 - \sum_{j=1}^{i+1} \pi_j\Big) u_{i+1}^*. \tag{3}$$

Using $\sum_{j=1}^{i+1}\pi_j u_j = \sum_{j=1}^{i}\pi_j u_j + \pi_{i+1}u_{i+1}$, equation (3) becomes

$$\sum_{j=1}^{i}\pi_j u_j + \left(1-\sum_{j=1}^{i}\pi_j\right)u_i^* = \sum_{j=1}^{i}\pi_j u_j + \pi_{i+1}u_{i+1} + \left(1-\sum_{j=1}^{i+1}\pi_j\right)u_{i+1}^*.$$

Cancel the common term $\sum_{j=1}^{i}\pi_j u_j$ and apply $\sum_{j=1}^{i+1}\pi_j = \sum_{j=1}^{i}\pi_j + \pi_{i+1}$, we have

$$\left[\left(1-\sum_{j=1}^{i+1}\pi_j\right)+\pi_{i+1}\right]u_i^* = \pi_{i+1}u_{i+1} + \left(1-\sum_{j=1}^{i+1}\pi_j\right)u_{i+1}^*,$$

which implies that

$$\left(1-\sum_{j=1}^{i+1}\pi_j\right)u_i^* - \left(1-\sum_{j=1}^{i+1}\pi_j\right)u_{i+1}^* = \pi_{i+1}u_{i+1} - \pi_{i+1}u_i^*. \tag{4}$$

This in turn gives:

$$\left(1-\sum_{j=1}^{i+1}\pi_j\right)(u_i^* - u_{i+1}^*) = \pi_{i+1}(u_{i+1} - u_i^*). \tag{5}$$

This implies that $u_{i+1}^* \geq u_i^*$ if and only if $u_i^* \geq u_{i+1}$. Substituting $\sum_{j=1}^{i+1}\pi_j = \sum_{j=1}^{i}\pi_j + \pi_{i+1}$ and regrouping terms shows that

$$\left(1-\sum_{j=1}^{i}\pi_j\right)(u_i^* - u_{i+1}^*) = \pi_{i+1}(u_{i+1} - u_{i+1}^*). \tag{6}$$

This implies that $u_{i+1} \geq u_{i+1}^*$ if and only if $u_i^* \geq u_{i+1}^*$. Similarly, $u_{i-1}^* \geq u_i^* \iff u_i \geq u_i^*$.

To show that there exists $i \in \{1, 2, ..., k\}$ such that $u_i \geq u_i^* \geq u_{i+1}$, we consider two cases. First, if $u_1^* \geq u_2^*$, then the fact that $u_k^* = +\infty$ implies that there exists i such that $\min\{u_{i+1}^*, u_{i-1}^*\} \geq u_i^*$. The first part of lemma 5 implies that $u_i \geq u_i^* \geq u_{i+1}$ for such i. Second, if $u_2^* > u_1^*$, then inequality (5) implies that $u_1^* \geq u_2$. To show that $i = 1$ satisfies our requirement, we only need to show that $u_1 \geq u_1^*$. Suppose by way of contradiction that $u_1 < u_1^*$, then given that $u_1\pi_1 + (1-\pi_1)u_1^* = u^*$, we have $u_1 < u^* < u_1^*$. This contradicts an implication that π is non-credible which is $u_1 > u^*$. □

Define strategy s^i to be the sender pure strategy (in the subgame following interim belief

π) where the sender conducts only one fully informative additional experiment and discloses the resulting outcome if and only if $\theta \in \{\theta_1, ..., \theta_i\}$. Recall that type n occurs with probability close to 1. Define strategy s^{nd} to be a sender pure strategy where the sender conducts n uninformative experiments and discloses all outcomes. By definition, strategy s^i is only feasible for types $t \geq 1$ and strategy s^{nd} is feasible only for $t \geq n$. Let $\mathbb{G} \subseteq \Delta(\Theta) \times \mathbb{R}$ be such that $(\pi, u) \in \mathbb{G}$ if and only if there exists $\alpha \in \Delta(A)$ such that α is the receiver's best reply at posterior belief π and $\mathbb{E}_{a\sim\alpha}[u^s(a)] = u$. By definition, $\mathbb{G}$ is closed and connected.

