ISSN (Print) 0473-453X
Discussion Paper No. 1247
ISSN (Online) 2435-0982

OPTIMAL FEEDBACK DYNAMICS AGAINST FREE-RIDING IN COLLECTIVE EXPERIMENTATION

Chia-Hui Chen
Hülya Eraslan
Junichiro Ishida
Takuro Yamashita

Revised September 2025
July 2024

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

Optimal Feedback Dynamics Against Free-Riding in Collective Experimentation*

Chia-Hui Chen† Hülya Eraslan ‡ Junichiro Ishida§
Takuro Yamashita¶

September 4, 2025

Abstract

We consider a dynamic model in which a principal decides what information to release about a product of unknown quality (e.g., a vaccine) to incentivize agents to experiment with the product. Assuming forward-looking agents, their incentive to wait and see others' experiences poses a significant obstacle to social learning, implying suboptimality of full transparency. We show that the optimal feedback mechanism to mitigate information free-riding takes a strikingly simple form: the principal makes a binary recommendation and recommends against adoption with some probability even when she is relatively optimistic; once she recommends against adoption, she never reverses her stance.

JEL Codes: D82, D83

Keywords: strategic experimentation, information design, free-rider problem, social learning, COVID-19 pandemic.

*We thank Nejat Anbarcı, Jimmy Chan, Yi-Chun Chen, Kim-Sau Chung, Daisuke Hirata, Fahad Khalil, Jan Knoepfle, Gilat Levy, Meg Meyer, Masaki Nakabayashi, Marek Pycia, Bruno Strulovici, Leeat Yariv, Yosuke Yasuda and participants at Workshop on Contracts, Incentives and Information (3rd Edition) at Turin, International Workshop on Theory and Experiments, 2024 AMES, 2024 LACEA-LAMES, Workshop on Mechanism and Measurement, CTW, OEIO, Hong Kong Baptist, Kobe, CMU-Pitt, Keio, Academia Sinica, National Taiwan, Zurich, Indiana Kelley, Yonsei for helpful feedback. We acknowledge financial support from National Science Foundation (Grant SES-1730636 for Eraslan), JSPS Kakenhi (23K01309 for Chen, 21H04398 and 23K01347 for Ishida, and 23K01311 for Yamashita), and the Joint Usage/Research Center for Behavioral Economics at ISER.

†Institute of Economic Research, Kyoto University; chchen@kier.kyoto-u.ac.jp.

‡Department of Economics, Rice University, NBER and Institute of Social and Economics Research, University of Osaka; eraslan@rice.edu.

§Institute of Social and Economic Research, University of Osaka; jishida@iser.osaka-u.ac.jp.

¶Osaka School of International Public Policy, University of Osaka; yamashita.takuro.osipp@osaka-u.ac.jp.

1 Introduction

The COVID-19 pandemic presented an unprecedented challenge, compelling governments around the world to navigate through uncertainties and introduce many new, often untested, measures to cope with it. Consider the uptake of mRNA vaccines as an example. While this radically new technology had been known for years and considered safe among medical scientists, it had never been widely used before the pandemic, and politicians and citizens alike were not entirely sure about its possible adverse consequences. Amid this uncertainty, a potential impediment to widespread adoption is citizens' unwillingness to be among the first to try out the new technology: Hamel, Lopes and Brodie (2021) document that "31% of the public say that when an FDA-approved vaccine for COVID-19 is available to them for free, they will 'wait until it has been available for a while to see how it is working for other people' before getting vaccinated themselves." Those who stand to gain less from the vaccine may choose to wait and see others' experiences, but this kind of wait-and-see attitude slows down the adoption process, resulting in a loss of surplus that can be substantial.

The presence of information free-riding is in fact ubiquitous in many instances of policy experimentation that require the voluntary participation of citizens; examples include technology adoption in a variety of sectors such as healthcare, education, and manufacturing. In close inspection, two aspects of this experimentation process are particularly noteworthy. First, experimentation of this kind is often "feedback-based" in that uncertainty regarding the net value of a new policy is revealed gradually through the feedback of the citizens who have embraced the experimentation. Second, it is also often "large-scale," involving a large number of citizens who interact anonymously, such that they do not directly observe what others have experienced on the whole, e.g., the frequencies of severe side effects, and must instead rely on the information released by central authorities. The first aspect makes incentivizing the citizens to experiment a central problem for the government, whereas the second aspect provides a potentially viable tool to achieve this goal. Along with the fact that contingent monetary transfers are often not feasible in this type of environment, these two aspects give rise to a new class of information-design problems in which the government determines what information to disseminate to the public, with the aim of overcoming information free-riding and facilitating social learning.

To address this issue, we consider a collective experimentation environment that consists of a principal (e.g., a government) and a continuum of long-lived and forward-looking agents (e.g., citizens). The principal has a product to be consumed (or a technology to be adopted) by the agents who are heterogeneous with respect to their valuations of the product. The quality

of the product is initially unknown to anyone and gradually revealed via the arrival of news whose arrival rate is proportional to the measure of agents who consume the product. However, reflecting the large-scale nature of the experimentation, news is observable only to the principal, who therefore faces a problem of what information to disseminate to the agents. Formally, the principal designs and commits to a feedback mechanism that specifies the distribution of the sequence of beliefs subject to Bayes plausibility.[1] Given some feedback mechanism, each agent independently decides when, if ever, to adopt the product based on his expectation of how the belief evolves as dictated by the mechanism in place.

We begin our analysis by observing that in the environment described above, the amount of adoption under full disclosure is generally insufficient compared to its first-best level. This observation stems from the free-rider problem that has been well documented in the strategic experimentation literature (Bolton and Harris, 1999; Keller, Rady and Cripps, 2005): because information is a public good whose benefits are not fully internalized, each agent has an excessive incentive to wait and see the experimentation outcomes of other agents. However, in large-scale experimentation where agents have no direct access to the outcomes, the extent of the free-rider problem depends crucially on what information the principal disseminates to the agents. This point can be seen most clearly if we consider an extreme measure of no disclosure: if the principal commits to revealing no information, there is no point in waiting for news, and the free-rider problem dissipates completely. Of course, revealing no information is never optimal because that would imply entirely giving up the gains from additional information gleaned from the experimentation. The principal thus needs to strike the right balance between exploiting the gains from the experimentation on one hand and managing the free-rider incentive on the other.

The principal's problem is potentially complicated because the set of mechanisms is large in our dynamic context, with a myriad of choices available for the principal. Specifically, the principal can choose not only what information to disclose in a given period but also when to do so: for instance, she may withhold favorable information by garbling signals in some period and deliver it later with a delay. Despite this potential complication, we show that the optimal feedback mechanism to mitigate information free-riding takes a strikingly simple form: (i) when the principal is "pessimistic," she always recommends against adoption; (ii) when she is "optimistic," she recommends adoption by fully disclosing all the information with some probability and recommends against adoption with the remaining probability; (iii) once

[1] In the baseline model, we consider a benevolent principal who maximizes the sum of the payoffs of all agents. We later extend the analysis by incorporating a more general objective function.

she recommends against adoption, she will never reverse her stance (permanent termination).[2] Owing to this result, the principal's problem effectively reduces to an optimal-stopping problem, and the resulting optimal mechanism can be characterized thoroughly by a sequence of the continuation probabilities conditional on having observed no news.

This characterization result is a consequence of two observations, which we label as *termination* and *caution* for clarity. The termination property suggests that if the principal suspends the adoption process for a period, she will never resume the process again, even if she has received no bad news. This property alternatively implies that there is no strategic gain from withholding information to slow down adoption, and the optimal mechanism entails no delayed information release, which enables us to substantially reduce the set of mechanisms we need to consider. The caution property suggests that the optimal mechanism should exercise caution by occasionally terminating the adoption process even when she is relatively optimistic. This essentially means that the principal garbles when issuing a warning, saying "I may have observed bad news," instead of clearly stating "I have definitely observed bad news." This latter result pertains to the fact that the value of information varies across agents with different valuations: information that distinguishes small risk and no risk is crucial for agents with lower valuations but irrelevant for those with higher valuations. The optimal mechanism can mitigate the free-rider problem and achieve the optimal outcome because it provides no useful information for agents with higher valuations, similar to the no-disclosure policy, but still provides enough information to improve the welfare of those with lower valuations. Both of these properties are largely detail-free and robust to various alterations to the underlying setup, as a consequence of which our main characterization result holds under a more general objective function that accommodates a wider range of social goals (Section 5.1) as well as under a more general information-generating process where both good news and bad news arrive (Section 5.2) and where news may arrive with some delay (Section 5.3).

Literature. Our model lies at the intersection of strategic experimentation (Bolton and Harris, 1999) and information design (Kamenica and Gentzkow, 2011). Our learning environment builds on the exponential-bandits model of Keller et al. (2005).[3] A crucial departure from this strand of literature is that we consider large-scale collective experimentation that involves a large number of agents, where each individual agent cannot directly observe the experimen-

[2] In the baseline model, we consider a bad-news ("breakdown") case where news arrives only when the state is bad. In this case, the principal is pessimistic when she has observed bad news and optimistic when she has not.

[3] More precisely, our baseline model follows the bad-news specification of Keller and Rady (2015), although we later extend our analysis to incorporate good news as well.

tation outcomes of other agents and must instead rely on the information released by the principal. This aspect of our model draws a clear contrast to canonical models of strategic experimentation that typically focus on a relatively small group of agents and assume that each agent can directly observe the experimentation outcomes of other agents (Keller et al., 2005; Bonatti and Hörner, 2011). The fact that only the principal can observe the experimentation outcomes gives her some leeway to control the agents' belief formation process, amounting to a new class of information-design problems as noted above.

Our analysis is in spirit most closely related to Kremer, Mansour and Perry (2014), Che and Hörner (2018), Knoepfle and Salmi (2024), and Lyu and Liao (2025) in that they all explore ways to facilitate social learning, with emphasis on agents as both consumers and generators of information.[4] Kremer et al. (2014) consider a social-learning environment à la Bikhchandani, Hirshleifer and Welch (1992) and characterize the optimal disclosure policy to mitigate information herding. Che and Hörner (2018) consider a similar feedback-based information structure to ours in which information arrives at a rate proportional to the amount of adoption and study how a recommender system can improve the incentives for early exploration. In both of these models, however, agents are assumed to be myopic and make once-and-for-all adoption decisions. By contrast, in our setting, agents are forward-looking and strategically choose their timing of adoption.[5] Within this framework, we focus on a different incentive issue—free-riding stemming from the option to wait and see others' experiences—and explore the optimal feedback mechanism to alleviate this problem. Knoepfle and Salmi (2024) consider forward-looking agents as we do but focus their attention on the optimal timing of disclosure (rather than optimal information design). They find that it is optimal to disclose bad news immediately while delaying the disclosure of good news, allowing the adoption process to continue unless bad news arrives.[6] Lyu and Liao (2025) study optimal information design for social learning with forward-looking agents. The key difference is that they consider sequentially arriving agents with homogeneous preferences.

[4]Baccara, Levy and Razin (2024) consider an environment in which there are two risky "fields" to choose from, and each agent can irreversibly join either field or wait for more information. In this framework, they also assume that the rate of information arrival in each field depends on the measure of agents who joined that field.

[5]Frick and Ishii (2024) also study social learning by forward-looking agents but without information design.

[6]By contrast, we find that it is optimal to terminate the adoption process with some probability even if no bad news has arrived. This *caution* property is key to incentivizing agents to contribute to social learning in earlier periods who otherwise may free-ride on others, which cannot be implemented simply by timing the disclosure of bad news. Aside from this aspect, another crucial difference is that they consider agents who are heterogeneous in their discount rates (time preferences), while we consider agents who are heterogeneous in their valuations. As a consequence, the key underlying mechanisms are qualitatively different. See Section 5.3 and footnote 20 for more discussion on this point.

Several works explore how to incentivize experimentation via monetary transfers. To name a few, Manso (2011) analyzes a moral-hazard problem with unknown success probability and shows that the optimal incentive contract exhibits substantial tolerance for early failure. Halac, Kartik and Liu (2016) consider a model of long-term contracting with both moral hazard and adverse selection and obtain a characterization of optimal menu contracts. Halac, Kartik and Liu (2017) consider both monetary rewards and information disclosure (about whether success had occurred or not) in a contest environment where multiple agents compete for a prize. We view our analysis as complementary to this strand of literature, where we focus exclusively on the use of information to motivate agents to take a risky alternative without monetary rewards. Our analysis is more applicable to situations where contingent monetary transfers are either unconventional or infeasible due to institutional factors, as is often the case in the context of policy experimentation.

