Gradual Learning from Incremental Actions

Tuomas Laiho, Pauli Murto, Julia Salmi*

March 29, 2021

Abstract

We introduce a collective experimentation problem where small agents choose the timing of an irreversible action under uncertainty and public feedback from their actions arrives gradually over time. We solve the decentralized equilibrium where agents maximize their own payoffs and the socially optimal policy, which internalizes the social value of information. The latter entails an informational tradeoff where acting today speeds up learning but postponing capitalizes on the option value of waiting. We show how different experimentation patterns – including the socially optimal policy – can be implemented as a decentralized equilibrium by using dynamic posted price mechanisms. Extending our analysis to revenue maximizing mechanisms, we study the monopoly pricing of a new durable good under gradual learning from user experience.

JEL classification: C61, D41, D42, D82, D83

Keywords: social learning, experimentation, optimal stopping, Brownian motion, dynamic posted-price mechanism, dynamic monopoly, innovation adoption

*Laiho: University of Oslo (tuomas.laiho@econ.uio.no), Murto: Aalto University School of Business (pauli.murto@aalto.fi), Salmi: University of Copenhagen (julia.salmi@econ.ku.dk). We thank Aislinn Bohren, Martin Cripps, Rahul Deb, Nils Christian Framstad, Dino Gerardi, Daniel Hauser, Jan Knoepfle, Matti Liski, Erik Madsen, Helene Mass, Konrad Mierendorff, Francesco Nava, Mariann Ollar, Marco Ottaviani, Sven Rady, Peter Norman Sørensen, Roland Strausz, Juuso Toikka, Christian Traeger, Juuso Välimäki, and various seminar and conference audiences for their helpful comments. This project has received financial support from ERC grant no. 683031 through PI Bård Harstad (Laiho) and from Emil Aaltonen Foundation, OP Group Research Foundation, Yrjö Jahnsson Foundation, and the ERC grant no. 682417 through PI Vasiliki Skreta (Salmi).

1 Introduction

Many economic decisions have irreversible effects but must be taken under uncertainty. A firm considering expanding its production capacity has to address uncertainty about future demand; a consumer buying a durable good today must evaluate her needs for it in the future; policy makers have to contemplate irreversible impacts of policies that may be hard to predict. If such uncertainties resolve exogenously over time, the decision maker's response is to postpone her actions in order to learn more. But when the actions themselves affect the learning process, there is a tradeoff between waiting for more information and acting now to speed up learning for later decisions.

This paper introduces a novel learning problem where an action taken today has a long-run impact on the flow of information. A continuum of small agents chooses when to stop – for example, when to adopt an innovation or enter a new market. An unknown binary state determines if stopping is profitable, so that the agents would rather stop sooner than later if the state is high. Upon stopping, each agent contributes to a public information flow about the state that guides further agents in their stopping decisions. Our main question is how the resulting path of stopping decisions – the expansion path – is determined on one hand when agents optimize individually and on the other hand when a planner, perhaps by means of an appropriate incentive scheme, coordinates the actions.

We identify two effects that shape the socially optimal expansion path. *The information generation effect* calls for aggressive expansion in order to improve information for future decisions. *The option value effect* calls for cautious expansion in order to have better information for the current decisions. A social planner balances these two effects, but individual agents internalize only the latter effect and thus the decentralized equilibrium suffers from informational free-riding. We characterize both individually and socially optimal behavior in our model and bridge the gap between the two by developing a dynamic posted-price mechanism that implements a desired collective outcome in a decentralized fashion.

Our main analytical innovation is to model the path of individual actions as a stock process, which controls the speed of learning. Each agent who has stopped produces a flow of i.i.d. signals conditional on the true state. In continuous time, the aggregate signal then follows a Brownian motion with an unknown drift, determined by the realised state, and a signal-to-noise ratio proportional to the stock of agents who have stopped. Each stopping decision thus has a long-run effect on information generation. Crucially, learning is *gradual* in contrast to the related experimentation literature where an action generates an *instantaneous* one-shot signal.

The techniques to solve the decentralized equilibrium and the socially optimal policy turn out to be quite different. The common challenge is that the problems are two-dimensional as both the belief and the stock affect the future. Furthermore, the stock and the belief processes are interlinked as the stock determines the flow of new information. A crucial observation to overcome these difficulties is that agents' behavior can be characterized by a belief-dependent threshold in both solutions. If the current stock is below the assigned level, more agents stop. If the current stock is above the assigned level, remaining agents wait for a potential increase in the belief.

We show that the decentralized equilibrium can be solved by analyzing "shortsighted" agents who optimize their stopping decisions against the assumption that no agent stops in the future. Intuitively, the result follows because future expansions only take place when the current marginal agent would be willing to stop too. The equivalence with shortsighted optimization and optimization under correct expectations is hence an equilibrium property, which holds only because other agents are solving similar optimal stopping problems, and may be violated with other (non-equilibrium) stock processes. One implication of the equivalence with shortsighted optimization is that only *past* actions affect decisions today. A natural benchmark for the decentralized equilibrium is a model without learning. Because of the option value effect, for any given belief fewer agents stop than in the no-learning benchmark.

Unlike the decentralized equilibrium, the socially optimal policy takes into account the social value of faster learning. The equivalence with shortsighted optimization breaks because the value of information depends on the expected future actions. Thus, both *past and future* actions affect socially optimal stopping. Because of the information effect, socially optimal policy favors earlier and more aggressive expansions than what happens in the decentralized equilibrium. Compared to the no-learning benchmark, gradual learning tends to first increase the optimal stock and then to decrease it.

After analyzing the decentralized and optimal solutions separately, we ask how to implement the optimal policy in a decentralized manner by using transfers. We show that policies in a broad class – including the social optimum – can be implemented using dynamic posted prices: anonymous prices that are adjusted continuously to reflect the news about the state, pinned down by the envelope theorem. We then show that in many cases the designer can even use *simple posted prices* that only depend on the stock of agents who have stopped. The designer does not need to keep track of beliefs: previous agents' decisions provide all the information she needs to set the simple posted prices.

The methodology we develop in this paper potentially applies to many economic problems such as pricing in markets, the design of environmental policies, the adoption of medical treatments and the roll-out of public policies. Towards the end, we develop two extensions to improve the applicability of the model. First, we show how to extend the mechanism design approach to revenue maximizing mechanisms and we apply the result to the durable goods monopolist who is selling a product of uncertain quality in the presence of learning from past buyers' experiences. We show that in such markets monopoly power can increase welfare over a competitive industry. Intuitively, the ability of the monopolist to internalize the informational externality can more than compensate for the loss of surplus caused by the traditional monopoly distortion. Finally, we extend our model to cover a case where different agent types are differently informative: high types may, for example, be experts who quickly discover the quality of an innovation or in the opposite case they may be fanatics whose informativeness is very low.

1.1 Related literature

Using the framework of our paper, the previous literature on learning can be divided based on whether the information generation effect or the option value effect is present in the model. The current paper is the first to analyze the dual effect of endogenous learning.

The information generation effect is present, and hence learning increases the optimal quantities, in papers analyzing classic single agent bandit problems and experimental consumption (Gittins and Jones 1974, Rothschild 1974, Prescott 1972 and Grossman, Kihlstrom and Mirman 1977). Introducing multiple agents to these models adds an informational externality that dampens the information generation effect. Bolton and Harris (1999), Keller, Rady and Cripps (2005) and Keller and Rady (2010) analyze such models under different assumptions on the learning technology. Applications include Bergemann and Välimäki (1997, 2000) and Bonatti (2011) who analyze dynamic pricing. In all of these papers, there is no option value effect because actions are reversible and hence learning always increases the level of optimal quantities relative to the no-learning benchmark.

When actions are irreversible but information arrives exogenously rather than endogenously, only the option value effect is present. Seminal papers in this literature include McDonald and Siegel (1986), Pindyck (1988), and Dixit (1989) and the ensuing literature on real options is summarized in Dixit and Pindyck (1994).

A few papers investigate social learning with irreversible actions, which bears similarities with informational free-riding in our decentralized solution. Frick and Ishii (2020) analyze the adoption of new technologies using a Poisson process with instantaneous feedback to model learning. The adoption rate of innovations is lower than without learning in their model because of the option value effect. An early paper by Rob (1991) makes a similar observation when analyzing sequential entry into a market of unknown size. Similarly, in the models of optimal timing under observational learning, the option value creates an incentive to wait causing socially inefficient delays (Chamley and Gale 1994, Murto and Välimäki 2011).

Introducing a large player can overturn the effect of social learning and irreversibility on optimal quantities because a large player internalizes the information generation effect. Che and Hörner (2017) study how a social planner, who designs a recommendation system for consumers, can mitigate informational free-riding. Laiho and Salmi (2020) analyze monopoly pricing in a similar setup. Both in Che and Hörner (2017) and in Laiho and Salmi (2020), the presence of a social planner or a monopolist induces learning to increase quantities. The crucial difference from the present paper is that these papers model instantaneous learning from each consumption decision: the planner and the monopolist do not face the option value effect since they get more information only by attracting new consumers.

Our assumption that learning is gradual implies that past actions matter for the current information flow. Two contemporaneous papers share this feature with us, although their models and key tradeoffs are otherwise different from ours. Liski and Salanié (2020) analyze a single-agent problem where a decision-maker controls the accumulation of a stock that triggers a one-time catastrophe at an unknown threshold level. The novel feature in their model is a random delay between the crossing of the threshold and the onset of the catastrophe. Martimort and Guillouet (2020) analyze a model with similar features focusing on a time-inconsistency problem under their assumptions. In these papers learning is about an unknown tipping point, whereas in our paper it is about a fixed unknown state.

Our paper is also related to the literature on dynamic mechanism design, especially in the context of optimal stopping. Board (2007), Kruse and Strack (2015) and Board and Skrzypacz (2016) analyze implementation of stopping rules for agents, whose private types evolve exogenously over time. In contrast, we analyze implementation in a game with externalities. In our model the private types are fixed over time, but agents optimize against a common state variable that evolves endogenously over time.

2 Model

2.1 Actions and payoffs

A unit mass of small agents choose when, if ever, to take an irreversible action (to stop). We index individual agents by their type θ and assume that θ is distributed according to a continuously differentiable distribution function F with a full support on $\Theta := [\underline{\theta}, \overline{\theta}]$. Time t is continuous and goes to infinity.

An agent's stopping payoff, $v_\omega(\theta)$, depends on the state of the world $\omega \in \{H, L\}$ such that the payoff is higher in the high state of the world for all types: $v_H(\theta) \geq 0 > v_L(\theta)$.[1] Payoffs are continuously differentiable and increasing in type: $v'_\omega(\theta) \geq 0$ for $\omega \in \{H, L\}$ where the inequality is strict for at least one $\omega = H$ or $\omega = L$. The realized payoff for an agent of type θ, who stops at time t, is $e^{-rt}v_\omega(\theta)$ where r is the common discount rate. An agent's outside option is zero and we normalize $v_H(\underline{\theta}) = 0$ so that $\underline{\theta}$ is the lowest type who would ever want to stop. The model is equivalent to a setting where agents receive a flow of state dependent payoffs $\pi_\omega(\theta) = rv_\omega(\theta)$ at every instant after stopping.

Agents are risk-neutral and maximize their expected discounted stopping payoffs. The agents do not know the state of the world ω but learn about it over time as we will describe next.

2.2 Learning

The key idea of *gradual* learning is that every agent who has stopped generates a flow of conditionally independent public signals. Therefore, we consider endogenous learning from the *stock* of stopped agents: let q_t denote the stock (measure) of agents who have stopped by time t.