Fix any i that satisfies $u_i \geq u_i^* \geq u_{i+1}$, which exists by Lemma 5. Let π_*^i denote the receiver's posterior belief that assigns probability $\frac{\pi_j}{\sum_{\tau=i+1}^k \pi_\tau}$ to state θ_j for $j \in \{i+1, ..., k\}$ and assigns zero probability to other states. When ε is sufficiently small, the fact that $\pi \in \Pi_\varepsilon^*$ implies that π is close to an interim belief π' under the sender's optimal experiment σ^* and that $a^*(\pi) = a^*(\pi')$. Because σ^* is the optimal experiment, we know that for every $a \in \arg\max_{a' \in A} u^r(\pi_*^i, a')$,

$$u^* > \sum_{j=1}^{i} \pi_j u_j + \Big(1 - \sum_{j=1}^{i} \pi_j\Big) u^s(a). \tag{7}$$

At interim belief π, let π^i denote the receiver's posterior belief about θ at a history that is induced by type 0 for every $\theta \in \{\theta_1, ..., \theta_k\}$, by types 1 to $n-1$ if and only if $\theta \in \{\theta_{i+1}, ..., \theta_k\}$, and is never induced by types n and above. Formally, let $p(t)$ denote the probability that the receiver's prior belief assigns to type t. Belief π^i is then defined as:

$$\pi^i(\theta_j) \equiv \frac{p(0)\pi_j}{p(0) + \sum_{t=1}^{n-1} p(t) \cdot \sum_{s=i+1}^{k} \pi_s} \text{ for every } j \leq i$$

and

$$\pi^i(\theta_j) \equiv \frac{\sum_{t=0}^{n-1} p(t)\pi_j}{p(0) + \sum_{t=1}^{n-1} p(t) \cdot \sum_{s=i+1}^{k} \pi_s} \text{ for every } j \geq i+1.$$

We consider two cases, depending on the receiver's best reply at π^i.

Case A: Suppose there exists $a \in \arg\max_{a' \in A} u^r(\pi^i, a')$ such that

$$u^* \leq \sum_{j=1}^{i} \pi_j u_j + \Big(1 - \sum_{j=1}^{i} \pi_j\Big) u^s(a). \tag{8}$$

Intuitively, this is the case where the probability of type 0 is high relative to the probability of types 1 to $n-1$. The definition of u_i^* in (2) implies that $u^s(a) \geq u_i^*$.

In equilibrium following interim belief π, sender types strictly greater than n (except the highest type N if it exists) conduct n uninformative experiments, fully disclose all outcomes, and then conduct another fully informative experiment and disclose its outcome if and only if the realized state $\theta = \theta_j$ satisfies $u_j > u^*$. Sender type N may use a different strategy, but it is different from s^i or s^{nd}. Sender types 1 to $n-1$ use strategy s^i.

Sender type n mixes between strategy s^{nd} and strategy s^i. If type n uses strategy s^{nd} for sure, the receiver's posterior when observing no additional experiment is π^i. If type n uses strategy s^i for sure, as it occurs with probability close to 1, the receiver's posterior $\widetilde{\pi}_*^i$ when observing no additional experiment is close to π_*^i, so there exists $a \in \arg\max_{a' \in A} u^r(\widetilde{\pi}_*^i, a')$ satisfying (7). Since graph $\mathbb{G}$ is closed and connected, type n can mix between strategies s^i and s^{nd} such that under the receiver's posterior belief when no additional outcome is disclosed, they have a (potentially mixed) best reply $\alpha \in \Delta(A)$ such that $\mathbb{E}_{a \sim \alpha}[u^s(a)] = u_i^*$, which makes type n indifferent between s^i and s^{nd}.

Case B: Suppose

$$u^* > \sum_{j=1}^{i} \pi_j u_j + \Big(1 - \sum_{j=1}^{i} \pi_j\Big) u^s(a)$$

for every $a \in \arg\max_{a' \in A} u^r(\pi^i, a')$. By (2), every such a satisfies $u^s(a) < u_i^*$; since $u_i \geq u_i^* \geq u_{i+1}$, this implies $u^s(a) < u_i$.

Following interim belief π, let type n use s^{nd}. Types above n first disclose n uninformative outcomes and then use any remaining capacity only to disclose fully revealing outcomes θ_j with $u_j > u^*$. Types $1, \ldots, n-1$ use a monotone disclosure rule: there is a cutoff disclosure probability q, so that all states above the cutoff are disclosed with probability one, the cutoff state may be mixed, and all lower-payoff states are concealed.

