Gradual Learning from Incremental Actions

Tuomas Laiho, Pauli Murto, Julia Salmi*

June 29, 2023

Abstract

We introduce a collective experimentation problem where agents choose the timing of irreversible actions under uncertainty and where public feedback from the actions arrives gradually over time. This kind of gradual learning where information is delayed and arrives over time is present in many real-life situations, such as adoption of new technologies, progressive market entry, and incremental roll-out of public policies. The socially optimal expansion path entails an informational trade-off where acting today speeds up learning but postponing capitalizes on the option value of waiting. We contrast the social optimum to the decentralized equilibrium where agents ignore the social value of information they generate. We show that the equilibrium can be obtained by assuming that agents treat the stock of past actions as constant, which yields a tractable closed-form solution to a complicated multi-dimensional problem.

JEL classification: C61, C73, D82, D83

Keywords: social learning, experimentation, optimal stopping, dynamic markets, innovation adoption

*Laiho: Ministry of Finance, Finland (tuomas.s.laiho@gmail.com), Murto: Aalto University School of Business (pauli.murto@aalto.fi, `http://www.aalto-econ.fi/murto/`), Salmi: University of Copenhagen (julia.salmi@econ.ku.dk, `https://sites.google.com/view/juliasalmi`).
We thank Aislinn Bohren, Martin Cripps, Rahul Deb, Francesc Dilme, Nils Christian Framstad, Dino Gerardi, Yingni Guo, 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), Academy of Finland (Murto), and from Emil Aaltonen Foundation, OP Group Research Foundation, Yrjö Jahnsson Foundation, and ERC grant no. 682417 through PI Vasiliki Skreta (Salmi).

1 Introduction

Many actions have long-run consequences that can be observed only gradually over time. As a concrete example, consider a decision to purchase the latest model of an electric car. When using the car after purchase, a new owner observes how well the car functions in different situations and whether any technical problems emerge. When more consumers buy the new model – the fleet of existing cars expands – learning on the aggregate level becomes more precise but remains gradual because each owner continues to make observations long after the buying decision. Similarly, the profitability of capital investments, environmental damage caused by pollution, or effects of public policies are realized gradually over time, potentially long after the critical decisions were made. While such situations abound, we lack a flexible modeling framework to address the welfare implications of this kind of endogenous gradual learning.

This paper introduces a novel learning problem to address this. The key feature of the model is that an irreversible 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, make an irreversible investment, or purchase a durable good. An unknown binary state determines if stopping is profitable for the agents. Crucially, learning is *gradual*: upon stopping, each agent initiates a persistent flow of information over time. This is in contrast to the standard experimentation models where an action generates an *instantaneous* one-time signal and further actions are needed to learn more.

Our main question is how the incremental path of stopping decisions – the expansion path – is determined on one hand when agents optimize individually and on the other hand when a central planner chooses the actions. We also ask how a planner can coordinate the actions of individual agents with a posted price mechanism. The contribution of this paper is twofold. First, we develop a novel methodological approach with suitable solution techniques. Second, we derive new economic insights that are caused by endogenous gradual learning.

Gradual learning creates a new trade-off for the socially optimal expansions:

the information generation effect calls for aggressive expansion in order to improve information for future decisions and *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. This informational trade-off does not arise when learning is instantaneous because there is then no social gain from waiting for more information.[1]

We approach experimentation under gradual learning by modeling 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, this leads to an aggregate signal that follows a Brownian motion with an unknown drift, determined by the true state, and a signal-to-noise ratio proportional to the stock of agents who have stopped. Each stopping decision thus affects information generation gradually over time.

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 that the state is high and the stock affect the future. Furthermore, the stock and the belief processes are interlinked as the stock determines the flow of new information. 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. The shortsighted approach works because information arrives smoothly over time under gradual learning.

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. Because of the information generation effect, socially optimal policy favors earlier and more aggressive expansions than what happens in the decentralized equilibrium. The difference between the two is especially pronounced when

[1]Papers studying instantaneous learning include Bonatti (2011), Che and Hörner (2017), Frick and Ishii (2020), and Laiho and Salmi (2023).

the learning technology is good and learning could potentially be fast. Compared to the no-learning benchmark, gradual learning tends to increase the socially optimal stock for low beliefs and to decrease it for high beliefs due to the informational trade-off between information generation and the option value of waiting.

Due to its technical tractability, the model with gradual learning has a potential to work as a workhorse model that can be further extended to analyze other phenomena. We provide one such extension in the companion paper, Laiho, Murto and Salmi (2023), by including the possibility that agents' payoffs depend directly on the other agents' stopping behavior.

In the last part of the paper, we show how to incorporate mechanism design techniques to implement the social optimum, or any other policy, as a decentralized equilibrium. We then demonstrate our theoretical results by solving the pricing problem of a monopolist selling a new durable good under social learning. We point out the potential benefit of market power to consumers: a large player internalizes the informational externality partly to the benefit of the consumers and this may result in a higher consumer welfare than in a competitive market equilibrium. The example highlights the quality of the learning process as an important determinant of welfare implications. An improvement in the learning technology – for example improved social learning caused by better communication technologies – amplifies the welfare loss of the decentralized equilibrium and hence puts more weight on the identified positive impact of market power. More generally, a better learning technology strengthens the case for policies to correct informational distortions highlighted by our model.

1.1 Related literature

Using the framework of our paper, the previous literature on learning can be organized 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 interaction of these effects.

The information generation effect is present 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. No option value effect exists in these papers 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). One can see our solution to the social planner's problem as extending the real options literature to endogenous learning.

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. Because feedback from past actions is instantaneous, endogenous learning does not create an option value effect for the social planner, and hence no informational trade-off on the social level, unlike in the present paper with gradual learning. In equilibrium there is no information generation effect, but free-riding on the information generated by others creates an option value effect resulting in inefficiently low rate of innovations. 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 (2023) analyze monopoly pricing in a similar setup. Both in Che and Hörner (2017) and in Laiho and Salmi (2023), the presence of a social planner or a monopolist induces learning to increase quantities. The crucial difference from the present paper is that the 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. More generally, there is no informational trade-off under instantaneous learning independent of whether actions are reversible or irreversible.

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 trade-offs 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.

Finally, our other paper Laiho, Murto and Salmi (2023) builds on the modeling techniques developed in this paper to analyze the joint effect of informational and payoff externalities. Laiho, Murto and Salmi (2023) shows that agents can be strictly better off and learning can be faster in an equilibrium under a worse learning technology if there are positive payoff externalities.

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: $v_H(\theta) \geq 0 > v_L(\theta)$.[2] Payoffs are continuously differentiable with bounded derivatives and increasing in type: for each $\theta \in \Theta$, $v'_\omega(\theta) \geq 0$ for both $\omega \in \{H, L\}$ and $v'_\omega(\theta) > 0$ 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:[3]

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

where we normalize $\mu_H = 1/2$ and $\mu_L = -1/2$, $\sigma > 0$ is the standard deviation of the process, and 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

[2]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.

[3]The process is otherwise equivalent to the learning processes in Bolton and Harris (1999) and in Moscarini and Smith (2001) but learning is from the stock of cumulative actions instead of being from the flow of new actions.

independent noisy signals over time and where the total informativeness per unit of q is normalized to stay constant. The signals can be for example interpreted as realized individual payoffs (see Appendix A).[4]

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:

$$dx_t = \frac{\sqrt{q_t}}{\sigma} x_t(1 - x_t) d\tilde{w}_t, \tag{2}$$

where $\tilde{w}_t$ is a standard Wiener process. 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 Online Appendix,[5] 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.

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

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

[4]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.

[5]Available at our home pages.

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. In a decentralized equilibrium agents optimize individually taking the policy as given:

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 $Q = Q^E$.*

The socially optimal policy maximizes the expected total welfare:

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)*

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

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

In section 3.5 we recast this as a control problem for the stock process $\{q_t\}$.

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, one can implement a policy Q by many different stopping profiles. However, because the stopping payoffs are increasing in θ, in equilibrium 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 $\theta > \theta'$*.*

Also socially optimal stopping order is monotone:

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. The lemmas mean that it is without loss of generality to restrict attention to monotone stopping profiles for which there is a one-to-one mapping between the stock q_t and the largest type θ_t who has not stopped: $q_t = 1 - F(\theta_t)$. Throughout the paper we use notation $q(\theta) := 1 - F(\theta)$ to denote the stock as a function of the current highest type, which has an inverse (current highest type): $\theta(q) := \{\theta : 1 - F(\theta) = q\}$. With a 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

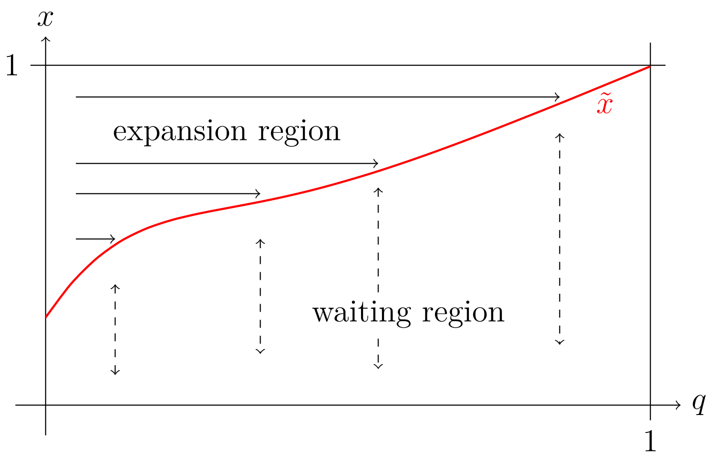

Figure 1: Dynamics in the waiting and expansion regions of the state space.