Finally, we also aim to contribute more broadly to the literature on dynamic information design. There are several recent works that analyze information disclosure in dynamic environments where an agent makes stopping decisions (Au, 2015; Ely, 2017; Nikandrova and Pancs, 2018; Che and Mierendorff, 2019; Ely and Szydlowski, 2020; Orlov, Skrzypacz and Zryumov, 2020).[7] There are two notable differences from this strand of literature. First, while those previous works generally focus on single-agent cases, we consider an environment where the principal attempts to persuade a continuum of agents who are heterogeneous with respect to their valuations.[8] Second, and more importantly, the principal's information structure in our setting is endogenous in that the precision of the information she acquires depends on the agents' adoption decisions. This aspect of our model stands in sharp contrast to the standard setting where the information structure is exogenously given and fixed over time.

2 Illustrative example

We first provide a simple example with two periods and two agent types to highlight the main ideas. Suppose that a government attempts to introduce a new vaccine to a continuum of citizens with unit measure. Each citizen can take the vaccine in either period or can choose to abstain from it altogether. The quality of the vaccine, which is initially unknown to anyone, can be either good ($\omega = 0$) or bad ($\omega = 1$), and the common prior probability that the vaccine

[7] Ball (2023) and Correia da Silva and Yamashita (2024) consider information disclosure in repeated interactions.

[8] Alonso and Camara (2016) and Chan, Gupta, Li and Wang (2019) consider persuasion of a group of heterogeneous agents in static environments. Also, see Kolotilin, Mylovanov, Zapechelnyuk and Li (2017), and Guo and Shmaya (2019) for works on persuasion of privately informed receivers.

is bad is $m \in (0,1)$. The citizens are divided into two preference types, where a half of them have a low valuation ($v = v_\ell$) and the other half have a high valuation ($v = v_h > v_\ell$). If a citizen with valuation v chooses to take the vaccine, his payoff is $v - \omega$; if not, his payoff is normalized to 0. We assume $v_h < 1$ so that even the high-valuation citizens would not take the vaccine if they knew that the vaccine is bad. Each citizen maximizes the sum of payoffs with no time discounting.

The quality of the vaccine is partially revealed via the arrival of news at the end of period 1. Here, we consider a bad-news scenario where news (e.g., the occurrence of severe side effects) arrives only when the state is bad. We make two crucial assumptions regarding the information structure. First, as news reflects the aggregate outcome, it is observable only to the government but not to each citizen. Second, conditional on the state being bad, the likelihood of news arriving depends on and is increasing in the measure of citizens who choose to take the vaccine. Specifically, after a measure n of citizens take the vaccine in period 1, the government observes bad news with probability $1 - e^{-\lambda\omega n}$. Let μ_2^p denote the government's belief that the state is bad in period 2. If bad news is observed, the government learns that the state is bad for sure, and hence $\mu_2^p = 1$. If not, taking n as given, the government's belief is updated to

$$\mu_2^p = \underline{\mu}(n) := \frac{me^{-\lambda n}}{me^{-\lambda n} + 1 - m},$$

which is lower than the initial prior for any $n > 0$, i.e., "no news is good news."

For a given n, the government can induce any distribution of the citizens' period-2 beliefs over $[\underline{\mu}(n), 1]$ subject to Bayes plausibility. The government's problem is to determine this distribution, in order to maximize the sum of the payoffs of all citizens. Let μ_2 denote the citizens' period-2 belief that the state is bad based on the information released by the government. Since period 2 is the last period, the problem faced by the citizens in that period is straightforward: given some belief μ_2, a citizen with valuation v takes the vaccine if $v \geq \mu_2$. Taking this as given, in period 1, a citizen with valuation v takes the vaccine if

$$v - m \geq \mathbb{E}_\mu[\max\{v - \mu, 0\}]. \tag{1}$$

Each citizen's decision in period 1 depends on the continuation payoff, which in turn depends on the distribution of period-2 beliefs. For illustration, assume $1 > v_h > m > v_\ell > \underline{\mu}(0.5)$, so that the low-valuation citizens would never take the vaccine in period 1 but could be persuaded to do so in period 2 if enough information were generated in period 1.

To illustrate the value of information design in this context, we begin with two extreme cases

as benchmarks. We first consider the full-disclosure policy where the government mechanically reveals everything it observes (so that $\mu_2 = \mu_2^p$ with probability 1). In this case, if a high-valuation citizen waits until period 2, he takes the vaccine in period 2 if and only if no news is observed. This event occurs with probability $me^{-\lambda n^*} + 1 - m$, and the expected payoff is $v_h - \underline{\mu}(n^*)$, where n^* is the equilibrium measure of citizens taking the vaccine. Since the expected payoff of taking the vaccine in period 1 is $v_h - m$, the IC constraint for the high-valuation citizens to take the vaccine in period 1 can be written as

$$\begin{aligned} v_h - m &\geq (me^{-\lambda n^*} + 1 - m)(v_h - \underline{\mu}(n^*)) \\ &= (me^{-\lambda n^*} + 1 - m)v_h - me^{-\lambda n^*}, \end{aligned}$$

taking n^* as given.[9] It is easy to verify that the IC constraint does not hold for any $n^* > 0$ for the high-valuation citizens; as such, no one chooses to take the vaccine in period 1 under full disclosure.[10] This is a manifestation of the free-rider problem, and as a consequence, the opportunity for social learning is entirely lost, leading to a suboptimal outcome.

In a sense, the full-disclosure policy fails because it is too informative, giving the high-valuation citizens an excessive incentive to wait for news. This implies that the optimal feedback mechanism must be somewhat more obscure than the full-disclosure policy. To illustrate this point, we now turn to the opposite case of no disclosure where the government commits to revealing no additional information (so that $\mu_2 = m$ with probability 1). In this case, there is clearly no point in waiting for news, and it is (weakly) optimal for the high-valuation citizens to take the vaccine in period 1,[11] resulting in the maximum amount of information that can be generated in this environment. However, this no-disclosure policy is also suboptimal because none of the low-valuation citizens could be induced to take the vaccine in period 2, given $\mu_2 = m$. This argument suggests that the government must find a middle ground between the two extreme policies.

So, what is the optimal way to disclose information in this environment? To achieve the optimal outcome, the government must induce (i) the high-valuation citizens to take the vaccine in period 1 and (ii) the low-valuation citizens to do so in period 2 occasionally (in case no news is observed). We have already observed that the no-disclosure policy achieves (i) but not (ii). Note that the no-disclosure policy is effective for (i) because it provides no valuable information

[9] Throughout the analysis, we assume that an agent chooses to adopt if he is indifferent, except for the case where no agents are induced to adopt.

[10] As noted above, the IC constraint does not hold for the low-valuation citizens by assumption.

[11] The high-valuation citizens are indifferent between taking the vaccine and waiting for news because of no time discounting. The incentive can be made strict if they discount future payoffs even slightly.

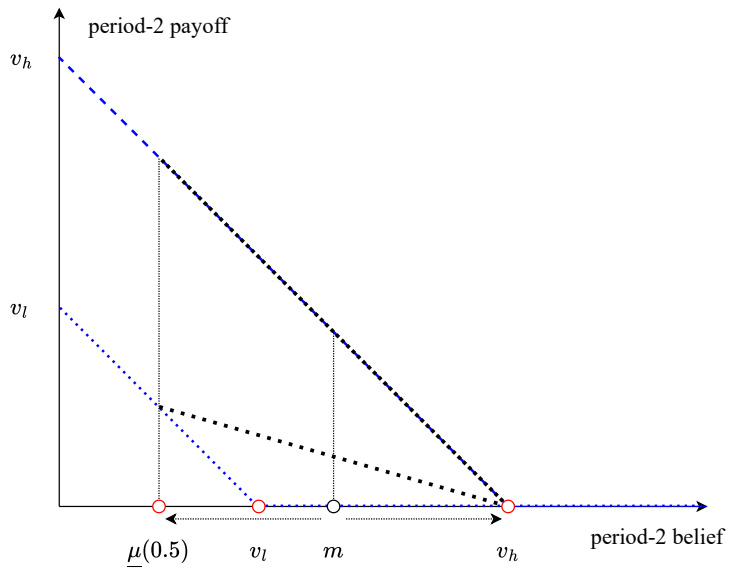

Figure 1: The dashed line at the top represents the value function of the high-valuation citizens, and the dotted line at the bottom represents the value function of the low-valuation citizens. A mean-preserving spread of m into $\underline{\mu}(0.5)$ and v_h raises the expected payoff of the low-valuation citizens while keeping the expected payoff of the high-valuation citizens constant.

to the high-valuation citizens in that their behavior does not depend on the information released in period 2. This reasoning suggests, however, that any mechanism that has this feature can achieve (i) as well. Specifically, we modify the no-disclosure mechanism by applying a mean-preserving spread to m and splitting it into $\underline{\mu}(0.5)$ and v_h. This modified mechanism gives the high-valuation citizens the same continuation payoff and hence continues to achieve (i) because the support of the belief distribution lies entirely to the left of their indifference point v_h; in other words, the extra information provided by the mean-preserving spread is irrelevant for the high-valuation citizens. It is, however, beneficial for the low-valuation citizens because the support spans over their indifference point $\mu_2 = v_\ell$, allowing them to make more informed decisions in period 2. Figure 1 graphically illustrates this situation.

This argument suggests that the government must provide the most accurate information, i.e., $\mu_2 = \underline{\mu}(0.5)$, whenever it chooses to recommend adoption to the low-valuation citizens in period 2. This in turn implies that the government must surely recommend against adoption if it has observed bad news. If it has observed no news, it may recommend adoption with some probability less than 1 (given that full disclosure is not optimal in this example). To sum up, the government can alleviate information free-riding and achieve the optimal outcome by the following mechanism:

- If the government observes bad news, it recommends against adoption with probability 1;
- If the government observes no news, it recommends adoption with some probability $\beta \in (0,1)$ and recommends against adoption with the remaining probability.

This proposed mechanism is characterized by the continuation probability β, which measures the degree of transparency with $\beta = 1$ ($\beta = 0$) corresponding to full (no) disclosure. It is fully revealing when the government is pessimistic (has observed bad news) but randomizes when it is optimistic (no news). The reason why such a mechanism works pertains to the fact that the value of information differs across different preference types: more precise information at the low end of the belief distribution is valuable for the low-valuation citizens but is irrelevant for the high-valuation agents. Since the government must persuade the high-valuation citizens to adopt early to generate the information necessary to persuade the low-valuation citizens, a mechanism that is precise at the low end but obscure at the high end works to achieve the optimal outcome. Remarkably, this observation can be extended to any (continuous) type distribution and any number of periods as we will detail below, even though the problem becomes substantially more complicated.

This illustrative example is simple enough to explicitly obtain the optimal mechanism. Since the low-valuation citizens always benefit from more information, the optimal mechanism is such that it provides the most accurate information to the extent that it does not affect the incentive of the high-valuation citizens. This implies that the high-valuation citizens must be held indifferent between adopting and waiting. If we let β^* be the optimal continuation probability, it must solve

$$v_h - m = \beta^*[(me^{-0.5\lambda} + 1 - m)v_h - me^{-0.5\lambda}].$$

Thus, the optimal continuation probability is given by

$$\beta^* = \frac{v_h - m}{(me^{-0.5\lambda} + 1 - m)v_h - me^{-0.5\lambda}}.$$

Some useful observations can be drawn from this exercise.

1. A lower v_h leads to less transparency.
2. A higher λ (more efficient feedback) leads to less transparency.
3. A higher m (more pessimistic outlook) leads to less transparency.

As is intuitively clear, anything that aggravates the free-rider problem—low valuations, more efficient feedback, and a more pessimistic outlook—makes the optimal mechanism less transparent (a lower β^*). Whenever we have $\beta^* < 1$, although lowering the continuation probability in the optimistic state necessarily entails a loss of surplus, the gain from alleviating information free-riding is more than enough to compensate for the loss.

This example highlights the potential welfare gains of a well-designed information feedback mechanism in a setting plagued by information free-riding. However, some important questions still remain unanswered, as the set of implementable feedback dynamics is inherently constrained in this simple example. When there are more than two periods, the principal can possibly design a more sophisticated mechanism: for instance, she may withhold information for some periods and deliver it later with a delay. Additionally, with more than two types, the principal must also determine which types to persuade into adoption and may therefore benefit from making non-binary recommendations. To address these issues and offer more robust insights, we extend the analysis below to a more general environment with many (continuous) types and arbitrarily many periods.