We now choose q and the receiver's best reply after no disclosure. For each q, let $V(q)$ be the set of sender payoffs from no disclosure, given the receiver's posterior when types $1, \ldots, n-1$ use the monotone rule with disclosure probability q. Let $\mathbb{V} \equiv \{(q, v) : v \in V(q)\}$, and let $\mathbb{U}$ be the graph of payoff levels that make the monotone disclosure rule incentive compatible for types $1, \ldots, n-1$: if $q = \sum_{j=1}^{\tau} \pi_j$, then $(q, v) \in \mathbb{U}$ iff $v \in [u_{\tau+1}, u_\tau]$, while if $\sum_{j=1}^{\tau} \pi_j < q < \sum_{j=1}^{\tau+1} \pi_j$, then $(q, v) \in \mathbb{U}$ iff $v = u_{\tau+1}$. Thus any intersection of $\mathbb{V}$ and $\mathbb{U}$ gives a monotone disclosure rule and a receiver best reply after no disclosure that satisfy the lower types' incentive constraints.

Let $q^* \equiv \sum_{j=1}^{i} \pi_j$. By the hypothesis of Case B, every payoff in $V(q^*)$ is strictly below u_i. Also $u^* \in V(1)$, while $u_k < u^*$ and $(1, u_k) \in \mathbb{U}$. Since $\mathbb{V}$ and $\mathbb{U}$ are closed and connected, and $\mathbb{U}$ is decreasing in q, there exists an intersection $(\hat{q}, \hat{v}) \in \mathbb{V} \cap \mathbb{U}$ with $\hat{q} \geq q^*$.

It remains only to check that type n does not want to imitate the lower types' monotone disclosure rule. If $\hat{q} = q^*$, this follows directly from the Case B inequality and $u_i^* \in [u_{i+1}, u_i]$. Suppose instead that $\hat{q} > q^*$. Write $\hat{q} = \sum_{j=1}^{\tau} \pi_j + \beta \pi_{\tau+1}$ for some $\tau \geq i$, $\beta \in [0, 1]$, and define $u^*(\tau, \beta)$ by

$$u^* = \sum_{j=1}^{\tau} \pi_j u_j + \beta \pi_{\tau+1} u_{\tau+1} + \Big(1 - \sum_{j=1}^{\tau} \pi_j - \beta \pi_{\tau+1}\Big) u^*(\tau, \beta). \tag{9}$$

This is the no-disclosure payoff that makes the sender indifferent between s^{nd} and the corresponding monotone disclosure rule. A direct differentiation of (9) gives

$$\frac{\partial u^*(\tau, \beta)}{\partial \beta} = \pi_{\tau+1} \frac{(1 - \sum_{j=1}^{i} \pi_j) u_i^* - \sum_{j=i+1}^{\tau} \pi_j u_j - \sum_{j=\tau+1}^{k} \pi_j u_{\tau+1}}{(1 - \sum_{j=1}^{\tau} \pi_j - \beta \pi_{\tau+1})^2} \geq 0,$$

where the inequality follows from $u_i^* \geq u_{i+1} \geq u_j$ for all $j \geq i+1$. Hence $u^*(\tau, \beta) \geq u^*(i, 0) = u_i^*$. Since $\mathbb{U}$ is decreasing, $\hat{q} > q^*$ implies $\hat{v} \leq u_{i+1} \leq u_i^* \leq u^*(\tau, \beta)$. Thus the payoff from the lower types' disclosure rule is no greater than the payoff from s^{nd}, so type n does not want to imitate them.

Therefore, the chosen monotone disclosure rule satisfies the lower types' incentive constraints, type n optimally uses s^{nd}; higher types also find it optimal to do so by the off-path-belief argument described above.

F Proof of Proposition 2

Let

$$\widehat{\pi} \equiv \frac{\pi^* \, \pi^{**}}{p \, \pi^{**} + (1-p) \, \pi^*} \quad \text{and} \quad a(\pi) \equiv 1 - \frac{\eta \pi}{\pi^{**} - \pi} \text{ for every } \pi \in [\pi^*, \widehat{\pi}). \tag{10}$$

By definition, $\widehat{\pi} \in (\pi^*, \pi^{**})$. Let $a^{\emptyset}(\pi) \in [0, 1+\eta]$ denote the receiver's expected action when their naive belief after the initial experiment is π and no additional outcome is disclosed.