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.

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 myopic. 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 myopic cutoff:

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

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 $\theta(\hat{q})$ assuming that q_t is constant and equal to $\hat{q}$. This one-dimensional stopping problem can be solved using standard techniques in the literature (see e.g. Dixit and Pindyck (1994) and the team problem in Bolton and Harris (1999)). We show that the equilibrium in the original problem is obtained by solving this "shortsighted" problem separately for each individual type and tying these solutions together.[6]

[6] Our method to solve the decentralized equilibrium is inspired by a model of industry level investments by Leahy (1993) who shows that under exogenous uncertainty the competitive equi-

The intuition for the equivalence between the shortsighted problem and the original problem is the following. Consider the problem of type θ who is considering whether or not to stop today. By Lemma 1, later expansions in the stock will only take place when some lower type $\theta' < \theta$ finds it optimal to stop, in which case it is also optimal for the higher type θ to stop. In other words, future expansions only take place under circumstances where θ wants to stop in any case, and therefore those expansions have no bearing on the marginal consideration for stopping today. Hence, today's continuation value of the *marginal* type is the same in equilibrium as in the shortsighted problem. As this intuition suggests, the optimality of shortsighted behavior is an equilibrium property and may well be violated against other (non-equilibrium) stock processes. The intuition does not rely on the properties of the learning process in any way, and therefore we expect the result to generalize to other processes as such. We also show in Laiho, Murto and Salmi (2023) that the result extends to a setting, where agents that stop later impose a direct payoff externality to those agents that have stopped earlier.[7] In Appendix B, we formalize the argument to get the following result:

Proposition 1. *There is a unique decentralized equilibrium, which is a boundary policy 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 $x^E(q(\theta))$, which is precisely the optimal stopping threshold for a shortsighted agent of type θ who assumes that the stock remains fixed at $q(\theta)$ forever. The

librium behavior coincides with that of 'myopic' investors who ignore the effect the future investments have on the price.

[7] One critical assumption for the result is that agents are infinitesimally small, which implies that an individual deviation will not influence the stock process. Suppose, to the contrary, that there are N players and by stopping a player causes a discrete jump in the stock. With incomplete information (i.e. if players' types are private information), the result would not carry over, which we can deduce from the model analyzed by Décamps and Mariotti (2004). With complete information, we believe that this property would continue to hold (see e.g. Cetemen et al. (2023) for a similar equilibrium property in another context), but contrary to our setup there might be multiple equilibria.

decentralized equilibrium is a boundary policy: whenever the belief is about to cross the boundary $x^E(q)$, 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. This is because the better the learning technology, the greater the option value of waiting and hence the higher the belief threshold at which an agent stops. Figure 2 depicts the policy for different values of σ. 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* $(x^{myop}(0), 1)$*, the stock of stopped agents is strictly smaller in the decentralized equilibrium than in the no-learning benchmark.*

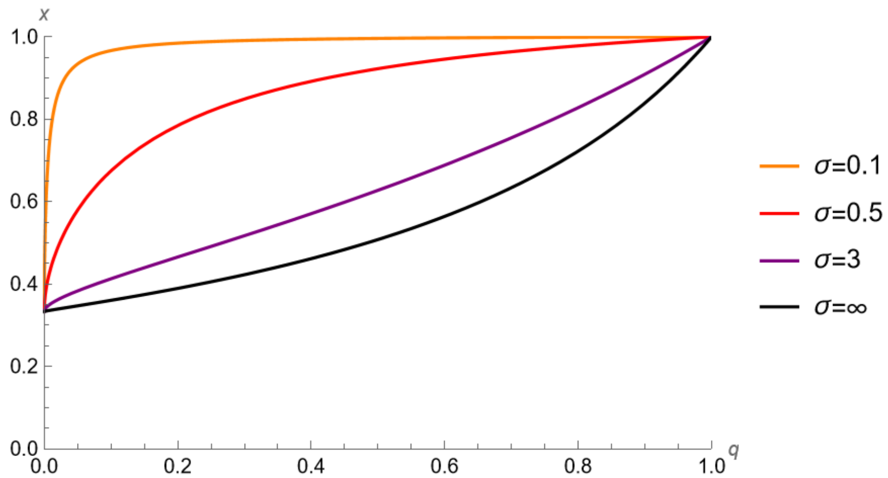

Figure 2: Equilibrium policy $x^E(q)$ for different σ when $v_H(q) = 1 - q$, $v_L(q) = -1/2$, and $r = 0.1$.

The intuition behind Corollary 1 is that agents want to fully utilize 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 agent's 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. Then both *past and future* stopping decisions affect the solution.

3.5 Social optimum

We now consider the problem in Definition 2 where a benevolent social planner seeks to maximize agents' expected joint payoff. The problem is identical to a problem of a single decision maker who controls a path of incremental expansions. Solving the problem is the main analytical contribution of this paper.

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

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

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 finding the value function that solves the HJB equation (5). In the waiting region below x^*, we have $q^* = q$ and hence (5) reduces to a differential equation. In the waiting region, the value consists of the value of potential future actions. Solving the partial differential equation yields:[8]

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

where $B(q)$ is a function to be determined and

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

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 equates 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

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

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

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^*:

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

[8] We have discarded the other root of the characteristic equation, $\widetilde{\Phi}(x,q) := x^{1-\beta(q)}(1-x)^{\beta(q)}$, as we must have that the value converges to the static solution as $x \to 0$ and $x \to 1$.

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 that satisfies $x^*(q) \leq x^E(q)$ for all $q \in [0, 1]$ (proof of Lemma 6 in Appendix C.2). This solution is our candidate for social optimum. We first verify that, together with the value function in (6), it solves the HJB equation, and we further verify that it also maximizes the original objective (4). In the process, we show that $x^*(q)$ is continuous and strictly increasing in q and hence satisfies the requirements for a boundary policy.

Proposition 2. *The socially optimal policy x^* is a boundary policy that satisfies $x^*(q) \leq x^E(q)$ for all $q \in [0, 1]$. It solves the initial value problem (9) with initial value $x^*(1) = 1$.*

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. The socially optimal policy balances the resulting trade-off between the efficient use of information (*option value effect*) and the efficient production of information (*information generation effect*).

Figure 3 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

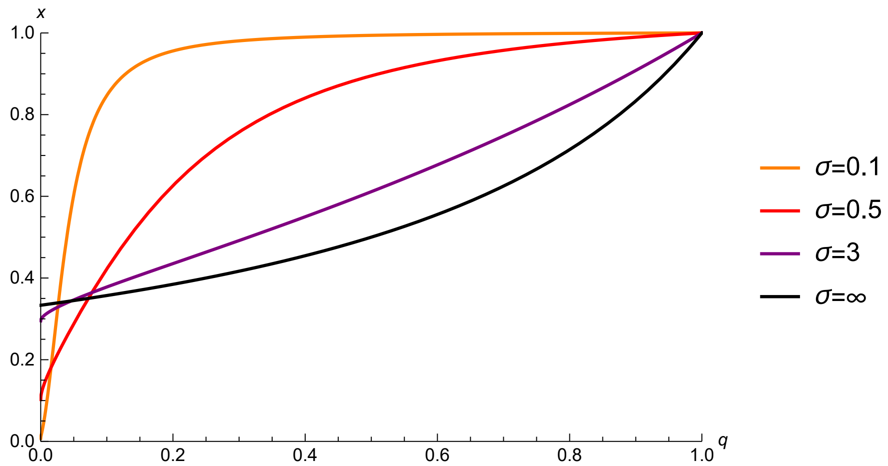

Figure 3: Socially optimal policy $x^*(q)$ for different σ when $v_H(q) = 1 - q$, $v_L(q) = -1/2$, and $r = 0.1$.

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 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 myopic policy without learning ($\sigma = \infty$). Hence, gradual learning may either increase or decrease expansions as the informational trade-off suggests. The following proposition generalizes this observation (see Appendix C.3 for the proof).

Proposition 3. *There exists* $\underline{x} \in (x^{myop}(0), 1)$ *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 4 illustrates the relationship between the solutions. Compared to the no-learning benchmark, gradual learning first increases and then decreases optimal expansions over time. The decentralized policy requires a higher belief for further expansions than the other policies.

Finally, it is illuminating to look at what happens to the actual speed of learning when the learning technology improves. To do that, let $q^*_\sigma(x)$ and $q^E_\sigma(x)$ denote the socially optimal and the decentralized stocks for signal precision σ.