3 Model

Environment. Consider the same experimentation environment in which a principal (e.g., a government) has a product to be consumed by a continuum of agents (e.g., citizens). Time is discrete and denoted by $t = 1, \dots, T$ where T is the terminal period. In each period t, each agent decides whether to "adopt" the product or "wait" for news; if an agent chooses to adopt, he immediately leaves the game. The cost of adopting the product is determined by the state of nature $\omega \in \{0, 1\}$, which is initially unknown to anyone, where $\omega = 0$ denotes the good (low-cost) state while $\omega = 1$ denotes the bad (high-cost) state. The common prior of the state being bad is $m \in (0, 1)$.

Information. The true state is gradually revealed via the arrival of a signal (news) observable only to the principal. We primarily consider a bad-news ("breakdown") case where a signal arrives only when the state is bad; in Section 5.3, we extend the analysis and consider the case where a signal (or a "breakthrough") also arrives when the state is good.[12] Let $s_t \in \{b, \varnothing\}$ be the signal realized at the end of period t, where $s_t = b$ indicates bad news while $s_t = \varnothing$

[12] We focus on the bad-news case because it is more relevant in many instances of policy experimentation. For instance, in the context of vaccine uptake, it is more likely that the government would receive feedback when the vaccine is of bad quality.

indicates no news. The probability of receiving signal b at the end of period t when the state is ω and the measure of agents adopting in period t is n_t is given by

$$\mathbb{P}[s_t = b \mid \omega, n_t] = 1 - e^{-\lambda \omega n_t},$$

which suggests "no news is good news." At the beginning of each period t, the principal updates her belief (that the state is bad) upon observing s_{t-1}, which we denote by μ_t^p with $\mu_1^p = m$. In the baseline model, we assume that the underlying information-generating process takes an exponential form, purely for simplicity and clarity; in Section 5.2, we consider a more general process and show that our main results hold in this case as well.

Mechanism. The principal designs experiments to generate information about the state of nature. Following the convention, we frame this as an information-design problem in which the principal commits to a feedback mechanism P at the outset of the game. The mechanism specifies the distribution of the sequence of beliefs μ_t subject to Bayes plausibility and her own information structure, where μ_t denotes the agents' period-t belief. The public history of the game at the beginning of period t can be summarized by the sequence of realized beliefs up to that point, which we denote by $\mu^{t-1} := (\mu_1, \dots, \mu_{t-1})$. The mechanism can then be written as $P = (P_t(\cdot \mid \mu^{t-1}))_{t=1}^T$, where $P_t(\cdot \mid \mu^{t-1})$ is the probability mass function of period-t beliefs conditional on the public history of the game.[13] Let $M_t(\mu^{t-1})$ denote the support of $P_t(\cdot \mid \mu^{t-1})$, i.e., $M_t(\mu^{t-1})$ is the smallest set M such that $\sum_{\mu_t \in M} P(\mu_t \mid \mu^{t-1}) = 1$.

Payoffs. Each agent is characterized by his valuation $v \in [0, 1]$, which is distributed according to some distribution F with full support over $[0, 1]$. Let f denote the corresponding density. If an agent with valuation v (hereafter, simply agent v) adopts in period t, he receives a payoff of $v - \omega c$ and leaves the game; if not, his payoff for the period is normalized to 0. To focus on the case where the free-riding incentive is maximized, we assume there is no time discounting. For now, we also assume that the principal is benevolent and maximizes the sum of the payoffs of all agents, which we call the total surplus for clarity; in Section 5.1, we extend the analysis to incorporate a more general objective function that can capture other factors such as payoff externalities and the welfare of third parties (e.g., firm profit).

[13] To be more precise, the mechanism in its most general form is a mapping from the principal's private history $(\mu^{p,t-1}, \mu^{t-1})$ to a distribution of beliefs, where $\mu^{p,t-1} := (\mu_1^p, \dots, \mu_{t-1}^p)$ is the sequence of her own beliefs. We use the current representation for brevity because μ^t is a sufficient statistic for the agents' incentives and welfare in this setting.

4 Analysis

In the two-period example, there is only one possible history, which is $\mu_1 = m$, and any disclosure mechanism is entirely summarized by a single distribution of period-2 beliefs $P_2(\cdot \mid m)$. By contrast, when there are more than two periods, the problem becomes substantially more complicated because the set of possible histories could expand exponentially over time. In the general setting, we need to consider how the current distribution of beliefs potentially affects their future evolution in the continuation game, thereby giving rise to a much wider array of feedback mechanisms.

4.1 Preliminaries

We begin by providing a more detailed account of three key constraints in this general setup: IC constraint, Bayes plausibility, and consistency. In the Appendix (Lemma A.1), we show that the standard skimming property holds, so that if agent v' prefers to wait in some period, all agents $v \in [0, v')$ strictly prefer to wait in that period. In each period t, therefore, there is a threshold type, denoted by v_t^*, who is indifferent between adopting and waiting, with $v_0^* = 1$. Once a mechanism P is fixed, we can pin down a non-increasing sequence of threshold types up to period t for any given history μ^{t-1}. For clarity, we say that the principal "recommends adoption" in period t if $v_t^* < v_{t-1}^*$ and "recommends against adoption" if $v_t^* = v_{t-1}^*$.

If the threshold type v_t^* adopts in period t, his expected payoff is $v_t^* - \mu_t c$. If he chooses to wait, the skimming property implies that he will adopt the next time the principal recommends adoption. Let H_t^s be the set of all subsequences $\mu^{t,t+s} := (\mu_{t+1}, \ldots, \mu_{t+s})$ such that $v_{t+s}^* < v_t^* = v_{t+s-1}^*$ under P, which represents the set of histories in which the principal suspends adoption until period $t+s-1$ and resumes in period $t+s$. Then, for a given history μ^t and the corresponding sequence of threshold types up to period t, the indifference condition is given by

$$v_t^* - \mu_t c = \sum_{s=1}^{T-t} \sum_{\mu^{t,t+s} \in H_t^s} P_{t+1}(\mu_{t+1}|\mu^t) \cdots P_{t+s}(\mu_{t+s}|\mu^{t+s-1})(v_t^* - \mu_{t+s} c), \tag{2}$$

if a solution $v_t^* \in [0, v_{t-1}^*)$ exists. If no solution exists in $[0, v_{t-1}^*)$, no agents are induced to adopt in period t, and we let $v_t^* = v_{t-1}^*$. In what follows, we call (2) the IC constraint.

The principal's problem is restricted by two conditions in our setting. First, as usual, Bayes plausibility requires that the expectation of posteriors must coincide with the prior conditional

on μ^{t-1}, i.e.,

$$\sum_{\mu_t \in M_t(\mu^{t-1})} \mu_t P_t(\mu_t \mid \mu^{t-1}) = \mu_{t-1}. \tag{3}$$

Second, any feasible mechanism must also be consistent with the principal's private information. Specifically, at any point in time, the principal is in either one of the two information states: *pessimistic* (if she has observed bad news) or *optimistic* (if she has observed no news). In the bad-news case, the principal's belief is always $\mu_t^p = 1$ in the pessimistic state, and

$$\mu_t^p = \underline{\mu}(v_{t-1}^*) := \frac{me^{-\lambda(1-F(v_{t-1}^*))}}{me^{-\lambda(1-F(v_{t-1}^*))} + 1 - m}$$

in the optimistic state. Since the principal mixes these two information states to generate a belief, any feasible belief must be bounded between $\underline{\mu}(v_{t-1}^*)$ and 1, i.e.,

$$M_t(\mu^{t-1}) \subset \left[\underline{\mu}(v_{t-1}^*), 1\right]. \tag{4}$$

Throughout the analysis, we maintain the following assumption to focus our attention on relevant cases.

Assumption 1 $c \geq 1 > mc$.

The first inequality ensures that learning is essential for all agents,[14] while the second inequality rules out the trivial case where no agents can be induced to adopt.

4.2 First-best allocation

Before analyzing the model, we first examine the problem of a social planner who can unilaterally determine a non-increasing sequence $(v_t^*)_{t=1}^T$ to maximize the total surplus, given the available information in each period. In this setting, the optimal response to news is straightforward: the social planner terminates the adoption process (by setting a constant threshold for all future periods) if and only if bad news arrives. The social planner's problem is then formulated as

$$\max_{(v_t^*)_{t=1}^T} \sum_{t=1}^{T} \left(me^{-\lambda(1-F(v_{t-1}^*))} \int_{v_t^*}^{v_{t-1}^*} (v-c)\mathrm{d}F(v) + (1-m) \int_{v_t^*}^{v_{t-1}^*} v\mathrm{d}F(v) \right).$$

[14] If $c < 1$, it is optimal for agents $v \in [c, 1]$ to adopt regardless of the state. The optimal decision for those agents is hence trivial in that they always adopt immediately in period 1 in any mechanism.

For the subsequent analysis, it is convenient to define

$$q(v_t^*, v_{t-s}^*) := \underline{\mu}(v_{t-s}^*)e^{-\lambda(F(v_{t-s}^*)-F(v_t^*))} + 1 - \underline{\mu}(v_{t-s}^*)$$

as the probability of observing no news between periods $t-s$ and t (when the threshold changes from v_{t-s}^* to v_t^*), conditional on having observed no news up to period $t-s$. Similarly, define

$$r(v_t^*, v_{t-s}^*) := \underline{\mu}(v_{t-s}^*)e^{-\lambda(F(v_{t-s}^*)-F(v_t^*))}$$

as the conditional joint probability of the state being bad and receiving no news between periods $t-s$ and t. Given these definitions, the first-order condition can be expressed as

$$\left[v_t^* - \underline{\mu}(v_{t-1}^*)c - \left(q(v_t^*, v_{t-1}^*)v_t^* - r(v_t^*, v_{t-1}^*)c\right)\right] f(v_t^*) + \lambda r(v_t^*, 1)\int_{v_{t+1}^*}^{v_t^*}(c-v)\mathrm{d}F(v) = 0, \quad (5)$$

for $t = 1, \ldots, T-1$, with $v_T^* = \underline{\mu}(v_{T-1}^*)c$. The first term on the left-hand side represents the individual net gain from adoption, while the second term represents the social gain from information generation through adoption.

To illustrate the source of inefficiency in this environment, consider the full-disclosure case as a no-intervention benchmark. Under full disclosure, the threshold type must satisfy

$$v_t^* - \underline{\mu}(v_{t-1}^*)c = q(v_t^*, v_{t-1}^*)v_t^* - r(v_t^*, v_{t-1}^*)c.$$

Note that this corresponds to the first term, meaning that the individual net gain is fully internalized by the threshold type and equals 0. This suggests that the social cost of adoption delay arises purely from the second term in (5). Since $c > v_t^*$ for all t by Assumption 1, this term is always positive, implying that the amount of adoption is generally insufficient under full disclosure. Alternatively, this suggests that to implement the first-best allocation, the individual net gain for the threshold type needs to be negative, but this is of course infeasible when the principal lacks the ability to force agents into action. We must therefore search for a second-best solution to mitigate this inefficiency, which we will discuss next.

4.3 Characterization

We now consider an environment where the principal can only control information accessible to the agents but not their actions directly. The principal's problem is to determine the distribution of the sequence of beliefs to maximize the expected total surplus subject to the

three constraints noted above. Define $W_t(\mu_t, v_t^*, v_{t-1}^*)$ as the continuation surplus in period t, when the current belief is μ_t and agents $v \in [v_t^*, v_{t-1}^*)$ choose to adopt. Given some mechanism P, the continuation surplus is given by

$$W_t(\mu_t, v_t^*, v_{t-1}^*) = \int_{v_t^*}^{v_{t-1}^*} (v - \mu_t c) \mathrm{d}F(v) + \sum_{s=1}^{T-t} \sum_{\mu^{t,t+s} \in H_t^s} P_{t+1}(\mu_{t+1}|\mu^t) \cdots P_{t+s}(\mu_{t+s}|\mu^{t+s-1}) W_{t+s}(\mu_{t+s}, v_{t+s}^*, v_t^*).$$

The principal's problem is defined as

$$\max_P \; W_1(m, v_1^*, 1),$$

subject to Bayes plausibility (3) and consistency (4), where $(v_t^*)_{t=1}^T$ is a non-increasing sequence determined by the IC constraint (2) whenever $v_t^* < v_{t-1}^*$. Recall that since the principal has no additional information over and above the prior in period 1, Bayes plausibility requires $P_1(m) = 1$. Our focus is therefore on period 2 and beyond.

