Bayesian Persuasion with Selective Disclosure*

Yifan Dai[†] Drew Fudenberg[‡] Harry Pei[§]

October 30, 2025

Abstract: We analyze a model where a sender first commits to a publicly observable experiment and then can secretly conduct some additional experiments and selectively disclose their outcomes. The sender has private information about the maximal number of additional experiments they can conduct (i.e., their *type*). We show that the sender's payoff is bounded below their commitment payoff in all equilibria when (i) the receiver has little information about the sender's type and (ii) the sender can benefit from selectively disclosing additional information at some beliefs induced by the optimal experiment under full commitment.

Keywords: Bayesian persuasion, lack of commitment, selective disclosure.

JEL Codes: C73, D82, D83.

*We thank Pak Hung Au, Jeffrey Ely, Ying Gao, Alexander Jakobsen, Navin Kartik, Xiao Lin, Elliot Lipnowski, Stephen Morris, Alessandro Pavan, Alexander Wolitzky, Kun Zhang, and Gabriel Ziegler for helpful conversations and 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. In contrast, when the prosecutor can commit to an information-gathering process and to fully disclosing its results—as in the Bayesian persuasion model of Kamenica and Gentzkow (2011)—their statements become informative and can influence the judge's decision.

In practice, however, even when experts can commit to acquiring and truthfully disclosing certain types of information, it may be hard for them to commit not to conduct further investigations and then selectively disclose their findings. One context where this issue arises is the pharmaceutical industry. When pharmaceutical companies seek approval for new drugs, they must register certain clinical trials with the FDA and are required to disclose the resulting outcomes. Nevertheless, some companies conduct additional trials after a drug has been approved. Although they are legally obligated to report all clinical trial results, they sometimes conceal or delay the publication of findings—particularly for post-approval trials that were not yet registered.[1] This lack of commitment can undermine the expert's credibility, as decision-makers may suspect that unfavorable evidence has been withheld and therefore discount the expert's advice.

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, as in Kamenica and Gentzkow (2011). Our modeling innovation is that after the initial experiment, the sender may privately conduct additional experiments and selectively disclose some of their outcomes to the receiver. The sender has private information about the maximum number of additional experiments they can conduct, which we refer to as their *type*. This heterogeneity captures differences in the sender's resources or investigative capabilities.

Our results examine how the sender's lack of commitment affects their credibility, and in particular, their ability to attain the payoff they would obtain under full commitment. Throughout this section, we illustrate our findings via an example with two states, *good* and *bad*, and three actions 0, 2, and 3. The sender's payoff is state-independent and equals the receiver's action. The receiver's optimal action is 0 when the good state occurs with probability less than $\frac{1}{3}$, is 3 when the good state occurs with probability more than $\frac{2}{3}$, and is 2 otherwise. Figure 1 depicts the sender's indirect utility as a function of the receiver's belief about the state.

Our first observation is that the sender's lack of commitment has bite only when the receiver faces

[1] DeVito, Bacon, and Goldacre (2020) found that of 4209 trials due to report results to the FDA between 2018 and 2019, only 722 (40.9%) did so within the 1-year deadline. 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.

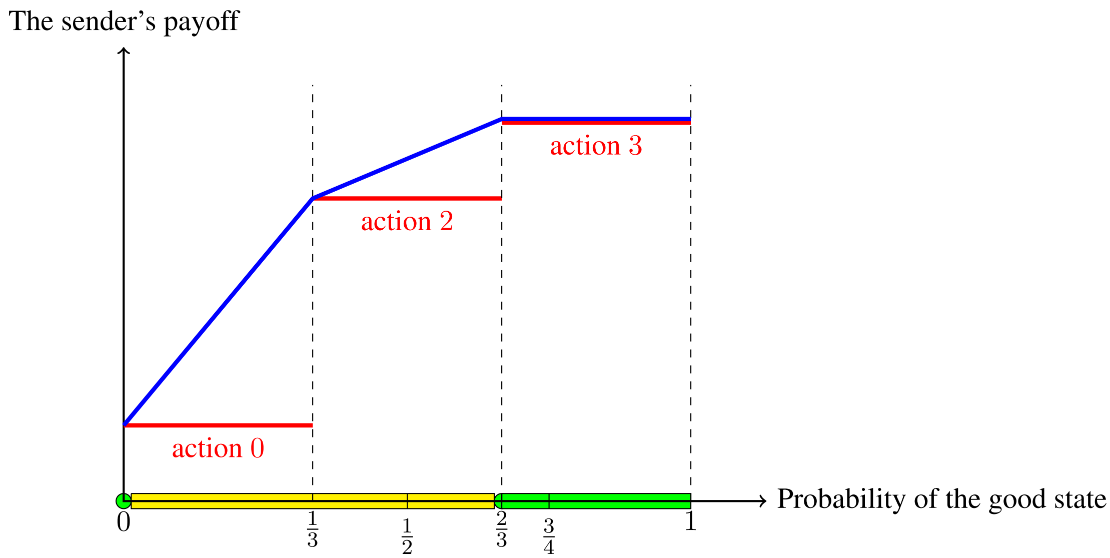

Figure 1: The sender's payoff under the receiver's best reply as a function of the receiver's belief (red) and its concave closure (blue), with non-credible beliefs in yellow and credible beliefs in green.

uncertainty about the sender's capacity. This is because when the receiver knows the sender's capacity t, the sender can prove to the receiver that no information is hidden by conducting t uninformative additional experiments and disclosing all the outcomes. Hence, the sender can secure their *commitment payoff* (i.e., their payoff when they can commit not to conduct additional experiments) by choosing the initial experiment to be the *optimal experiment* (i.e., the one that is optimal for the sender if they have full commitment).

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 our 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*, which are the ones 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 obtain their commitment payoff under all type distributions *if and only if* the optimal experiment does not induce any non-credible belief.[2] If the optimal experiment induces at least one non-credible belief, then under an open set of type distributions, e.g., distributions that are close to the uniform distribution on the set $\{0, 1, 2, ..., n\}$ for large enough n, all of the sender's equilibrium payoffs are bounded below their commitment payoff. Under a more demanding requirement, which in the example is that the prior probability of the good state is strictly positive but is no more than $\frac{1}{3}$, the sender's payoff is bounded below their commitment payoff under all type distributions that assign positive probability to sufficiently many types. These results suggest that the sender cannot attain their commitment payoff when sufficiently many types occur with positive probability and the optimal experiment induces at least one non-credible belief. In contrast, when one sender type occurs with probability close to 1, Theorem 2 shows that there is always an equilibrium in which the sender's payoff is close to their commitment payoff.

To explain the intuition for Theorem 1, suppose toward a contradiction that there is an equilibrium in which the sender's payoff is close to the commitment payoff. If many types have positive probability, there will be many types of the sender under which the receiver will hold at least one non-credible belief induced by the optimal experiment. Take any two such types. The higher-type sender can imitate the lower type's equilibrium strategy up to the point where that non-credible belief would arise, then perform a fully informative additional experiment and disclose its outcome if and only if it reveals a state that leads to a higher payoff for the sender. Under this deviation, the higher type guarantees itself the lower type's equilibrium payoff plus a strictly positive constant. A contradiction obtains once under sufficiently many types, at least one non-credible belief occurs with probability bounded above zero.

Our paper contributes to the literature on Bayesian persuasion by introducing a novel form of commitment problem that arises naturally in several applications, such as pharmaceutical companies conducting clinical drug trials. The form of lack of commitment we considered stands in contrast to those in existing models such as the sender can commit to an information structure but cannot commit to truthfully reveal what they learn (Ivanov 2010, Lipnowski, Ravid and Shishkin 2022), cannot commit to information structures but can commit not to falsify information (Henry and Ottaviani 2019, Titova and Zhang 2025), or cannot commit not to deviate to strategies that are non-detectable, i.e., inducing the same message distribution (Lin and Liu 2024), or cannot commit in the stage game but interacts repeatedly with a sequence of receivers (e.g., Pei 2023, Best and Quigley 2024, Mathevet, Pearce and Stacchetti 2024).[3]

[2]The payoff environment is monotone if there are complete orders on the states and actions such that (i) the sender's payoff increases in the receiver's action and (ii) the receiver's payoff has strictly increasing differences. In non-monotone environments, we obtain conditions under which the sender can or cannot obtain their commitment payoff, but there is a gap between our necessary conditions and our sufficient conditions.

[3]Lipnowski and Ravid (2020) and Kamenica and Lin (2025) examine the value of commitment by comparing the sender's payoff in the Bayesian persuasion game with their equilibrium payoff in a cheap talk game where the sender directly observes the state.

More closely related is the work of Felgenhauer and Loerke (2017), in which the sender has state-independent payoffs and cannot commit to a public experiment, but can covertly choose a sequence of binary experiments and then select a subset of the resulting outcomes which they disclose to a receiver.[4] They show that the sender's ability to conduct secret experiments increases the precision of the receiver's information compared to the Bayesian persuasion benchmark. In contrast, the sender in our model can commit to an initial experiment and to publicly disclose its outcome before conducting additional experiments. Our model is motivated by applications of persuasion models where experts can commit to certain experiments but cannot commit not to conduct additional experiments and to selectively reveal their outcomes.

In our model, the sender receives private information about their type before conducting additional experiments. This is related to the literature on information design with an informed sender, which includes the works of Perez-Richet (2014), Koessler and Skreta (2023), and Zapechelnyuk (2023), where the sender's private information directly affects the receiver's payoff. In our model, the sender's type in our model is payoff-irrelevant for the receiver, although it affects the receiver's inference about the payoff-relevant state once the sender cannot commit not to conceal the outcomes of the additional experiments.