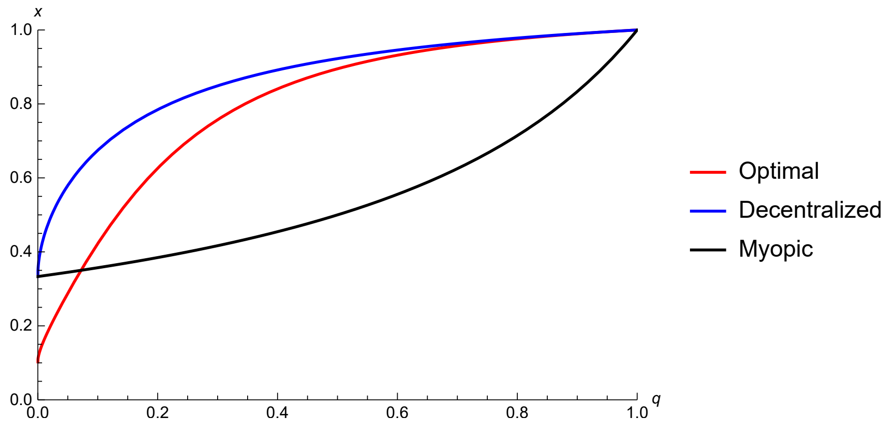

Figure 4: Different policies when $v_H(q) = 1 - q$, $v_L(q) = -1/2$, $\sigma = 0.5$, and $r = 0.1$.

Proposition 4. *(a) The socially optimal signal-to-noise ratio* $\sqrt{q_\sigma^*(x)}/\sigma \to \infty$ *as* $\sigma \to 0$ *for all* $x \in (0, 1)$*. (b) The decentralized signal-to-noise ratio* $\sqrt{q_\sigma^E(x)}/\sigma \to a(x)$ *as* $\sigma \to 0$ *where* $a(x) = 0$ *for all* $x \leq x^{stat}(0)$ *and* $a(x) \in (0, \infty)$ *for all* $x \in (x^{stat}(0), 1)$*.*

Learning gets arbitrarily fast in the socially optimal solution when the learning technology improves, whereas learning remains slow in the decentralized equilibrium. The latter is caused by informational free-riding: no-one wants to be the first one to stop if information arrives fast. This result suggests that the signal precision σ is an important determinant of welfare implications of the model as we will highlight in section 4.3. In Appendix C.4, we prove Proposition 4 and derive the functional form for $a(x)$.

3.6 Long-run distribution of the stock

One can characterize the probability distribution of the stock in the long-run for any boundary policy, including the decentralized equilibrium and social optimum. Since the long-run stock $q_\infty := \lim_{t\to\infty} q_t$ is equal to the value of the boundary policy $\widetilde{q}\,(\cdot)$ evaluated at the historical maximum value of the process x_t, we can do this by analyzing the distribution for the maximum value of the belief process x_t. Here we utilize the belief process being a martingale with a continuous path that eventually converges to truth. Note that if $\omega = H$, then the stock q_t must converge to 1 as the agents learn that stopping is profitable, whereas if $\omega = L$, the long run stock remains random as some fraction of the agents will have stopped by mistake.

Proposition 5. *Take an arbitrary boundary policy $\widetilde{q}(x)$ with inverse $\widetilde{x}(q)$ and assume that the intial stock satisfies $q_{0+} := \max(q_0, \widetilde{q}(x_0)) > 0$.*[9] *The probability distribution of the long-run stock is given by:*

$$\begin{aligned} \Pr(q_\infty \le q \,|\omega = L) &= \begin{cases} 0 & \textit{if } q < q_{0+} \\ \frac{\widetilde{x}(q)-x_0}{\widetilde{x}(q)(1-x_0)} & \textit{if } q_{0+} \le q < \widetilde{q}(1) \\ 1 & \textit{if } q \ge \widetilde{q}(1) \end{cases}, \\ \Pr(q_\infty \le q \,|\omega = H) &= \begin{cases} 0 \; \textit{if } q < \widetilde{q}(1) \\ 1 \; \textit{if } q \ge \widetilde{q}(1) \end{cases}. \end{aligned}$$

Since the socially optimal policy function is always below the decentralized equilibrium policy function, i.e. $x^*(q) < x^E(q)$ for all q, the long-run stock tends to be higher in social optimum than in equilibrium:

Corollary 2. *The long-run stock in social optimum dominates the long-run stock in decentralized equilibrium in the sense of first-order stochastic dominance.*

4 Mechanism design

We have seen that the social planner's solution and the decentralized equilibrium differ from each other. In this complementary section we bridge this gap by showing how to implement a given policy in a decentralized manner with anonymous posted prices. We complement the analysis by deriving a revenue maximizing mechanism. Lastly, we demonstrate the results of this paper in the context of a market for a new durable good.

[9] If $q_{0+} = 0$, i.e. $x_0 \le \widetilde{x}(0)$ and $q_0 = 0$, then $q_t \equiv 0$ for all $t \ge 0$ and no learning will ever take place.

4.1 Posted prices

We now show how a designer can implement a boundary policy by using anonymous posted prices.[10] A *posted price rule* $P : [0,1] \times [0,1] \to \mathbb{R}$ defines a transfer payment that an agent has to pay to the designer if willing to stop at a given state. With such a policy in place, the stopping payoff for an agent of type θ who decides to stop in state (x, q) is

$$u_\theta(x,q) = x v_H(\theta) + (1-x) v_L(\theta) - P(x,q). \tag{10}$$

Let Q be an arbitrary boundary policy with policy function $\widetilde{x}$ satisfying $\widetilde{x}(q_1) = 1$ and its inverse $\widetilde{q}\,(x) : [0,1] \longrightarrow [0, q_1]$ with the convention $\widetilde{q}(x) = 0$ for $x \leq \widetilde{x}(0)$. Note that we allow here for the possibility that the designer wants to implement a restricted maximal stock $q_1 < 1$. The following proposition characterizes a posted price rule that implements Q:

Proposition 6. *Fix a boundary policy Q with policy function $\widetilde{x}$. Then there exists a posted price rule P such that Q is a decentralized equilibrium of the game where the stopping payoff is given by* (10)*. For states along the stopping boundary, the posted price rule is uniquely pinned down by*

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

For states away from the boundary, the posted price rule is not uniquely determined, but one valid rule is the following:

- *For states $x > \widetilde{x}(q)$, set $P(x,q) = P(x, \widetilde{q}(x))$.*
- *For states $x < \widetilde{x}(q)$, set $P(x,q) = v_H(\overline{\theta})$.*

We prove the proposition in Appendix D. The posted price at the boundary is pinned down by the envelope theorem of Milgrom and Segal (2002) and we verify that incentive compatibility holds globally along the stopping boundary. Below the boundary, the designer only needs to make sure that the transfer payment is high

[10] Depending on the context and sign, these can also be interpreted as taxes or subsidies.

enough to make the cost of stopping prohibitive, which is trivially accomplished by setting the payment at $v_H(\overline{\theta})$. Above the boundary, one must guarantee that those agents needed for the state to move immediately to the boundary do indeed want to stop, and $P(x,q) = P(x,\widetilde{q}(x))$ is one way to accomplish that.

The posted price rule considered here responds to changes in both x and q. In Online Appendix, we investigate posted price rules that only depend on one of the two variables. Such rules are of special interest because they are often easier to use in practice. For example, a seller of a new durable good can set the price based on the cumulative past sales instead of both sales and reviews or other feedback from past buyers. We show that a given policy Q can always be implemented by a posted price that is only a function of the current belief x and can often (but not always) be implemented by a posted price that is only a function of the current stock.[11]

4.2 Revenue maximizing designer

In some applications a designer maximizes an objective other than social surplus. Here we show how we can adapt our techniques to find a mechanism that maximizes the revenues from the transfers. For this subsection 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. For a given policy, 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 Appendix D for the proof):[12]

[11] We show in Online Appendix that a policy Q can be implemented with a posted price that depends only on the stock if and only if the prospective transfer rule, defined as $P^S(q) := P(\widetilde{x}(q), q)$ is everywhere increasing in q.

[12] The approach extends a result in Laiho and Salmi (2023) and shares similar features with Board (2007) who analyzes the optimal sale of options.

Lemma 3. *Incentive compatibility implies that the designer's expected revenue is:*

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

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

Using this result we can solve the revenue maximizing designer's problem by using the planner's solution in Proposition 2. We only need to replace the stopping payoffs with virtual stopping payoffs and use the revenue maximizing quantity under complete information as the initial value. We demonstrate this in the next subsection where we analyze the problem of a durable goods monopolist.

4.3 Application: markets for new durable goods

We now pull together our theoretical results to discuss an application. Consider a new durable good with uncertainty about the product quality. Gradual social learning naturally arises: after an individual buyer purchases the product, he starts using it and observes how well it functions over time. As a result, potential future buyers learn gradually from the fleet of existing users. This is in contrast to experience goods, such as movies, where consumption is one-shot and hence instantaneous learning is a natural modeling approach. The existing literature on experimentation has focused solely on experience goods.

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 F is twice continuously differentiable and has monotone hazard rate. The monopolist faces marginal cost of production $c > 0$ and commits to a pricing scheme, P_t. Furthermore, we assume $\underline{\theta} \geq c$, so that it is always profitable to serve all types.[13]

[13]This assumption also ensures that the monopolists stopping payoff in high state of the world is always positive: $v_H(q) \geq 0$.