A few more notations would be helpful in formally stating the main result. First, for $t \geq 2$, we divide the support $M_t(\mu^{t-1})$ into two disjoint sets $M_t^A(\mu^{t-1})$ and $M_t^N(\mu^{t-1})$, where

$$M_t^A(\mu^{t-1}) := \{\mu_t \in M_t(\mu^{t-1}) : v_t^* < v_{t-1}^*\},$$
$$M_t^N(\mu^{t-1}) := \{\mu_t \in M_t(\mu^{t-1}) : v_t^* = v_{t-1}^*\},$$

with $M_1(\mu^0) = M_1^A(\mu^0) = \{m\}$. We say that the game is in the *adoption phase* in period t if $\mu_t \in M_t^A(\mu^{t-1})$ and in the *no-adoption phase* if $\mu_t \in M_t^N(\mu^{t-1})$. Also, for any $\beta \in [0,1]$, let $\tilde{\mu}(v_{t-1}^*, v_{t-2}^*; \beta)$, or simply $\tilde{\mu}_t$, denote the belief that satisfies Bayes plausibility in period t when the mechanism assigns $\mu_t = \underline{\mu}(v_{t-1}^*)$ with probability $\beta q(v_{t-1}^*, v_{t-2}^*)$ and $\mu_t = \tilde{\mu}_t$ with the remaining probability conditional on $\mu_{t-1} = \underline{\mu}(v_{t-2}^*)$. Formally, for $t = 2, \ldots, T$, it is given by the solution to

$$\beta q(v_{t-1}^*, v_{t-2}^*) \underline{\mu}(v_{t-1}^*) + (1 - \beta q(v_{t-1}^*, v_{t-2}^*)) \tilde{\mu}_t = \underline{\mu}(v_{t-2}^*),$$

with $\tilde{\mu}_1 = 1$.[15]

[15] In period 1, we have $\mu_1 = \underline{\mu}(v_0^*) = m$ with probability 1. The value of $\tilde{\mu}_1$ can be anything other than m, which we need only for part (b) of Definition 1 below.

Definition 1 *A mechanism P is a binary-message termination mechanism (or, simply, a binary-message mechanism) if for each $t = 2, \dots, T$,*

(a) when $\mu_{t-1} = \underline{\mu}(v^*_{t-2})$, there is $\beta_t \in [0,1]$ such that

$$P_t(\mu_t \mid \mu^{t-1}) = \begin{cases} \beta_t q(v^*_{t-1}, v^*_{t-2}) & \text{if } \mu_t = \underline{\mu}(v^*_{t-1}), \\ 1 - \beta_t q(v^*_{t-1}, v^*_{t-2}) & \text{if } \mu_t = \tilde{\mu}_t, \\ 0 & \text{otherwise}, \end{cases}$$

(b) when $\mu_{t-1} = \tilde{\mu}_{t-1}$, the mechanism reveals no information, i.e.,

$$P_t(\mu_t \mid \mu^{t-1}) = \begin{cases} 1 & \text{if } \mu_t = \mu_{t-1}, \\ 0 & \text{otherwise}, \end{cases}$$

where $(v^*_t)_{t=1}^T$ is the induced sequence of thresholds when $\mu_t = \underline{\mu}(v^*_{t-1})$ for all t.[16]

The following facts are useful for interpreting binary-message mechanisms. Throughout the paper, all the proofs are relegated to the Appendix.

Lemma 1 *In a binary-message termination mechanism,*

(i) $\tilde{\mu}_t = \frac{v^*_{t-1}}{c}$,

(ii) $M^A_t(\mu^{t-1}) = \{\underline{\mu}(v^*_{t-1})\}$ *and* $M^N_t(\mu^{t-1}) = \{\frac{v^*_{t-1}}{c}\}$ *if* $\mu_{t-1} = \underline{\mu}(v^*_{t-2}) \in M^A_{t-1}(\mu^{t-2})$,

(iii) $M^N_t(\mu^{t-1}) = \{\frac{v^*_{t-2}}{c}\}$ *and* $\frac{v^*_{t-1}}{c} = \frac{v^*_{t-2}}{c}$ *if* $\mu_{t-1} = \frac{v^*_{t-2}}{c} \in M^N_{t-1}(\mu^{t-2})$,

where $(v^*_t)_{t=1}^T$ is the induced sequence of thresholds when the game stays in the adoption phase for all t.

Along with Lemma 1, Definition 1 provides two essential properties, which we call *termination* and *caution* for clarity. First, part (b) of the definition states the termination property where if the game is in the no-adoption phase for a period, it will never revert back to the adoption phase; as such, $\mu_t \in M^N_t(\mu^{t-1})$ indicates permanent termination of the adoption process. Since the principal will never recommend adoption thereafter, we often say that the adoption

[16] Under Definition 1, there is a unique sequence of thresholds as long as the game stays in the adoption phase.

process "terminates" once she recommends against adoption.[17] Second, part (a) requires the principal to provide the most accurate information $\mu_t = \underline{\mu}(v^*_{t-1})$ with some probability in period t if the game is in the adoption phase in period $t-1$. This alternatively implies that the principal must be optimistic whenever the game is in the adoption phase.[18] Given this, when the game is in the adoption phase in period $t-1$ (implying that the principal is optimistic in period $t-1$), she remains optimistic (by observing no news) with probability $q(v^*_{t-1}, v^*_{t-2})$, in which case she randomizes between continuation (with probability β_t) and termination (with probability $1-\beta_t$). With the remaining probability, the principal turns pessimistic (by observing bad news), in which case the adoption process surely terminates. We refer to this as the caution property, because such a mechanism is biased towards termination in that it terminates surely in the pessimistic state but may also terminate with some probability in the optimistic state (where the principal has observed no bad news).

In a binary-message mechanism, the state transition follows a Markov process, given in Table 1, with the no-adoption phase as an absorbing state. Note that any binary-message mechanism is characterized solely by a sequence $(\beta_t)_{t=1}^T$, where each β_t admits an economically meaningful interpretation: $\beta_t = 1$ corresponds to full disclosure while $\beta_t = 0$ corresponds to no disclosure. The following statement is the main characterization result of this paper, stating that an optimal mechanism can always be found within the class of binary-message mechanisms.

	adoption phase in period $t+1$	no-adoption phase in period $t+1$
adoption phase in period t	$\beta_{t+1} q(v^*_t, v^*_{t-1})$	$1 - \beta_{t+1} q(v^*_t, v^*_{t-1})$
no-adoption phase in period t	0	1

Table 1: State transition matrix

Theorem 1 *There exists an optimal mechanism that is a binary-message termination mechanism.*

[17] It is important to note, however, that the principal only makes recommendations and does not literally have the authority to terminate the adoption process. The phrasing is simply a shorthand for indicating that the principal commits to never recommending adoption again.

[18] To realize the lowest possible belief, the principal must send a fully separating message in the optimistic state. The conclusion then follows since $\mu_t \in M^A_t(\mu^{t-1})$ if and only if $\mu_t = \underline{\mu}(v^*_{t-1})$ by Lemma 1 .

4.4 Discussion

The proof of Theorem 1 is lengthy, and we relegate its technical details to the Appendix. Below, we provide the underlying intuition for the two essential properties of binary-message mechanisms, labeled as *termination* and *caution*.

First, the termination property states that there is no gain from adopting a more complicated "stop-and-go" mechanism that switches back and forth between the adoption phase and the no-adoption phase. This can be seen most clearly by inspecting a three-period example. Consider a mechanism P that induces the lowest possible belief, which we denote by $\underline{\mu}$, with probability $\beta_2 q(v_1^*, 1)$ in period 2 and the belief $\tilde{\mu}_2$ leading to no adoption with the remaining probability. Suppose further that in the history with $v_1^* = v_2^*$ (no adoption in period 2), P induces the lowest possible belief $\mu_3 = \underline{\mu}$ with probability $\gamma_3 q(v_1^*, 1)$ in period 3. The total surplus attained by this mechanism is then given by

$$\begin{aligned} W_1(m, v_1^*, 1) = &\int_{v_1^*}^1 (v - mc)\mathrm{d}F(v) + \beta_2 q(v_1^*, 1) W_2(\underline{\mu}, v_2^*, v_1^*) \\ &+ (1 - \beta_2 q(v_1^*, 1))\gamma_3 q(v_1^*, 1) W_3(\underline{\mu}, v_3^*, v_1^*), \end{aligned}$$

where the IC constraint in period 1 is

$$v_1^* - mc = \beta_2 q(v_1^*, 1)(v_1^* - \underline{\mu}c) + (1 - \beta_2 q(v_1^*, 1))\gamma_3 q(v_1^*, 1)(v_1^* - \underline{\mu}c).$$

On one hand, the IC constraint is unaffected as long as $\beta_2 q(v_1^*, 1) + (1 - \beta_2 q(v_1^*, 1))\gamma_3 q(v_1^*, 1)$ is fixed. On the other hand, we have $W_2(\underline{\mu}, v_2^*, v_1^*) \geq W_3(\underline{\mu}, v_3^*, v_1^*)$ because in period 2, the principal could always replicate what she would do in period 3 (by providing no information in period 3). This means that we can improve the original mechanism by reducing γ_3 and increasing β_2 in a way to keep $\beta_2 q(v_1^*, 1) + (1 - \beta_2 q(v_1^*, 1))\gamma_3 q(v_1^*, 1)$ constant. In the end, this implies $\gamma_3 = 0$, i.e., it is optimal to front-load all the information. This argument directly implies the termination property that substantially reduces the class of mechanisms we need to consider.

Given this, the next step is to show that it is always possible to weakly improve the total surplus by applying a mean-preserving spread to any interior belief, which ensures the caution property. Intuitively, such an operation is beneficial for agents if their value functions are convex; it is in fact relatively straightforward to establish this when there are only two periods. However, the situation becomes more complicated when there are more than two periods, because a mean-preserving spread may affect future beliefs and hence the payoffs of

agents who may adopt in future. To illustrate this point, suppose we start from some interior belief $\mu_t' \in (\underline{\mu}(v_{t-1}^*), \frac{v_{t-1}^*}{c})$ at which a positive measure of agents adopt in period $t < T$. To improve the payoffs of those agents by providing more information, their behavior must be contingent on the new information. A natural way to do this is to split μ_t' into $\underline{\mu}(v_{t-1}^*)$ and $\frac{v_{t-1}^*}{c}$ such that they choose to adopt at $\mu_t = \underline{\mu}(v_{t-1}^*)$ but not to adopt at $\mu_t = \frac{v_{t-1}^*}{c}$. However, while such an operation clearly benefits those agents who are induced to adopt in period t, it necessarily reduces the continuation probability and hence adversely affects agents who may adopt in period $t+1$ and beyond.

Despite this complication, we can still show that the benefit of more precise information always dominates the cost of more frequent termination. Specifically, we establish this result by constructing a payoff-equivalent mechanism that sends either $\underline{\mu}(v_{t-1}^*)$ or $\frac{v_t^*}{c}$ instead of sending μ_t' for sure, where v_t^* is the threshold following $\mu_t = \underline{\mu}(v_{t-1}^*)$. As this argument can be applied to any interior belief μ_t', we conclude that for any non-caution mechanism, there exists a caution mechanism that is weakly superior. This observation directly implies that the principal must provide the most accurate information in the adoption phase by terminating surely in the pessimistic state while randomizing in the optimistic state. The reason why this caution property holds is the same as in the two-period example: more precise information at the low end of the belief distribution helps agents with lower valuations make more informed decisions but is irrelevant for those with higher valuations. The principal can exploit this fact to alleviate the free-rider problem and achieve the optimal outcome.

In any period t, the principal is either pessimistic with $\mu_t^p = 1$ or optimistic with $\mu_t^p = \underline{\mu}(v_{t-1}^*)$. The caution property suggests that whenever the game is in the adoption phase, it must induce the lowest possible belief, i.e., $M_t^A(\mu^{t-1}) = \{\underline{\mu}(v_{t-1}^*)\}$, which implies that the principal surely terminates in the pessimistic state. When the game switches to the no-adoption phase, the mechanism specifies $\tilde{\mu}_t = \frac{v_{t-1}^*}{c}$. The resulting mechanism uses only two messages, one indicating the continuation of the adoption process and the other indicating its permanent termination, which amounts to a simple binary-message termination mechanism.