Our paper is also related to the literature following Dye (1985) on the disclosure of evidence when the amount of evidence available to the sender is unknown to the receiver, and especially to Dzuida (2011) and Gao (2025) where a sender has multiple pieces of evidence and decides which subset of them to disclose. In contrast, the set of evidence available to our sender is endogenous, which is determined by the sender's choice of additional experiments.[5]

2 The Baseline Model

A sender (s) interacts with a receiver (r). The sender's payoff function $u^s(a)$ 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 function $u^r(\theta, a)$ depends on both a and the state of the world $\theta \in \Theta$. We assume that Θ and A are finite sets. 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 introduce the following generic assumption on the receiver's payoff function:

Assumption 1. *The receiver's payoff function $u^r(\theta, a)$ satisfies:*

1. *For every $\theta \in \Theta$, there exists a unique $a \in A$ that maximizes $u^r(\theta, a)$.*

[4] Felgenhauer and Schulte (2014) study a similar model except that the experiments are exogenously fixed.
[5] Pei (2025) studies a repeated game where each player can selectively disclose the signals generated by their actions.

2. *For every* $a \in A$ *and* $\pi \in \Delta(\Theta)$ *such that* $a \in \arg\max_{a' \in A} \sum_{\theta \in \Theta} \pi(\theta) u^r(\theta, a')$, *there exists an open set of beliefs* $\Pi \subseteq \Delta(\Theta)$ *such that* $\{a\} = \arg\max_{a' \in A} \sum_{\theta \in \Theta} \pi'(\theta) u^r(\theta, a')$ *for every* $\pi' \in \Pi$.

Assumption 1 requires that (i) the receiver has a strict best reply when their belief is degenerate and (ii) for every action a^* that is one of the receiver's best replies under *some* belief π about the state, there exists an open set of beliefs under which a^* is the receiver's strict best reply. Since Θ and A are finite, both requirements are satisfied for a set of receiver-payoff functions of full Lebesgue measure in $\mathbb{R}^{|\Theta| \times |A|}$.

An *experiment*, denoted by σ, is a mapping from Θ to the set of probability distributions over outcomes $s \in S$, where S is a countably infinite set. For any σ, we use $\sigma(s \mid \theta)$ to denote the probability of outcome s conditional on state θ. Let Σ denote the collection of all experiments. We refer to an experiment σ together with one of its resulting outcomes s as a *pair*. Let $\mathcal{H}_k$ denote the set of sequences that consist of k pairs.

At the beginning of the game, the sender chooses an *initial experiment* σ_0, and then both players observe σ_0 and its realized outcome $s_0 \in S$. Our modeling innovation is that after observing (σ_0, s_0) but before the receiver chooses their action, the sender may conduct some *additional experiments*. 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.[6] 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 $\{\sigma_i, s_i\}_{i=1}^{k-1}$. After observing the outcomes of all the additional experiments they conducted, the sender chooses a subset of outcomes to disclose to the receiver. 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 additional experiments that are not disclosed. The receiver then chooses $a \in A$.

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 set of experiments the sender has revealed.[7] The receiver's strategy is $\alpha : \mathcal{H} \to \Delta(A)$, which maps the sequence of pairs they observe to their actions. The sender's information sets after the initial node belong to $\mathcal{H} \times \mathbb{N}$. A pure strategy for the sender consists of (i) an initial experiment $\sigma_0 \in \Sigma$, (ii) a mapping from $\mathbb{N} \times \mathcal{H}$ to $\Sigma \cup \{\emptyset\}$ that captures the sender's choice of the next additional experiment conditional on their type t and the outcomes of the experiments already conducted, where $\emptyset$ stands for conducting no further experiment,[8] and (iii) a mapping from every type $t \in \mathbb{N}$ and every $h \in \mathcal{H}_k$

[6] Section 4 extends our results to the case where the sender observes t before choosing their initial experiment.

[7] Whether the receiver observes the order of the disclosed additional experiments does not affect any of our results.

[8] If the sender's type is t, then they can only choose $\emptyset$ (i.e., conduct no further experiment) for any $h \in \mathcal{H}_k$ with $k \geq t + 1$.

with $k \geq 1$ to subsets of $\{1, 2, ..., k-1\}$, where i belongs to the chosen subset when the outcome of the ith additional experiment is disclosed. The sender's *mixed strategy* is a distribution of their pure strategies.

For every $h \in \mathcal{H}$, the receiver's *naive belief* at h, denoted by $\widehat{\pi}_h$, is their posterior belief about θ if they observe h and believe that the sender has disclosed the outcomes of all the experiments that have been conducted. By definition, the receiver's naive belief depends only on the pairs in h but not on the order of these pairs and the sender's strategy. 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 equilibrium 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 and (ii) the receiver's belief is derived from Bayes rule on the equilibrium path, and is consistent with the conditional probability system induced by the sender's strategy at every information set.[9] Consistency with respect to the conditional probability system implies that at every information set $h \in \mathcal{H}$ with length k, the support of the receiver's posterior belief (i) only assigns positive probability to t that is at least $k-1$ and (ii) only assigns positive probability to states that belong to the support of the receiver's naive belief $\widehat{\pi}_h$.[10]

3 Analysis

Our results compare the sender's equilibrium payoffs with those in the Bayesian persuasion game, which we call their commitment payoff. We show that the sender's lack of commitment matters only when the receiver faces nontrivial uncertainty about the sender's type, and that the sender's commitment payoff strictly exceeds their expected payoff from fully disclosing the state. We establish conditions under which, in all equilibria, lack of commitment results in payoffs that are bounded below the sender's commitment payoff.

[9] In general games, PEBE imposes additional restrictions that are always satisfied here. A conditional probability system on set $\Omega \equiv \Theta \times \mathbb{N} \times \mathcal{H}$ is a function $\mu(\cdot|\cdot) : 2^{\Omega} \times 2^{\Omega} \backslash \{\emptyset\} \to [0, 1]$ such that (i) for all non-empty $\Omega_0 \subseteq \Omega$, $\mu(\cdot|\Omega_0)$ 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 $\mu(\Omega_3|\Omega_2)\mu(\Omega_2|\Omega_1) = \mu(\Omega_3|\Omega_1)$.

[10] When t is bounded above by t^*, PEBE does not imply that the receiver's belief conditional on observing any $h \in \mathcal{H}_{t^*+1}$ equals their naive belief. Nevertheless, all the equilibria constructed in our examples and proofs satisfy this additional requirement.

3.1 Preliminaries

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 when their belief about the state is $\pi \in \Delta(\Theta)$. Let

$$\overline{u}(\pi) \equiv \max_{a \in A^*(\pi)} u^s(a) \quad \text{and} \quad \underline{u}(\pi) \equiv \min_{a \in A^*(\pi)} u^s(a)$$

denote the sender's highest and lowest payoffs under π, respectively. We will use $u(\theta)$ to denote the sender's payoff when the receiver plays their (unique) best reply in state θ, and denote the sender's *full-disclosure payoff* by

$$\underline{V}(\pi_0) \equiv \sum_{\theta \in \Theta} \pi_0(\theta) u(\theta).$$

As in the Bayesian persuasion model of Kamenica and Gentzkow (2011), once the prior π_0 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. In an abuse of notation, we use $\sigma[\pi]$ to denote the probability with which belief π occurs in the distribution over beliefs that characterizes σ. We 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. The sender's *commitment payoff* is

$$\overline{V}(\pi_0) \equiv \max_{\sigma \in \Sigma[\pi_0]} \sum_{\pi \in \Delta(\Theta)} \sigma[\pi] \overline{u}(\pi),$$

which is the payoff they receive when they can commit not to conduct any additional experiment. By definition, $\underline{V}(\pi_0) \leq \overline{V}(\pi_0)$. An experiment is *optimal* if it belongs to $\arg\max_{\sigma \in \Sigma[\pi_0]} \sum_{\pi \in \Delta(\Theta)} \sigma[\pi] \overline{u}(\pi)$.

Assumption 2. *The primitives (π_0, u^s, u^r) are such that there is a unique optimal experiment.*

In the case where $|\Theta| = 2$, Assumption 2 is satisfied for all (π_0, u^s, u^r) except for those where $\overline{V}(\pi_0) = \max_{\pi \in \Delta(\Theta)} \overline{u}(\pi)$ and a zero-measure set of payoff functions (u^s, u^r). We use σ^* to denote the unique optimal experiment. Assumptions 1 and 2 together ensure that there are experiments that are close to σ^* that generate strict receiver incentives (see e.g. Lipnowski, Ravid and Shishkin 2024 and Ali, Kleiner and Zhang 2024), from which the sender can secure payoffs arbitrarily close to $\overline{V}(\pi_0)$ once they can commit not to conduct additional experiments. Proposition 1 shows that in our game, the sender's equilibrium payoff must lie between $\underline{V}(\pi_0)$ and $\overline{V}(\pi_0)$, and that their payoff is $\overline{V}(\pi_0)$ in all equilibria once their type t is common knowledge.