The monopolist's problem is a mechanism designer's problem whose objective is to maximize the expected sum of transfers net of the cost of production, $\mathbb{E}\left[\int_0^\infty e^{-rt}(P_t - c)dq_t\right]$. Using Lemma 3, the monopolist's objective can be written as

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

where ϕ_ω is the virtual valuation: $\phi_H(\theta) := \theta - \frac{1-F(\theta)}{f(\theta)} - c$ and $\phi_L(\theta) := -c$.

The monopolist seeks to maximize the expected virtual surplus. With this in mind, we can use the planner's solution in Proposition 2 to characterize the monopolist's solution:[14]

Corollary 3. *The monopolist's policy x^M is characterized by $x^M(\overline{q}^M) = 1$ and $x^{M\prime}(q) = g(q, x^M(q))$, 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) *for $v_H(\theta) = \theta - (1 - F(\theta))/f(\theta) - c$ and $v_L(\theta) = -c$.*

The monopolist sells more whenever the belief reaches the boundary $x^M(q)$ and waits otherwise. The posted prices that implement the solution satisfy (11) in Proposition 6.

We contrast the monopoly solution with the planner's socially optimal solution and the competitive market equilibrium. The planner's solution can be found also by applying Proposition 2 but with different stopping payoffs. Because the planner internalizes the value the agents receive, she evaluates policies based on the total surplus (instead of virtual surplus): $v_H(q) = \theta(q) - c, v_L(q) = -c$. In particular, notice that the planner's stopping payoff in high state of the world is always greater than the monopolist's corresponding stopping payoff. Using these payoffs for the planner then gives the following Corollary:

Corollary 4. *The social planner's policy x^P is characterized by $x^P(\overline{q}^P) = 1$ and $x^{P\prime}(q) = g(q, x^P(q))$, where $\overline{q}^P$ solves $\theta(\overline{q}^P) = c$ and g is given in* (9) *for $v_H(q) = \theta(q) - c$ and $v_L(q) = -c$.*

[14] Notice that 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. In addition, because the solution in Proposition 2 is a boundary policy, Proposition 6 guarantees that posted prices are without loss of generality and hence the revenue maximizing policy can be implemented.

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

Corollary 5. *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.

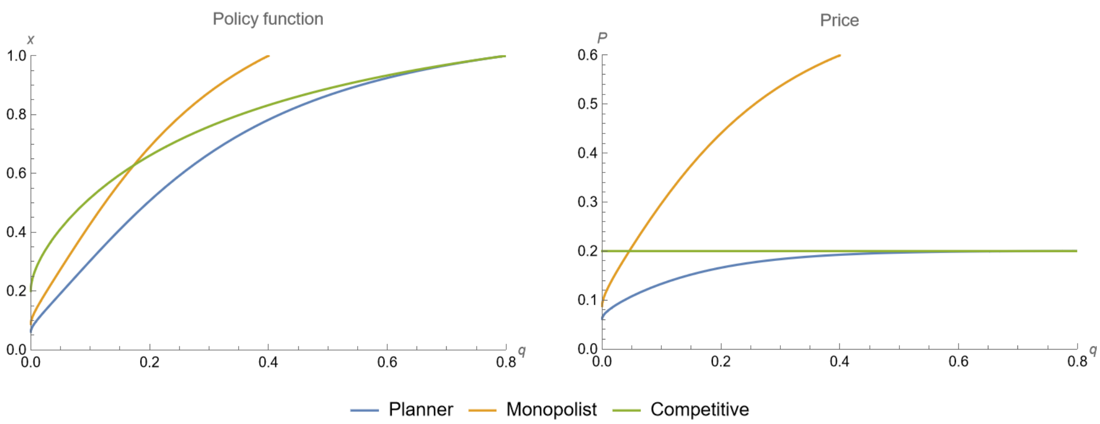

Figure 5: Different solutions for uniform $(0, 1)$ types, $c = 0.2$, $r = 0.1$, and $\sigma = 0.5$.

Figure 5 gives a numerical example of the solutions. The left panel shows the policy functions. 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. This comparison in fact holds for all type distributions and parameter values (see Online Appendix).

The right panel shows the corresponding posted prices (11) at the stopping boundary. The monopoly starts at a price below marginal cost in order to initially boost information generation, but then increases the price steeply to extract more revenues. The social planner prices everywhere below the marginal cost in order to fully internalize the informational externality.

The monopolist's incentives to generate information are weaker than the planner's because the monopolist cannot capture all 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: $x^M(0) > x^P(0)$.

With endogenous learning, the quality of the learning technology is a key determinant of welfare differences between different market structures. The parameter that captures this in our model is the noise parameter σ. The left panel in Figure 6 shows the difference in total welfare between the planner's solution and the competitive solution as a function of initial belief and with different values of σ.[15] The better the learning technology, i.e. the lower the value of σ, the greater the welfare loss in competitive equilibrium. This is because the social planner takes full advantage of improved learning technology, whereas informational free-riding limits how much the actual speed of learning in equilibrium increases when the learning technology improves (see Proposition 4).[16]

A striking implication of endogenous learning in the current context is that market power can benefit consumers. Although a monopoly harms consumers by distorting the allocation in the usual way, it also spills benefits to the consumers

[15] The kink at initial belief 0.2 is the threshold at which the first unit is sold in competitive equilibrium.

[16] The monopolist might be able to influence the speed of social learning through e.g. product design or development of communication channels such as user forums. Since the monopolist's problem is isomorphic to that of a planner maximizing the total welfare, it is clear that monopoly profit is increasing in the quality of the learning technology and so the monopolist has every incentive to make the learning technology as good as possible. Our numerical results suggest that also the consumers are better off with a better learning technology.

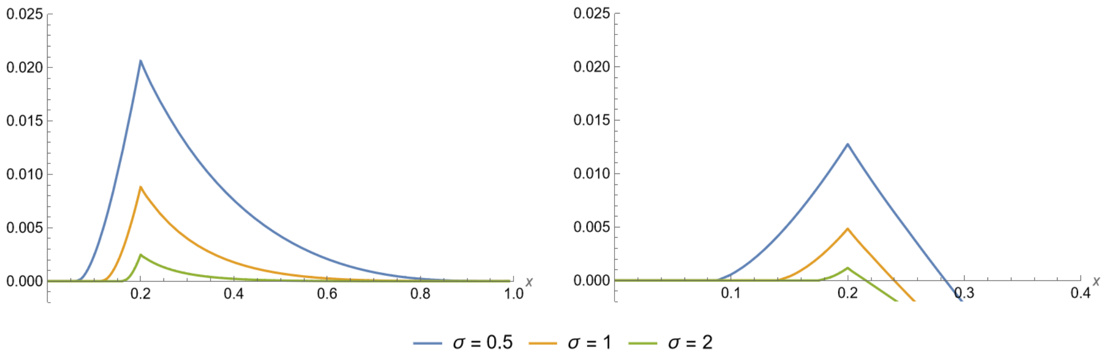

Figure 6: Difference in total welfare in the social planner vs. competitive solution (left panel) and in monopoly vs. competitive solution (right panel) as a function of the initial belief for uniform $(0, 1)$ types, $c = 0.2$, and $r = 0.1$.

from its ability to internalize the informational externality. The right panel in Figure 6 shows the difference in total welfare between the monopoly solution and the competitive equilibrium. Positive values indicate that the monopoly solution has a greater total welfare. Since consumers get a positive share of the total surplus in information rents and since consumer surplus is zero in the competitive solution for initial beliefs below 0.2, this implies that for a range of initial beliefs the consumers are better off in the monopoly solution than in the competitive solution. Again, we see that the quality of the learning technology plays a key role: the better the learning technology, the greater the scope for the consumers to benefit from market power.

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 there is a crucial difference between the two different competitive environments. 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 the difference between the social planner's price and marginal cost in the right panel of Figure 5).

5 Concluding remarks

We conclude by summarizing a few compelling reasons why the gradual arrival of endogenous information matters. First, gradual learning enables the analysis of various real-life situations where the long-run consequences of a decision determine its profitability. Because gradual learning creates a novel informational trade-off on the social level between information generation and the option value of waiting, it dramatically shapes the incentives of experimentation.

A second important motive to model the gradual arrival of information is technical. As demonstrated in this paper, the decentralized equilibrium can be solved in a closed form under gradual learning. We further show how mechanism design techniques can be utilized to conduct policy analysis in our environment. The solution method extends to richer environments, such as to models with payoff externalities (see Laiho, Murto and Salmi 2023).

An important takeaway from the paper is that the signal precision has subtle implications for learning and welfare. We show that even if signals get arbitrarily precise, learning remains slow in the equilibrium. This contrasts with the socially optimal solution, in which the true state is learned arbitrarily fast as the learning technology improves. As a result, the equilibrium welfare loss is particularly severe if the learning technology is good.