4.5 Optimality of partial disclosure

In light of Theorem 1, the principal's problem can now be substantially simplified because we only need to specify the continuation probabilities instead of the whole distribution of the

sequence of beliefs. Note that the principal's objective function can be written as

$$W_1(m, v_1^*, 1) = \sum_{t=1}^{T} B_t \left(m e^{-\lambda(1-F(v_{t-1}^*))} \int_{v_t^*}^{v_{t-1}^*} (v-c) \mathrm{d}F(v) + (1-m) \int_{v_t^*}^{v_{t-1}^*} v \mathrm{d}F(v) \right),$$

where $B_t := \Pi_{s=1}^{t} \beta_s$ is the cumulative continuation probability. The IC constraint in period t is given by

$$v_t^* - \underline{\mu}(v_{t-1}^*)c = \beta_{t+1} \left(q(v_t^*, v_{t-1}^*) v_t^* - r(v_t^*, v_{t-1}^*) c \right), \tag{6}$$

with $\beta_{T+1} = 0$. From these, the principal's problem is redefined as

$$\max_{(\beta_t)_{t=1}^T} \; W_1(m, v_1^*, 1),$$

subject to (6).

Recall that the principal has no private information and has no choice but to set $P_1(m) = 1$, which implies $\beta_1 = 1$. For $t \geq 2$, a change in β_t affects v_{t-1}^* and also $v_t^*, \ldots, v_T^*$ through its effect on v_{t-1}^*. Taking derivative with respect to β_t yields

$$\begin{aligned} \frac{\partial W_1}{\partial \beta_t} = &- \sum_{s=t}^{T} B_s \lambda r(v_{s-1}^*, 1) f(v_{s-1}^*) \frac{\partial v_{s-1}^*}{\partial \beta_t} \int_{v_s^*}^{v_{s-1}^*} (c-v) \mathrm{d}F(v) \\ &+ \sum_{s=t}^{T} \frac{B_{s+1}}{\beta_t} \int_{v_s^*}^{v_{s-1}^*} \left(q(v_{s-1}^*, 1) v - r(v_{s-1}^*, 1) c \right) \mathrm{d}F(v). \end{aligned} \tag{7}$$

The individual net gain from adoption disappears by the envelope theorem. The first term of (7) represents the social loss of information, induced by an increase in β_t, which is always negative. The second term represents the gain from continuation, which is always positive. This illustrates the key trade-off faced by the principal: lower transparency, i.e., a lower β_t, makes agents more willing to experiment now and generates more information (as captured by the first term) but leads to more premature termination (as captured by the second term). It is easy to see that the first term disappears as $\beta_t \to 0$ while the second term is unaffected. It immediately follows from this that it is never optimal to set $\beta_t = 0$ (no disclosure) for any t.

At the other end, we can also show that full disclosure is never optimal. To see this, suppose the principal sets $\beta_{t+1} = 1$ for some t. The IC constraint in period t then becomes

$$v_t^* - \underline{\mu}(v_{t-1}^*)c = q(v_t^*, v_{t-1}^*) v_t^* - r(v_t^*, v_{t-1}^*) c. \tag{8}$$

Provided that $c > v_t^*$, the only solution that can satisfy this is $v_t^* = v_{t-1}^*$,[19] i.e., no agents adopt in period t, resulting in an extreme form of procrastination that produces no useful information along the way. We can then show that setting $\beta_{t+1} = 1$ is not optimal for the same reason employed in establishing the termination property. This observation suggests that there is always room for information design to improve welfare when agents are sufficiently patient and forward-looking. At the same time, though, the fact that $\beta_t < 1$ implies that the adoption process must terminate prematurely with some positive probability, indicating that the first-best allocation can never be achieved in this environment.

Theorem 2 *Neither full disclosure nor no disclosure is ever optimal, i.e., $\beta_t \in (0,1)$ for all $t = 2, \ldots, T$. The first-best allocation can never be achieved.*

5 Extensions

In the main body of the analysis, we deliberately consider the simplest setting to illustrate the key insights in a transparent manner. As it turns out, though, our main characterization result, Theorem 1, is robust to various alterations to the underlying setup and hence holds more generally. Below, we discuss three possible extensions—one on the principal's objectives and two on the information-generating process—to ensure that our framework can be applied widely to a range of situations of practical importance.

5.1 More general objective function

So far, we have considered a benevolent principal whose goal is to maximize the sum of the payoffs of all agents. In practice, however, the principal's objective can be more diverse and include benefits that are not fully internalized by the agents. For instance, in the case of vaccine uptake, one major social goal is to achieve herd immunity, the benefit of which can be shared equally among all citizens irrespective of their vaccination decisions. The vaccination process also involves third parties, such as pharmaceutical companies, and the government may have some interest in their well-being as well.

To capture this situation, we now suppose that the principal's objective function includes an additional term that depends on the sequence of past thresholds $v^{*t} := (v_s^*)_{s=1}^t$. Specifically,

[19] To see, note that $r(v_t^*, v_{t-1}^*) = q(v_t^*, v_{t-1}^*) + \mu(v_{t-1}^*) - 1$. Substituting for $r(v_t^*, v_{t-1}^*)$ and rearranging, the only solution that can satisfy this equation when $c > v_t^*$ must have $q(v_t^*, v_{t-1}^*) = 1$. This is only possible when $v_t^* = v_{t-1}^*$.

suppose that the principal's payoff in period t is now more generally given by

$$\int_{v_t^*}^{v_{t-1}^*} (v - \mu_t c) \mathrm{d}F(v) + \alpha H_t(v^{*t}),$$

where $H_t : [0,1]^t \to \mathbb{R}_+$. In this specification, $H_t(v^{*t})$ measures the external benefit that is contingent possibly on the entire sequence of past thresholds, and $\alpha \geq 0$ is the welfare weight given to the external benefit with our baseline model corresponding to a special case with $\alpha = 0$. This extended specification of the objective function can accommodate a range of goals that depend on the stock and the current flow in a flexible manner: for instance, the level of herd immunity should be positively correlated with the stock of citizens who have been vaccinated, whereas a pharmaceutical firm's profit in a given period should be positively correlated with the flow of citizens who get vaccinated in that period. We assume that this additional term satisfies the following monotonicity condition.

Assumption 2 *$H_t(v^{*t}) \leq H_t(\hat{v}^{*t})$ if $v_s^* \geq \hat{v}_s^*$ for all $s \leq t$.*

Assumption 2 ensures that if we have two sequences that start from and end at the same thresholds, the one that is consistently lower (i.e., the one that has a faster adoption rate) yields a weakly higher surplus. For instance, Assumption 2 is satisfied if the external benefit is given by

$$H_t(v^{*t}) = v_{t-1}^* - v_t^* + \eta(1 - v_t^*),$$

where $\eta > 0$ is a weight given to the stock of agents who have adopted. We can show that our main characterization result holds in this extended setup under mild conditions.

Theorem 3 *Theorem 1 holds under Assumption 2 if H_T is weakly convex in v_T^* or α is sufficiently small.*

5.2 Good news and bad news

In the baseline model, we focus on the case where the principal may observe bad news (a "breakdown") when the state is bad. Here, we extend this setup by introducing a possibility that the principal may also observe good news (a "breakthrough") when the state is good.

Specifically, let $s_t \in \{\varnothing, b, g\}$, and assume

$$\mathbb{P}[s_t = b \mid \omega, n_t] = 1 - e^{-\lambda^b \omega n_t},$$
$$\mathbb{P}[s_t = g \mid \omega, n_t] = 1 - e^{-\lambda^g (1-\omega) n_t}.$$

The principal's belief jumps down to 0 upon observing good news and jumps up to 1 upon observing bad news. In the absence of news, the principal's period-t belief is given by

$$\mu_t^p = \underline{\mu}(v_{t-1}^*) := \frac{m e^{-\lambda^b (1 - F(v_{t-1}^*))}}{m e^{-\lambda^b (1 - F(v_{t-1}^*))} + (1-m) e^{-\lambda^g (1 - F(v_{t-1}^*))}}.$$

In this extended setup, the distribution of the principal's beliefs has a three-point support. We represent this distribution by $(\pi_t^{\varnothing}, \pi_t^g)$, where

$$\mu_t^p = \begin{cases} 0 & \text{with probability } \pi_t^g, \\ \underline{\mu}(v_{t-1}^*) & \text{with probability } \pi_t^{\varnothing}, \\ 1 & \text{with probability } 1 - \pi_t^g - \pi_t^{\varnothing}. \end{cases}$$

With the possibility of good news, the optimal mechanism may no longer be binary, and our definition of binary-message mechanism in the baseline model needs to be modified. Consider an extended version of Definition 1.

Definition 2 *A mechanism is a trinary-message termination mechanism if for each* $t = 2, \ldots, T$,

(a) when $\mu_{t-1} = \underline{\mu}(v_{t-2}^*)$, *there is* $\beta_t \in [0,1]$ *such that either*

1. $M_t^A(\mu^{t-1}) = \{0, \hat{\mu}^p(v_{t-1}^*)\}$ *and*

$$P_t(\mu_t \mid \mu^{t-1}) = \begin{cases} \pi_t^g & \text{if } \mu_t = 0, \\ \beta_t \pi_t^{\varnothing} & \text{if } \mu_t = \underline{\mu}(v_{t-1}^*), \\ 1 - \pi_t^g - \beta_t \pi_t^{\varnothing} & \text{if } \mu_t = \tilde{\mu}_t, \end{cases}$$

where $\tilde{\mu}_t$ *solves* $\beta_t \mu_t^{\varnothing} \underline{\mu}(v_{t-1}^*) + (1 - \pi_t^g - \beta_t \pi_t^{\varnothing}) \tilde{\mu} = \mu_{t-1}$, *or*

2. $M_t^A(\mu^{t-1}) = \{0\}$ *and*

$$P_t(\mu_t \mid \mu^{t-1}) = \begin{cases} \beta_t \pi_t^g & \text{if } \mu_t = 0, \\ 1 - \beta_t \pi_t^g & \text{if } \mu_t = \tilde{\mu}_t, \end{cases}$$

where $\tilde{\mu}_t$ *solves* $(1 - \beta_t \pi_t^g)\tilde{\mu}_t = \mu_{t-1}$*;*

(b) when $\mu_{t-1} = \tilde{\mu}_{t-1}$*, the mechanism assigns all the probability on* μ_{t-1}*, i.e.,*

$$P_t(\mu_t \mid \mu_{t-1}) = \begin{cases} 1 & \text{if } \mu_t = \mu_{t-1}, \\ 0 & \text{otherwise.} \end{cases}$$

Part (b) of the definition states the termination property, which is exactly the same as in Definition 1. The difference thus lies in part (a), which states that the mechanism may possibly use three messages. This is what happens in case 1 of part (a), where the mechanism fully reveals good news and mixes (some of) no news with bad news: when good news is observed and revealed, all remaining agents adopt; otherwise, the mechanism works just like a binary-message mechanism. In case 2, the mechanism is all or nothing—either to induce everyone to adopt or no one—where it mixes (some of) good news and (all of) no news with bad news. In either case, therefore, an extended version of the caution property holds in that the principal sometimes terminates the adoption process even when she is relatively optimistic while providing the most accurate information when she continues.

Theorem 4 *There is an optimal mechanism that is a trinary-message termination mechanism.*

The theorem demonstrates that the key insight from our baseline model with only bad news continues to hold in this more general environment. In addition, by incorporating good news, this general setting highlights an important distinction from Che and Hörner (2018), whose baseline model considers only good news. In their setting, the optimal mechanism mixes good news with (some of) no news to keep agents more optimistic in the absence of news.[20] While this strategy is clearly not feasible in our baseline model, it is feasible in the extended setup that also features good news. However, Theorem 4 establishes that even in this richer environment, the optimal mechanism consistently mixes (some or all of) no news with bad

[20]The solution proposed by Knoepfle and Salmi (2024) also partly builds on this intuition: they show that it is optimal to delay the disclosure of good news, which effectively mixes good news with no news. Since they consider forward-looking agents with different discount rates, delayed disclosure generates additional benefit because of the heterogenous cost of waiting. Note that this mechanism is qualitatively different from ours.

news, and mixes some of good news with no news only when no news and bad news are fully pooled. This distinction stems from the heterogeneity of agents in our model: since the agents possess different valuations for the product to be experimented, they have different adoption thresholds and hence respond differently to disclosed information. Our solution exploits this feature to persuade high-valuation agents to adopt early (or not to free-ride) to generate information useful for low-valuation agents.