Proposition 1. *Fix any $\pi_0 \in \Delta(\Theta)$.*

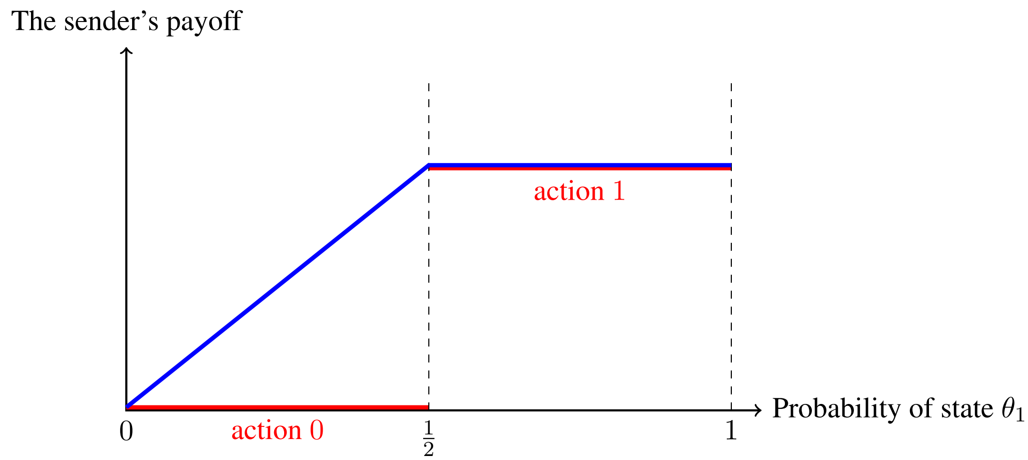

Figure 2: The sender's payoff function $\overline{u}(\pi)$ (red) and its concave closure (blue) in Example 1.

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

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

Proposition 1 shows that the sender's lack of commitment lowers their equilibrium payoff only if (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. If either condition fails, the sender attains their commitment payoff in all equilibria.

The proof is in Appendix A. Intuitively, the sender can secure their full-disclosure payoff by fully disclosing the state in the initial experiment. In settings where the receiver also knows t, the sender can secure a payoff that is arbitrarily close to their commitment payoff by (i) choosing an initial experiment that is close to σ^* under which the receiver has strict incentives at each induced belief and (ii) conducting t additional experiments and fully disclosing their outcomes.

3.2 Results: Attaining the Commitment Payoff

We focus on settings where the type distribution p is not degenerate and (u^s, u^r, π_0) satisfies $\overline{V}(\pi_0) > \underline{V}(\pi_0)$. We start with an example where the sender attains their commitment payoff in some equilibrium even though p is non-degenerate and the sender's commitment payoff exceeds their full disclosure payoff, which is also depicted graphically in Figure 2.

Example 1: *Suppose $a \in A \equiv \{0, 1\}$ and $\theta \in \Theta \equiv \{\theta_0, \theta_1\}$. The sender's payoff is a and the receiver prefers $a = 0$ if and only if their belief assigns probability less than $\frac{1}{2}$ to state θ_1. Let π_0 denote the* *probability of θ_1 under the prior belief. When $\pi_0 \in (0, \frac{1}{2})$, the sender's optimal experiment σ^* induces belief $\frac{1}{2}$ with probability $2\pi_0$ and belief 0 with probability $1 - 2\pi_0$. Regardless of p, there always exists an equilibrium where the sender chooses σ^* as their initial experiment, conducts no additional experiment, and obtains their commitment payoff. Bayes rule implies that the receiver's posterior belief coincides with their naive belief on the equilibrium path; we specify off-path beliefs that have the same property. The sender will choose σ^* as their initial experiment and will conduct no additional experiment on the equilibrium path. The receiver's posterior belief will coincide with their naive belief at every information set.*

A key feature of Example 1 is that σ^* only induces beliefs that are strongly credible in the following sense:

Definition 1. *Belief $\pi \in \Delta(\Theta)$ is* strongly credible *if $\overline{u}(\pi) \geq \underline{u}(\pi')$ for every π' such that supp$(\pi') \subseteq$ supp(π).*

By definition, if a belief is strongly credible, the sender cannot strictly benefit from revealing more information. In particular, degenerate beliefs are strongly credible. In Example 1, a belief is strongly credible if and only if the probability it assigns to state θ_1 belongs to $\{0\} \cup [\frac{1}{2}, 1]$. If all beliefs induced by the optimal experiment are strongly credible, the sender can attain their commitment payoff in an equilibrium where (i) they choose the optimal experiment σ^* as their initial experiment and conduct no additional experiment on the equilibrium path and (ii) at every information set, the receiver's posterior belief coincides with their naive belief.

Proposition 2. *If all the beliefs induced by σ^* are strongly credible, then for every type distribution $p \in \Delta(\mathbb{N})$, there exists a PEBE in which the sender's payoff is $\overline{V}(\pi_0)$.*

The proof is omitted since it follows directly from the description in the preceding paragraph.

When σ^* induces some beliefs that are not strongly credible, the sender might be tempted to disclose some additional outcomes in order to induce the receiver to take more favorable actions. As a result, the receiver may get suspicious when not enough additional outcomes are disclosed. Their suspicion may hurt the low-type senders who cannot conduct enough additional experiments. This can be seen from Example 2, which is depicted graphically in Figure 3.

Example 2: *Suppose $A \equiv \{0, 2, 3\}$, $\Theta \equiv \{\theta_0, \theta_1\}$, and $t \in \{0, 1\}$. The receiver's prior belief assigns probability $\pi_0 \in (0, \frac{1}{3})$ to state θ_1 and probability $\frac{1}{2}$ to $t = 1$. The sender's payoff equals a. Action 0 is optimal for the receiver if $\pi \in [0, \frac{1}{3}]$, 2 is optimal if $\pi \in [\frac{1}{3}, \frac{2}{3}]$, and 3 is optimal if $\pi \in [\frac{2}{3}, 1]$.*

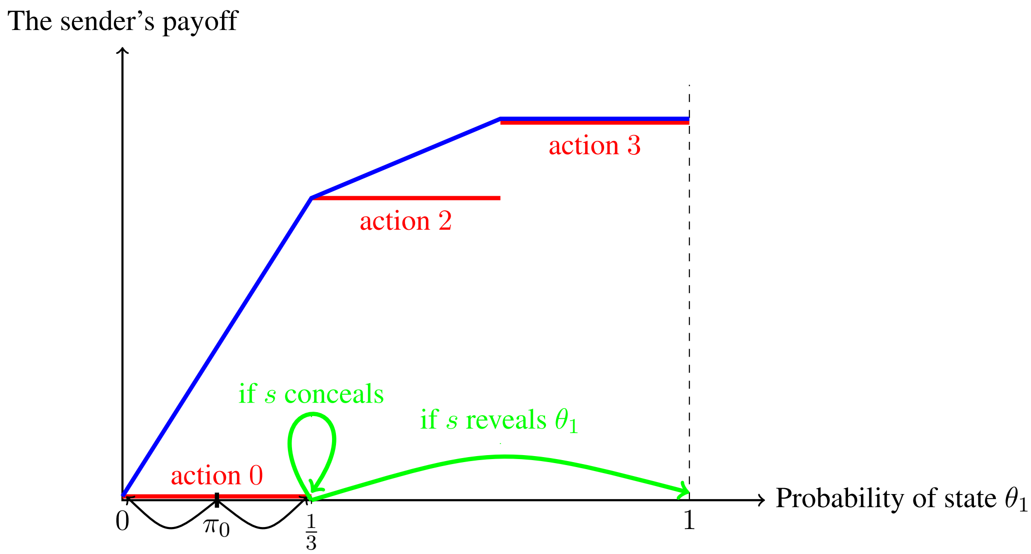

Figure 3: The sender's payoff function $\overline{u}(\pi)$ (red) and its concave closure (blue) in Example 2. The green arrows represent type 1 sender's profitable deviation after observing the outcome of the initial experiment that leads to naive belief $\frac{1}{3}$.

We claim that then the sender's equilibrium payoff is bounded below $\overline{V}(\pi_0)$*, which here is* $6\pi_0$*. To see why, suppose by way of contradiction that there exists an equilibrium in which the sender's payoff is* $6\pi_0$*. Then conditional on state* θ_1*, the receiver's posterior belief must be* $\frac{1}{3}$ *conditional on each sender type. Since types* 0 *and* 1 *both occur with positive probability, after conducting the initial experiment, there must exist* $h \in \mathcal{H}_1$ *at which the receiver's naive belief is* $\frac{1}{3}$*. However, if this is the case, type* 1 *sender will have an incentive to conduct an additional fully informative experiment at* h*, and reveal its outcome if and only if* $\theta = \theta_1$*. This contradicts the requirement that the receiver's posterior belief is* $\frac{1}{3}$ *when the state is* θ_1*.*

The fact that the sender can have an incentive to disclose additional outcomes at beliefs that are not strongly credible does not necessarily preclude them from obtaining their commitment payoff, as shown by the next example, which is depicted in Figure 4.