As a final point note that irreversibility of actions is a crucial assumption in our model. The conclusions in models with reversible actions such as Bonatti (2011) are significantly different. A natural extension to our model would be to analyze what happens if stopping decisions would be partially reversible. While we believe that many of qualitative properties of our results stay the same, pursuing such an extension is beyond the scope of this paper. If agents can exit the stock, we will naturally have three instead of two state variables to keep track of. It is an interesting avenue for future research to see to what extent partial reversibility changes the results in this paper. A concrete example would be to analyze a model similar to Dixit (1989) in which the agents can reverse their decision by paying a fixed cost.

Appendix

A Additional material for Sections 2 and 3.1

A.1 Learning process as the continuous limit

Consider a discrete model where the number of agents is n and where the period length is dt. Let the signal process be such that in each period, each agent who has stopped generates a normally distributed conditionally iid. signal:

$$y_t^i \sim N\left(\frac{\mu_\omega dt}{n}, \frac{\sigma^2 dt}{n}\right).$$

This normalization keeps the informativeness of the aggregate signal constant while letting the number of small agents to grow as in Bergemann and Välimäki (1997).

When the number of agents who has stopped is $k \leq n$, this implies the following aggregate signal:

$$\sum_{i=1}^{k} y_t^i \sim N\left(\mu_\omega dt\frac{k}{n}, \sigma^2 dt\frac{k}{n}\right).$$

Let $q = k/n$ denote the fraction of agents who have stopped. Now, the signal process (1) follows once we take the limit when $n \to \infty$ (and $k \to \infty$ so that k/n stays fixed) and $dt \to 0$.

Notice that the limiting distribution for the aggregate signal depends only on the mean and the variance of y_t^i (the central limit theorem). Hence, the signal process (1) is also the limiting process for the case where y_t^i is not normally distributed, including the case where agents communicate through binary signals.

Furthermore, we can rewrite the model so that the individual signals represent realized payoffs in a model where agents start receiving a stochastic flow payoff after stopping: $\pi_t(\theta) = \pi_\omega(\theta) + \epsilon_t(\theta)$ where $\epsilon_t(\theta) \sim N\left(0, \sigma^2(\pi_H(\theta) - \pi_L(\theta))^2\right)$. The noise term is scaled so that every increment in q is equally informative. This assumption is not necessary: we analyze in Online Appendix the case of heterogeneous informativeness and show that both the analysis and the qualitative results remain unchanged if the stopping profile is monotone. When we set

$\pi_\omega(\theta) = rv_\omega(\theta)$, the expected stopping payoff is $x_t v_H(\theta) + (1 - x_t)v_L(\theta)$ just like in the main text. Since there are no further actions after stopping, it does not matter how fast the agents learn privately after they have stopped: the parameter σ can be interpreted to capture both the noise in the private learning and the noise in communication.

A.2 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. □

A.3 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

B.1 Proof of Proposition 1

We will show that the policy in Proposition 1 is a decentralized equilibrium. Fix policy Q to be the boundary policy in Proposition 1 and consider optimal stopping of type θ against it. Except possibly at the initial time $t = 0$, the state (x_t, q_t) will remain in set $\mathbf{X}$ that we call the feasible region:

$$\mathbf{X} := \left\{(x, q) : 0 \leq q \leq 1,\, 0 < x \leq x^E(q)\right\}.$$

Since Q is a Markovian process, we can express the stopping problem of type θ as a Markovian problem, where the task it to choose optimally a stopping set $S_\theta \subseteq \mathbf{X}$ in the feasible region. (We will also check at the end that the optimal behavior outside of $\mathbf{X}$ is consistent with the initial jump at time $t = 0$.) Denote by $F_\theta(x, q)$ the value function under optimally chosen stopping set 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 first hitting time of S_θ and $u_\theta(x) := x v_H(\theta) + (1 - x) v_L(\theta)$ is the stopping value at belief x.

Before analyzing the shape of the optimal stopping set, we can already conclude some basic properties of $F_\theta(x, q)$. In the stopping set, $(x, q) \in S_\theta$, we must have $F_\theta(x, q) = u_\theta(x)$. In the continuation set, $(x, q) \in \mathbf{X} \backslash S_\theta$, the properties of $F_\theta(x, q)$ are determined by the infinitesimal generator of the process $(x_t, q_t)_{t \geq 0}$. Althought the process is two-dimensional, q_t increases only when x_t hits new historical record values and the set of such times is of zero measure. The process q_t is hence constant almost everywhere and the infinitesimal generator of (x_t, q_t) in the interior of $\mathbf{X} \backslash S_\theta$ reduces to that of the process x_t as if q_t is fixed. We can write the infinitesimal generator of x_t as (see e.g. Peskir, Shiryaev and Shirayev (2006)):

$$\frac{x^2 (1 - x)^2 q}{2\sigma^2} \frac{\partial^2}{\partial x^2}.$$

It follows that the Hamilton-Jacobi-Bellman equation for the agent's value in the interior of $\mathbf{X} \backslash S_\theta$ takes the form:

$$rF_\theta(x, q) = \frac{x^2 (1 - x)^2}{2\sigma^2} q \frac{\partial^2 F_\theta(x, q)}{\partial x^2}.$$

This is a partial-differential equation of $F_\theta(x,q)$, but it only involves derivatives with respect to x, and we can write its general solution in closed form as:

$$F_\theta(x,q) = A_\theta(q)\,\Phi(x,q) + B_\theta(q)\,\widetilde{\Phi}(x,q)\,, \tag{12}$$

where $A_\theta(q)$ and $B_\theta(q)$ are functions of q (we index by θ to emphasize where dependence on type enters), and where

$$\begin{aligned}
\Phi(x,q) &= x^{\beta(q)}(1-x)^{1-\beta(q)}\,, \\
\widetilde{\Phi}(x,q) &= x^{1-\beta(q)}(1-x)^{\beta(q)}\,, \\
\beta(q) &= \frac{1}{2}\left(1+\sqrt{1+\frac{8r\sigma^2}{q}}\right).
\end{aligned}$$

At the boundary $x^E(q)$, the stock q_t increases instantaneously as x hits new record values. Whenever such a boundary point is in the continuation region, the following condition of *normal reflection* must hold (Peskir, Shiryaev and Shirayev (2006)):[17]

$$\frac{\partial}{\partial q}\left[F_\theta(x,q)\right]_{x=x^E(q)} = 0. \tag{13}$$

As a preliminary step, we solve an auxiliary optimal stopping problem, where the stock is assumed to be fixed at $q_t \equiv q$ forever:

Lemma 4. *Assume that the stock is fixed at $q_t \equiv q$ forever. Then, it is optimal for θ to stop if and only if $x_t \geq \widehat{x}_\theta(q)$, where*

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

The corresponding value function is

$$\overline{F}_\theta(x;q) = \begin{cases} u_\theta(x) \;\; \text{if } x \geq \widehat{x}_\theta(q), \\ \left(\frac{x}{\widehat{x}_\theta(q)}\right)^{\beta(q)}\left(\frac{1-x}{1-\widehat{x}_\theta(q)}\right)^{1-\beta(q)} u_\theta(\widehat{x}_\theta(q)) \;\; \text{if } x < \widehat{x}_\theta(q), \end{cases}$$

where $u_\theta(x) := xv_H(\theta) + (1-x)\,v_L(\theta)$ is the stopping value at belief x.

[17] When the boundary $x^E(q)$ is hit, the time path of q_t is not differentiable; the time derivative dq/dt is unbounded. Therefore, if it were to be the case that $\frac{\partial}{\partial q}\left[F_\theta(x,q)\right]_{x=x^E(q)} \neq 0$, then the expected instantaneous rate of change in the value function, $\mathbb{E}\left[dF_\theta(x,q)\right]/dt$, would explode at the moment of hitting the boundary.

Proof. This is a standard one-dimensional optimal stopping problem and it is well known that the solution is some stopping threshold that we denote $\widehat{x}_\theta(q)$ (see e.g. Dixit and Pindyck (1994) or the team problem in Bolton and Harris (1999)). The value function, denoted $\overline{F}_\theta(x;q)$, must take the form (12) when $x < \widehat{x}_\theta(q)$. If it is certain that $\omega = L$, then the option to stop is worthless and we get the boundary condition $\overline{F}_\theta(0;q) = 0$. This implies $B_\theta(q) = 0$. The value-matching condition $\overline{F}_\theta(\widehat{x}_\theta(q);q) = u_\theta(\widehat{x}_\theta(q))$ and the smooth-pasting condition $\frac{\partial}{\partial x}\overline{F}_\theta(\widehat{x}_\theta(q);q) = \frac{\partial}{\partial x}u_\theta(\widehat{x}_\theta(q))$ uniquely determine the remaining constant $A_\theta(q)$ and the stopping threshold $\widehat{x}_\theta(q)$ and we get the formulas given in the Lemma. □

The lemma says that it is optimal to wait below $\widehat{x}_\theta(q_t)$ if q_s is assumed fixed for all $s > t$. If we relax this assumption and allow q_s to increase arbitrarily for $s > t$, then waiting at time t becomes even more desirable. This is because higher future values of q_s means improved future learning, which in turn will increase the value of waiting relative to immediate stopping. The lemma therefore implies that no matter what policy we have, it can never be optimal for θ to stop if $x_t < \widehat{x}_\theta(q_t)$:

Lemma 5. *If the current belief satisfies $x_t < \widehat{x}_\theta(q_t)$, then stopping immediately is strictly dominated for type θ.*

Proof. Assume that the current belief is $(x_t, q_t) = (x, q)$, where $x < \widehat{x}_\theta(q)$. Consider a simple strategy such that θ stops as soon as x_t hits $\widehat{x}_\theta(q)$ (no matter how q_t evolves). For a moment, assume that the stock is fixed at $q_s = q$ for all $s > t$. Then by Lemma 4, this simple strategy gives value

$$\overline{F}_\theta(x;q) = \left(\frac{x}{\widehat{x}_\theta(q)}\right)^{\beta(q)} \left(\frac{1-x}{1-\widehat{x}_\theta(q)}\right)^{1-\beta(q)} u_\theta(\widehat{x}_\theta(q)).$$

On the other hand, if we keep the threshold $\widehat{x}_\theta(q)$ as above, but assume that the current stock is fixed at a higher level, $q_s = q' > q$ for all $s > t$, then the value of this simple strategy gives

$$\begin{aligned} & \left(\frac{x}{\widehat{x}_\theta(q)}\right)^{\beta(q')} \left(\frac{1-x}{1-\widehat{x}_\theta(q)}\right)^{1-\beta(q')} u_\theta(\widehat{x}_\theta(q)) \\ > & \left(\frac{x}{\widehat{x}_\theta(q)}\right)^{\beta(q)} \left(\frac{1-x}{1-\widehat{x}_\theta(q)}\right)^{1-\beta(q)} u_\theta(\widehat{x}_\theta(q)) = \overline{F}_\theta(x;q), \end{aligned}$$

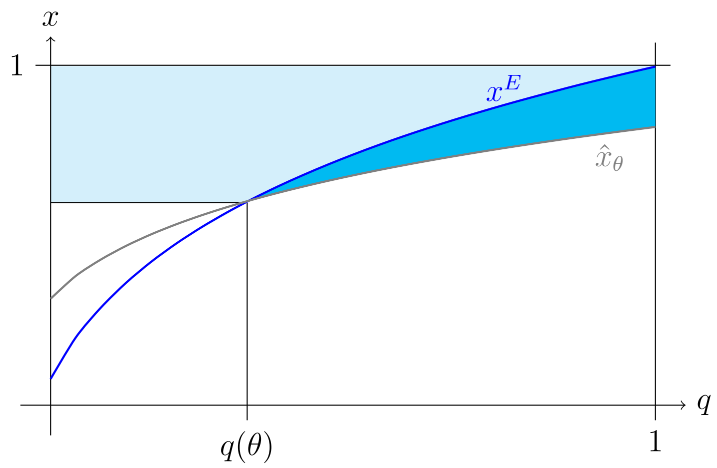

Figure 7: Optimal stopping for type θ.

where the inequality follows from $\beta(q)$ being decreasing in q. In other words, the value of such a simple threshold strategy is increasing in the learning speed determined by q. It then follows that for an arbitrary Q (where $q_t = q$ and $q_s \geq q$ for $s > t$), the value of the simple strategy of stopping at threshold $\widehat{x}_\theta(q)$ is weakly higher than $\overline{F}_\theta(x;q)$. This means that $F_\theta(x,q) \geq \overline{F}_\theta(x;q)$, where $F_\theta(x,q)$ is the value of θ under *optimal* stopping rule (instead of the simple strategy). Since we assumed $x < \widehat{x}_\theta(q)$, we have $\overline{F}_\theta(x;q) > u_\theta(x)$ by Lemma 4, and therefore also $F_\theta(x,q) > u_\theta(x)$. Hence, stopping immediately cannot be optimal for θ. $\square$

With these preliminary results in place, we now consider the optimal stopping policy of θ against Q. Our plan is to show that the optimal stopping region S_θ is the dark blue shaded region in Figure 7, i.e.

$$S_\theta = \left\{(x,q) : q \geq q(\theta), x \in \left[\widehat{x}_\theta(q), x^E(q)\right]\right\}. \tag{14}$$

As a first step, we note that it cannot be optimal for θ to stop at any $(x,q) \in \mathbf{X}$ with $q < q(\theta)$. This follows directly from Lemma 5 above. Since all $(x,q) \in \mathbf{X}$ with $q < q(\theta)$ satisfy $x < \widehat{x}_\theta(q)$, it is strictly dominant for θ to wait.

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

$F_\theta(x^E(q), q) > u_\theta(x^E(q))$. We will show below that this implies

$$\frac{\partial}{\partial x}\left[F_\theta(x,q)\right]_{x=x^E(q)} \geq \frac{\partial}{\partial x}\left[u_\theta(x)\right]_{x=x^E(q)}, \tag{15}$$

which, as we will further show below, leads to a contradiction.

There are two possible cases that we consider separately. First, suppose that even though $\left(x^E(q), q\right) \notin S_\theta$, it is optimal to stop at some lower belief, i.e. there is some $x' < x^E(q)$ such that $(x', q) \in S_\theta$ (let x' denote the highest such belief). In that case $F_\theta(x', q) = u_\theta(x')$. The continuation value $F_\theta(x,q)$ takes the form (12) in the interval $\left(x', x^E(q)\right)$ with boundary condition $F_\theta(x', q) = u_\theta(x')$. Direct calculations show that $F_\theta(x,q)$ is convex in x on the interval. Since we also necessarily have $F_\theta(x,q) \geq u_\theta(x)$ for all $x \in (x', x^E(q))$, (15) follows.

Second, 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 $x < x^E(q)$. The continuation value must vanish as $x \to 0$, and the corresponding boundary condition $F_\theta(0,q) = 0$ implies that the term B_θ in (12) vanishes. Hence, the value function $F_\theta(x,q)$ takes 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

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

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

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

which is the case if and only if $q \geq q(\theta)$. Noting that $\frac{\partial}{\partial x}\left[u_\theta(x)\right]_{x=x^E(q)} = v_H(\theta) - v_L(\theta)$, we may conclude that (15) holds in this case too.

Given that (15) holds, the rate of change in $F_\theta(x,q)$ along the boundary is

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

where the last term of the first line disappears by (13) and where the inequality follows from (15).

We have now shown that $F_\theta(x^E(q),q) > u_\theta(x^E(q))$ implies $\frac{d}{dq}F_\theta\left(x^E(q),q\right) \geq \frac{d}{dq}u_\theta\left(x^E(q)\right)$. Applying this iteratively to all $q' > q$, we conclude that this implies further that $F_\theta(x^E(q'),q') > u_\theta(x^E(q'))$ for all $q' \in [q,1]$, and in particular $F_\theta(x^E(1),1) > u_\theta(x^E(1))$. We know that $x^E(1) = 1$, so this yields $F_\theta(1,1) > v_H(\theta)$. This is a contradiction, because $v_H(\theta)$ is the stopping payoff under certainty of state $\omega = H$, which is clearly an upper bound for the value function for θ.

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 dark blue 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 achieve that value, θ should optimize as if q is fixed, i.e. stop at all points $\left[\widehat{x}_\theta(q), x^E(q)\right]$.

We have now shown that the stopping rule defined in (14) 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. Since the initial state point (x_0, q_0) may be above the boundary, we must also check the optimal behavior of θ for initial state points $(x_0, q_0) \notin \mathbf{X}$. If $(x_0, q_0) \notin \mathbf{X}$, then Q commands the stock to jump instantenously to point $q_{0+} := \left\{q : x^E(q) = x_0\right\}$. The point (x_0, q_{0+}) is in the optimal stopping region of θ if and only if $x_0 \geq \widehat{x}_\theta(q(\theta))$, and therefore it is optimal for θ to stop at time $t = 0$ if $(x_0, q_0) \notin \mathbf{X}$ and $x_0 \geq \widehat{x}_\theta(q(\theta))$. We conclude that the optimal stopping region of θ contains also the light shaded region in Figure 7. This means that the initial jump from (x_0, q_0) to (x_0, q_{0+}) is consistent with all types $\theta \geq \theta(q_{0+})$ optimally stopping at $t = 0$. Collecting all this together, we can conclude that Q is a decentralized equilibrium.

It remains to prove the uniqueness part of the proposition, i.e. that no other equilibrium policies exist than the boundary policy defined in the proposition. For this it suffices to show that in any equilibrium q_t cannot increase at state points where $x_t < x^E(q_t)$ and q_t cannot stay put at state points where $x_t > x^E(q_t)$.

Take some decentralized equilibrium policy and some arbitrary history h_t with current state (x_t, q_t). Since by Lemma 1 optimized stopping times are monotone in θ, it must be that types $\theta > \theta(q_t)$ have stopped while types $\theta < \theta(q_t)$ have not yet stopped at h_t. We now show that for the cutoff type $\theta(q_t)$ both waiting above the boundary $x^E(q_t)$, and stopping below the boundary $x^E(q_t)$, are inconsistent with Q being an equilibrium.