5.3 More general information-generating process

The final extension concerns the arrival rate of news, which we have thus far assumed to take an exponential form and, more importantly, to depend only on the flow of agents who adopt in that period. Under this assumption, the state of the game at the beginning of each period t is thoroughly summarized by the current threshold v^*_{t-1}, irrespective of all past thresholds. Aside from the restriction on its functional form, however, this specification rules out the possibility that some news may arrive with a delay. In reality, the arrival of news can be stochastic and span over periods.

To capture this possibility, we now extend the baseline specification and suppose that the probability of bad news arriving is given by a more general function of past thresholds. Specifically, we now assume that the probability of observing news depends possibly on the entire sequence of thresholds $v^{*t} := (v^*_s)_{s=1}^t$:

$$\mathbb{P}[s_t = b \mid \omega, v^{*t}] = \begin{cases} 1 - G_t(v^{*t}) & \text{if } \omega = 1, \\ 0 & \text{if } \omega = 0. \end{cases}$$

Let

$$\underline{\mu}(v^{*t-1}) := \frac{m\Pi_{s=2}^{t-1} G_s(v^{*s-1})}{m\Pi_{s=2}^{t-1} G_s(v^{*s-1}) + 1 - m}.$$

be the lowest possible belief in period t as a function of past thresholds. We assume that this function satisfies a monotonicity condition analogous to Assumption 2.

Assumption 3 *$\underline{\mu}(v^{*t-1}) \geq \underline{\mu}(\hat{v}^{*t-1})$ if $v^*_s \geq \hat{v}^*_s$ for all $s \leq t-1$.*

As it turns out, with appropriate modifications,[21] our main characterization result holds independently of the details about the arrival rate of news. Even under the general information-

[21] In the extended specification, $\underline{\mu}(\cdot)$ and $q(\cdot)$ in Definition 1 need to be modified as functions of the entire sequence of thresholds. The definition of $\tilde{\mu}_t$ also needs to be modified accordingly.

generating process, the optimal mechanism always has a two-point support and sends either $\mu_t = \underline{\mu}(v^{*t-1})$ or $\mu_t = \frac{v^*_{t-1}}{c}$. The only caveat is that Lemma 1 may not always hold under the general information-generating process. In the baseline specification, the termination property essentially states that no adoption in one period implies no adoption in all future periods. This property may not always hold in the extended model. For instance, consider a situation where information arrives with a one-period delay rather than immediately. Specifically, the probability of receiving bad news at the end of period t depends only the measure of agents who adopted in period $t-1$ (as opposed to period t as in the baseline specification), and is given by

$$G_t(v^{*t}) = e^{-\lambda(F(v^*_{t-2})-F(v^*_{t-1}))},$$

for all $t \geq 2$. In period 1, there is no "period $t-1$," and hence no information arrives (i.e., $G_1(v^*_1) = 0$ for any $v^*_1 > 0$), which immediately implies $v^*_2 = v^*_1$. However, it does not necessarily mean $v^*_t = v^*_2$ for all $t > 2$, because the new information arriving in period 2 can potentially induce new adoption (and further information generation in future). To account for this possibility and obtain a further characterization of this extended model, define $V^{\varnothing}$ as the set of sequences v^{*t-1} such that $\underline{\mu}(v^{*T-1}) = \underline{\mu}(v^{*t-1})$ if $v^*_{T-1} = v^*_{t-1}$.[22] That is, $V^{\varnothing}$ collects all sequences of thresholds such that no future information would arrive if no additional adoption were induced after that sequence; note that any arbitrary sequence belongs to $V^{\varnothing}$ in the baseline specification. Given this definition, we obtain an extended version of Theorem 1.

Theorem 5 *Theorem 1 holds if the information-generating process satisfies Assumption 3. In the optimal mechanism,*

(i) $v^*_T = v^*_t$ *if and only if* $\mu_t = \frac{v^*_{t-1}}{c}$*;*

(ii) $\underline{\mu}(v^{*t-1}) \in M^N_t(\mu^{t-1})$ *only if* $v^{*t-1} \notin V^{\varnothing}$*.*

The first property states that the adoption process terminates in period t if and only if $\mu_t = \frac{v^*_{t-1}}{c}$, which is also the case in the baseline specification. The difference thus lies in the second property, which states that no adoption in one period does not necessarily imply permanent termination: if $\underline{\mu}(v^{*t-1}) \in M^N_t(\mu^{t-1})$, there is no adoption in period t, but we could still have $v^*_T < v^*_t = v^*_{t-1}$ by the first statement. Note that since any arbitrary sequence belongs to $V^{\varnothing}$, $v^{*t-1} \notin V^{\varnothing}$ never occurs in the baseline specification. In this sense, Theorem 5 is a generalization of Theorem 1 which nests the baseline specification.

[22] Since the threshold and the lowest possible belief are both weakly decreasing, we must have $v^*_{t+k-1} = v^*_{t-1}$ and $\underline{\mu}(v^{*t+k-1}) = \underline{\mu}(v^{*t-1})$ for all $k = 1, \ldots, T-t-1$ if $v^{*t-1} \in V^{\varnothing}$ and $v^*_{T-1} = v^*_{t-1}$.

6 Conclusion

Social learning is often hampered by citizens' desire to wait and see others' experiences. In this paper, we explore optimal ways to resolve this issue via information design and characterize the optimal feedback mechanism to deter information free-riding. The optimal mechanism is cautious in the sense that it certainly terminates when the principal is pessimistic and terminates with some probability even when she is optimistic. The key driving force of this mechanism is the observation that the value of information varies across agents with different valuations: more precise information at the low end of the belief distribution is beneficial for agents with lower valuations but irrelevant for those with higher valuations. As such, a mechanism that is consistently more precise at the low end but obscure at the high end can induce agents with higher valuations to adopt earlier, thereby alleviating the free-rider problem, while still enabling those with lower valuations to make more informed decisions. We show that the optimal mechanism takes a simple form, which effectively reduces the principal's problem to an optimal stopping problem. We also show that the key principles identified in this paper are robust to various alterations in the underlying setup and can be applied to a range of social situations.

Appendix: Proofs

Lemma A.1 (Skimming Property) *Given any mechanism P, if type v prefers adopting to waiting, then all types above v strictly prefer adopting as well; if type v prefers waiting to adopting, then all types below v strictly prefer waiting.*

Proof. Without loss of generality, assume that an agent adopts if he is indifferent between adopting and waiting. It is obvious that given any μ^{T-1}, the statement holds in period T. Let $\overline{v}_T(\mu^T)$ be the least upper bound of types who strictly prefers waiting to adopting in period T. Suppose that the statement holds for $t = \tau+1, \cdots, T$, and let $\overline{v}_t(\mu^t)$ be the least upper bound of types who strictly prefer waiting in period t. We now prove that the statement also holds for period τ. Let $S^s_\tau(v)$ be the set of all subsequences $\mu^{\tau,\tau+s} := (\mu_{\tau+1}, \ldots, \mu_{\tau+s})$ such that $v < \overline{v}_\tau(\mu^\tau)$ and $v \geq \overline{v}_{t+s}(\mu^{t+s})$ for all $t = \tau+1, \cdots, \tau+s-1$ in P. If type v prefers adopting to waiting in period τ, we must have

$$v - \mu_\tau c \geq \sum_{s=1}^{T-\tau} \sum_{\mu^{\tau,\tau+s} \in S^s_\tau(v)} P_{\tau+1}(\mu_{\tau+1} \mid \mu^\tau) \cdots P_{\tau+s}(\mu_{\tau+s} \mid \mu^{\tau+s-1})(v - \mu_{\tau+s} c).$$

The right derivative with respect to v of the left-hand side is 1, which is strictly greater than the right derivative of the right-hand side given by

$$\sum_{s=1}^{T-\tau} \sum_{\mu^{\tau,\tau+s} \in S_\tau^s(v)} P_{\tau+1}(\mu_{\tau+1} \mid \mu^\tau) \cdots P_{\tau+s}(\mu_{\tau+s} \mid \mu^{\tau+s-1}).$$

This implies that for $v' > v$,

$$v' - \mu_\tau c > \sum_{s=1}^{T-\tau} \sum_{\mu^{\tau,\tau+s} \in S_\tau^s(v')} P_{\tau+1}(\mu_{\tau+1} \mid \mu^\tau) \cdots P_{\tau+s}(\mu_{\tau+s} \mid \mu^{\tau+s-1})(v' - \mu_{\tau+s} c).$$

Conversely, if type v prefers waiting to adopting, then

$$v - \mu_\tau c \leq \sum_{s=1}^{T-\tau} \sum_{\mu^{\tau,\tau+s} \in S_\tau^s(v)} P_{\tau+1}(\mu_{\tau+1} \mid \mu^\tau) \cdots P_{\tau+s}(\mu_{\tau+s} \mid \mu^{\tau+s-1})(v - \mu_{\tau+s} c).$$

The left derivative with respect to v on the left-hand side is 1, which is strictly greater than the left derivative on the right-hand side given by

$$\sum_{s=1}^{T-\tau} \sum_{\mu^{\tau,\tau+s} \in \tilde{S}_\tau^s(v)} P_{\tau+1}(\mu_{\tau+1} \mid \mu^\tau) \cdots P_{\tau+s}(\mu_{\tau+s} \mid \mu^{\tau+s-1}),$$

where $\tilde{S}_\tau^s(v) \subset S_\tau^s(v)$ is the set of all subsequences $\mu^{\tau,\tau+s} := (\mu_{\tau+1}, \ldots, \mu_{\tau+s})$ such that $v < \overline{v}_\tau(\mu^\tau)$ and $v > \overline{v}_{t+s}(\mu^{t+s})$ for all $t = \tau+1, \cdots, \tau+s-1$ in P. Therefore, for $v'' < v$,

$$v'' - \mu_{T-1} c < \sum_{s=1}^{T-\tau} \sum_{\mu^{\tau,\tau+s} \in S_\tau^s(v'')} P_{\tau+1}(\mu_{\tau+1} \mid \mu^\tau) \cdots P_{\tau+s}(\mu_{\tau+s} \mid \mu^{\tau+s-1})(v'' - \mu_{\tau+s} c).$$

This shows that the statement holds in period τ. By induction, the statement holds for all periods. ■

Proof of Lemma 1. In a binary-message mechanism, the threshold type must satisfy

$$v_t^* - \underline{\mu}(v_{t-1}^*)c = \beta_{t+1} q(v_t^*, v_{t-1}^*)(v_t^* - \underline{\mu}(v_t^*)c) + (1 - \beta_{t+1} q(v_t^*, v_{t-1}^*)) \max\{v_t^* - \tilde{\mu}_{t+1} c, 0\}. \quad (9)$$

We first show that $v_t^* - \tilde{\mu}_{t+1} c \leq 0$. Suppose otherwise. Then, (9) holds for $v \in [\tilde{\mu}_{t+1} c, v_t^*)$, but this violates the skimming property established in Lemma A.1. As this means $\max\{v_t^* -$

$\tilde{\mu}_{t+1}c, 0\} = 0$, (9) can now be written as

$$v^*_{t-1} - \underline{\mu}(v^*_{t-2})c = \beta_t q(v^*_{t-1}, v^*_{t-2})(v^*_{t-1} - \underline{\mu}(v^*_{t-1})c). \tag{10}$$

Recall that

$$\beta_t q(v^*_{t-1}, v^*_{t-2})\underline{\mu}(v^*_{t-1}) + (1 - \beta_t q(v^*_{t-1}, v^*_{t-2}))\tilde{\mu}_t = \underline{\mu}(v^*_{t-2}),$$

by Bayes plausibility. Substituting this into (10) yields

$$(1 - \beta_t q(v^*_{t-1}, v^*_{t-2}))(v^*_{t-1} - \tilde{\mu}_t c) = 0,$$

which establishes property (i). Properties (ii) and (iii) follow immediately from this. ■

Proof of Theorem 1. To prove the theorem, we need to establish that the following two properties hold in an optimal mechanism.

1. If $\mu_t = \frac{v^*_{t-1}}{c}$, then $\mu_s = \mu_t = \frac{v^*_{t-1}}{c}$ and $v^*_s = v^*_{t-1}$ with probability 1 for all $s > t$.
2. If $\mu_t < \frac{v^*_{t-1}}{c}$, then $\mu_t = \underline{\mu}(v^*_{t-1})$.

Taking v^*_{t-1} as given, we consider two cases.