Example 3: *Suppose* $A \equiv \{0, 2, 3, \frac{7}{2}\}$*,* $\Theta \equiv \{\theta_0, \theta_1\}$*, and* $t \in \{0, 1\}$*. As in Example 2, we use the probability of state* θ_1 *to represent the receiver's belief. 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* $\pi \in [\frac{3}{4}, 1]$*. Suppose* $\pi_0 = \frac{3}{8}$ *and* $t = 0$ *with probability* $\frac{1}{3}$*.*[11] *We construct an equilibrium in which the sender's expected payoff is* $\overline{V}(\pi_0) = \frac{5}{2}$*, which is attained by inducing*

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

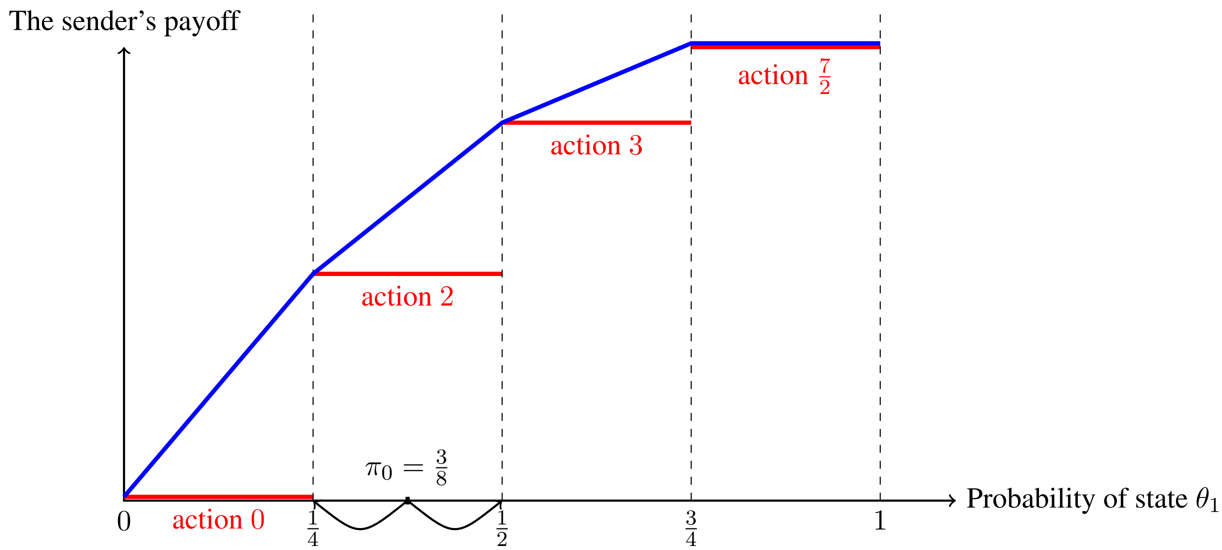

Figure 4: The sender's payoff function $\overline{u}(\pi)$ (red) and its concave closure (blue) in Example 3.

posteriors $\frac{1}{4}$ and $\frac{1}{2}$ with equal probability.

- *The sender chooses an uninformative initial experiment.*
- *Type 1 sender conducts an additional experiment where outcome $\overline{s}$ occurs with probability 1 when $\theta = \theta_1$, and occurs with probability $\frac{3}{5}$ when $\theta = \theta_0$. They reveal the outcome if and only if it is $\overline{s}$.*
- *The receiver's posterior belief is $\frac{1}{2}$ after observing $\overline{s}$ and is $\frac{1}{4}$ after observing no additional signal. Their posterior belief coincides with their naive belief at any off-path information set.*

In Example 3, type 1 sender has no incentive to fully reveal state θ_1 via additional experiments. Instead, they prefer to disclose an additional signal that partially reveals the state and induces posterior belief $\frac{1}{2}$. Under certain type distributions, this leads to equilibria where the sender obtains their commitment payoff.

Motivated by Example 3, we examine how the sender's type distribution affects their ability to attain their commitment payoff. We start by introducing a condition on beliefs that is weaker than strong credibility:

Definition 2. *Belief $\pi \in \Delta(\Theta)$ is* credible *if $\overline{u}(\pi) \geq u(\theta)$ for every $\theta \in supp(\pi)$. We say π is non-credible if it is not credible.*

By definition, π is credible as long as the sender cannot strictly benefit from fully revealing any state. This condition is less demanding than a belief being strongly credible, which requires the sender not being able to benefit from disclosing any information, including but not limited to fully revealing certain states.

Proposition 3 shows that when σ^* induces at least one belief that is not credible, the sender cannot attain their commitment payoff in any equilibrium under some distribution of their type.

Proposition 3. *If the optimal experiment σ^* induces a non-credible belief, then there exist $\eta > 0$ and an open set of type distributions (in the product topology) such that, for every type distribution p in this set, the sender's payoff in any PEBE is at most $\overline{V}(\pi_0) - \eta$.*

The proof of Proposition 3, which is in Appendix B, shows that for any sufficiently large integer n, the conclusion holds in an open neighborhood around the uniform distribution on $\{0, 1, ...n\}$. The key step of our proof bounds type $t+1$ sender's payoff from below by computing their payoff from initially using the equilibrium strategy of type t and then at every receiver posterior belief π induced by type t that is not credible, conducting an additional experiment that fully reveals the state, and disclosing the resulting outcome if and only if the corresponding state θ satisfies $u(\theta) > \overline{u}(\pi)$. 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 posterior beliefs. Suppose by way of contradiction that there exists an equilibrium where the sender's ex ante expected payoff is close to their commitment payoff, then there will be sufficiently many types that will induce non-credible posterior beliefs with probability bounded above zero. Since the difference between the highest type's payoff and the lowest type's payoff is bounded above, this will lead to a contradiction once n is large enough.

In general, credibility is less demanding than strong credibility, so Propositions 2 and 3 do not cover cases where all beliefs induced by σ^* are credible but some of them are not strongly credible. Nevertheless, a belief is credible if and only if it is strongly credible when the payoff environment is *monotone*, which is the case in all of our examples and has been a primary focus of the literature on Bayesian persuasion.

Definition 3. *A payoff environment (u^s, u^r) is monotone if there exist complete orders on Θ and A under which $u^s(a)$ is strictly increasing in a and $u^r(\theta, a)$ has strictly increasing differences.*

Theorem 1. *Suppose the payoff environment (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 $\overline{V}(\pi_0)$.*

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

We omit the proof of Theorem 1 since it is an immediate consequence of Propositions 2 and 3, as well as the fact that all credible beliefs are strongly credible in monotone environments.

The next proposition shows that when the payoff environment satisfies a more demanding requirement than that in Proposition 3, the sender's payoff is bounded below their commitment payoff for any type distribution placing positive probability on sufficiently many types. For each $\theta \in \Theta$, let Π^*_θ denote the set of state distributions that (i) have θ in their support and (ii) are induced by the optimal experiment σ^*.

Proposition 4. *Suppose there exists $\theta \in \Theta$ such that*

$$u(\theta) > \max_{\pi \in \Pi^*_\theta} \overline{u}(\pi).^{12}$$

Then there exists $\lambda > 0$ such that, for every sufficiently small $\eta > 0$, there exists $n \in \mathbb{N}$ with the following property: for any type distribution p assigning probability more than $\lambda\eta$ to at least n types, the sender's payoff in every PEBE is strictly less than $\overline{V}(\pi_0) - \eta$.

Proposition 4 assumes the existence of a state θ such that, when the true state is θ, no belief $\pi \in \Pi^*_\theta$ induced by the optimal experiment σ^* is credible: each yields a sender payoff strictly below that from fully revealing θ, i.e. $\overline{u}(\pi) < u(\theta)$. This assumption is satisfied in Examples 2 and 3, although it is more demanding than the hypothesis of Proposition 3, which only requires the existence of a non-credible belief induced by the optimal experiment σ^*.

The proof is in Appendix C. Its logic parallels that of Proposition 3: we derive a lower bound on the higher-type sender's payoff by evaluating the payoff obtained from imitating the equilibrium strategy of a lower type and then performing an additional experiment that fully reveals the state, disclosing the result only if it yields a strictly higher payoff. When $u(\theta) > \max_{\pi \in \Pi^*_\theta} \overline{u}(\pi)$, the higher type can guarantee a strictly larger payoff than any lower type's equilibrium payoff, regardless of the probability that the lower type's strategy induces non-credible beliefs. The contrapositive of the proposition—namely, that 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.

A common theme of Propositions 3 and 4 is that when some beliefs induced by σ^* are non-credible, the sender cannot attain their commitment payoff once sufficiently many types occur with probability bounded above zero. In contrast, if the type distribution p assigns probability sufficiently close to 1 to one type, there always exists an equilibrium in which the sender's payoff is close to their commitment payoff.

[12] This requirement is stronger than that in Proposition 3, which only requires that σ^* induces a non-credible belief. Appendix D provides an example showing that the stronger condition in Proposition 4 is not redundant.

Theorem 2. *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 $\overline{V}(\pi_0) - \delta$.*

Theorems 1 and 2 together imply that in environments that are monotone, as long as the optimal experiment induces some non-credible beliefs, the highest payoff the sender can obtain depends on the dispersion of the receiver's belief about their type. If this belief is sufficiently concentrated on one point (so that the receiver faces little uncertainty about the sender's capacity to conduct additional experiments), then there exist equilibria in which the sender's payoff is approximately $\overline{V}(\pi_0)$. If the receiver's belief assigns positive probability to many sender types and each of those types occur with probability at least η, then all of the sender's equilibrium payoffs are bounded below their commitment payoff.

Theorem 2 and Proposition 4 are complementary, as they provide lower and upper bounds on the gap between the sender's commitment payoff and their highest equilibrium payoff. The proof of Proposition 4 implies that the gap is bounded below by a linear function of the probability of the nth highest 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 inducing those non-credible beliefs. Theorem 2 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 2, 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 $u(\theta)$. The sender's ex ante expected payoff is close to $\overline{V}(\pi_0)$ 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. The equilibrium strategies constructed in the proof depend on the relative probabilities of type 0 and types 1 to $n-1$, such that type n uses the equilibrium strategy of types 1 to $1-n$ with positive probability when this relative probability is high, and types 1 to

$n-1$ mix between disclosing and not disclosing in some state when this relative probability is low.

3.3 The Sender's Worst Equilibrium Payoff

Section 3.2 focuses on whether the sender attains their commitment payoff in the best equilibrium. Our next result examines the sender's lowest equilibrium payoff. It shows that as long as the sender's type t is drawn from a distribution that is unbounded from above, there always exists an equilibrium where the sender obtains their full disclosure payoff $\underline{V}(\pi_0)$, which is the lower bound derived in Proposition 1.