Specifically, the public learns about the state by observing a Brownian diffusion

$$dy_t = q_t\mu_\omega dt + \sigma\sqrt{q_t}dw_t, \tag{1}$$

[1] The analysis easily extends to the case where $v_L(\theta) > 0$ for some types. The only change is that all types, who get a positive stopping payoff in both states of the world, stop immediately.

where we normalize $\mu_H = 1/2$ and $\mu_L = -1/2$ and where w_t is a standard Wiener process. Signal process (1) is the limit of a model where q_t is composed of discrete units that produce conditionally independent noisy signals over time and where the total informativeness is normalized to stay constant. The signals can be interpreted as realized individual payoffs (see Online Appendix).[2]

We denote by x_t the public posterior belief $x_t = Pr(\omega = H|\mathcal{F}_t)$, where $\mathcal{F}_t$ is the natural filtration generated by the signal process (1). The unconditional law of motion for the public belief follows from Bayes' rule:

$$dx_t = \frac{\sqrt{q_t}}{\sigma} x_t(1 - x_t)dw_t. \tag{2}$$

In equation (2), the term $\frac{\sqrt{q_t}}{\sigma}$ is the signal-to-noise ratio of the process (1) and determines how fast the belief converges to the truth. Hence, the higher the stock of stopped agents the more informative the public signals. In Section 6, we extend the model to allow for a more general relationship between the stock q_t and the signal-to-noise ratio.

2.3 Solution concepts

We use the term *policy* for a description of how the stock q_t evolves over time. A policy $Q = \{q_t\}_{t\geq 0}$ is an increasing stochastic process adapted to $\mathcal{F}_t$. Notice that the signal process itself depends on the evolution of q_t, so that in effect we are defining policy Q jointly with signal process Y.

Individual agents take the policy Q as given when they choose their stopping strategies. A strategy for an agent of type θ is a stopping time $\tau(\theta)$ adapted to $\mathcal{F}_t$. The payoff to type θ adopting $\tau(\theta)$ under Q is

$$\mathbb{E}\left[e^{-r\tau(\theta)} v_\omega(\theta) \middle| Q\right], \tag{3}$$

where the vertical line notation means that the expectation is for some fixed process Q.

[2]See Bergemann and Välimäki (1997, 2000), Bolton and Harris (1999), Moscarini and Smith (2001), and Bonatti (2011) for other applications and further discussion. The difference to these papers is that they do not consider learning from the stock but from the flow of new actions.

We say that a stopping profile $\mathcal{T} = \{\tau(\theta)\}_{\theta\in\Theta}$ is *consistent with* Q if

$$Pr\left[\int_{\Theta} \mathbf{1}(\tau(\theta) \leq t) dF(\theta) = q_t \middle| Q\right] = 1$$

for all t. In other words, $\mathcal{T}$ is consistent with Q if the measure of agents that it commands to stop always matches the policy.

It is convenient to define solution concepts directly in terms of a policy rather than in terms of a stopping profile. We consider two solution concepts:

Definition 1. *A policy Q^E is a decentralized equilibrium if there exists a profile $\mathcal{T}^E$ such that i) it is consistent with Q^E and ii) $\tau^E(\theta)$ maximizes* (3) *for each θ when $Q = Q^E$.*

Definition 2. *A policy Q^* is socially optimal if there exists a profile $\mathcal{T}^*$ such that i) it is consistent with Q^* and ii)*

$$\mathbb{E}\left[\int_{\underline{\theta}}^{\overline{\theta}} e^{-r\tau^*(\theta)} v_\omega(\theta) dF(\theta) \middle| Q^*\right] \geq \mathbb{E}\left[\int_{\underline{\theta}}^{\overline{\theta}} e^{-r\tau(\theta)} v_\omega(\theta) dF(\theta) \middle| Q\right],$$

for any policy Q and profile $\mathcal{T} = \{\tau(\theta)\}_{\theta\in\Theta}$ consistent with Q.

3 Analysis

Our objective is to analyze how gradual learning affects stopping decisions. First, we discuss some common properties that hold regardless of whether stopping times are individually or socially optimal and present the no-learning benchmark. Then, we solve both the (unique) decentralized equilibrium and the socially optimal policy. Lastly, we compare the decentralized equilibrium and the socially optimal solution to the no-learning benchmark and provide comparative statics results on the effects of learning.

3.1 Higher types stop first

In principle, we can implement a policy Q by many different stopping profiles. However, because the stopping payoffs are increasing in θ, higher type agents want to stop whenever a lower type agent wants to stop, which leads to monotone stopping profiles:

Lemma 1. *If $\mathcal{T} = \{\tau(\theta)\}_{\theta\in\Theta}$ maximizes* (3) *for each θ for given process Q, then*

$$\Pr\left[\tau(\theta) \leq \tau(\theta') \,\middle|\, \mathcal{F}_t; Q\right] = 1$$

whenever $\theta > \theta'$.

Monotone stopping times are also socially optimal:

Lemma 2. *Any stopping profile $\mathcal{T} = \{\tau(\theta)\}_{\theta\in\Theta}$ consistent with Q satisfies:*

$$\mathbb{E}\left[\int_{\underline{\theta}}^{\overline{\theta}} e^{-r\tau(\theta)} v_\omega(\theta)\, dF(\theta) \,\middle|\, \mathcal{F}_t; Q\right] \leq \mathbb{E}\left[\int_{\underline{\theta}}^{\overline{\theta}} e^{-r\tau^{mon}(\theta)} v_\omega(\theta)\, dF(\theta) \,\middle|\, \mathcal{F}_t; Q\right],$$

where $\tau^{mon}(\theta) := \inf\{t : q_t \geq 1 - F(\theta)\}$.

We prove both Lemma 1 and Lemma 2 in Appendix A.

As it is without loss of generality to restrict attention to monotone stopping profiles, there is a one-to-one mapping between the stock q_t and the largest type θ_t who has not stopped: $q_t = 1 - F(\theta_t)$. It is useful to let the stock be a function of the current highest type, $q(\theta) := 1 - F(\theta)$, which has an inverse (current highest type): $\theta(q) := \{\theta : 1 - F(\theta) = q\}$. With slight notational abuse, we use $v_\omega(q)$ to denote the stopping payoff of type $\theta(q)$.

3.2 Boundary policies

This subsection discusses the dynamics in our model. It turns out that both solutions can be characterized as *boundary policies*:

Definition 3. *A policy Q is a boundary policy if there exists a continuous function $\tilde{q} : [0,1] \to [0,1]$ such that $q_t = \tilde{q}(\max_{s\in[0,t]} x_s)$ where $\tilde{q}$ is strictly increasing for all x such that $\tilde{q}(x) > 0$.*

A boundary policy is Markovian: agents' stopping decisions depend only on the stock and the belief. Because stopping is irreversible, the stock at time t is determined by the highest belief reached up to t. A boundary policy hence divides the stock-belief state space into two regions: in *the expansion region*, more agents stop until the stock equals $\tilde{q}(x)$ and in *the waiting region*, everyone waits.

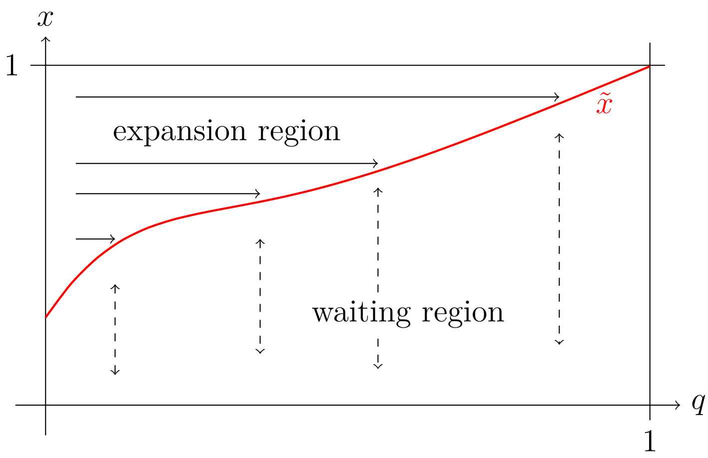

Figure 1: Dynamics in the waiting and expansion regions of the state space.

A boundary policy is fully characterized by the inverse of $\tilde{q}$, a *policy function* $\tilde{x} : [0, 1] \to [0, 1]$, which maps the stock to the cutoff belief. It turns out that it is easier to use policy functions to characterize our solutions than functions $\tilde{q}$. Figure 1 illustrates a boundary policy and the implied dynamics in the state space. Above the boundary, the stock increases (horizontal movement in the figure) and below it, the stock stays constant and only the belief moves (vertical movement). As soon as the belief hits the boundary from below, the quantity is pushed towards right along the boundary. The expansions in the stock are immediate (depicted by solid arrows in the figure), whereas the belief fluctuates according to the diffusion process (2) (dashed arrows). Apart from the possible initial jump, the stock process stays below the boundary and is continuous almost surely.

It is useful to note that since a boundary policy is Markovian in the stock-belief state space, we can express an individual agent's best-response to such a policy as an optimally chosen stopping region in the state space. We utilize this in establishing the existence and uniqueness of a decentralized equilibrium.

3.3 No-learning benchmark

We start our analysis with the benchmark case without learning, which allows us to disentangle how learning affects the decentralized equilibrium and the socially optimal solution.

When there is no learning but the common belief stays constant, the agents' stopping problem is static. An agent stops if and only if his type is so high that the expected payoff is positive: $xv_H(\theta) + (1-x)v_L(\theta) \geq 0$. Hence, the no-learning policy is characterized by the following cutoff:

$$x^{stat}(q) = \frac{-v_L(q)}{v_H(q) - v_L(q)},$$

where $v_\omega(q) \coloneqq v_\omega(\theta(q))$.

Individual optimization and socially optimal policies coincide when there is no learning.

3.4 Decentralized equilibrium

We next characterize the decentralized equilibrium defined in Definition 1. An optimal stopping time for an individual agent trades off the cost of waiting with the option value of waiting. Because the belief process changes endogenously as the stock of stopped agents increases, waiting not only brings more information but also faster learning. Despite this, we show that we can solve *equilibrium* stopping times by first solving a sequence of stopping problems where each agent finds the optimal time to stop when the stock is fixed. That is, we fix $q_t = \hat{q}$ for all t and find the optimal stopping time for type $\theta(\hat{q})$ (pinned down by Lemma 1), assuming that q_t is constant and equal to $\hat{q}$. This problem is a one-dimensional stopping problem and can be solved using standard techniques in the literature (Dixit and Pindyck (1994), see Appendix B for details). We show that the equilibrium in the original problem corresponds to this "shortsighted" problem in which agents do not take future stopping decisions by other agents into account.[3]

The intuition for the equivalence between the shortsighted problem and the original problem is the following: because later expansions in the stock happen only when it is optimal for lower type agents to stop, all higher type agents strictly prefer stopping always when the stock expands (skimming property). Hence,

[3]Our method to solve the decentralized equilibrium is inspired by Leahy (1993) who shows that under exogenous uncertainty the competitive equilibrium behavior coincides with that of 'myopic' investors who ignore the effect the future investments have on the price.

higher type agents want to stop even before the expansion, which means that future expansions do not change the optimal stopping times. In Appendix B, we formalize this argument to get the following result:

Proposition 1. *There is a unique decentralized equilibrium, which is characterized by an increasing policy function x^E:*

$$x^E(q) := \frac{-\beta(q)v_L(q)}{(\beta(q)-1)\,v_H(q) - \beta(q)v_L(q)},$$

where $\beta(q) := \frac{1}{2}\left(1 + \sqrt{1 + \frac{8r\sigma^2}{q}}\right)$.

According to Proposition 1, an agent of type θ waits until the belief reaches the cutoff $x^E(q(\theta))$. The decentralized equilibrium is thus a boundary policy: the policy function x^E defines a boundary so that whenever the belief is about to cross the boundary, more agents stop.

Notice that the cutoff $x^E(q)$ is increasing in the signal precision (decreasing in σ), which means that a better learning technology decreases the stock of agents who are willing to stop at any given belief. The no-learning benchmark is a special case of the decentralized equilibrium as we take $\sigma \to \infty$, which directly gives:

Corollary 1. *For all beliefs in $(x^{stat}(0), 1)$, the stock of stopped agents is strictly smaller in the decentralized equilibrium than in the no-learning benchmark.*

The intuition behind Corollary 1 is that agents want to free-ride on the information provided by other agents. Proposition 1 implies that in the decentralized equilibrium only *past* stopping decisions affect individual agents' behavior because from an individual agents perspective past actions determine the speed of learning. In the next section, we analyze the socially optimal policy which takes into account the informational externality between agents. The solution then takes into account how both *past and future* stopping decisions affect learning.

3.5 Social optimum

Next, we consider the problem in Definition 2 where a benevolent social planner seeks to maximize agents' expected joint payoff. This problem is identical to a problem of a single decision maker who controls a path of incremental expansions.