Consider first the case where the state after history h_t satisfies $x_t < x^E(q_t)$. But then $x_t < \widehat{x}_\theta(q_t)$ for all types $\theta \leq \theta(q_t)$ (this is because $\widehat{x}_{\theta(q_t)}(q_t) = x^E(q_t)$ and $\widehat{x}_\theta(q_t)$ is decreasing in θ). By Lemma 5 it is strictly dominant for all types who have not yet stopped to wait. We conclude that q_t cannot increase at h_t.

Consider next the case where the state after history h_t satisfies $x_t > x^E(q_t)$. For contradiction, suppose that it is optimal for the cut-off type $\theta(q_t)$ to wait, i.e. it is optimal to choose some stopping time τ that gives

$$\mathbb{E}\left(e^{-r\tau}u_{\theta(q_t)}(x_\tau)\,|h_t\right) > u_{\theta(q_t)}(x_t)\,.$$

By Lemma 1, equilibrium stopping times are monotone in θ and so the lower types must wait even longer, i.e. optimal stopping times for types $\theta < \theta(q_t)$ satisfy $\tau(\theta) \geq \tau$ a.s. But this means that q_t stays fixed until τ, and hence the same stopping time τ would give type $\theta(q_t)$ a payoff strictly higher than $u_{\theta(q_t)}(x_t)$ also in the auxiliary problem analyzed in Lemma 4, where q_t is fixed *by assumption.* Since we have $x_t > x^E(q_t) = \widehat{x}_{\theta(q_t)}(q_t)$, this is a contradiction with Lemma 4. We conclude that it cannot be optimal for the cut-off type $\theta(q_t)$ to delay stopping. Since this conclusion holds for the cut-off type in any state value (x_t, q_t) satisfying $x_t > x^E(q_t)$, the only policy consistent with players choosing optimally their stopping times is the one where q_t jumps immediately to the boundary point q satisfying $x^E(q) = x_t$. $\square$

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

$$B_q(q) = A^1(x^*(q),q)\,x^*(q) + A^2(x^*(q),q)\,, \tag{16}$$

$$B(q) = U^1(x^*(q),q)\,x^*(q) + U^2(x^*(q),q)\,, \tag{17}$$

where

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

Differentiating (17) with respect to q and using the chain rule gives

$$\begin{aligned}
B_q(q) = &\left[U_x^1(x^*(q),q)\,x^{*\prime}(q)+U_q^1(x^*(q),q)\right]x^*(q)+U^1(x^*(q),q)\,x^{*\prime}(q)\\
&+U_x^2(x^*(q),q)\,x^{*\prime}(q)+U_q^2(x^*(q),q)
\end{aligned} \tag{18}$$

Equating (16) and (18), 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.

C.1 Proof of Proposition 2

The proof contains three parts. In part 1, we show that the initial value problem (9) has a solution $x^*(q)$ that we take as our candidate for socially optimal policy. This solution has the property $x^*(q) < x^E(q)$ for all $q < 1$ and it 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) (all proofs of the lemmas are in Appendix C.2):

Lemma 6. *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* $\lim_{q \to 1} g(x^E(q), q) = x^{E'}(1)$.

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 Part 1 and let $q^*(x)$ be its inverse with the convention $q^*(x) = 0$ for $x \leq x^*(0)$. Its associated value function is

$$V(x,q) = \begin{cases} \int_q^{q^*(x)} (xv_H(s) + (1-x)v_L(s))\, ds + V(x, q^*(x)), \text{ for } q < q^*(q) \\ B(q)\,\Phi(x,q), \text{ for } q \geq q^*(x), \end{cases} \tag{19}$$

where $B(q)$ is given by (17). By construction, for $q \geq q^*(x)$, $V(x,q)$ satisfies

$$rV(x,q) = \frac{1}{2}V_{xx}(x,q)\frac{x^2(1-x)^2}{\sigma^2}q \tag{20}$$

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

$$V_q(x, q^*(x)) + xv_H(q^*(x)) + (1-x)v_L(q^*(x)) = 0, \tag{21}$$

$$V_{qx}(x, q^*(x)) + v_H(q^*(x)) - v_L(q^*(x)) = 0. \tag{22}$$

Differentiating (20) with respect to q, we have

$$rV_q(x,q) = \frac{1}{2}V_{xx}(x,q)\frac{x^2(1-x)^2}{\sigma^2} + \frac{1}{2}V_{xxq}(x,q)\frac{x^2(1-x)^2}{\sigma^2}q, \tag{23}$$

which allows us to re-write (21) as:

$$r[xv_H(q^*(x)) + (1-x)v_L(q^*(x))] + \frac{1}{2}V_{xx}(x, q^*(x))\frac{x^2(1-x)^2}{\sigma^2} + \frac{1}{2}V_{xxq}(x, q^*(x))\frac{x^2(1-x)^2}{\sigma^2}q = 0. \tag{24}$$

We next state three lemmas that concern the partials of the value function below, above, and at the boundary, respectively. Their proofs are in a separate section, Appendix C.2.

Lemma 7. *For $q \geq q^*(x)$, we have $V_q(x,q) + xv_H(q) + (1-x)v_L(q) \leq 0$.*

Lemma 8. *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 9. *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 $q^*(x)$):

$$rV(x,q) = \max_{q' \geq q} \left(r \int_q^{q'} \left(xv_H(s) + (1-x) v_L(s) \right) ds + \frac{1}{2} V_{xx}(x,q') \frac{x^2 (1-x)^2}{\sigma^2} q' \right). \tag{25}$$

The term in the parenthesis is a continuous function in q' and its derivative with respect to q' is

$$r\left(xv_H(q') + (1-x) v_L(q')\right) + \frac{1}{2} V_{xx}(x,q') \frac{x^2(1-x)^2}{\sigma^2} + \frac{1}{2} V_{xxq}(x,q') \frac{x^2(1-x)^2}{\sigma^2} q'. \tag{26}$$

Let us inspect the sign of this for different values of q'. For $q' \geq q^*(x)$, we can use (23) to write (26) as

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

which is negative by lemma 7. It follows that whenever $q \geq q^*(x)$, the right-hand side of (25) is maximized by choosing $q' = q$, i.e. keeping q fixed.

For $q' < q^*(x)$, we can use lemma 8 to write (26) as

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

which is decreasing in q'. Moreover, combining (24) and Lemma 9 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 (25) 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 (20) for $q \geq q^*(x)$, the left- and right-hand sides of (25) coincide with this choice of q'. Hence, we have shown that $V(x,q)$ defined in (19) 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 candidate value function is evaluated. From generalized Itô's formula we have[18]

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

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:

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

Taking conditional expectations , multiplying by $-e^{rt}$ and simplifying then gives

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

The candidate value function is bounded and therefore clearly satisfies the transversality condition: $\lim_{T\to\infty}\mathbb{E}[e^{-r(T-t)}V^*(x_T,q_T)] = 0$. Hence, taking the limit $T \to \infty$ gives that $V^*(x,q) \geq \max_Q U(Q;x,q)$.

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.

[18]To see that $V \in C^2$ check V_x at the boundary. The continuity of V_{xx} and V_{qq} follows from Lemma 8 and the continuity of V_q and V_{qx} are implied by the value matching and smooth pasting conditions.

C.2 Proof of Lemmas 6, 7, 8, and 9

Proof of Lemma 6. Taking the derivative of $g(x,q)$ with respect x gives:

$$
\begin{aligned}
g_x(x,q) = -\, & \Big[\beta''(q)\Big(x^2(1-2x)(\beta(q)-1)^3 v_H(q)^2 - 2(1-x)x\beta(q)(\beta(q)-1) \\
& \times v_H(q)v_L(q)((1-2x)\beta(q)-x) + (1-x)^2(1-2x)\beta(q)^3 v_L(q)^2\Big) \\
& + \beta'(q)\Big(2x^2(2x-1)(\beta(q)-1)^2 v_H(q)^2\beta'(q) + (1-x)^2\beta(q)^2 v_L(q) \\
& \times \Big(2(1-2x)v_L(q)\beta'(q) - 2x(\beta(q)-1)v_H'(q) - (1-2x)\beta(q)v_L'(q)\Big) \\
& + x v_H(q)\Big(4(1-x)v_L(q)\beta'(q)\Big((1-2x)\beta(q)^2 + 2x\beta(q) + x\Big) \\
& - x(\beta(q)-1)^2\Big((1-2x)(\beta(q)-1)v_H'(q) + 2(1-x)\beta(q)v_L'(q)\Big)\Big)\Big)\Big]/ \\
& \Big[(x(\beta(q)-1)^2 v_H(q) + (1-x)(\beta(q))^2 v_L(q))^2\beta'(q)\Big].
\end{aligned}
$$

Both $g(x,q)$ and $g_x(x,q)$ are bounded if their denominators are bounded away from zero. We show that this is true if $q < 1$ and $x \le x^E(q)$ by showing that it holds at $x = x^E$. First for the denominator of $g(x,q)$ we have:

$$x^E(q)(\beta(q)-1)^2 v_H(q) + (1-x^E(q))(\beta(q))^2 v_L(q) < 0, \tag{27}$$

for all $q \in [0,1)$. Notice that the left-side is increasing in x and hence (27) implies the same inequality for all lower x. The condition (27) is equivalent with

$$\frac{-(\beta(q)-1)\beta(q)v_H(q)v_L(q)}{\beta(q)v_L(q) - (\beta(q)-1)v_H(q)} < 0$$

which is true because the numerator is positive (other terms are positive except $v_L(q) < 0$) and the denominator is negative. Together with $\beta'(q) < 0$, this implies that the denominator of g is strictly positive and bounded away from zero. We can also conclude that both g and g_x are bounded and continuous in both x and q for all (x,q) such that $q < 1$ and $x \le x^E(q)$. Hence g is Lipschitz continuous for all $q < 1$.