Case 1: $\mu_t \geq \frac{v^*_{t-1}}{c}$. In this case, it is clear that no agents adopt in period t. Moreover, we also claim $v^*_s = v^*_{t-1}$ for all $s \geq t$. Suppose on the contrary that $v^*_s < v^*_{t-1}$ for some $s \geq t$. Then, we must have $\mu_s < \frac{v^*_{t-1}}{c}$ with some positive probability, but that violates the IC constraint of agent v^*_{t-1} in period $t-1$. As this is a contradiction, no adoption will be induced once we have $\mu_t \geq \frac{v^*_{t-1}}{c}$. This argument then rules out any $\mu_t > \frac{v^*_{t-1}}{c}$ by Lemma 1. Therefore, if $\mu_t \geq \frac{v^*_{t-1}}{c}$, it must be that $\mu_s = \frac{v^*_{t-1}}{c}$ for all $s \geq t$.

Case 2: $\mu_t < \frac{v^*_{t-1}}{c}$. We show that we can increase the total surplus by stretching out any interior belief $\mu_t \in (\underline{\mu}(v^*_{t-1}), \frac{v^*_{t-1}}{c})$ into two outer points. Alternatively, this suggests that it is not optimal to assign a positive probability to any interior belief. We establish this result via induction.

We begin with the period-T surplus, which can be written as

$$W_T(\mu_T, \mu_T c, v^*_{T-1}) = \int_{\mu_T c}^{v^*_{T-1}} (v - \mu_T c)\mathrm{d}F(v).$$

Taking derivative of W_T with respect to μ_T then yields $-c(F(v^*)-F(\mu_T c))$, which is increasing in μ_T, i.e., the period-T surplus is convex in μ_T. This ensures the caution property for period T.

If the mechanism does not possess the caution property, there is some $\tau < T$ and $\mu^{\tau-1}$ such that $P_\tau(\mu'_\tau \mid \mu^{\tau-1}) > 0$ for some $\mu'_\tau \in (\underline{\mu}(v^*_{\tau-1}), \frac{v^*_{\tau-1}}{c})$. Fix $v^*_{\tau-1}$ and let $v^*_\tau \in (0, v^*_{\tau-1}]$ be the threshold following $\mu^\tau = (\mu^{\tau-1}, \mu'_\tau)$. Note that the IC constraint requires $\mu'_\tau \leq \frac{v^*_\tau}{c}$. However, if $\mu'_\tau = \frac{v^*_\tau}{c}$, the continuation payoff must be equal to 0, but that contradicts the induction hypothesis. We thus assume $\mu'_\tau < \frac{v^*_\tau}{c}$ without loss of generality. Given this, in the next period, the agents' belief is either $\underline{\mu}(v^*_\tau)$ or $\frac{v^*_\tau}{c}$ by the induction hypothesis, where Bayes plausibility requires

$$\mu'_\tau = \phi \frac{v^*_\tau}{c} + (1-\phi)\underline{\mu}(v^*_\tau),$$

for some $\phi \in (0,1)$. Call the event that $\mu_{\tau+1} = \underline{\mu}(v^*_t)$ following $\mu^\tau = (\mu^{\tau-1}, \mu'_\tau)$ event E, and let $v^*_{\tau+1}$ be the corresponding threshold. Also, define $\hat{\phi}$ such that

$$\mu'_\tau = \hat{\phi} \frac{v^*_\tau}{c} + (1-\hat{\phi})\underline{\mu}(v^*_{\tau-1}).$$

Note that $v^*_\tau \leq v^*_{\tau-1}$ implies $\hat{\phi} \leq \phi$.

Now consider a modified mechanism $\hat{P}$ (where we denote the agents' belief in $\hat{P}$ generically by $\hat{\mu}_t$ throughout the proofs). Specifically,

- in period τ, it sends either $\hat{\mu}_\tau = \frac{v^*_\tau}{c}$ with probability $\hat{\phi}$ or $\hat{\mu}_\tau = \underline{\mu}(v^*_{\tau-1})$ with probability $1-\hat{\phi}$;
- in period $\tau+1$, following $\hat{\mu}_\tau = \underline{\mu}(v^*_{\tau-1})$, it sends either $\hat{\mu}_{\tau+1} = \frac{v^*_\tau}{c}$ with probability ψ or $\hat{\mu}_{\tau+1} = \underline{\mu}(v^*_\tau)$ with probability $1-\psi$, where

$$\underline{\mu}(v^*_{\tau-1}) = \psi \frac{v^*_\tau}{c} + (1-\psi)\underline{\mu}(v^*_\tau).$$

Call the event that $\hat{\mu}_{\tau+1} = \underline{\mu}(v^*_\tau)$ following $\mu^\tau = (\mu^{\tau-1}, \underline{\mu}(v^*_{\tau-1}))$ event $\hat{E}$. Notice that the agents' belief conditional on event $\hat{E}$ in the modified mechanism is the same as that conditional on E in the original mechanism. Moreover, the probability of event E given history $\mu^{\tau-1}$ is the

same as the probability of event $\hat{E}$ condition on $\mu^{\tau-1}$, because

$$\mathbb{P}(E \mid \mu^{\tau-1}) = 1 - \phi = \frac{v_\tau^* - \mu'_\tau c}{v_\tau^* - \underline{\mu}(v_\tau^*)c},$$
$$\mathbb{P}(\hat{E} \mid \mu^{\tau-1}) = (1-\hat{\phi})(1-\psi) = \frac{v_\tau^* - \mu'_\tau c}{v_\tau^* - \underline{\mu}(v_{\tau-1}^*)c} \frac{v_\tau^* - \underline{\mu}(v_{\tau-1}^*)c}{v_\tau^* - \underline{\mu}(v_\tau^*)c}.$$

Given this, it is easy to verify that the two mechanisms achieve the same total surplus. First, the modification does not affect the payoffs of agents $v \in [v_{\tau-1}^*, 1]$. For agents $v \in [v_\tau^*, v_{\tau-1}^*)$, the payoffs are the same because they adopt in period t with the same expected belief in either mechanism. For agents $v \in [v_{\tau+1}^*, v_\tau^*)$, their payoff is 0 unless event E (event $\hat{E}$) occurs in the original (modified) mechanism; otherwise, these two events occur with the same probability, and their payoffs are again the same. Finally, for agents $v \in [0, v_{\tau+1}^*)$, their behavior remains exactly the same after the modification. This means that this modified mechanism attains a weakly higher total surplus.

This argument implies both the termination and caution properties and hence completes the proof of Theorem 1. First, if $v_\tau^* < v_{\tau-1}^*$, we split μ_τ to $\underline{\mu}(v_{\tau-1}^*)$ and $\frac{v_\tau^*}{c}$, which is what the caution property suggests. Moreover, following $\hat{\mu}_t = \frac{v_\tau^*}{c}$, we have Case 1, and no more adoption will be induced. Second, if $v_\tau^* = v_{\tau-1}^*$, we again split μ_τ to $\underline{\mu}(v_{\tau-1}^*)$ and $\frac{v_\tau^*}{c} = \frac{v_{\tau-1}^*}{c}$, which establishes both the termination and caution properties. Note that in the latter case, the modified mechanism privies no information in period τ (as $\hat{\mu}_\tau = \hat{\mu}_{\tau+1}$ with probability 1), it effectively implements the period $\tau+1$ allocation of the original mechanism a period earlier. ■

Proof of Theorem 2. Since we have already established $\beta_t > 0$ in the main text, here we focus on showing $\beta_t < 1$. We first establish $\beta_T < 1$. To this end, observe that as $\beta_T \to 1$, v_{T-1}^* converges to v_{T-2}^*, i.e., no agents adopt in period $T-1$, and agents $v \in [\mu_{T-1}c, v_{T-2}^*)$ adopt in period T. However, the same allocation can be implemented a period earlier by setting $\beta_T = 0$, in which case agents $v \in [\mu_{T-1}c, v_{T-2}^*)$ adopt in period $T-1$ (while no agents adopt in period T). These two allocations are equivalent and yield the same continuation surplus. The existence of an interior solution ($\beta_T < 1$) is then implied by the fact that it is never optimal to set $\beta_T = 0$.

This argument suggests that v_{T-1}^* is bounded away from v_{T-2}^*. Let $v_{T-1}^* = v_a^*$ and $v_T^* = v_b^*$ be the optimal thresholds where $v_{T-2}^* > v_a^* > v_b^*$. Now consider period $T-1$ and let $\beta_{T-1} \to 1$. Then, in the limit, v_{T-2}^* converges to v_{T-3}^* (where $v_0^* = 1$ if $T = 3$), and hence the period-$T-2$ surplus converges to 0. However, the principal can instead implement the same allocation a period earlier by inducing $v_{T-2}^* = v_a^*$ and $v_{T-1}^* = v_T^* = v_b^*$. Observe that the implemented

allocation is suboptimal because we must have $v_T^* < v_{T-1}^*$ in the optimal mechanism. We thus conclude $\beta_{T-1} < 1$ in the optimal mechanism. We can then apply this argument repeatedly to establish the proposition.

Finally, the fact that $\beta_t < 1$ means that the adoption process terminates prematurely with some positive probability even in the good state. Therefore, the first-best allocation is not achievable in this environment. ■

Proof of Theorem 3. We can apply the same argument as in Theorem 1. Observe first that the argument in Case 1 is independent of the principal's objective function, as it only relies on the IC constraint of the threshold type. Now suppose that the objective function is convex and the caution property holds in period T. Consider some period $\tau < T$ and $\mu'_\tau \in (\underline{\mu}(v^{*\tau-1}), \frac{v^*_{\tau-1}}{c})$. If $v^*_\tau < v^*_{\tau-1}$, we can apply the same modification to replicate the same allocation and achieves the same payoff for the principal. If $v^*_\tau = v^*_{\tau-1}$, the modification allows us to implement the period-$\tau+1$ allocation a period earlier, which is a weak improvement under Assumption 2.

Given this, we only need to show that the period-T surplus is convex to invoke the induction hypothesis. Since we know that the surplus is strictly convex when $\alpha = 0$, this is the case if $H_T(\cdot)$ is weakly convex in v_T^* or α is sufficiently small. ■

Proof of Theorem 4. In the proof, we first claim that if $M_t^A(\mu^{t-1})\backslash\{0\}$ is nonempty, then $P_t(0 \mid \mu^{t-1}) = \pi_t^g$, meaning that the mechanism fully reveals good news. This essentially leaves us with two beliefs, $\underline{\mu}(v^*_{t-1})$ and 1, as in the baseline model with only bad news. We can then apply the same argument as in the proof of Theorem 1 to show that $M_t^A(\mu^{t-1}) \cap (\underline{\mu}(v^*_{t-1}), 1) = \varnothing$, which corresponds to part (a)-1 of Definition 2. Since part (a)-2 is the only possibility if $M_t^A(\mu^{t-1})\backslash\{0\}$ is empty, this completes the proof that the optimal outcome can be implemented by a trinary-message mechanism.

We establish the claim by induction. Suppose there is $\mu'_T \in M_T^A(\mu^{T-1})\backslash\{0\}$ and $P_T(0 \mid \mu^{T-1}) < \pi_T^g$. We can then consider a modified mechanism that splits μ'_T into 0 and $\frac{v^*_{T-1}}{c}$. This modification does not change the thresholds in earlier periods but improves the payoffs of all remaining agents in period T, showing that $P_T(0 \mid \mu^{T-1}) = \pi_T^g$ if $M_T^A(\mu^{T-1})\backslash\{0\}$ is nonempty. Moreover, by applying the same argument as in the proof of Theorem 1, we can confirm that $M_T^A(\mu^{T-1}) \cap (\underline{\mu}(v^*_{T-1}), 1) = \varnothing$. This establishes the claim for period T.

Next, consider an arbitrary period $\tau < T$, given that the claim holds for all periods $t > \tau$. Suppose there is μ'_τ such that $\mu'_\tau \in M_\tau^A(\mu^{\tau-1})\backslash\{0\}$ and $P_\tau(0 \mid \mu^{s-1}) < \pi_\tau^g$. Under the induction hypothesis, we have either $M_{\tau+1}^A(\mu^{\tau-1}, \mu'_\tau) = \{0, \underline{\mu}(v^*_\tau)\}$ or $M_{\tau+1}^A(\mu^{\tau-1}, \mu'_\tau) = \{0\}$. We consider these two cases in turn.