Proposition 5. *If p assigns positive probability to infinitely many types, then there exists a PEBE in which the sender fully reveals θ and conducts no additional experiment on the equilibrium path.*

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 2 and Proposition 5 together imply that there are equilibria in which the sender approximately attains their commitment payoff and there also equilibria in which the sender obtains their full disclosure payoff. Hence, compared to settings where the receiver knows the sender's type, in settings where the receiver is almost certain about the sender's type, the sender's lack of commitment has a significant effect on their payoff in the worst equilibrium.

Our proof of Proposition 5 constructs an equilibrium where the receiver forms skeptical beliefs at histories where the state is not fully revealed: At every information set $h \in \mathcal{H}$, the receiver's posterior belief will assign probability 1 to the worst state in the support of their naive belief $\widehat{\pi}_h$ (i.e., the state under which the receiver's best reply minimizes the sender's payoff). When the sender's type distribution p is unbounded, no matter how many additional experiments the sender has revealed, it is always plausible that the sender has concealed the outcome of at least one additional experiment and that this outcome reveals the worst state in the support of the receiver's naive belief.

Proof of Proposition 5: Let $\Theta \equiv \{\theta_1, \theta_2, ..., \theta_n\}$. Without loss of generality, we enumerate the states so that $u(\theta_n) \geq ... \geq u(\theta_1)$. Consider the following strategy profile and assessment: The sender chooses an initial experiment that fully reveals the state and conducts no additional experiment on the equilibrium path. For every $t \geq 1$ and every off-path history $\tilde{h}_t \in \mathcal{H}_t$, let $i(\tilde{h}_t)$ denote the lowest i such that $\theta_i \in \text{supp}(\widehat{\pi}_{\tilde{h}_t})$. Sender types t and above will conduct one fully informative additional experiment at $\tilde{h}_t$ and disclose the outcome unless the state is $\theta_{i(\tilde{h}_t)}$. At every off-path information set, the receiver's posterior belief assigns probability 1 to state $\theta_{i(\tilde{h}_t)}$, 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 and the sender's expected payoff is $\underline{V}(\pi_0)$. □

4 Discussions

We discuss the robustness of our results to several alternative variations of our baseline model.

The Sender Observes Their Type Before Conducting the Initial Experiment: In our baseline model, the sender observes the realized t after choosing the initial experiment. All of our results extend to a setting where the sender observes t before choosing the initial experiment, once we focus on the sender's expected payoff before observing t, that is, their ex ante expected payoff.

Proposition 1 extends to this alternative setting. The argument has three steps. First, regardless of the time at which the sender observes their type, they can secure their full disclosure payoff by committing to an initial experiment that fully reveals θ, after which the receiver's skepticism is no longer relevant. Second, in equilibrium, the sender's strategy will induce a distribution over the receiver's beliefs with expectation π_0. Therefore, their ex ante expected payoff in any equilibrium cannot exceed $\overline{V}(\pi_0)$. Third, when the type distribution assigns probability 1 to t, the sender can secure any payoff that is lower but is arbitrarily close to $\overline{V}(\pi_0)$ by (i) committing to an initial experiment σ^ε that is close to the optimal experiment σ^* in which the receiver has strict incentives at every belief induced by σ^ε and (ii) conducting t additional experiments, all of them are uninformative, and disclose all the outcomes.

Proposition 5 also extends, since when p assigns positive probability to infinitely many types, there is still an equilibrium in which (i) all types of the sender choose an initial experiment that fully reveals the state and (ii) the receiver's posterior belief at every information set assigns probability 1 to the worst state for the sender among the states that occur with positive probability under the receiver's naive belief.

Proposition 2 and Theorem 2 extend since there still exists an equilibrium where all types choose the same initial experiment. Given that each type of the sender in our baseline model obtains a payoff that is weakly greater than their full disclosure payoff, their incentive to conduct the same initial experiment can be sustained by the receiver's belief that as long as the sender has deviated to a different initial experiment, the receiver will entertain skeptical beliefs at every subsequent information set.

Propositions 3 and 4 are based on arguments that bound the difference between a higher-type sender and a lower-type sender's equilibrium payoffs, by considering a higher-type sender's payoff from a deviation that initially uses the lower type's equilibrium strategy, after which they conduct an additional experiment and selectively disclose the resulting outcome. Our argument does not rely on the assumption that different types of the sender need to conduct the same initial experiment, which implies that our conclusions extend to the case where the sender observes their type before choosing their initial experiment.

The Sender Faces a Cost of Running Additional Experiments: Consider another extension of the baseline model in which the sender observes t, the maximal number of additional experiments they can conduct, after choosing the initial experiment σ_0, but they face an additive cost $c > 0$ for conducting each additional experiment. Minor variations of our results extend to the case where c is sufficiently small.

For Propositions 1 and 5, the sender can secure their full disclosure payoff regardless of c since they can do so by choosing a fully informative initial experiment, and when p is unbounded and c is small, they can obtain their full disclosure payoff in equilibrium since choosing a fully informative initial experiment and conducting no additional experiment is an equilibrium as long as the receiver entertains skeptical beliefs. For any degenerate p and $\varepsilon > 0$, there exists $\overline{c} > 0$ such that when $c < \overline{c}$, the sender's payoff is more than $\overline{V}(\pi_0) - \varepsilon$ in all equilibria. This is because when c is small enough, the sender can secure a payoff that is ε-close to $\overline{V}(\pi_0)$ by choosing an initial experiment that is close to σ^*, conducting t uninformative additional experiments, and fully disclosing all the outcomes.

Proposition 2 extends to all c since our constructive proof does not require the sender to conduct additional experiments on the equilibrium path. Propositions 3 and 4 extend as long as the cost of conducting each additional experiment c is small relative to the bounds we derived on the difference between a higher-type and a lower-type sender's equilibrium payoffs, which depend only on the payoff environment (π_0, u^s, u^r) and the gap η between the sender's equilibrium payoff and their commitment payoff. One can also show the following version of Theorem 2: For every $\delta > 0$, there exists $\varepsilon > 0$ such that for every $c < \varepsilon$ and $p \in \Delta(\mathbb{N})$ where $p(n) > 1 - \varepsilon$ for some $n \in \mathbb{N}$, there exists a PEBE in which the sender's payoff is more than $\overline{V}(\pi_0) - \delta$.

5 Conclusion

We study a novel form of limited commitment in Bayesian persuasion models, where the sender cannot commit not to conduct additional experiments and selectively disclose their outcomes. We show that the sender's ability to attain their commitment payoff depends on two key factors: (i) whether the sender can benefit from selectively disclosing additional information at any belief induced by the optimal experiment, and (ii) the precision of the receiver's information about the sender's ability to conduct additional experiments. When the optimal experiment induces at least one belief under which the sender can benefit from revealing more information and the receiver is sufficiently uncertain about the sender's ability to conduct additional experiments, the sender cannot attain anything close to their commitment payoff in any equilibrium. However, when the sender cannot benefit from selective disclosure at any belief induced by the optimal experiment, or the receiver is almost certain how many experiments the sender can conduct, the sender can approximately attain their commitment payoff.

A Proof of Proposition 1

In any equilibrium, the sender's strategy will induce a distribution over the receiver's posterior beliefs about θ with expected value π_0. Hence, the sender's equilibrium payoff is no more than their payoff in an auxiliary game where they can choose any distribution over the receiver's posterior beliefs about θ subject to the constraint that the expectation equals π_0. By definition, this upper bound is $\overline{V}(\pi_0)$. The sender also has the option to deviate by choosing an initial experiment that fully reveals θ and conducting no additional experiment regardless of the realized θ. Under such a deviation, for every $\theta \in \Theta$, the sender obtains payoff $u(\theta)$ when the realized state is θ. Hence, the sender obtains an expected payoff $\underline{V}(\pi_0)$ from this deviation, which implies that their equilibrium payoff must be at least $\underline{V}(\pi_0)$.

In order to show that the sender's payoff is $\overline{V}(\pi_0)$ in all equilibria when p is degenerate, let

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

which are the sender's expected payoffs when the receiver's information structure is σ, who breaks ties in favor of and against the sender, respectively. Since Θ and A are finite, the set $\arg\max_{\sigma \in \Sigma} \overline{V}^{\sigma}$ is non-empty. Recall that σ^* is the optimal experiment. By definition, $\overline{V}(\pi_0) = \overline{V}^{\sigma^*}$. When the receiver's payoff function satisfies the second part of Assumption 1, the results in Lipnowski, Ravid and Shishkin (2024) on finite environments imply that for every $\varepsilon > 0$, there exists an experiment σ^{ε} such that $\underline{V}^{\sigma^{\varepsilon}} > \overline{V}^{\sigma^*} - \varepsilon$.

Suppose p is degenerate and assigns probability 1 to type t^*. Fix any equilibrium and $\varepsilon > 0$. Consider the sender's payoff under the following deviation: Conduct an initial experiment σ^{ε} that satisfies $\underline{V}^{\sigma^{\varepsilon}} > \overline{V}^{\sigma^*} - \varepsilon = \overline{V}(\pi_0) - \varepsilon$. The sender then conducts t^* additional experiments, all of which are uninformative, regardless of the realized outcome of the initial experiment. They then disclose the outcomes of all these additional experiments to the receiver. After observing t^* additional outcomes, the receiver's posterior belief assigns probability 1 to the sender disclosing all the outcomes, which implies that their posterior belief about θ must coincide with their naive belief. This implies that the sender's payoff is at least $\overline{V}(\pi_0) - \varepsilon$ under such a deviation. Since $\varepsilon > 0$ can be arbitrary, the sender's payoff in every equilibrium is at least $\overline{V}(\pi_0)$.