From Lemma 2, we know that the skimming property holds for the social optimum and hence the problem is reduced to finding the policy Q that maximizes the expected social welfare. We denote the planner's payoff in state (x, q) as

$$U(Q; x, q) = \mathbb{E}\Big[\int_q^1 e^{-r\tau(s)}(xv_H(s) + (1-x)v_L(s))ds \Big| x, q; Q\Big]. \tag{4}$$

The planner's problem is then to find $\sup_Q U(Q; x, q)$. By applying Itô's lemma and using the properties of the Brownian motion, we have the following Hamilton-Jacobi-Bellman (HJB) equation for the planner's problem:

$$rV(x, q) = \max_{q^* \geq q} \left(r \int_q^{q^*} (xv_H(s) + (1-x)v_L(s))ds + \frac{1}{2} V_{xx}(x, q^*) \frac{x^2(1-x)^2}{\sigma^2} q^* \right). \tag{5}$$

We solve the planner's problem by showing that the HJB equation is solved by a boundary policy that cuts the state space into an expansion region and a waiting region. A verification argument then shows that our candidate solution also maximizes the original objective (4).

The optimal policy could in principle consist of several waiting and expansion regions. We proceed by guessing that there is only one expansion and only one waiting region and then later verify this guess (in Appendix C). Let $x^* : [0, 1] \to [0, 1]$ denote our candidate for the socially optimal policy, which we derive next. Function x^* splits the state space in two so that for a given q the planner waits for beliefs $x < x^*(q)$ and expands for beliefs $x \geq x^*(q)$. Since the planner internalizes the value of information for further decisions, we should intuitively expect the socially optimal expansion region to be larger than in the case of decentralized equilibrium, i.e. $x^*(q) < x^E(q)$. We shall verify that this property indeed holds.

We start by solving the value function that solves the HJB equation (5). In the waiting region below x^*, the value consists of the value of potential future actions and can be solved as:[4]

$$V(x, q) = B(q)\Phi(x, q), \tag{6}$$

[4]We have discarded the other root of the characteristic equation as we must have that the value converges to the static solution as $x \to 0$ and $x \to 1$.

where

$$\Phi(x,q) := x^{\beta(q)}(1-x)^{1-\beta(q)} \text{ and } \beta(q) := \frac{1}{2}\left(1+\sqrt{1+\frac{8r\sigma^2}{q}}\right).$$

The next step is to find functions B and x^* that maximize the right-side of the HJB equation. To characterize these, we apply value matching and smooth pasting conditions, which are necessary for the optimality of policy x^*. Because the planner controls the intensity of experimentation, the conditions apply to a marginal increase of the stock q. The value matching condition is thus $V_q(x^*(q),q) = -x^*(q)v_H(q) - (1-x^*(q))v_L(q)$ and the smooth pasting condition is $V_{qx}(x^*(q),q) = -v_H(q) + v_L(q)$. Notice that the HJB equation consists of only future, not past, stopping payoffs and therefore the value matching condition equals the marginal value of increasing the stock with the lost stopping payoff.

Using Equation (6), we can write the value matching and smooth pasting conditions as

$$x^*(q)v_H(q) + (1-x^*(q))v_L(q) + B_q(q)\Phi(x^*(q),q) + B(q)\Phi_q(x^*(q),q) = 0, \quad (7)$$

$$v_H(q) - v_L(q) + B_q(q)\Phi_x(x^*(q),q) + B(q)\Phi_{qx}(x^*(q),q) = 0. \quad (8)$$

Our candidate policy x^* must balance the direct payoff effect, the first term in both equations, against both the option value of waiting and the value of information generation. The last two show up in the latter terms of each equation as the derivatives of the value function.

We show in Appendix C that the system (7) - (8) can be transformed into a non-linear differential equation that defines our candidate policy x^*:

$$x^{*\prime}(q) = g(x^*(q),q), \quad (9)$$

where

$$\begin{aligned} g(x,q) =& x(1-x)\Big[x\Big(\beta'(q)(\beta(q)-1)v_H'(q) - ((\beta(q)-1)\beta''(q) - 2(\beta'(q))^2)v_H(q)\Big) \\ &+ (1-x)\Big(\beta'(q)\beta(q)v_L'(q) - (\beta(q)\beta''(q) - 2(\beta'(q))^2)v_L(q)\Big)\Big] / \\ &\Big[\Big(x(\beta(q)-1)^2 v_H(q) + (1-x)(\beta(q))^2 v_L(q)\Big)\beta'(q)\Big]. \end{aligned}$$

The appropriate initial condition for the differential equation is $x^*(1) = 1$ because the solution must equal the no-learning benchmark when the belief equals one.

The denominator of function g is zero at $(1, 1)$ and hence a potential singularity problem arises. However, we show in Appendix C that the initial value problem has a unique solution below the decentralized solution i.e. a solution satisfying $x^*(q) \leq x^E(q)$ for all $q \in [0, 1]$.[5] We then verify that together with the value function in (6) this candidate policy x^* solves the HJB equation, and we further verify that it also maximizes the original objective (4). In the process, we show that the policy function x^* is continuous and strictly increasing in q and hence satisfies the requirements for a boundary policy.

Proposition 2. *The socially optimal policy is a boundary policy characterized by the unique solution to the initial value problem* $x^{*\prime}(q) = g(x^*(q), q)$ *and* $x^*(1) = 1$ *such that the solution is always below the decentralized solution:* $x^*(q) \leq x^E(q)$ *for all* $q \in [0, 1]$.

Proposition 2 confirms that we can solve the potentially complicated history-dependent problem with a simple boundary policy. However, unlike the decentralized equilibrium, we cannot solve the planner's problem in closed form because the planner is truly forward-looking. For the socially optimal policy, *both past and future* actions are relevant. The past generates information that is useful in evaluating the right decision today, whereas future decisions can be based on information generated by today's action. A socially optimal policy balances the resulting tradeoff between the efficient use of information (option value effect) and the efficient production of information (information generation effect).

Figure 2 provides a numerical example of the effects of the signal precision. The smaller the noise parameter σ is, the more precise the signals are. Better learning technology decreases the cutoff belief $x^*(q)$ when the stock is small and increases it when the stock is high. This arises because improved learning amplifies both information generation and option value effects. The former dominates in the beginning, when the existing stock is low and there are many uncommitted

[5] To see that the uniqueness can only hold in a restricted domain, note that $g(1, q) = 0$ and hence the initial value problem has a trivial solution $x(q) = 1$ for all $q \leq 1$.

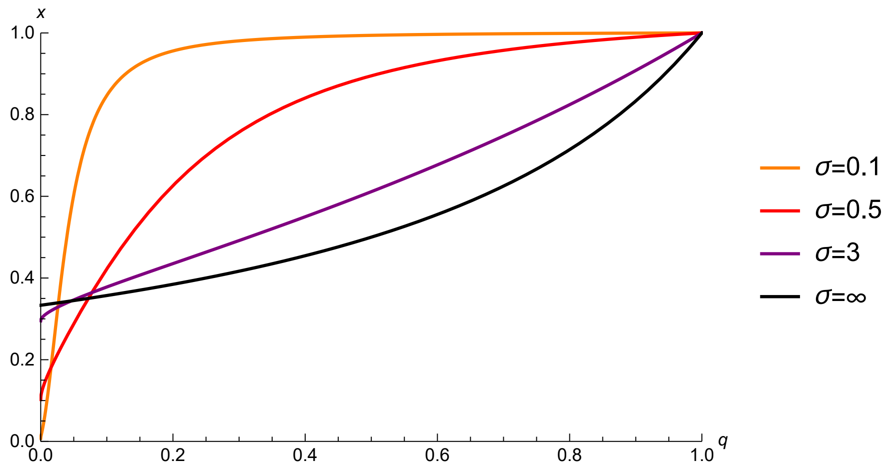

Figure 2: Socially optimal policy $x^*(q)$ for different σ when $v_H(q) = 1 - q$, $v_L(q) = -1/2$, and $r = 0.1$.

agents who benefit from more information. Conversely, the option value effect dominates later when there are few such agents. Notice that the policies with learning (finite σ) are first below and later above the policy without learning ($\sigma = \infty$). Hence, gradual learning may either increase or decrease expansions as the informational tradeoff suggests. The following proposition generalizes this observation (see Appendix C for the proof).

Proposition 3. *There exists $\underline{x} \in (x^{stat}(0), 1)$ and $\overline{x} \in [\underline{x}, 1)$ such that the optimal stock is strictly larger than the no-learning benchmark for all beliefs in $(x^*(0), \underline{x})$ and strictly lower for all beliefs in $(\overline{x}, 1)$.*

Figure 3 illustrates the relationship between the solutions. Gradual learning first increases and then decreases optimal expansions. The decentralized policy is everywhere above the other policies.

4 Mechanism design

In this section, we bridge the gap between individual optimization and the socially optimal policy by analyzing how a designer can implement policies – including the socially optimal policy – with transfers. We present our analysis in steps starting with static implementation before moving on to dynamic implementation. We

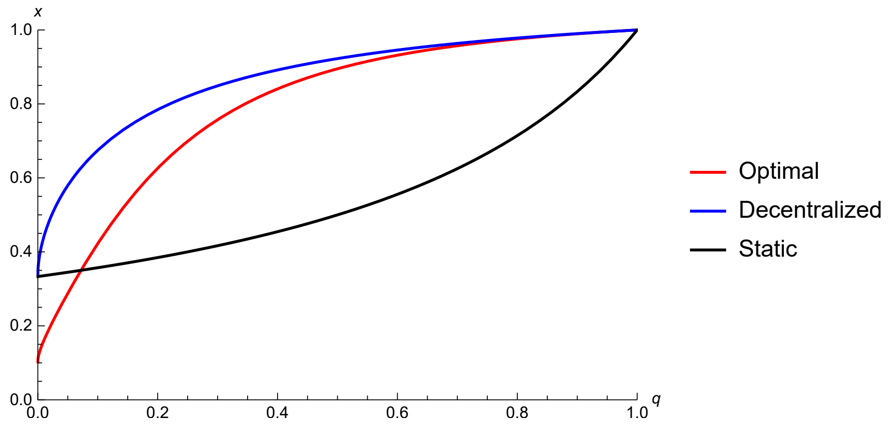

Figure 3: Different policies when $v_H(q) = 1 - q$, $v_L(q) = -1/2$, $\sigma = 0.5$, and $r = 0.1$.

conclude by analyzing revenue maximizing mechanisms.

4.1 Static implementation by a direct mechanism

As a building block towards dynamic implementation, we first consider static direct implementation. This means that all transactions between the designer and the agents take place at time 0: the designer announces a policy Q to be implemented and offers a menu $\{\tau(\theta), P_0(\theta)\}_{\theta\in\Theta}$, where an agent who reports θ pays an upfront transfer $P_0(\theta)$ and gets as an allocation an obligation to stop at time $\tau(\theta) = \inf\{t : q_t = 1 - F(\theta)\}$.

We now outline the steps to pin down transfers that implement a given boundary policy Q and the associated monotone stopping profile as a Bayes-Nash equilibrium.[6] Let $W_0(\theta)$ be the time-0 value of an agent θ under policy Q. Requiring truthful reporting to be optimal, we can use the envelope theorem of Milgrom and Segal (2002) to write out an agent's value (and setting $W_0(\underline{\theta}) = 0$):

$$W_0(\theta) = \int_{\underline{\theta}}^{\theta} W_0'(s)ds = \mathbb{E}\left[\int_{\underline{\theta}}^{\theta} e^{-r\tau(s)}\Big(x_{\tau(s)}v_H'(s) + (1 - x_{\tau(s)})v_L'(s)\Big)ds\right],$$

where $\tau(\theta) = \inf(t : q_t = 1 - F(\theta))$ and the expectation is taken over belief process X induced by policy Q. An *ex-ante* transfer rule, $P_0 : [\underline{\theta}, \overline{\theta}] \to \mathbb{R}$, where each type

[6]This ex-ante implementation applies even if the policy Q is not a boundary policy.