Case 1: $M^A_{\tau+1}(\mu^{\tau-1}, \mu'_\tau) = \{0, \underline{\mu}(v^*_\tau)\}$. In this case, the IC constraint can be written as

$$\begin{aligned} v^*_\tau - \mu'_\tau c &= P_{\tau+1}(0 \mid \mu^{\tau-1}, \mu'_\tau) v^*_\tau + P_{\tau+1}(\underline{\mu}(v^*_\tau) \mid \mu^{\tau-1}, \mu'_\tau)(v^*_\tau - \underline{\mu}(v^*_\tau)c) \\ &= g(0)v^*_\tau + \beta_{\tau+1} g(\underline{\mu}(v^*_\tau))(v^*_\tau - \underline{\mu}(v^*_\tau)c), \end{aligned} \tag{11}$$

where $g(\cdot)$ is the probability mass function for the principal's belief after agents $v \in [v^*_\tau, v^*_{\tau-1})$ adopt. Given this, we now consider a modified mechanism $\hat{P}$ that sends either $\hat{\mu}_\tau = 0$ with probability ε or $\hat{\mu}_\tau = \frac{\mu'_\tau}{1-\varepsilon}$ with probability $1-\varepsilon$, instead of sending $\hat{\mu}_\tau = \mu'_\tau$ for sure, where $\varepsilon > 0$ is an arbitrarily small positive number. If $\hat{\mu}_\tau = 0$, the state is good for sure, and all the remaining agents will adopt. If $\hat{\mu}_\tau = \frac{\mu'_\tau}{1-\varepsilon}$, the IC constraint is

$$v^*_\tau - \frac{\mu'_\tau}{1-\varepsilon} c = \hat{g}(0)v^*_\tau + \beta_{\tau+1}\hat{g}(\underline{\mu}(v^*_\tau))(v^*_\tau - \underline{\mu}(v^*_\tau)c), \tag{12}$$

where $\hat{g}(\cdot)$ is the corresponding probability mass function in the modified mechanism. Observe that the joint probability of $\mu^p_\tau = 0$ and $\hat{\mu}_\tau = \frac{\mu'_\tau}{1-\varepsilon}$ is $g(0) - \varepsilon$, while the probability of $\hat{\mu}_\tau = \frac{\mu'_\tau}{1-\varepsilon}$ is $1-\varepsilon$. It thus follows that $\hat{g}(0) = \frac{g(0)-\varepsilon}{1-\varepsilon}$. Similarly, we obtain $\hat{g}(\underline{\mu}(v^*_\tau)) = \frac{g(\underline{\mu}(v^*_{\tau-1}))}{1-\varepsilon}$. Plugging these into (12), we obtain the same condition as in (11), suggesting that the threshold v^*_τ is unaffected by this modification.

It is easy to see that this modified mechanism achieves the same total surplus. First, this modification does not affect the payoffs of agents $v \in [v^*_\tau, 1]$. Consider agents $v \in [v^*_{\tau+1}, v^*_\tau)$, where $v^*_{\tau+1}$ denotes the threshold following $\mu^{\tau+1} = (\mu^{\tau-1}, \mu_\tau, \underline{\mu}(v^*_\tau))$. The continuation payoff of those agents in period τ in the original mechanism is

$$g(0)v + \beta_{\tau+1} g(\underline{\mu}(v^*_\tau))(v - \underline{\mu}(v^*_\tau)c),$$

whereas that in the modified mechanism is

$$\varepsilon v + (g(0) - \varepsilon)v + \beta_{\tau+1} g(\underline{\mu}(v^*_\tau))(v - \underline{\mu}(v^*_\tau)c),$$

so that their payoffs are also identical. We thus conclude that there is a modified mechanism that weakly improves the original mechanism.

Case 2: $M_{\tau+1}(\mu^{\tau-1}, \mu'_\tau) = \{0\}$. In this case, the IC constraint can be written as

$$v^*_\tau - \mu'_\tau c = P_{\tau+1}(0 \mid \mu^{\tau-1}, \mu'_\tau) v^*_\tau,$$

from which we obtain

$$P_{\tau+1}(0 \mid \mu^{\tau-1}, \mu'_\tau) = \frac{v^*_\tau - \mu'_\tau c}{v^*_\tau}.$$

Note that the continuation payoff of agent $v \in (0, v^*_\tau)$ is

$$P_{\tau+1}(0 \mid \mu^{\tau-1}, \mu'_\tau)v = \frac{v^*_\tau - \mu'_\tau c}{v^*_\tau}v.$$

Now consider a modified mechanism $\hat{P}$ that sends $\hat{\mu}_\tau = \frac{v^*_{\tau-1}}{c}$ with probability $\frac{\mu'_\tau c}{v^*_{\tau-1}}$ and $\hat{\mu}_\tau = 0$ with the remaining probability instead of sending μ'_τ for sure. Note that this mean-preserving spread does not affect the payoff of agent $v^*_{\tau-1}$ but raises the payoffs of agents $v \in [v^*_\tau, v^*_{\tau-1})$. Also, with this modification, all agents adopt in period τ when the belief is 0. As this event occurs with probability $\frac{v^*_{\tau-1}-\mu'_\tau c}{v^*_{\tau-1}}$, the expected payoff is $\frac{v^*_{\tau-1}-\mu'_\tau c}{v^*_{\tau-1}}v$, which is larger than the payoff in the original mechanism because $v^*_{\tau-1} > v^*_\tau$. This proves $M^A_\tau(\mu^{\tau-1}) = \{0\}$.

By induction, the preceding argument establishes that if $M^A_t(\mu^{t-1}) \setminus \{0\}$ is nonempty, $P_t(0 \mid \mu^{t-1}) = \pi^g_t$ for all t. In words, this means that if there is an interior continuation belief, the mechanism must reveal good news with probability 1. When the mechanism does not reveal good news, there are two possibilities: the principal has either observed no news or bad news. Note that this situation is exactly the same as in the baseline model with only bad news. We can thus apply the same argument as in the proof of Theorem 1 to establish the extended caution property as stated in part (a)-1, i.e., $M^A_t(\mu^{t-1}) \cap (\underline{\mu}(v^*_{t-1}), 1) = \varnothing$, which completes the proof. ■

Proof of Theorem 5. To show that Theorem 1 holds in the extended setup, we follow essentially the same approach as in the proof of Theorem 3. Observe first that the argument in Case 1 (of the proof of Theorem 1) holds under any information-generating process, as it only relies on the IC constraint of the threshold type. Observe also that the caution property holds in period T under any information-generating process. This allows us to establish the theorem by induction. Consider some period $\tau < T$ and $\mu'_\tau \in (\underline{\mu}(v^{*\tau-1}), \frac{v^*_{\tau-1}}{c})$. If $v^*_\tau < v^*_{\tau-1}$, we can apply the same modification to replicate the same allocation, which is therefore independent of the information-generating process. If $v^*_\tau = v^*_{\tau-1}$, the modification allows us to implement the period-$\tau+1$ allocation a period earlier, which is a weak improvement under Assumption 3. This proves that Theorem 1 holds in the extended setup as well.

We next show that $v^*_T = v^*_t$ if and only if $\mu_t = \frac{v^*_{t-1}}{c}$. The sufficiency is obvious and implied

by the previous argument: once $\mu_t = \frac{v^*_{t-1}}{c}$, then $\mu_{t+k} = \frac{v^*_{t-1}}{c}$ for all $k = 1, \ldots, T-t$, as it would otherwise violate the IC constraint of the threshold agent v^*_{t-1}. To establish the necessity, suppose on the contrary that $v^*_T = v^*_t$ but $\mu_t > \frac{v^*_{t-1}}{c}$. However, to have $v^*_T = v^*_t$, we must have $\mu_T = \frac{v^*_{t-1}}{c}$ with probability 1, which is not feasible if $\mu_t > \frac{v^*_{t-1}}{c}$.

Finally, we show that $\underline{\mu}(v^{*t-1}) \in M^N_t(\mu^{t-1})$ only if $v^{*t-1} \notin V^{\varnothing}$. Suppose on the contrary that $\underline{\mu}(v^{*t-1}) \in M^N_t(\mu^{t-1})$ but $v^{*t-1} \in V^{\varnothing}$. If $\underline{\mu}(v^{*t-1}) \in M^N_t(\mu^{t-1})$, no agents adopt in period t. Given that $v^{*t-1} \in V^{\varnothing}$, this implies $\underline{\mu}(v^{*t}) = \underline{\mu}(v^{*t-1})$, so that no additional information can be provided and no agents adopt in period $t+1$. Applying this argument repeatedly, we conclude $\mu_T = \underline{\mu}(v^{*t-1})$ with probability 1. This is, however, a contradiction because agents $v \in [\underline{\mu}(v^{*t-1}), \frac{v^*_{t-1}}{c})$ would then have an incentive to adopt in period t. ■

References

Alonso, Ricardo and Odilon Camara, "Persuading Voters," *American Economic Review*, 2016, *106* (11), 3590–3605.

Au, Pak Hung, "Dynamic Information Disclosure," *Rand Journal of Economics*, 2015, *46* (4), 791–823.

Baccara, Mariagiovanna, Gilat Levy, and Ronny Razin, "Research Waves," *mimeo*, 2024.

Ball, Ian, "Dynamic Information Provision: Rewarding the Past and Guiding the Future," *Econometrica*, 2023, *91* (4), 1363–1391.

Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch, "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of Political Economy*, 1992, *100* (5), 992–1026.

Bolton, Patrick and Christopher Harris, "Strategic Experimentation," *Econometrica*, 1999, *67* (2), 349–374.

Bonatti, Alessandro and Johannes Hörner, "Collaborating," *American Economic Review*, 2011, *101*, 632–663.

Chan, Jimmy, Seher Gupta, Fei Li, and Yun Wang, "Pivotal Persuasion," *Journal of Economic Theory*, 2019, *180*, 178–202.

Che, Yeon-Koo and Johannes Hörner, "Recommender Systems as Mechanisms for Social Learning," *The Quarterly Journal of Economics*, 2018, *133* (2), 871–925.

_ **and Konrad Mierendorff**, "Optiaml Dynamic Allocation of Attention," *American Economic Review*, 2019, *109* (8), 2993–3029.

Correia da Silva, Joao and Takuro Yamashita, "Information Design in Repeated Interaction," *mimeo*, 2024.

Ely, Jeffrey C., "Beeps," *American Economic Review*, 2017, *107* (1), 31–53.

_ **and Martin Szydlowski**, "Moving the Goalposts," *Journal of Political Economy*, 2020, *128* (2), 393–782.

Frick, Mira and Yuhta Ishii, "Innovation Adoption by Forward-Looking Social Learners," *Theoretical Economics*, 2024, *forthcoming.*

Guo, Yingni and Eran Shmaya, "The Interval Structure of Optimal Disclosure," *Econometrica*, 2019, *87* (2), 653–675.

Halac, Marina, Navin Kartik, and Qingmin Liu, "Optimal Contracts for Experimentation," *Review of Economic Studies*, 2016, *83* (3), 1040–1091.

_ , _ , and _ , "Contests for Experimentation," *Journal of Political Economy*, 2017, *125* (5), 1523–1569.

Hamel, Liz, Lunna Lopes, and Mollyann Brodie, "KFF COVID-19 Vaccine Monitor: What Do We Know About Those Who Want to "Wait and See" Before Getting a COVID-19 Vaccine?," *KFF Poll*, 2021.

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

Keller, Godfrey and Sven Rady, "Breakdowns," *Theoretical Economics*, 2015, *10* (1), 175–202.

_ , _ , and Martin Cripps, "Strategic Experimentation with Exponential Bandits," *Econometrica*, 2005, *73* (1), 39–68.

Knoepfle, Jan and Julia Salmi, "Dynamic Evidence Disclosure: Delaying the Good to Accelerate the Bad," *arXiv: 2406.11728*, 2024.

Kolotilin, Anton, Tymofiy Mylovanov, Andriy Zapechelnyuk, and Ming Li, "Persuasion of a Privately Informed Receiver," *Econometrica*, 2017, *85* (6), 1949–1964.

Kremer, Ilan, Yishay Mansour, and Motty Perry, "Implementing the "Wisdom of the Crowd"," *Journal of Political Economy*, 2014, *122* (5), 988–1012.

Lyu, Chen and Lu Liao, "Information Design for Social Learning with Patient Agents," *mimeo*, 2025.

Manso, Gustavo, "Motivating Innovation," *Journal of Finance*, 2011, *66* (5), 1823–1860.

Nikandrova, Arina and Romans Pancs, "Dynamic Project Selection," *Theoretical Economics*, 2018, *13* (1), 115–143.

Orlov, Dmitry, Andrzej Skrzypacz, and Pavel Zryumov, "Persuading the Principal to Wait," *Journal of Political Economy*, 2020, *128* (7), 2473–2837.