B Proof of Proposition 3

Since every degenerate belief is credible and σ^* is assumed to be the unique optimal experiment, the hypothesis that σ^* inducing a belief that is non-credible implies that $\overline{V}(\pi_0) > \underline{V}(\pi_0)$. 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 $u(\theta) > \overline{u}(\underline{\pi})$. Let

$B(\pi, d)$ denote 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 1. *Suppose r is small enough 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 $\mathbb{E}_{\pi\sim\sigma}\overline{u}(\pi) > \overline{V}(\pi_0) - \eta$, then $\sigma^\theta[B(\underline{\pi}, r)] \geq \varepsilon$.*

Proof of Lemma 1. 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), $\sup_{\sigma\in\mathcal{C}} \mathbb{E}_{\pi\sim\sigma}(\overline{u}(\pi)) < \overline{V}(\pi_0)$. Let $\eta \equiv \min\{\overline{V}(\pi_0) - \sup_{\sigma\in\mathcal{C}} \mathbb{E}_\sigma(\overline{u}), m/2\} > 0$ and $\varepsilon \equiv \frac{\underline{\pi}(\theta)}{2\pi_0(\theta)}\eta > 0$. It follows that for any $\sigma \in \Sigma[\pi_0]$ with $\mathbb{E}_{\pi\sim\sigma}(\overline{u}(\pi)) > \overline{V}(\pi_0) - \eta$, $\sigma \notin \mathcal{C}$ and hence $\sigma[B(\underline{\pi}, r)] > m/2 \geq \eta$. By definition,

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

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

Let $\overline{v}$ and $\underline{v}$ denote the highest and lowest feasible payoffs for the sender, respectively. Since $\overline{u}(\pi)$ is an upper semi-continuous function with respect to π, there exists $r > 0$ such that for every $\pi \in B(\underline{\pi}, r)$, we have $u(\theta) > \overline{u}(\pi)$ and $\pi(\theta) > \underline{\pi}(\theta)/2 \geq 0$. By Lemma 1, there exists $\eta, \varepsilon > 0$ such that for every $\sigma \in \Sigma[\pi_0]$, if $\mathbb{E}_{\pi\sim\sigma}\overline{u}(\pi) > \overline{V}(\pi_0) - \eta$, then $\sigma^\theta[B(\underline{\pi}, r)] \geq \varepsilon$. Fix this selection of r, η, ε and choose $N \in \mathbb{N}$ with

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

Endow the set of type distributions $\Delta(\mathbb{N})$ 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 $\overline{V}(\pi_0) - \eta$.

Suppose by way of contradiction that there exists an equilibrium in which the sender's payoff is strictly greater than $\overline{V}(\pi_0) - \eta$. Let $\widetilde{\sigma}$ denote the equilibrium distribution of the receiver's posterior beliefs.

For any $t \in \{1, 2, ..., N\}$, pick any pure strategy that is used by type t with positive probability in equilibrium and let $\mathcal{H}_t^*$ denote the set of information sets such that for every $h \in \mathcal{H}_t^*$, under that pure strategy used by type t, (i) h will be reached with positive probability, and that type t will conduct no additional experiment afterwards and (ii) the equilibrium probabilities with which they disclose the outcomes of the additional experiments at h will induce a receiver posterior belief that belongs to $B(\underline{\pi}, r)$.

For every $t \geq 0$, consider type $t+1$'s payoff when they deviate to the following strategy: Use the same strategy as type t except for histories that belong to $\mathcal{H}_t^*$. At every $h \in \mathcal{H}_t^*$, 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+1} denote the equilibrium payoff of type $t+1$ and $\widetilde{\sigma}_{t+1}^{\theta}[\cdot]$ denote the induced distribution of receiver posterior beliefs given type $t+1$'s equilibrium strategy and conditional on state θ. Type $t+1$'s payoff from the above deviation is at least

$$V_t + \widetilde{\sigma}_t^{\theta}[B(\underline{\pi}, r)]\pi_0(\theta)(u(\theta) - \overline{u}(\underline{\pi})),$$

since they will induce the same receiver-action as one of the pure strategies type t uses with positive probability in equilibrium, except that when the state is θ and type t reaches information sets that belong to $\mathcal{H}_t^*$, type $t+1$'s deviation will induce the receiver to take their optimal action for state θ. Hence, for every $t \geq 0$, we have

$$V_{t+1} - V_t \geq \widetilde{\sigma}_t^{\theta}[B(\underline{\pi}, r)]\pi_0(\theta)\Big(u(\theta) - \overline{u}(\underline{\pi})\Big).$$

It follows that

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

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

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

By Lemma 1, 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,

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

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

C Proof of Proposition 4

Since $u(\theta) > \max_{\pi \in \Pi^*_\theta} \overline{u}(\pi)$ and the best reply correspondence is upper-hemi-continuous, there exists $\varepsilon > 0$ such that for every π' such that the Hausdorff distance between π' and Π^*_θ is less than ε, we have $\overline{u}(\pi') \leq \max_{\pi \in \Pi^*_\theta} \overline{u}(\pi)$. Since the optimal experiment is unique and the sender's payoffs are bounded, we know that there exists $\lambda_0 > 0$ such that for every $\eta > 0$ and every equilibrium where the sender's payoff is more than $\overline{V}(\pi_0) - \eta$, we have that conditional on every $\theta \in \Theta$, with probability at least $1 - \lambda_0 \eta$, the receiver will reach information sets where their posterior beliefs will have Hausdorff distance less than η to Π^*_θ. Let $\overline{v}$ and $\underline{v}$ denote the sender's highest and lowest feasible payoffs, respectively.

Let $\lambda \equiv 2\lambda_0$ and

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

and pick any η such that

$$\eta < \min\{\varepsilon, \frac{1}{\lambda n}\}.$$

Suppose by way of contradiction that p assigns probability more than $\lambda \cdot \eta$ to at least n types yet there exists an equilibrium where the sender's payoff is more than $\overline{V}(\pi_0) - \eta$. By construction, the receiver's posterior belief will have Hausdorff distance less than η to set Π^*_θ with probability more than $1 - \lambda\eta/2$ conditional on θ. Let $t_1 < ... < t_n$ denote n types that occur with probability at least $\lambda\eta$. Let $\mathbb{E}_t$ and $\Pr_t$ denote the expectation and probability induced by the equilibrium strategy of type t. Since $\mathbb{E}_t\left[\Pr(D \geq \eta \mid \theta, t)\right] < \lambda\eta/2$, Markov's inequality gives $\Pr_t\left(\Pr(D \geq \eta \mid \theta, t) \geq 1/2\right) < \lambda\eta$. 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 $\Pr(D < \eta \mid \theta, t_i) > 1/2$. For every $i \in \{1, 2, ..., n\}$, pick any pure strategy that is used with positive probability in equilibrium by type t_i and let $\mathcal{H}^*_i$ denote the set of sender histories such that for every $h \in \mathcal{H}^*_i$, under that pure strategy used by type t_i, (i) they will reach h with positive probability after which they will conduct no additional experiment and (ii) the equilibrium probabilities with which they disclose the outcomes of the additional experiments at h will lead to a receiver posterior belief that has Hausdorff distance less than η to Π^*_θ.

For every $i \geq 2$, consider type t_i's payoff when they deviate to the following strategy: Use the same strategy as type t_{i-1} except for histories that belong to $\mathcal{H}^*_i$. At every $h \in \mathcal{H}^*_i$, 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

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

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

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

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

D The Requirement in Proposition 4 is not Redundant

We use an example to show that the requirement in Proposition 4 cannot be weakened to that in Proposition 3. Suppose $A \equiv \{0, 2, 3\}$, $\Theta \equiv \{\theta_0, \theta_1\}$, and $t \in \{0, 1\}$. The receiver's prior belief assigns probability π_0 to state θ_1. The sender's payoff equals a. The receiver's optimal action is 0 if $\pi \in [0, 1/3]$, is 2 if $\pi \in [1/3, 2/3]$, and is 3 if $\pi \in [2/3, 1]$, and the prior state distribution is $\pi_0 = 1/2$, so the optimal experiment induces beliefs $1/3$ and $2/3$. Belief $1/3$ is not credible, but the sender's payoff under state θ_1 is the same as their payoff when they induce belief $2/3$. Suppose the prior type distribution $p \in \Delta(\mathbb{N})$ is given by $p(0) = 1/3$ and $p(t) = \frac{1}{3}\frac{1}{2^{t-1}}$ for all $t \in \mathbb{N} \setminus \{0\}$.

We will show that this example has an equilibrium where (i) the receiver's posterior belief is $1/3$ at every information set where no additional outcome is disclosed and is $2/3$ at every on-path information set where one additional outcome is disclosed, (ii) the sender conducts an uninformative initial experiment, (iii) if the sender's realized type t is at least 1, the sender will conduct t additional experiments, all of which have two signal realizations $\{\underline{s}, \overline{s}\}$ with $\sigma(\underline{s}|\theta_0) = 2/3$ and $\sigma(\overline{s}|\theta_1) = 1$, and will only disclose the first realized outcome that is $\overline{s}$, and (iv) the receiver's posterior belief is 0 at every off-path information set with a non-degenerate naive belief.