Lemma 6. *In any equilibrium, we have (i) $a^{\emptyset}(\pi) = 0$ for every $\pi \in [0, \pi^*)$, (ii) $a^{\emptyset}(\pi) \in (0, a(\pi)] \cup \{a(\pi^*)\}$ for every $\pi \in [\pi^*, \widehat{\pi})$, and (iii) $a^{\emptyset}(\pi) \leq 1$ for every $\pi < \pi^{**}$.*

Proof of Lemma 6: Since type 1 sender has the option not to disclose the additional outcome, in equilibrium, their payoff is no less than $a^{\emptyset}(\pi)$ when the receiver's naive belief following the initial experiment is π. If $a^{\emptyset}(\pi) > 1$, then the receiver's posterior belief at every subsequent information set is at least π^{**}, which cannot happen when the naive belief satisfies $\pi < \pi^{**}$ since belief is a martingale so the receiver's average belief equals π. Similarly, if $\pi < \pi^*$, it cannot be the case that $a^{\emptyset}(\pi) > 0$. This is because otherwise, the receiver's posterior belief at every subsequent information set is at least π^*, which cannot happen when $\pi < \pi^*$.

Next, suppose by way of contradiction that $\pi \in [\pi^*, \widehat{\pi})$ but $a^{\emptyset}(\pi) \notin (0, a(\pi)] \cup \{a(\pi^*)\}$. Type 1's uniquely optimal strategy at π is to (i) conduct an additional binary experiment such that the receiver's posterior belief is π^{**} when the outcome is good and their naive belief is 0 when the outcome is bad and (ii) disclose the good outcome and conceal the bad outcome. By Bayes rule, the receiver's posterior belief without any additional disclosure is strictly less than π^* when $\pi < \widehat{\pi}$. This contradicts the hypothesis that $a^{\emptyset}(\pi) > a(\pi) > 0$. □

We solve for the sender-optimal equilibrium. Since $\pi_0 \in (0, \pi^*)$, by Lemma 6, it is without loss to focus on binary initial experiments where (i) one induced naive belief is 0, (ii) one induced naive belief is in $[\pi^*, \pi^{**}]$, and (iii) $a^{\emptyset}(\pi) = a(\pi)$ for any $\pi \in [\pi^*, \widehat{\pi})$, $a^{\emptyset}(\pi) = 1$ for any $\pi \in [\widehat{\pi}, \pi^{**})$, and $a^{\emptyset}(\pi^{**}) = 1 + \eta$. This leads to three candidates for the optimal equilibrium:

1. When the higher naive belief $\pi \in [\pi^*, \widehat{\pi})$ and $a^{\emptyset}(\pi) = a(\pi)$, type 1 mixes between (i) conducting an uninformative experiment and discloses its outcome and (ii) conducting an experiment that induces beliefs π^{**} and 0 and only discloses the good outcome. In

such an equilibrium, the sender's ex ante expected payoff is $p\frac{\pi_0 a(\pi)}{\pi} + (1-p)\frac{\pi_0}{\pi}$, so the optimal equilibrium within this class leads to payoff

$$V_1(\pi_0) \equiv \max_{\pi \in [\pi^*, \widehat{\pi}]} \Big\{ p\frac{\pi_0 a(\pi)}{\pi} + (1-p)\frac{\pi_0}{\pi} \Big\},$$

In equilibria where $a^{\emptyset}(\pi) = a(\pi^*)$ for some $\pi \in (\pi^*, \widehat{\pi})$, it is optimal for type 1 to conduct an additional experiment that induces beliefs 0, π^*, and π^{**} each with positive probability. However, such an equilibrium leads to a lower expected payoff compared to the one where the sender induces naive beliefs 0 and π^* in the initial experiment.

2. When the higher naive belief $\pi \in [\widehat{\pi}, \pi^{**})$, type 1 conducts an experiment that induces beliefs π^{**} and 0 and only discloses the good outcome. The optimal equilibrium within this class gives the sender ex ante expected payoff

$$V_2(\pi_0) \equiv \frac{\pi_0}{\widehat{\pi}} \Big\{ p + (1-p)\Big(\frac{\widehat{\pi}}{\pi^{**}}(1+\eta) + (1 - \frac{\widehat{\pi}}{\pi^{**}}) \Big) \Big\}$$

3. When the higher naive belief is π^{**}, type 1 does not conduct any additional experiment and the sender's ex ante expected payoff is

$$V_3(\pi_0) \equiv \frac{\pi_0}{\pi^{**}}(1+\eta).$$

Therefore, $V_{\max}(\pi_0) = \max\{V_1(\pi_0), V_2(\pi_0), V_3(\pi_0)\}$ and some algebra manipulations lead to the formulas in Proposition 2.

References

Ali, S. N., A. Kleiner, and K. Zhang (2024): "From Design to Disclosure," Working Paper.

Antic, N. and H. Pei (2026): "Selective Disclosure in Overlapping Generations," *arXiv preprint arXiv:2602.09406*.