θ is assigned a given stopping rule at time 0, is hence pinned down as

$$\begin{aligned} P_0(\theta) \quad &= \quad \mathbb{E}\Big[e^{-r\tau(\theta)}(x_{\tau(\theta)}v_H(\theta) + (1 - x_{\tau(\theta)})v_L(\theta))\Big] \\ &\quad -\mathbb{E}\left[\int_{\underline{\theta}}^{\theta} e^{-r\tau(s)}\Big(x_{\tau(s)}v'_H(s) + (1 - x_{\tau(s)})v'_L(s))\Big)ds\right], \end{aligned} \tag{10}$$

an expression that is easy to compute numerically for a fixed Q.

To finish the argument, we have to check that incentive compatibility holds globally. This can be done by noting that we associate policy Q with a monotone stopping profile where higher type agents always stop before lower type agents. We show formally in the Online Appendix that this implies an increasing differences property that guarantees global incentive compatibility.

Lemma 3. *Given a boundary policy Q, a direct mechanism $\{\tau(\theta), P_0(\theta)\}_{\theta\in\Theta}$, where an agent reporting type θ pays transfer $P_0(\theta)$ defined in (10) and commits to stopping at time $\tau(\theta) = \inf\{t : q_t = 1 - F(\theta)\}$, satisfies incentive compatibility and participation constraints for every type.*

4.2 Dynamic implementation by posted prices

In the previous section, we reduced the dynamic problem into a static one by assuming that a report made at time 0 involves a commitment to follow a pre-specified stopping rule. However, committing to such behavior is not likely to be feasible in practice. We next investigate how to implement the same policy by a posted price that postpones the transaction and type revelation to the moment when the agent stops.

As a preliminary step, consider a restricted dynamic implementation where the agents are only allowed to choose amongst the stopping times $\{\tau(\theta)\}_{\theta\in\Theta}$ where the state hits boundary points $(\widetilde{x}(q), q)$ for some q. Denote by $P(\widetilde{x}(q), q)$ the payment requested at the moment of stopping from an agent who chooses to stop at the boundary point $(\widetilde{x}(q), q)$. To guarantee that agent $\theta(q)$ wants to stop exactly at $(\widetilde{x}(q), q)$ rather than some other boundary point $(\widetilde{x}(q'), q')$, we can design the payment $P(\widetilde{x}(q), q)$ so that from an ex-ante perspective it replicates the corresponding static payment $P_0(\theta(q))$ for type $\theta(q)$:

$$\begin{aligned} &P(\widetilde{x}(q), q) = \widetilde{x}(q)(v_H(\theta(q)) + (1-\widetilde{x}(q))v_L(\theta(q))) \\ &- \mathbb{E}\Big[\int_{\underline{\theta}}^{\theta(q)} e^{-r(\tau(s)-\tau(\theta(q)))}\Big(\widetilde{x}_{\tau(s)}v_H'(s) + (1-\widetilde{x}_{\tau(s)})v_L'(s))\Big)ds\Big|x(q), q\Big]. \end{aligned} \tag{11}$$

Since this satisfies

$$\mathbb{E}\Big[e^{-r\tau(\theta)}P(\widetilde{x}(q), q)\Big|x_0, q_0\Big] = P_0(\theta(q)),$$

an agent who intends to stop at $(\widetilde{x}(q), q)$ is ex-ante indifferent between paying $P_0(\theta(q))$ at time zero and paying $P(\widetilde{x}(q), q)$ at the moment of stopping. As long as the agents are not allowed to stop outside of the stopping boundary, incentive compatibility continues to hold in this restricted dynamic implementation.

It remains to be shown that the agents do not have an incentive to stop below the boundary when allowed to do so.[7] We consider here two natural transfer schemes that coincide with $P(\widetilde{x}(q), q)$ at the boundary but differ below it.

By a *dynamic posted price*, we refer to a transfer that is a function of the current belief and hence continuously responds to news about the state. The dynamic posted price, $P^D(x)$, is fully pinned down by (11) as the two transfer schemes must coincide at boundary points:

$$P^D(x) := \begin{cases} P(x, \widetilde{q}(x)) & \text{for } x \geq \widetilde{x}(0) \\ P(\widetilde{x}(0), 0) & \text{for } x < \widetilde{x}(0) \end{cases} \tag{12}$$

where $\widetilde{q}(x) : [\widetilde{x}(0), 1] \longrightarrow [0, q_1]$ is the inverse of $\widetilde{x}(\cdot)$. We allow for the possibility that the designer wants to implement a restricted maximal stock, i.e. $q_1 < 1$.

If agent θ stops at state (x, q), his stopping payoff is

$$u_\theta^D(x) = xv_H(\theta) + (1-x)v_L(\theta) - P^D(x).$$

The term $P^D(x)$ makes the optimal stopping problem more complicated than the corresponding problem in the context of the decentralized equilibrium. Yet, without even explicitly solving the individual agents' stopping problems we can

[7] A trivial but non-practical way to do this is to fix the transfer payment to be arbitrarily high whenever $x < \widetilde{x}(q)$ making the cost of stopping prohibitive below the boundary.

prove that this pricing rule is immune to deviations to stopping in state (x, q) that is not on the intended boundary and hence it dynamically implements the intended policy.

When using the dynamic posted price rule (12), the designer needs to continuously observe the news process and carry out detailed Bayesian calculation to adjust the transfer. To make the job of the designer easier and the commitment assumption more palatable we consider as an alternative *simple posted prices* that depend only on the stock q. For example, a seller of a new durable good could set the price based on the cumulative past sales instead of reviews or other feedback from past buyers. As with the dynamic posted price, we set the simple posted price $P^S(q)$ to coincide with (11) at boundary points:

$$P^S(q) := P(\widetilde{x}(q), q) \text{ for } q \in [0, q_1]. \tag{13}$$

Now agents face an optimal stopping problem where the stopping payoff depends on q as well as x:

$$u^S_\theta(x, q) = xv_H(\theta) + (1-x)\, v_L(\theta) - P^S(q).$$

An important property of this scheme is that the stopping payoff changes abruptly at the boundary. Whether the transfer is increasing or decreasing turns out to be critical for providing sufficient stopping incentive at the boundary without inviting deviations to stop below it. If the transfer is increasing, even a slight postponement would make stopping more expensive for the agent at the boundary, providing an additional incentive to stop at the boundary relative to the states below it. However, if the transfer is decreasing, stopping at the boundary is less attractive as a delay would be rewarded with a reduction in transfer payment. As a result, if the decreasing simple posted price scheme is such that an agent is willing to stop at the boundary, he wants to stop even before reaching it.

We summarize our findings in the proposition below:

Proposition 4. *Let Q denote a boundary policy with a strictly increasing policy function $\widetilde{x}$ and $\widetilde{x}(q_1) = 1$. Then:*

- *Q can be implemented by a dynamic posted price $P^D(x)$ in (12).*

- *Q can be implemented by a simple posted price $P^S(q)$ in (13) if and only if $\left(P^S\right)'(q) \geq 0$ for all $q \in [0, q_1]$.*

The proof of Proposition 4 is in the Online Appendix. The key issue is to rule out deviations to stop too early below the intended stopping boundary. We show that we can always rule such deviations out for the dynamic posted price, but in the case of the simple posted price they are ruled out if and only if $P^S(q)$ is everywhere increasing in q. As a part of the next subsection, we give an example where $P^S(q)$ is not everywhere increasing.

4.3 Socially optimal transfers

Now we are ready to connect decentralized optimization and the socially optimal policy and demonstrate the properties of socially optimal transfers. The left panel of Figure 4 depicts the socially optimal boundary policy $\widetilde{x}$ for stopping payoffs $v_H(q) = 1 - q$, $v_L(q) = -\eta - q$, where the three cases correspond to different values for parameter η which increases the cost of stopping in the low state. The right panel shows the corresponding socially optimal transfer function $P(\widetilde{x}(q), q)$ as a function of q. We see that $P(\widetilde{x}(q), q)$ is always negative: the designer pays each agent so that they internalize the information generation effect. The optimal transfer goes smoothly to zero when q approaches 1 because the information generation effect disappears and the decentralized and the optimal policies coincide even without transfers. These properties hold generally for the socially optimal transfer.

We know from Proposition 4 that we can always use a dynamic posted price to implement the policy, but a simple posted price works only if $P^S(q) := P(\widetilde{x}(q), q)$ is monotone. In this example, the optimal policy can only be implemented with simple posted prices when the risk parameter is large enough. That is, transfers are non-monotone when $\eta = 0.2$, but motone when $\eta = 0.4$ or $\eta = 0.6$. The transfer rule becomes non-monotone When η is small because the highest types are almost willing to stop even without transfers whereas lower types need larger transfers as they get a large negative payoff if $\omega = L$ and need to be compensated

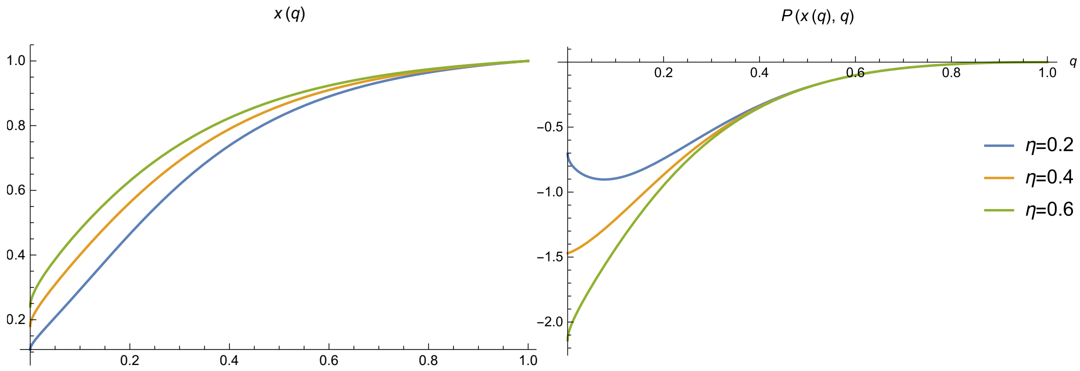

Figure 4: The left panel shows the socially optimal policy for different values of η. The right panel shows the corresponding transfers. $v_H(q) = 1 - q, v_L(q) = -\eta - q, r = 0.1, \sigma = 1$.

for the option value of waiting.

4.4 Revenue maximizing designer

In this section, we show how the techniques that we developed extend to the case of a revenue maximizing designer. We then use the results in Section 5 to solve the problem of a durable good monopolist. Throughout we assume that the type distribution F is twice continuously differentiable and has monotone hazard rate.

Consider a designer whose objective is to maximize the expected sum of transfers, $\mathbb{E}\left[\int_0^\infty e^{-rt} P_t dq_t\right]$, where P_t is the transfer that the agent pays if he stops at time t. Using the techniques in the previous section, the incentive compatible posted price is pinned down by (11). To back out the incentive compatible revenue, we change the order of integration to get a virtual surplus representation for the designer's payoff (see Online Appendix for the proof):[8]

Proposition 5. *Incentive compatibility implies that the designer's expected revenue is:*

$$\mathbb{E}\left[\int_0^\infty e^{-rt} P_t dq_t\right] = \mathbb{E}\left[\int_{\underline{\theta}}^{\overline{\theta}} e^{-r\tau(\theta)} \phi_\omega(\theta) d\theta\right], \tag{14}$$

where $\phi(\theta)$ is the virtual stopping payoff: $\phi_\omega(\theta) := v_\omega(\theta) - v'_\omega(\theta)\frac{1-F(\theta)}{f(\theta)}$.

[8] The approach extends a result in our earlier paper Laiho and Salmi (2020) and shares similar features with Board (2007) who analyzes the optimal sale of options.

Now, we can solve the revenue maximizing designer's problem by using Proposition 2 if we replace the stopping payoffs with virtual stopping payoffs and use a different initial value: $x^*(q_1) = 1$ where q_1 solves $\phi_H(\theta(q_1)) = 1$. Here the monotone hazard rate condition is important as it guarantees that the virtual valuation is increasing in θ and hence the problem satisfies all our assumptions in Section 2. Because we know that the planner's solution is a boundary policy, Proposition 4 guarantees that posted prices are without loss of generality and hence the revenue maximizing mechanism can be implemented.

5 Durable goods monopoly

We apply our results to analyze durable good monopoly when there is uncertainty about the product quality. We first solve the monopolist's problem under commitment and then compare it to the socially optimal solution and the perfectly competitive solution.