To see that such an equilibrium exists, we first verify that the receiver's posterior beliefs on the equilibrium path are consistent with Bayes rule. After observing an additional outcome $(\sigma, \overline{s})$, the receiver's posterior belief is

$$\frac{\pi_0(1 - p(0))}{\sum_{t=1}^{\infty} p(t)(1 - (1 - \pi_0)(\frac{2}{3})^t)} = \frac{\frac{1}{3}}{\frac{1}{3}\sum_{t=1}^{\infty} \frac{1}{2^{t-1}}(1 - \frac{1}{2}(\frac{2}{3})^t)} = \frac{2}{3}.$$

After observing no additional outcome, the receiver's posterior belief is

$$\frac{\pi_0 p(0)}{p(0)+\sum_{t=1}^{\infty} p(t)(1-\pi_0)(\frac{2}{3})^t} = \frac{\frac{1}{6}}{\frac{1}{3}+\frac{1}{3}\sum_{t=1}^{\infty}\frac{1}{2^t}(\frac{2}{3})^t} = \frac{1}{3}.$$

Next, we check that no sender type has an incentive to deviate. For type t, the on-path payoff can be calculated recursively: first conduct the experiment σ, reveal the outcome $(\sigma, \overline{s})$ if $\overline{s}$ realizes; otherwise, continue experimenting until a signal $\overline{s}$ realizes in the remaining $t-1$ experiments. $\overline{s}$ realizes with probability $2/3$, leading to payoff $\overline{u}(2/3)$, and otherwise the sender gets a continuation payoff $V_{t-1} = (1-(2/3)^{t-1})\overline{u}(2/3) + (2/3)^{t-1}\overline{u}(1/3)$, which is the expected payoff from repeating σ for $t-1$ times given $\theta = \theta_0$. Hence, type t's on-path payoff is $\frac{2}{3}\overline{u}(2/3) + \frac{1}{3}V_{t-1}$. The most profitable deviation for type t is to first conduct a fully informative experiment, reveal the outcome if $\theta = \theta_1$, and otherwise repeat the experiment σ for $t-1$ times and only disclose the first realized outcome that is $\overline{s}$. We may also calculate the payoff recursively, which is $\frac{1}{2}\overline{u}(1) + \frac{1}{2}V_{t-1}$. The on-path payoff dominates the best deviation as $V_{t-1} < 3 = \overline{u}(2/3)$.

E Proof of Theorem 2

When the optimal experiment σ^* fully reveals θ, we have $\overline{V}(\pi_0) = \underline{V}(\pi_0)$, in which case Proposition 1 will imply that there exists an equilibrium that attains payoff $\overline{V}(\pi_0)$. In what follows, we focus on the case where σ^* does not fully reveal θ. Recall that $\Pi^* \subseteq \Delta(\Theta)$ denotes the set of posterior beliefs induced by σ^*. By Lemma 1 of Lipnowski, Ravid and Shishkin (2025) and the uniqueness of σ^* (see Assumption 2), we have (i) $|\Pi^*| \leq |\Theta|$, (ii) the beliefs in Π^* are affinely independent, and (iii) every belief in Π^* occurs with strictly positive probability under experiment σ^*.

For any $\pi \in \Delta(\Theta)$ and $\varepsilon > 0$, let $a^*(\pi)$ denote the receiver's best reply under belief π (if there are multiple best replies, then pick the one that maximizes the sender's payoff) and let $B(\pi, \varepsilon)$ denote the set of beliefs that has distance less than ε to π. If the triple (u^s, u^r, π_0) satisfies Assumptions 1 and 2, we have the following lemma:

Lemma 2. *For any $\varepsilon > 0$ and optimal experiment σ^* that satisfies our regularity conditions with support Π^*, there exists 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 $\Pi^*_\varepsilon \equiv \{\pi_\varepsilon(\pi)\}_{\pi \in \Pi^*}$.*

By construction, for every $\delta > 0$, there exists $\varepsilon > 0$ such that the sender's expected payoff from committing to any such experiment σ^*_ε is at least $\overline{V}(\pi_0) - \delta$.

Suppose the receiver's prior belief about the sender's type $p \in \Delta(\mathbb{N})$ assigns probability close to 1 to $t = n \in \mathbb{N}$. Consider an equilibrium in which the sender conducts initial experiment σ^*_ε. The receiver's naive belief after observing the outcome of this initial experiment is called their *interim belief*. Following any interim belief, if the sender reveals any additional signal that is informative and induces a non-degenerate naive belief π, then the receiver's posterior belief assigns probability 1 to some $\theta \in \text{supp}(\pi)$ that satisfies

$$u(\theta) \leq u(\theta') \text{ for every } \theta' \in \text{supp}(\pi).$$

Similarly, if the sender chooses any initial experiment other than σ^*_ε, then at every interim belief π, the receiver's posterior belief will assign probability 1 to some $\theta \in \text{supp}(\pi)$ that satisfies

$$u(\theta) \leq u(\theta') \text{ for every } \theta' \in \text{supp}(\pi).$$

Hence, it is without loss to focus on subgames where the sender does not deviate to other initial experiments.

Recall the definition of credible beliefs. Following any credible interim belief π that belongs to Π^*_ε, all types of the sender conduct no additional experiment in equilibrium. Doing so is optimal for all types of the sender given the receiver's off-path belief and the facts that (i) $\pi \in \Pi^*_\varepsilon$ and (ii) σ^* does not fully reveal θ, which implies that it is also suboptimal at any interim belief in Π^* as long as the receiver breaks ties in favor of the sender, and hence, is also suboptimal at any posterior belief in Π^*_ε since by construction, the receiver has strict incentives at those posterior beliefs. The rest of our proof focuses on subgames following interim posterior beliefs in Π^*_ε that are non-credible. Interim posterior beliefs outside of Π^*_ε occurs with probability less than ε under σ^*_ε, which will have negligible impact on the sender's ex ante expected payoff.

Consider the subgame following interim belief $\pi \in \Pi^*_\varepsilon$ that is non-credible. Let $\text{supp}(\pi) \equiv \{\theta_1, ..., \theta_k\}$, $\pi_k \equiv \pi(\theta_k)$, and $u_k \equiv u(\theta_k)$. 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:

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

Intuitively, u^*_i is the sender's payoff from no disclosure so that he is 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 3 shows that this incentive constraint is equivalent to u_i^* reaching a local minimum at i.

Lemma 3. *$u_i \geq u_i^* \geq u_{i+1}$ if and only if $\min\{u_{i+1}^*, u_{i-1}^*\} \geq u_i^*$.*

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

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

Equation (2) is equivalent to

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

which can be rewritten as

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

This implies that $u_{i+1}^* \geq u_i^*$ if and only if $u_i^* \geq u_{i+1}$. Another equivalent way to write (2) is

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

from which we obtain that

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

This implies that $u_{i+1} \geq u_{i+1}^*$ if and only if $u_i^* \geq u_{i+1}^*$. An analogous argument shows $u_{i-1}^* \geq u_i^* \iff u_i \geq u_{i-1}^*$. □

Lemma 4. *There exists $i \in \{1, 2, ..., k\}$ such that $u_i \geq u_i^* \geq u_{i+1}$.*

Proof. Consider two cases. First, if $u_1^* \geq u_2^*$, then the fact that $u_k^* = +\infty$ implies that there exists i such that $\min\{u_{i+1}^*, u_{i-1}^*\} \geq u_i^*$. Lemma 3 implies that $u_i \geq u_i^* \geq u_{i+1}$ for such i. Second, if $u_2^* > u_1^*$, then inequality (3) 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 of at least 1 and strategy s^{nd} is feasible only for types of at least 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 $u^s(\alpha) = u$. 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 4. 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 t σ^* is the optimal experiment, we know that for every $a \in \arg\max_{a' \in A} u^r(\pi_*^i, a')$,

$$u^* > \sum_{j=1}^{i} \pi_j u_j + \Big(1 - \sum_{j=1}^{i} \pi_j\Big) u^s(a). \tag{5}$$

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. 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

$$u^* \leq \sum_{j=1}^{i} \pi_j u_j + \Big(1 - \sum_{j=1}^{i} \pi_j\Big) u^s(a). \tag{6}$$

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 (1) implies that $u^s(a) \geq u_i^*$. In equilibrium following interim posterior belief π, sender types strictly greater than n 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 $\theta = \theta_j$ that satisfies $u_j > u^*$. Sender types 1 to $n-1$ use strategy s^i. Sender type n mixes between strategy s^{nd} and strategy s^i. The probability with which type n uses strategy s^i is 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 $u^s(\alpha) = u_i^*$, which makes the sender of type n indifferent between strategy s^i and strategy s^{nd}. Such a mixing probability exists since type n occurs with probability close to 1, there exists

$a \in \arg\max_{a' \in A} u^r(\pi^i_*, a')$ that satisfies (5) under belief π^i_*, and that Graph $\mathbb{G}$ is closed and connected.

When the probability of type n is close enough to 1, type n will use strategy s^{nd} with probability close to 1. This implies that after observing the additional outcomes from n uninformative experiments and nothing else, the receiver's posterior belief will be close to π. The fact that $\pi \in \Pi^*_\varepsilon$ implies that $a^*(\pi)$ remains strictly optimal for the receiver at such a posterior belief.

Case B: Suppose next that $u^* > \sum_{j=1}^{i} \pi_j u_j + \left(1 - \sum_{j=1}^{i} \pi_j\right) u^s(a)$ for every $a \in \arg\max_{a' \in A} u^r(\pi^i, a')$. Intuitively, this is the case where the probability of type 0 is low relative to the probability of types 1 to $n-1$. By definition, $u^s(a) < u^*_i$. Since the construction of i requires that $u^*_i \in [u_{i+1}, u_i]$, we know that $u^s(a) < u_i$. On the equilibrium path following interim belief π, sender types strictly greater than n will 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 is $\theta = \theta_j$ that satisfies $u_j > u^*$. Sender types 1 to $n-1$ mix between strategy s^j and strategy s^{j+1} for some $j \geq i$. Sender type n uses strategy s^{nd} for sure.