Arieli, I. and C. Stewart (2025): "Bayesian Persuasion without Commitment," Working Paper.

Ben-Porath, E., E. Dekel, and B. L. Lipman (2018): "Disclosure and Choice," *Review of Economic Studies*, 85, 1471–1501.

Best, J. and D. Quigley (2024): "Persuasion for the Long Run," *Journal of Political Economy*, 132, 1740–1791.

Corrao, R. and Y. Dai (2025): "The Bounds of Mediated Communication," Working Paper.

Dai, Y., D. Fudenberg, and H. Pei (2026): "Bayesian Persuasion with Selective Disclosure," *arXiv preprint arXiv:2601.05914*.

DeVito, N. J., S. Bacon, and B. Goldacre (2020): "Compliance with Legal Requirement to Report Clinical Trial Results on ClinicalTrials.gov: A Cohort Study," *The Lancet*, 395, 361–369.

Dye, R. A. (1985): "Disclosure of Nonproprietary Information," *Journal of Accounting Research*, 123–145.

Dziuda, W. (2011): "Strategic Argumentation," *Journal of Economic Theory*, 146, 1362–1397.

Felgenhauer, M. and P. Loerke (2017): "Bayesian Persuasion with Private Experimentation," *International Economic Review*, 58, 829–856.

Felgenhauer, M. and E. Schulte (2014): "Strategic Private Experimentation," *American Economic Journal: Microeconomics*, 6, 74–105.

Fudenberg, D. and J. Tirole (1991): "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 53, 236–260.

Gao, Y. (2025): "Inference from Selectively Disclosed Data," Working Paper.

Goltsman, M., J. Hörner, G. Pavlov, and F. Squintani (2009): "Mediation, Arbitration and Negotiation," *Journal of Economic Theory*, 144, 1397–1420.

Guttman, I., I. Kremer, and A. Skrzypacz (2014): "Not Only What But Also When: A Theory of Dynamic Voluntary Disclosure," *American Economic Review*, 104, 2400–2420.

Hart, S., I. Kremer, and M. Perry (2017): "Evidence Games: Truth and Commitment," *American Economic Review*, 107, 690–713.

Henry, E. and M. Ottaviani (2019): "Research and the Approval Process: The Organization of Persuasion," *American Economic Review*, 109, 911–955.

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

Koessler, F. and V. Skreta (2023): "Informed Information Design," *Journal of Political Economy*, 131, 3186–3232.

Kreps, D. and R. Wilson (1982): "Sequential Equilibria," *Econometrica*, 863–894.

Kreutzkamp, S. and Y. Lou (2025): "Persuasion without Ex-Post Commitment," *Journal of Economic Theory*, 106058.

Lin, X. and C. Liu (2024): "Credible Persuasion," *Journal of Political Economy*, 132, 2228–2273.

Lipnowski, E. and D. Ravid (2020): "Cheap Talk with Transparent Motives," *Econometrica*, 88, 1631–1660.

Lipnowski, E., D. Ravid, and D. Shishkin (2022): "Persuasion via Weak Institutions," *Journal of Political Economy*, 130, 2705–2730.

——— (2025): "Perfect Bayesian Persuasion," *JPE: Microeconomics*, forthcoming.

Lou, Y. (2023): "Private Experimentation, Data Truncation, and Verifiable Disclosure," *arXiv preprint arXiv:2305.04231*.

Lyu, Q. and W. Suen (2025): "Information Design in Cheap Talk," Working Paper.

Mathevet, L., D. Pearce, and E. Stacchetti (2024): "Reputation and Information Design," Working Paper.

Matthews, S. and A. Postlewaite (1985): "Quality Testing and Disclosure," *RAND Journal of Economics*, 328–340.

Pei, H. (2023): "Repeated Communication with Private Lying Costs," *Journal of Economic Theory*, 210, 105668.

——— (2025): "Community Enforcement with Endogenous Records," *Review of Economic Studies*, forthcoming.

Perez-Richet, E. (2014): "Interim Bayesian Persuasion: First Steps," *American Economic Review*, 104, 469–474.

Rappoport, D. (2025): "Evidence and skepticism in verifiable disclosure games," *Theoretical Economics*, 20, 1213–1246.

Shishkin, D., M. Titova, and K. Zhang (2025): "Withholding Verifiable Information," Working Paper.

Titova, M. and K. Zhang (2025): "Persuasion with Verifiable Information," *Journal of Economic Theory*, 106102.

Zapechelnyuk, A. (2023): "On the Equivalence of Information Design by Uninformed and Informed Principals," *Economic Theory*, 76, 1051–1067.