The model is as follows. Neither the monopolist nor the buyers know the true quality of the product, $\omega \in \{H, L\}$, but they observe a public signal process (1), generated by past sales, q_t. Each buyer wants to purchase one unit and exits after purchase. Similar to the general model a buyer's utility from consumption depends on his private type, $\theta \in [\underline{\theta}, \overline{\theta}]$, and the common quality: $\mathbb{E}[u(\theta, \omega)] = \mathbb{E}[\mathbf{1}_{\omega=H} \cdot \theta] = x_t\theta$, where x_t is the current belief that the quality is high. In addition, we assume that the type distribution satisfies the conditions in Section 4.4. The monopolist faces marginal cost of production $c > 0$ and commits to a pricing scheme, P_t.

We can use Proposition 5 from Section 4.4 to write the monopolist's objective by using the expected virtual valuation net of the cost of production. Then, posted prices from Equation (11) can be used to implement the desired policy. The monopolist's problem becomes a special case of the model in Section 2 where $v_H(\theta) = \theta - (1 - F(\theta))/f(\theta) - c$ and $v_L(\theta) = -c$. This allows us to use Proposition 2 to characterize the profit maximizing policy:

Corollary 2. *The monopolist's policy is to sell when the belief is above $x^M(q)$ and to wait when it is below. The policy x^M is characterized by $x^M(\overline{q}^M) = 1$ and*

${x^M}'(q) = g(q, x^M(q))$, *where* $\overline{q}^M$ *solves* $\theta(\overline{q}^M) - (1 - F(\theta(\overline{q}^M)))/f(\theta(\overline{q}^M)) = c$ *and* g *is given in* (9).

We contrast the monopoly solution with the planner's socially optimal solution and the competitive market equilibrium. The planner's solution can be found by applying Proposition 2 for the case where $v_H(q) = \theta(q) - c$ and $v_L(q) = -c$:

Corollary 3. *The social planner's policy* x^P *is characterized by* $x^P(\overline{q}^P) = 1$ *and* ${x^P}'(q) = g(q, x^P(q))$, *where* $\overline{q}^P$ *solves* $\theta(\overline{q}^P) = c$ *and* g *is given in* (9).

Suppose next that there are no barriers of entry to the market so that the price equals the marginal cost: $P_t = c$. An individual buyer's purchasing problem then coincides with the decentralized equilibrium in Section 3.4, with $v_H(q) = \theta(q) - c$ and $v_L(q) = -c$, and we have the following corollary to Proposition 1:

Corollary 4. *The competitive market policy is*

$$x^C(q) = \frac{\beta(q)c}{(\beta(q) - 1)\theta(q) + c}.$$

The planner's solution and the competitive equilibrium are the socially optimal and the decentralized solutions of the same problem, whereas the monopoly solution uses different stopping payoffs. This difference leads to different inefficiencies in monopoly and competitive markets.

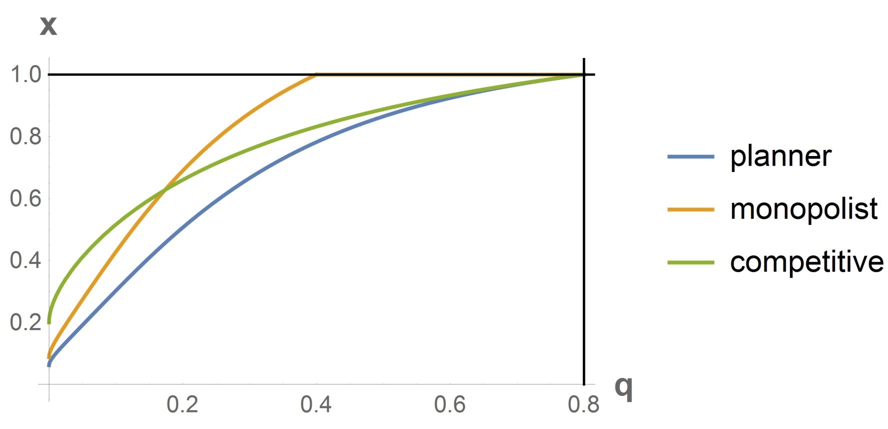

Figure 5: Different solutions for uniform $(0, 1)$ types, $c = 0.2$, $r = 0.1$, and $\sigma = 0.5$.

In the numerical example of Figure 5, the monopolist's and the competitive policies are everywhere above the planner's policy: both markets require inefficiently high belief for new consumers to purchase the product. The monopoly policy is first below and then above the competitive policy. This is because the information generation effect encourages the monopolist to sell in the beginning and because early sales do not generate large information rents to other buyers. Later on, the monopolist reduces sales as the option value effect gets stronger and because later sales impose information rents to higher type buyers. The competitive market ignores the information generation effect but is otherwise efficient and therefore competitive sales are larger for high beliefs. We show in Online Appendix that this comparison holds for all type distributions and parameter values.

Notice that because the monopolist sells to lower type buyers only if the belief increases, the implied prices can be non-monotone. Endogenous learning favors introductory offers, and even pricing below the marginal cost because incentives to generate information are the strongest in the beginning. However, prices are still inefficiently high. The monopolist's incentives to generate information are weaker than the planner's because the monopolist cannot capture all the value from the buyers. This creates a distortion at the top of the type distribution, which is not present when the quality is known. In other words, a higher initial belief is needed for the monopolist to be willing to launch the product: $x^M(0) > x^P(0)$.

As a final remark, notice that a regulator can implement the socially efficient consumption in both monopoly and competitive markets by using appropriate subsidies but the subsidy schemes differ qualitatively. To encourage the monopolist to sell more, the regulator should use *back-loaded* subsidy that increases over sales: $s(x, q) = x(1 - F(\theta(q)))/f(\theta(q))$. A back-loaded subsidy scheme incentivizes the monopolist to sell the socially optimal amount because she internalizes the benefits of information generation. If the market is competitive, however, the subsidy must be *front-loaded* because competition eliminates dynamic incentives (see Figure 4 in Section 4.3).

6 Type-dependent informativeness: experts versus fanatics

Our baseline model assumes that all agents are equally informative: one unit of the stock q produces the same (marginal) amount of information. However, this might not necessarily be true in many applications of our model. For example, first buyers might be fans of the product whose experience matters less for the general population. Or conversely, the first units might be acquired by experts who are able to deduce the true value of the product much more quickly than average users. In this subsection, we show that as long as we can ensure that agents' stopping decision are monotone so that higher types stop first, the analysis in the baseline model is still valid.

Let the marginal informativeness of agent θ be $i(\theta) > 0$ so that the total "information stock" at time t is $z_t := \int_{\theta \in S_t} i(\theta) dF(\theta)$ where S_t is the set of agents who have stopped by time t. The evolution of the news process Y_t is then given by $dy_t = z_t \mu_\omega dt + \sigma \sqrt{z_t} dw_t$. The two especially interesting cases are when the high types are more informative, $i'(\theta) > 0$ ("experts"), and when the low types are more informative, $i'(\theta) < 0$ ("fanatics"). Throughout we assume that function i is continuously differentiable.

Notice first that the stopping profile in the decentralized equilibrium must be monotone in type because individual agents ignore the effect their stopping has on information, and thus Lemma 1 holds as in the baseline model. Because of this, we have a one-to-one relationship between the stock and the information stock: $z_t = h(q_t)$ where h is an "informativeness" function that satisfies $h'(q) = i(\theta(q))$. The analysis of the decentralized equilibrium then stays essentially the same as before.[9]

The socially optimal stopping profile is monotone in the experts environment as both the marginal informativeness and the stopping payoff are increasing in the type. It is also monotone in the fanatics environments when marginal informative-

[9] We only need to adjust β function (plug in $h(q)$ instead of q). With this change Proposition 1 still characterizes decentralized behavior.

ness is not changing too extremely (see Online Appendix). In these cases, we can solve the socially optimal solution with a change of variables. Consider a problem that is otherwise identical to the problem solved in Section 3.5 but where the stock process Q is replaced with the information stock process Z and where the stopping payoffs are scaled so that they take into account how many agents need to stop to increase the information stock by one unit: $\hat{v}_\omega(z) := v_\omega(h^{-1}(z))h^{-1\prime}(z)$. We show in Online Appendix that solving this problem solves the planner's original problem when infomativeness of stopping is type-dependent:[10]

Proposition 6. *Suppose either condition (i) or condition (ii) holds:*

(i) Marginal informativeness is increasing in type (experts).

(ii) Marginal informativeness is decreasing in type (fanatics) but informativeness does not change too extremely: $\frac{-i'(\theta)}{i(\theta)} \leq \frac{-v_L'(\theta)}{v_L(\theta)}$ holds for all θ.

Then the socially optimal policy is characterized by $x^*(z(q))$ where x^* is as defined in Proposition 2.

When high types are extremely fanatical, we may run into trouble: because low types are more informative, the social planner may want to use them first for experimentation. If we restrict to monotone allocations over types, the policy function may be non-increasing because the scaled payoffs are non-monotone. Essentially the social planner may want to bunch some agents to stop together because lower types may have a larger social value of stopping.

Figure 6 illustrates how heterogeneous informativeness affects the socially optimal policy. As before, the comparison between optimal quantities depends on the informational tradeoff between option value and information generation effects. The information generation effect is more pronounced for low q in the expert environment, while it is more pronounced for higher q in the fanatics environment. We see this in Figure 6 as the solution in the expert environment is first below and then quickly rises above the fanatics solution. The expert environment favors

[10] The key steps are to show that the optimal stopping profile is monotone in type and that the policy function x^* is monotone in z.

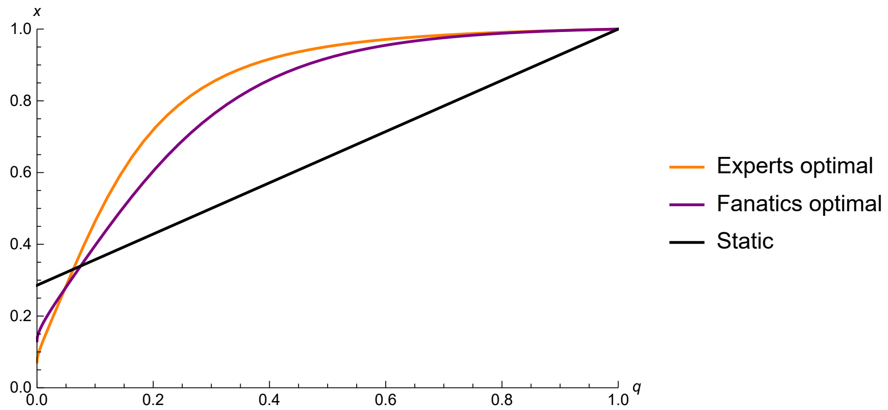

Figure 6: Socially optimal policies for experts (concave $h(q) = 3/2q - 1/2q^2$) and fanatics (convex $h(q) = 1/2q + 1/2q^2$) when $v_H(q) = 1 - q$, $v_L(q) = -0.4 - q$, $\sigma = 0.5$, and $r = 0.1$.

relatively early expansions because the high types produce more information than in the fanatics environment.

7 Concluding remarks

The main contribution of this paper is to develop a methodology for analyzing implications of irreversibility and gradual learning. We demonstrate this in the context of pricing new durable goods, but gradual learning is present in many other contexts that we do not explicitly model in this paper. We conclude by pointing out some potential avenues for further applications.

In markets for new products and services, firms investing in productive capital face uncertainty about market demand. Learning the true long-term demand takes time because of noise in consumer behavior and because the initial capacity may be too low to capture all important market segments. In the language of our model, the total capital installed is the stock variable that facilitates market experimentation and generates revenue. If learning were *exogenous*, the decentralized, or competitive, equilibrium corresponds to the social optimum, as established by Leahy (1993). With endogenous learning, this correspondence is no longer true because of the informational externality (see the difference between Proposition 1 and Proposition 2). Studying interactions between payoff externalities and informational externalities in the context of an oligopolistic market structure points to another avenue to extend our work.