We show that there exist $j \geq i$ as well as a mixing probability for types 1 to $n-1$ between strategy s^j and strategy s^{j+1} such that under the receiver's posterior belief after observing no additional outcome being disclosed, there exists $\alpha \in \Delta(A)$ that is optimal for the receiver such that if the receiver takes action α when there is no disclosure, then the following two incentive constraints are satisfied. First,

- When both strategy s^j and s^{j+1} are played with positive probability, the sender's payoffs from Strategies s^j and s^{j+1} are weakly higher than that from strategy s^τ for every $\tau \in \{1, 2, ..., k\}$.
- When only strategy s^j is played with positive probability, the sender's payoff from strategy s^j is weakly higher than that from strategy s^τ for every $\tau \in \{1, 2, ..., k\}$.

This ensures that sender types 1 to $n-1$ have no incentive to deviate to alternative strategies. Second, the sender's expected payoffs from Strategies s^j and s^{j+1} are no more than u^*. This ensures that sender type n has no incentive to imitate types 1 to $n-1$.

A *disclosure rule* is a mapping from $\{\theta_1, ..., \theta_k\}$ to the probabilities with which each state is disclosed. A *monotone disclosure rule* is one in which for every θ_j that is disclosed with positive probability, every state θ_τ with $\tau < j$ is disclosed with probability 1. By definition, each monotone disclosure rule can be pinned down by its probability of disclosure, denoted by q.

Let $V(q) \subseteq \mathbb{R}$ denote the set of sender's expected payoff from disclosing no additional signal when the receiver believes that sender types 1 to $n-1$ use monotone disclosure rule with disclosure probability q.

By definition, $u^* \in V(1)$ since only type 0 discloses no additional signal when types 1 to $n-1$'s disclosure probability is 1. Let $\mathbb{V} \subseteq [0,1] \times \mathbb{R}$ be such that $(q,v) \in \mathbb{V}$ if and only if $v \in V(q)$. Let $\mathbb{U} \subseteq [0,1] \times \mathbb{R}$ be such that (i) for every q that can be written as $q = \sum_{j=1}^{\tau} \pi_j$ for some integer τ, we have $(q,v) \in \mathbb{U}$ if and only if $v \in [u_{\tau+1}, u_\tau]$ and (ii) for every q that satisfies $\sum_{j=1}^{\tau} \pi_j < q < \sum_{j=1}^{\tau+1} \pi_j$, we have $(q,v) \in \mathbb{U}$ if and only if $v = u_{\tau+1}$. Both $\mathbb{V}$ and $\mathbb{U}$ are continuous, and moreover, graph $\mathbb{U}$ is decreasing in q with $(1, u_k) \in \mathbb{U}$. The hypothesis for this case requires that when $q = q^* \equiv \sum_{j=1}^{i} \pi_j$, every payoff in $V(q^*)$ is less than u_i. Since $u_k < u^*$, there exists an intersection (q,v) between $\mathbb{V}$ and $\mathbb{U}$ with $q \geq q^*$. By definition, any monotone disclosure rule with disclosure probability q and the receiver's action that leads to non-disclosure payoff v will satisfy the first incentive constraint.

In the last step, we verify that the intersection (q,v) also satisfies the second incentive constraint. If the intersection is at $q = q^*$, then the second incentive constraint is trivially satisfied since $u_i^* \in [u_{i+1}, u_i]$ and $u^* > \sum_{j=1}^{i} \pi_j u_j + \left(1 - \sum_{j=1}^{i} \pi_j\right) u^s(a)$ for every $a \in \arg\max_{a' \in A} u^r(\pi^i, a')$. In what follows, we focus on the case where the intersection is at $q > q^*$. For every $\tau \in \{1, 2, ..., k\}$ and $\beta \in [0,1]$, let $u^*(\tau, \beta)$ be defined as

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

or equivalently

$$u^* = \sum_{j=1}^{\tau} \pi_j u_j + \beta \pi_{\tau+1} u_{\tau+1} + \left(1 - \sum_{j=1}^{\tau} \pi_j - \beta \pi_{\tau+1}\right) u^*(\tau, \beta). \tag{7}$$

Intuitively, $u^*(\tau, \beta)$ is the sender's payoff from disclosing no additional signal that makes them indifferent between strategy s^{nd} (from which they receive payoff u^*) and randomizing between strategies s^τ and $s^{\tau+1}$ with probabilities $1-\beta$ and β, respectively. By definition, $u^*(\tau, 0) = u_\tau^*$ as defined in (1).

Recall that i is chosen such that $u_i^* \equiv u^*(i, 0) \geq u_{i+1}$. Equation (7) together with the fact that $u_{\tau+1} < u^*(\tau, \beta)$ before the intersection q implies that for every $\tau \geq i$ and $\beta \in [0,1]$ with $\sum_{j=1}^{i} \pi_j + \beta \pi_{i+1} \leq q$, $u^*(\tau, \beta)$ is increasing in both τ and β. For monotone disclosure rules, it implies that the function $u^*(\tau, \beta)$, which captures the sender's no-disclosure payoff that makes them indifferent, is strictly increasing in the disclosure probability when it is between $q^* \equiv \sum_{j=1}^{i} \pi_j$ and the intersection q. Since $\mathbb{U}$ is decreasing in q and $u_i^* \in [u_{i+1}, u_i]$ and $v \leq u_{i+1}$ for every $(q,v) \in \mathbb{U}$ with $q > q^*$, we know that for every $(q,v) \in \mathbb{U}$ with $q \equiv \sum_{j=1}^{\tau} \pi_\tau + \beta \pi_{\tau+1} > q^*$, we have $v \leq u^*(\tau, \beta)$. This implies that any intersection between $\mathbb{U}$ and $\mathbb{V}$ where $q > q^*$ will satisfy the second incentive constraint.

References

[1] Ali, S. Nageeb, Andreas Kleiner and Kun Zhang (2024) "From Design to Disclosure," Working Paper.

[2] DeVito, Nicholas, Seb Bacon, and Ben Goldacre (2020) "Compliance with Legal Requirement to Report Clinical Trial Results on ClinicalTrials.gov: A Cohort Study." *Lancet* 395.10221, 361-369.

[3] Best, James and Daniel Quigley (2024) "Persuasion in the Long Run," *Journal of Political Economy*, 132(5), 1740-1791.

[4] Dye, Ronald (1985) "Disclosure of Nonproprietary Information," *Journal of Accounting Research*, 23(1), 123-145.

[5] Dzuida, Wioletta (2011) "Strategic Argumentation," *Journal of Economic Theory*, 146, 1362-1397.

[6] Felgenhauer, Mike and Elisabeth Schulte (2014) "Strategic Private Experimentation," *American Economic Journal: Microeconomics*, 6(4), 74-105.

[7] Felgenhauer, Mike and Petra Loerke (2017) "Bayesian Persuasion with Private Experimentation," *International Economic Review*, 58(3), 829-856.

[8] Fudenberg, Drew and Jean Tirole (1991) "Perfect Bayesian Equilibrium and Sequential Equilibrium," *Journal of Economic Theory*, 53(2), 236-260.

[9] Gao, Ying (2025) "Inference from Selectively Disclosed Data," Working Paper.

[10] Henry, Emeric and Marco Ottaviani (2019) "Research and the Approval Process: The Organization of Persuasion," *American Economic Review*, 109(3), 911-955.

[11] Kamenica, Emir and Matthew Gentzkow (2011) "Bayesian Persuasion," *American Economic Review*, 101(6), 2590-2615.

[12] Kamenica, Emir and Xiao Lin (2025) "Commitment and Randomization in Communication," Working Paper.

[13] Koessler, Frederic and Vasiliki Skreta (2023) "Informed Information Design," *Journal of Political Economy*, 131(11), 3186-3232.

[14] Kreps, David and Robert Wilson (1982) "Sequential Equilibria," *Econometrica*, 50(4), 863-894.

[15] Ivanov, Maxim (2010) "Informational Control and Organizational Design," *Journal of Economic Theory*, 145(2), 721-751.

[16] Lin, Xiao and Ce Liu (2024) "Credible Persuasion," *Journal of Political Economy*, 132(7), 2228-2273.

[17] Lipnowski, Elliot and Doron Ravid (2020) "Cheap Talk With Transparent Motives," *Econometrica*, 88(4), 1631-1660.

[18] Lipnowski, Elliot, Doron Ravid and Denis Shishkin (2022) "Persuasion via Weak Institutions," *Journal of Political Economy*, 130(10), 2705-2730.

[19] Lipnowski, Elliot, Doron Ravid and Denis Shishkin (2024) "Perfect Bayesian Persuasion," *JPE: Microeconomics*, forthcoming.

[20] Mathevet, Laurent, David Pearce and Ennio Stacchetti (2024) “Reputation and Information Design,” Working Paper.

[21] Pei, Harry (2023) “Repeated Communication with Private Lying Costs,” *Journal of Economic Theory*, 210, 105668.

[22] Pei, Harry (2025) “Community Enforcement with Endogenous Records,” *Review of Economic Studies*, forthcoming.

[23] Perez-Richet, Eduardo (2014) “Interim Bayesian Persuasion: First Steps,” *American Economic Review: Papers & Proceedings*, 104(5), 469-474.

[24] Titova, Maria and Kun Zhang (2025) “Persuasion with Verifiable Information,” *Journal of Economic Theory*, forthcoming.

[25] Zapechelnyuk, Andriy (2023) “On the equivalence of information design by uninformed and informed principals,” *Economic Theory*, 76, 1051-1067.