Our methodology can be applied to analyze adoption patterns of new technologies in a much broader sense than industry investment. Think of, for example, new medical devices and treatments. While they go through rigorous testing before approved for use, there are numerous examples of products which have been withdrawn due to concerns for the well-being of patients.[11] In many cases, the true effectiveness of new products is only learned gradually from experience. Similarly, R&D or development aid projects tie resources for many years and information about the outcomes arrives gradually over different stages of the project.

Many actions beyond adopting new technology have uncertain and irreversible consequences. One example is the emissions of pollution which affect the environment. Once created, it is hard to reduce a stock of emissions. A prominent example is the regulation of anthropogenic greenhouse gases whose effects on the climate and society remain uncertain despite considerable research effort. Arguably, we can learn the true effects only gradually from experience. Our mechanism design approach is a useful starting point for designing environmental regulation that takes into account both informational and environmental externalities.[12]

There are many sources of irreversibility in public policy making. Policies themselves may be hard to change because of political uncertainty, potential legal consequences, or legislative lags. But even if changed, a policy may have already affected many people. As a concrete example, consider educational policy, such as the maximal size of a class room, which has an irreversible impact on children. Later labor market outcomes and other information we collect from each cohort helps to evaluate the educational policy at the time when they went to school, independent of whether the norms have changed since then. The trade-off that our framework addresses is how to balance the value of information generation and the risk of irreversible negative consequences.

[11] See for example DePuy hip replacement recall or for example the FDA list of recalled medical devices for 2018.

[12] Liski and Salanié (2020) analyze the optimal policy in a different model that focuses on the possibility of a one-time catastrophe that is triggered once the emission stock exceeds an unknown threshold.

Appendix

A Monotonicity

Proof of Lemma 1

Proof. Let policy Q be fixed. Type θ wants to stop at time t if

$$x_t v_H(\theta) + (1 - x_t) v_L(\theta) \geq \mathbb{E}[e^{-r(\tau-t)}(x_\tau v_H(\theta) + (1 - x_\tau) v_L(\theta))|\mathcal{F}_t; Q],$$

for all stopping rules τ. Or equivalently,

$$v_L(\theta)(1 - x_t - \mathbb{E}[e^{-r(\tau-t)}(1 - x_\tau)|\mathcal{F}_t; Q]) + v_H(\theta)(x_t - \mathbb{E}[e^{-r(\tau-t)} x_\tau|\mathcal{F}_t; Q]) \geq 0.$$

The left-hand side is increasing in θ because expressions $(1 - x_t - \mathbb{E}[e^{-r(\tau-t)}(1 - x_\tau)])$ and $(x_t - \mathbb{E}[e^{-r(\tau-t)} x_\tau])$ are positive (follows from that x_τ is a martingale and $e^{-r(\tau-t)} < 1$) and v_ω is increasing. Therefore, if type θ wants to stop, type $\theta' > \theta$ wants to stop too. □

Proof of Lemma 2

Proof. $\mathcal{T}$ and $\mathcal{T}^{mon}$ are both consistent with Q. We show that monotone stopping ordering maximizes *ex post* welfare for all realized paths of (X, Q). The claim follows once we show that for all types $\theta, \theta' \in [\underline{\theta}, \overline{\theta}]$ such that $\theta > \theta'$ and for all realized stopping times $t, t' \in \mathbb{R}_+$ such that $t \leq t'$,

$$e^{-rt} v_\omega(\theta) + e^{-rt'} v_\omega(\theta') \geq e^{-rt'} v_\omega(\theta) + e^{-rt} v_\omega(\theta').$$

The above condition is equivalent with $(e^{-rt} - e^{-rt'})(v_\omega(\theta) - v_\omega(\theta')) \geq 0$, which necessarily holds as $t \leq t'$ and $v_\omega(\theta) \geq v_\omega(\theta')$ by assumption if $\theta > \theta'$. □

B Decentralized equilibrium

Proof of Proposition 1

Proof. As a preliminary step, we analyze the optimal stopping problem of type θ when the stock q_t is fixed at some q forever. We denote by $\overline{F}_\theta(x; q)$ the value function of this problem where the bar indicates that the stock is fixed. By applying Itô's lemma and using the properties of the Brownian motion, we get the following Hamilton-Jacobi-Bellman (HJB) equation for the agent's value before stopping:

$$r\overline{F}_\theta(x;q) = \frac{\partial^2}{\partial x^2}\overline{F}_\theta(x;q)\frac{x^2(1-x)^2}{2\sigma^2}q.$$

We can solve this second order ODE with a parameter q in closed form: $\overline{F}_\theta(x;q) = B_\theta(q)\Phi(x,q)$, where $B_\theta(q)$ is an unknown constant that depends on the parameter q and $\Phi(x,q) := x^{\beta(q)}(1-x)^{1-\beta(q)}$ and $\beta(q) := \frac{1}{2}\left(1+\sqrt{1+\frac{8r\sigma^2}{q}}\right)$.[13]

The unknown constant $B_\theta(q)$ can be solved together with the optimal stopping belief $\widehat{x}_\theta(q)$. At the time of stopping, the standard optimality conditions *value matching* and *smooth pasting* must hold:

$$\widehat{x}_\theta v_H(\theta) + (1-\widehat{x}_\theta)v_L(\theta) = B_\theta(q)\Phi(\widehat{x}_\theta, q), \tag{15}$$

$$v_H(\theta) - v_L(\theta) = B_\theta(q)\Phi_x(\widehat{x}_\theta, q). \tag{16}$$

Solving the pair of equations yields the following cutoff belief:

$$\widehat{x}_\theta(q) = \frac{\beta(q)v_L(\theta)}{\beta(q)v_L(\theta) + (1-\beta(q))v_H(\theta)}.$$

When $q_t = q$ forever, it is optimal for θ to stop if and only if $x_t \in [\widehat{x}_\theta(q), 1]$. Note that for all $\theta \in \left[\underline{\theta}, \overline{\theta}\right]$, $\widehat{x}_\theta(q)$ is continuous and strictly increasing in q. The relationship between x^E in Proposition 1 and $\widehat{x}_\theta$ satisfies $\widehat{x}_\theta(q) > x^E(q)$ for $q < q(\theta)$, $\widehat{x}_\theta(q) = x^E(q)$ for $q = q(\theta)$, and $\widehat{x}_\theta(q) < x^E(q)$ for $q > q(\theta)$. See Figure 7.

We now start with the uniqueness part of Proposition 1. Assume that Q is a decentralized equilibrium. By Lemma 1 the optimized stopping times are

[13] We have discarded the other root of the proposed solution as we must have that the value converges to the static solution as $x \to 0$ and $x \to 1$.

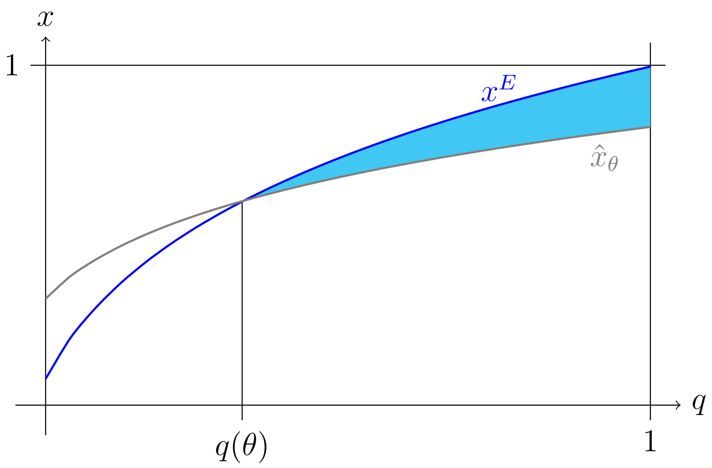

Figure 7: Optimal stopping for type θ.

monotone in θ. Hence, in equilibrium, θ cannot stop before q_t reaches $q(\theta)$, nor can she delay stopping until q_t has reached a level higher than $q(\theta)$. Take some history h_t such that $q_t = q(\theta)$ and denote by $F_\theta(h_t)$ the equilibrium continuation value of type θ. It follows that the optimized continuation value for θ at $q(\theta)$ cannot exceed the continuation value of the constrained problem where q_t stays fixed at q forever, i.e. we have $F_\theta(h_t) = \overline{F}_\theta(x; q(\theta))$. Any policy other than stopping at threshold $\widehat{x}_\theta(q(\theta)) = x^E(q(\theta))$ would give $F_\theta(h_t) < \overline{F}_\theta(x; q)$ and so stopping at the threshold is a necessary condition in a decentralized equilibrium. Since this holds for all types θ, we can conclude that if a decentralized equilibrium exists, then it must be the cutoff policy in Proposition 1.

Next, we show that the policy in Proposition 1 is indeed a decentralized equilibrium. Fix policy Q to be the cutoff policy in Proposition 1, and consider optimal stopping for θ against it. Note that Q defines a feasible region $\mathbf{X}$ in (x, q)-space such that $(x_t, q_t) \in \mathbf{X}$ for all $t > 0$: $\mathbf{X} = \left\{(x, q) : 0 \le q \le 1,\, 0 < x \le x^E(q)\right\}$. Since Q is a Markovian process, an optimal stopping rule against it can be expressed as some stopping region $S_\theta \subseteq \mathbf{X}$ in the feasible region. Denote by $F_\theta(x, q)$ the value function under optimally chosen stopping region S_θ:

$$F_\theta(x, q) = \mathbb{E}\left(e^{-r\tau(S_\theta)} u_\theta\left(x_{\tau(S_\theta)}\right) \middle| x, q\right),$$

where $\tau(S_\theta) = \inf(t : (x_t, q_t) \in S_\theta)$ is the time of hitting S_θ and $u_\theta(x) := x v_H(\theta) + (1 - x) v_L(\theta)$ is the stopping value. Clearly, $F_\theta(x, q) = u_\theta(x)$ whenever $(x_t, q_t) \in$

S_θ. At and above the boundary $x^E(q)$, where q_t increases, we must have:

$$\frac{\partial}{\partial q}\left[F_\theta(x,q)\right]_{x\geq x^E(q)}=0. \tag{17}$$

The plan for the rest of the proof is to show that the optimal stopping region S_θ is the shaded region in Figure 7, i.e.

$$S_\theta=\left\{(x,q):q\geq q(\theta),\,x\in\left[\widehat{x}_\theta(q),x^E(q)\right]\right\}. \tag{18}$$

As a first step, we note that it can never be optimal to stop for $q<q(\theta)$. This follows from the observation that for $q<q(\theta)$ any $(x,q)\in\mathbf{X}$ satisfies $x<\widehat{x}_\theta(q)$, i.e. it is not optimal to stop even assuming q_t to stay fixed forever. The continuation value in the equilibrium where q_t may increase in the future must be at least as high as the continuation value in the problem where q_t stays fixed forever. Hence, $\overline{F}_\theta(x;q)>u_\theta(x,q)$ implies $F_\theta(x,q)>u_\theta(x,q)$ for all $x<\widehat{x}_\theta(q)$ and so it cannot be optimal to stop.

As a second step, we will show that when $q\geq q(\theta)$, it is always optimal to stop at the boundary of $\mathbf{X}$, i.e. at $x=x^E(q)$. Suppose, to the contrary, that there is some $(x,q)\notin S_\theta$, where $x=x^E(q)$ and $q\geq q(\theta)$. This amounts to assuming that $F_\theta(x^E(q),q)>u_\theta(x^E(q))$, and we will show that this leads to a contradiction. First, suppose that it is optimal to stop at some (x',q), where $x'<x^E(q)$. In that case $F_\theta(x',q)=u_\theta(x')$. Since $F_\theta(x,q)$ is convex in x and we necessarily have $F_\theta(x,q)\geq u_\theta(x)$ for all $x\in(x',x^E(q))$, we get that

$$\frac{\partial}{\partial x}\left[F_\theta(x,q)\right]_{x=x^E(q)}\geq v_H(\theta)-v_L(\theta)=\frac{\partial}{\partial x}\left[u_\theta(x)\right]_{x=x^E(q)} \tag{19}$$

Now, suppose that it is optimal to wait for all (x,q), where $x<x^E(q)$, in which case $F_\theta(x,q)>u_\theta(x)$ for all $x<x^E(q)$ and $F_\theta(0,q)=0$. In that case, following the same argument as with the value function $\overline{F}_\theta(x;q)$ for fixed q, the value function $F_\theta(x,q)$ must take the form $F_\theta(x,q)=A_\theta(q)\,\Phi(x,q)$ for some function $A_\theta(q)$ and hence $\frac{\partial}{\partial x}\left[F_\theta(x,q)\right]_{x=x^E(q)}=A_\theta(q)\,\Phi_x\left(x^E(q),q\right)$. Our assumption $F_\theta(x^E(q),q)>u_\theta(x^E(q))$ is equivalent to

$$A_\theta(q)\,\Phi\left(x^E(q),q\right)>x^E(q)\,v_H(\theta)+\left(1-x^E(q)\right)v_L(\theta),$$

which further implies

$$\begin{aligned}\frac{\partial}{\partial x}\left[F_\theta(x,q)\right]_{x=x^E(q)} &> \frac{\Phi_x(x(q),q)}{\Phi(x(q),q)}\left[x^E(q)\,v_H(\theta)+\left(1-x^E(q)\right)v_L(\theta)\right]\\ &= \frac{\beta(q)-x^E(q)}{\left(1-x^E(q)\right)}v_H(\theta)+\frac{\beta(q)-x^E(q)}{x^E(q)}v_L(\theta).\end{aligned}$$

The last expression is greater than $v_H(\theta)-v_L(\theta)$ if and only if

$$x'\geq\frac{\beta(q)\,v_L(\theta)}{\beta(q)\,v_L(\theta)+(1-\beta(q))\,v_H(\theta)}=\widehat{x}_\theta(q),$$

which is the case if and only if $q\geq q(\theta)$. Therefore, we may conclude that (19) holds in this case too.

Given that (19) holds, the rate of change in $F_\theta(x,q)$ along the boundary is

$$\begin{aligned}&\frac{d}{dq}F_\theta\left(x^E(q),q\right)=\frac{\partial}{\partial x}\left[F_\theta(x,q)\right]_{x=x^E(q)}\frac{d}{dq}x^E(q)+\frac{\partial}{\partial q}\left[F_\theta(x,q)\right]_{x=x^E(q)}\\ &=\frac{\partial}{\partial x}\left[F_\theta(x,q)\right]_{x=x^E(q)}\frac{d}{dq}x^E(q)\geq\left[v_H(\theta)-v_L(\theta)\right]\frac{d}{dq}x^E(q)=\frac{d}{dq}u_\theta\left(x^E(q)\right),\end{aligned}$$

where the second term of the first line disappears by (17) and where the inequality is implied by (19).

But if $F_\theta(x^E(q),q)>u_\theta(x^E(q))$ implies $\frac{d}{dq}F_\theta\left(x^E(q),q\right)\geq\frac{d}{dq}u_\theta\left(x^E(q)\right)$, then it must further imply $F_\theta(x^E(q'),q')>u_\theta(x^E(q'))$ for all $q'\in[q,1]$, and in particular $F_\theta(x^E(1),1)>u_\theta(x^E(1))$. But $x^E(1)=1$, so this yields $F_\theta(x^E(1),1)>v_H(\theta)$, which is a contradiction since the value cannot be higher than the stopping payoff under certainty of state $\omega=H$.

We conclude that it is optimal to stop at all boundary points for $q>q(\theta)$. To see that this implies that it is also optimal to stop within the whole shaded region in Figure 7, i.e. $\left\{(x,q):q\geq q(\theta),x\in\left[\widehat{x}_\theta(q),x^E(q)\right]\right\}\in S_\theta$, note that q_t can only increase if x_t reaches $x^E(q)$. Since θ stops at latest when x_t reaches $x^E(q)$, the optimal continuation value $F_\theta(x,q)$ cannot exceed the corresponding value with q fixed, i.e. $\overline{F}_\theta(x;q)$. To reach that value, θ must stop at all points $\left[\widehat{x}_\theta(q),x^E(q)\right]$.

We have now shown that the stopping rule defined in (18) maximizes (3) for policy Q. Since q_t can only increase at the boundary points $x^E(q)$, the first point in S_θ ever reached is $(\widehat{x}_\theta(q(\theta)),q(\theta))$ and so the optimal stopping rule commands θ to stop exactly when q_t reaches $1-F(\theta)$ and is therefore consistent with Q. We can conclude that Q is a decentralized equilibrium. □

C Socially optimal policy

We use the derivatives of $\Phi(x,q)$ in many proofs of this section:

$$\begin{aligned}
\Phi =& \left(\frac{x}{1-x}\right)^{\beta(q)}(1-x), \Phi_q = \Phi\beta'(q)\ln\left(\frac{x}{1-x}\right),\\
\Phi_x =& \Phi\frac{(\beta(q)-x)}{x(1-x)}, \Phi_{xx} = \Phi\beta(q)\frac{(\beta(q)-1)}{x^2(1-x)^2},\\
\Phi_{qx} =& \Phi\beta'(q)x^{-1}(1-x)^{-1}\left[1+(\beta(q)-x)\ln\left(\frac{x}{1-x}\right)\right],\\
\Phi_{xxq} =& \Phi\frac{\beta'(q)}{x^2(1-x^2)}\left[\beta(q)+(\beta(q)-1)(1+\beta(q)\ln\left(\frac{x}{1-x}\right))\right].
\end{aligned}$$

Deriving the differential equation

We first show that the value matching and smooth pasting conditions, (7) and (8), imply the differential equation in (9). Solving (7) and (8) for $B_q(q)$ and $B(q)$ yields

$$\begin{aligned}
B_q(q) &= A^1(x^*(q),q)\,x^*(q) + A^2(x^*(q),q)\,, && (20)\\
B(q) &= U^1(x^*(q),q)\,x^*(q) + U^2(x^*(q),q)\,, && (21)
\end{aligned}$$

where

$$\begin{aligned}
A^1(x,q) &:= \frac{-\Phi_{qx}(x,q)\,(v_H(q)-v_L(q))}{\Phi(x,q)\,\Phi_{qx}(x,q)-\Phi_q(x,q)\,\Phi_x(x,q)},\\
A^2(x,q) &:= \frac{\Phi_{qx}(x,q)(-v_L(q))+\Phi_q(x,q)\,(v_H(q)-v_L(q))}{\Phi(x,q)\,\Phi_{qx}(x,q)-\Phi_q(x,q)\,\Phi_x(x,q)},\\
U^1(x,q) &:= \frac{\Phi_x(x,q)\,(v_H(q)-v_L)}{\Phi(x,q)\,\Phi_{qx}(x,q)-\Phi_q(x,q)\,\Phi_x(x,q)},\\
U^2(x,q) &:= \frac{-\Phi_x(x,q)\,(-v_L(q))-\Phi(x,q)\,(v_H(q)-v_L(q))}{\Phi(x,q)\,\Phi_{qx}(x,q)-\Phi_q(x,q)\,\Phi_x(x,q)}.
\end{aligned}$$

Differentiating (21) with respect to q and using the chain rule gives

$$\begin{aligned}
B_q(q) = &\left[U^1_x(x^*(q),q)\,x^{*\prime}(q) + U^1_q(x^*(q),q)\right]x^*(q) + U^1(x^*(q),q)\,x^{*\prime}(q)\\
&+ U^2_x(x^*(q),q)\,x^{*\prime}(q) + U^2_q(x^*(q),q) \qquad (22)
\end{aligned}$$

Equating (20) and (22), solving for $x^{*\prime}(q)$, and simplifying yields the expression (9) in the text.

Any solution that satisfies the differential equation (9) must be continuous.

Proof of Proposition 2

The proof contains three parts. In part 1, we show that the initial value problem (9) has a unique solution $x^*(q)$ with the property $x^*(q) < x^E(q)$ for all $q < 1$. We also show that $x^*(q)$ is continuous and strictly increasing and hence defines a boundary policy. In part 2, we show that our candidate policy $x^*(q)$ satisfies the HJB equation (5). In part 3, we verify that the solution to the HJB equation solves the original problem.

Part 1: solution to the initial value problem (9)

We first establish some key properties of function g in (9):

Lemma 4. *For all (x, q) such that $q < 1$ and $x \leq x^E(q)$, function $g(x, q)$ in* (9) *is strictly positive and strictly increasing in x and it is Lipschitz continuous for all $q \in [0, q_1]$ if $q_1 < 1$ and for all $x \leq x^E(q)$. Furthermore, $g(x^E(q), q) > x^{E'}(q)$ for $q < 1$ and $\lim_{q\to 1} g(x^E(q), q) = x^{E'}(1)$.*

See Online Appendix for the proof.

The singularity at (1,1) prevents us from directly applying the Picard-Lindelöf theorem to show the existence and uniqueness of a solution to the initial value problem (9). Instead, we note that the requirements for the Picard-Lindelöf theorem are satisfied for all initial conditions $x(q_1) = x_1$ where $x(q_1) \leq x^E(q_1)$ and $q_1 < 1$, and hence each such initial value problem defines a unique solution. Since g is increasing in x, these solutions diverge when approaching $(1, 1)$ and hence at most one path can approach $(1, 1)$ from below the decentralized policy. The fact that $\lim_{q\to 1} g(x^E(q), q) = x^{E'}(1)$ implies that there is a path that approaches $(1, 1)$ from the same direction as the decentralized policy $x^E(q)$ and the fact that $g(x^E(q), q) > x^{E'}(q)$ for $q < 1$ implies that such a path must be strictly below the decentralized solution for all $q < 1$. It follows that the initial value problem has a unique solution below the decentralized solution.

We have now shown that the initial value problem (9) has a unique solution x^* such that $x^*(q) \leq x^E(q)$ for all $x \leq q$. This solution $x^*(q)$ is continuous and strictly increasing in q, and it is our candidate policy.

Part 2: our candidate x^* solves the HJB equation

Fix $x^*(q)$ to be the candidate policy defined in the proposition and let $q^*(x)$ be its inverse with the convention $q^*(x) = 0$ for $x \leq x^*(0)$. Its associated value function is

$$V(x,q) = \begin{cases} \int_q^{q^*(x)} (xv_H(s) + (1-x)v_L(s))\,ds + V(x, q^*(x)), \text{ for } q < q^*(q) \\ B(q)\Phi(x,q), \text{ for } q \geq q^*(x), \end{cases} \tag{23}$$

where $B(q)$ is given by (21). By construction, for $q \geq q^*(x)$, $V(x,q)$ satisfies

$$rV(x,q) = \frac{1}{2}V_{xx}(x,q)\frac{x^2(1-x)^2}{\sigma^2}q \tag{24}$$

and at the boundary $q = q^*(x)$, the value matching and smooth pasting conditions (7) and (8) hold:

$$V_q(x, q^*(x)) + xv_H(q^*(x)) + (1-x)v_L(q^*(x)) = 0, \tag{25}$$

$$V_{qx}(x, q^*(x)) + v_H(q^*(x)) - v_L(q^*(x)) = 0. \tag{26}$$

Differentiating (24) with respect to q, we have

$$rV_q(x,q) = \frac{1}{2}V_{xx}(x,q)\frac{x^2(1-x)^2}{\sigma^2} + \frac{1}{2}V_{xxq}(x,q)\frac{x^2(1-x)^2}{\sigma^2}q, \tag{27}$$

which allows us to re-write (25) as:

$$\begin{aligned} r\left[xv_H(q^*(x)) + (1-x)v_L(q^*(x))\right] + \frac{1}{2}V_{xx}(x,q^*(x))\frac{x^2(1-x)^2}{\sigma^2} \\ + \frac{1}{2}V_{xxq}(x,q^*(x))\frac{x^2(1-x)^2}{\sigma^2}q = 0. \end{aligned} \tag{28}$$

We next state three lemmas that concern the partials of the value function below, above, and at the boundary, respectively. All proofs are in Online Appendix.

Lemma 5. *For $q \geq q^*(x)$, we have $V_q(x,q) + xv_H(q) + (1-x)v_L(q) \leq 0$.*

Lemma 6. *For $q < q^*(x)$, we have $V_{xx}(x,q) = V_{xx}(x,q^*(x))$, $V_{qq}(x,q) = V_{qq}(x,q^*(x))$, and $V_{xxq}(x,q) = 0$.*

Lemma 7. *For $q = q^*(x)$, we have $V_{xxq}(x,q) < 0$.*

We are now ready to show that our candidate policy satisfies the HJB-equation (5), which we re-write here using notation q' instead of q^* for the maximizer (this is to avoid confusion with boundary $q^*(x)$):

$$rV(x,q) = \max_{q'>q}\left(r\int_q^{q'}(xv_H(s) + (1-x)v_L(s))\,ds + \frac{1}{2}V_{xx}(x,q')\frac{x^2(1-x)^2}{\sigma^2}q'\right). \tag{29}$$

The right-hand side of this equation is a continuous function in q' and its derivative with respect to q' is

$$r(xv_H(q') + (1-x)v_L(q')) + \frac{1}{2}V_{xx}(x,q')\frac{x^2(1-x)^2}{\sigma^2} + \frac{1}{2}V_{xxq}(x,q')\frac{x^2(1-x)^2}{\sigma^2}q'. \tag{30}$$

Let us inspect the sign of this for different values of q'. For $q' \geq q^*(x)$, we can use (27) to write (30) as

$$r(xv_H(q') + (1-x)v_L(q')) + rV_q(x,q),$$

which is negative by lemma 5. It follows that whenever $q \geq q^*(x)$, the right-hand side of (29) is maximized by choosing $q' = q$, i.e. keeping q fixed.

For $q' < q^*(x)$, we can use lemma 6 to write (30) as

$$r(xv_H(q') + (1-x)v_L(q')) + \frac{1}{2}V_{xx}(x,q^*(x))\frac{x^2(1-x)^2}{\sigma^2},$$

which is decreasing in q'. Moreover, combining (28) and Lemma 7 we can conclude that it is positive in the limit $q' \to q^*(x)$, and hence it is positive for all $q' < q^*(x)$. Since the right-hand side of (29) is continuous, and its derivative is positive (negative) for $q' < q^*(x)$ ($q' \geq q^*(x)$), it is maximized at $q' = q^*(x)$ if $q < q^*(x)$.

We have now shown that for any $x \in (0,1)$, the right-hand side of the HJB equation is maximized by choosing $q' = \max\{q, q^*(x)\}$. Furthermore, since $V(x,q)$ satisfies (24) for $q \geq q^*(x)$, the left- and right-hand sides of (29) coincide with this choice of q'. Hence, we have shown that $V(x,q)$ defined in (23) satisfies the HJB-equation.

Part 3: verification

The verification of the solution follows from the standard arguments in the literature (see e.g. Fleming and Soner (2006)). Let V^* solve the HJB equation (5) and let $q^*(x,q) = \max\{q, q^*(x)\}$ be the corresponding q^*. Then, let $T \geq t$ be the time at which the game ends (we will take the limit as T goes to infinity). From generalized Itô's formula we have[14]

$$
\begin{aligned}
e^{-rT}V^*(x_T, q_T) &= e^{-rt}V^*(x_t, q_t) - \int_t^T e^{-rs} rV^*(x_s, q_s)ds + \int_t^T e^{-rs}V_x^*(x_s, q_s)dx_s \\
&+ \int_t^T e^{-rs}V_q^*(x_s, q_s)dq_s + \frac{1}{2}\int_t^T e^{-rs}V_{xx}^*(x_s, q_s)d[x]_s + \frac{1}{2}\int_t^T e^{-rs}V_{qq}^*(x_s, q_s)d[q]_s \\
&+ \int_t^T e^{-rs}V_{qs}^*(x_s, q_s)d[q,x]_s
\end{aligned}
$$

where $d[x]_t$ and $d[q]_t$ are the quadratic variations of x and q and $d[x,y]_t$ is their quadratic covariation. The process Q_t has bounded variation and hence $d[q]_t = d[x,y]_t = 0$. Notice also that $dx_t = x_t(1-x_t)\sigma^{-1}\sqrt{q_t}dw_t$ and $d[x]_t = x_t^2(1-x_t)^2\sigma^{-2}q_t dt$. We can further simplify the equation by noting that $V_q^* dq = -(xv_H(q) + (1-x)v_L(q))dq$. The HJB equation gives an upper bound for $\frac{q_s}{\sigma^2}x_s^2(1-x_s)^2V_{xx}^*(x_s, q_s) - rV^*(x_s, q_s) \leq \int_{q_s}^{q^*(x_s,q_s)}(xv_H(q) + (1-x)v_L(q))dq$, which equals zero for almost all s. Combining gives:

$$
\begin{aligned}
e^{-rT}V^*(x_T, q_T) &\leq e^{-rt}V^*(x_t, q_t) - \int_t^T e^{-rs}(x_s v_H(q_s) + (1-x_s)v_L(q_s)))dq_s \\
&+ \int_t^T e^{-rs}V_x{}^*(x_s, q_s)\frac{\sqrt{q_s}}{\sigma}x_s(1-x_s)dw_s. \qquad (31)
\end{aligned}
$$

Because the game ends at time T, $V^*(x_T, q_T) = 0$. Taking conditional expectations , multiplying by $-e^{rt}$ and simplifying then gives

$$
V^*(x_t, q_t) \geq \mathbb{E}\left[\int_t^T e^{-r(t-s)}(x_s\pi_H(q_s) + (1-x_s)\pi_L(q_s))ds|\mathcal{F}_t\right]. \qquad (32)
$$

The value is bounded and therefore clearly satisfies the transversality condition: $\lim_{T\to\infty}\mathbb{E}[e^{-rT}V^*(x_T, q_T)] = 0$. Hence, the limit of the right-hand side of (31) is well defined as $T \to \infty$. Therefore, we have that $V^*(x,q) \geq \max_Q U(Q; x, q)$ even as $T \to \infty$.

[14]To see that $V \in C^2$ check V_x at the boundary. The continuity of V_{xx} and V_{qq} follows from Lemma 6 and the continuity of V_q and V_{qx} are implied by the value matching and smooth pasting conditions.

The last step is to use the fact that Q, induced by policy x^*, achieves the pointwise maximum of the HJB-equation and thus the inequalities above become equalities: $V^*(x,q) = \max_Q U(Q;x,q)$. Our solution solves the original problem.

Proof of Proposition 3

Proof. First, we can use that we have proved that $x^*(0) < x^E(0) = x^{stat}(0)$ in the proof of Proposition 2 together with the continuity of the policy functions to get that there exists $\underline{q} > 0$ such that $x^{stat}(q) > x^*(q)$ for all $q < \underline{q}$. As the policy functions are strictly increasing and continuous, the stocks $q^*(x)$ and $q^{stat}(x)$ are pinned down as the inverse of the policy functions for all $x \geq x^*(0)$ and $x \geq x^{stat}(0)$ respectively. In addition, $q^*(x) = 0$ for all $x \leq x^*(0)$ and $q^{stat}(x) = 0$ for all $x \leq x^{stat}(0)$, and hence q^* and q^{stat} are continuous.

Let $\underline{x} := x^{stat}(\underline{q}) > x^{stat}(0)$ where the inequality follows from x^{stat} being strictly increasing. Then, $q^{stat}(x) < q^*(x)$ for all $x \in [x^{stat}(0), \underline{x})$ by that q^* and q^{stat} are the inverse functions of x^* and x^{stat}. Furthermore, $q^{stat}(x) = 0 < q^*(x)$ for all $x \in [x^*(0), x^{stat}(0)]$, which completes the proof.

Next, we show the other direction by showing that $x^*(1) = x^{stat}(1) = 1$ and $x_q^*(1) < x_q^{stat}(1)$. The first part is immediate. For the second part, use Lemma 4 and the uniqueness of the solution to get $x_q^*(1) = x_q^E(1)$. Now it is enough to show that the derivative of the equilibrium is smaller than of the static solution:

$$x^E{}_q(1) - x^{stat}{}_q(1) = \frac{(\beta-1)\beta v_L v_H'}{(\beta v_L)^2} - \frac{v_L v_H'}{(v_L)^2} = -\frac{v_L v_H'}{\beta v_L^2} < 0.$$

The static and optimal solutions meet at $q = 1$ but the optimal solution reaches the point above the static solutions. Hence, by continuity there must exist $\overline{q} < 1$ such that $x^*(q) > x^{stat}(q)$ for all $q \in (\overline{q}, 1)$, which then further implies the existence of $\overline{x} < 1$ by the same argument as used above for $\underline{x}$. $\square$

References

Bergemann, Dirk and Juuso Välimäki, "Market diffusion with two-sided learning," *RAND Journal of Economics*, 1997, *28* (4), 773–795.

_ **and** _ , "Experimentation in markets," *Review of Economic Studies*, 2000, *67* (2), 213–234.

Board, Simon, "Selling options," *Journal of Economic Theory*, 2007, *136* (1), 324–340.

_ **and Andrzej Skrzypacz**, "Revenue management with forward-looking buyers," *Journal of Political Economy*, 2016, *124* (4), 1046–1087.

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

Bonatti, Alessandro, "Menu pricing and learning," *American Economic Journal: Microeconomics*, 2011, *3* (3), 124–163.

Chamley, Christophe and Douglas Gale, "Information Revelation and Strategic Delay in a Model of Investment," *Econometrica*, 1994, *62* (5), 1065–1085.

Che, Yeon-Koo and Johannes Hörner, "Recommender Systems as Mechanisms for Social Learning," *The Quarterly Journal of Economics*, 2017, pp. 871–925.

Dixit, Avinash, "Entry and Exit Decisions under Uncertainty," *Journal of Political Economy*, 1989, *97* (3), 620–638.

Dixit, Avinash K. and Robert S. Pindyck, *Investment under Uncertainty*, Princeton: Princeton University Press, 1994.

Fleming, Wendell H and Halil Mete Soner, *Controlled Markov processes and viscosity solutions*, Vol. 25, Springer Science & Business Media, 2006.

Frick, Mira and Yuhta Ishii, "Innovation Adoption by Forward-Looking Social Learners," *Working Paper*, 2020.

Gittins, J. and D. Jones, "A Dynamic Allocation Index for the Sequential Allocation of Experiments," *Progress in Statistics, ed. by J. Gani*, 1974, pp. 241––266.

Grossman, Sanford J., Richard E. Kihlstrom, and Leonard J. Mirman, "A Bayesian Approach to the Production of Information and Learning by Doing," *The Review of Economic Studies*, 1977, *44* (3), 533–547.

Keller, Godfrey and Sven Rady, "Strategic experimentation with Poisson bandits," *Theoretical Economics*, 2010, *5* (2), 275–311.

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

Kruse, Thomas and Philipp Strack, "Optimal stopping with private information," *Journal of Economic Theory*, 2015, *159*, 702–727.

Laiho, Tuomas and Julia Salmi, "Coasian Dynamics and Learning," *Working paper*, 2020.

Leahy, John V, "Investment in competitive equilibrium: The optimality of myopic behavior," *The Quarterly Journal of Economics*, 1993, *108* (4), 1105–1133.

Liski, Matti and Francois Salanié, "Catastrophes, delays, and learning," *Toulouse School of Economics Working Papers No 1148*, 2020.

Martimort, David and Louise Guillouet, "Precaution, information and time-inconsistency: on the value of the precautionary principle," *CEPR Discussion Paper Series DP15266*, 2020.

McDonald, Robert and Daniel Siegel, "The value of waiting to invest," *The quarterly journal of economics*, 1986, *101* (4), 707–727.

Milgrom, Paul and Ilya Segal, "Envelope theorems for arbitrary choice sets," *Econometrica*, 2002, *70* (2), 583–601.

Moscarini, Giuseppe and Lones Smith, "The Optimal Level of Experimentation," *Econometrica*, 2001, *69* (6), 1629–1644.

Murto, Pauli and Juuso Välimäki, "Learning and Information Aggregation in an Exit Game," *The Review of Economic Studies*, 2011.

Pindyck, Robert S., "Irreversible Investment, Capacity Choice, and the Value of the Firm," *The American Economic Review*, 1988, *78* (5), 969–985.

Prescott, Edward C., "The Multi-Period Control Problem Under Uncertainty," *Econometrica*, 1972, *40* (6), 1043–1058.

Rob, Rafael, "Learning and Capacity Expansion under Demand Uncertainty," *The Review of Economic Studies*, 1991, *58* (4), 655–675.

Rothschild, Michael, "A two-armed bandit theory of market pricing," *Journal of Economic Theory*, 1974, *9* (2), 185–202.