To see that $g(x,q) > 0$, it is now enough to show that the numerator of (9) is strictly positive. First notice that the second term inside the brackets is always positive but the first term can be negative.[19] The first term is scaled by x, while

[19]This follows from $v_L(q) < 0, v_\omega'(q) < 0, \beta'(q) < 0, \beta(q) > 1$ and that $\beta(q)\beta''(q) > 2(\beta'(q))^2$.

the second therm is scaled by $(1-x)$. Therefore, if the numerator is positive at a belief above the boundary, it must be positive for the belief at the boundary as well. Since the decentralized belief, $x^E(q)$, is always above the fully optimal boundary, we can use it to show that the numerator is positive.

Plugging in $x^E(q)$ to the numerator of (9) and dividing by $x(1-x)$ gives:

$$\frac{\beta(q)v_L(q)\left(\beta'(q)\left(\beta(q)-1\right)v'_H(q)-\left(\left(\beta(q)-1\right)\beta''(q)-2\left(\beta'(q)\right)^2\right)v_H(q)\right)}{\beta(q)v_L(q)+(1-\beta(q))v_H(q)}$$
$$+\frac{\left(1-\beta(q)\right)v_H(q)\left(\beta'(q)\beta(q)v'_L(q)-\left(\beta(q)\beta''(q)-2\left(\beta'(q)\right)^2\right)v_L\left(q\right)\right)}{\beta\left(q\right)v_L\left(q\right)+\left(1-\beta\left(q\right)\right)v_H\left(q\right)}.$$

Since the denominator is negative ($v_L < 0$ and $\beta > 1$), this is proportional to

$$[v_H(q)v'_L(q)-v'_H(q)v_L(q)]\beta'(q)\beta(q)(\beta(q)-1)-2v_H(q)v_L(q)(\beta'(q))^2,$$

which is always positive because $v_H(q) > 0$ and $v_L(q), v'_H(q), v'_L(q) < 0$. Hence, $q(x,q) > 0$ for all $q \in [0,1)$ and $x \leq x^E(q)$.

Similar direct calculations show that $g_x > 0$ for all (x,q) such that $q < 1$ and $x \leq x^E(q)$.

Next, insert $x^E(q)$ to (dropping all dependencies) (9):

$$\begin{aligned}
g(x^E(q),q) =& \frac{\frac{-\beta(1-\beta)v_Lv_H}{(\beta v_L+(1-\beta)v_H)^2}}{\frac{\beta'\beta(1-\beta)v_Lv_H}{\beta v_L+(1-\beta)v_H}}\Bigg(\frac{\beta'\beta(1-\beta)(v_Lv'_H-v'_Lv_H)}{\beta v_L+(1-\beta)v_H} \\
&+\frac{\beta v_Lv_H(-2\beta'^2+(\beta-1)\beta'')}{\beta v_L+(1-\beta)v_H}+\frac{(\beta-1)v_Lv_H(-2\beta'^2+\beta\beta'')}{\beta v_L+(1-\beta)v_H}\Bigg) \\
=&\frac{v_H\left(2v_L\beta'-(\beta-1)\beta v'_L\right)+(\beta-1)\beta v_Lv'_H}{((\beta-1)v_H-\beta v_L)^2}.
\end{aligned}$$

The derivative of the decentralized policy x^E is

$$x^{E'}(q) = \frac{v_H\left(v_L\beta'-(\beta-1)\beta v'_L\right)+(\beta-1)\beta v_Lv'_H}{((\beta-1)v_H-\beta v_L)^2}.$$

By subtracting $x^{E'}(q)$ from $g(x^E(q),q)$, we get

$$g(x^E(q),q)-x^{E'}(q)=\frac{\beta'(q)v_L(q)v_H(q)}{(\beta(q)v_L(q)+(1-\beta(q))v_H(q))^2}.$$

This expression is strictly positive for $q < 1$ and goes to zero as q goes to 1 (since $v_H(q) \to 0$). □

Proof of Lemma 7. If the claim is not true, there must be some x and $q > q^*(x)$ such that

$$V_q(x,q) + xv_H(q) + (1-x)v_L(q) > 0. \tag{28}$$

We show that this leads to a contradiction by showing that (28) implies $V_{qx}(x,q) + v_H(q) - v_L(q) > 0$, which further implies that (28) holds also for all beliefs in $[x, x^*(q)]$, including $V_q(x^*(q), q) + x^*(q)v_H(q) + (1 - x^*(q))v_L(q) > 0$, which contradicts the value matching condition (21).

It remains to show that (28) implies $V_{qx}(x,q) + v_H(q) - v_L(q) > 0$. First notice that $V_q(x,q) = B_q(q)\Phi(x,q) + B(q)\Phi_q(x,q)$, which then together with (28) implies

$$B_q > -\frac{\Phi_q}{\Phi}B - \frac{xv_H + (1-x)v_L}{\Phi}$$

where we have left out all dependencies to simplify notation. We now get the following lower bound:

$$\begin{aligned}
V_{qx} + v_H - v_L &= B_q\Phi_x + B\Phi_{qx} + v_H - v_L \\
&> -\frac{\Phi_q\Phi_x}{\Phi}B - \frac{\Phi_x}{\Phi}(xv_H + (1-x)v_L) + B\Phi_{qx} + v_H - v_L \\
&= \Phi^{-1}[B(\Phi_{qx}\Phi - \Phi_q\Phi_x) + \Phi(v_H - v_L) - \Phi_x(xv_H + (1-x)v_L)].
\end{aligned} \tag{29}$$

The first term can be simplified as

$$\begin{aligned}
\Phi^{-1}B(\Phi_{qx}\Phi - \Phi_q\Phi_x) &= \frac{B\Phi\beta'}{x(1-x)} = \frac{\Phi\beta'}{x(1-x)}\frac{\Phi_x^*(x^*v_H + (1-x^*)v_L) - \Phi^*(v_H - v_L)}{\Phi_{qx}^*\Phi^* - \Phi_q^*\Phi_x^*} \\
&= \frac{x^*(1-x^*)}{x(1-x)}\frac{\Phi}{\Phi^*\Phi^*}[\Phi_x^*(x^*v_H + (1-x^*)v_L) - \Phi^*(v_H - v_L)],
\end{aligned}$$

where the notation Φ^* refers to $\Phi(x^*(q), q)$.

Now, (29) becomes

$$\begin{aligned}
&\frac{x^*(1-x^*)}{x(1-x)}\frac{\Phi}{\Phi^*\Phi^*}[\Phi_x^*(x^*v_H + (1-x^*)v_L) - \Phi^*(v_H - v_L)] \\
&- \frac{1}{\Phi}[\Phi_x(xv_H + (1-x)v_L) - \Phi(v_H - v_L)] \\
&= \frac{1}{x(1-x)}\Big(\frac{\Phi}{\Phi^*}((\beta-1)x^*v_H + \beta(1-x^*)v_L) - ((\beta-1)xv_H + \beta(1-x)v_L)\Big),
\end{aligned} \tag{30}$$

where we have used the following for both terms inside the brackets:

$$\Phi(v_H - v_L) - \Phi_x(xv_H + (1-x)v_L) = \Phi(v_H - v_L) - \Phi\frac{\beta - x}{x(1-x)}(xv_H + (1-x)v_L)$$
$$= \frac{-\Phi}{x(1-x)}((\beta-1)xv_H + \beta(1-x)v_L).$$

To conclude that (30) is larger than 0, notice first that $(\beta-1)xv_H+\beta(1-x)v_L < 0$ whenever $x < x^E(q)$ and that it is increasing in x. Then observe that $\Phi/\Phi^* \in (0,1)$ and hence $(\beta-1)xv_H + \beta(1-x)v_L < (\Phi/\Phi^*)((\beta-1)x^*v_H + \beta(1-x^*)v_L)$.

We conclude that $V_q + xv_H + (1-x)v_L > 0$ implies $V_{qx} + v_H - v_L > 0$ and the proof is complete. $\square$

Proof of Lemma 8. Fixing some (x,q) such that $q < q^*(x)$, differentiating (19) twice with respect to x, and simplifying gives:

$$\begin{aligned} V_{xx}(x,q) &= V_{xx}(x,q^*(x)) \\ &\quad +2(q^*)'(x)\left(V_{qx}(x,q^*(x)) + v_H(q^*(x)) - v_L(q^*(x))\right) \\ &\quad +(q^*)''(x)\left(V_q(x,q^*(x)) + xv_H'(q^*(x)) + (1-x)v_L'(q^*(x))\right) \\ &\quad +\left((q^*)'(x)\right)^2\left(V_{qq}(x,q^*(x)) + xv_H'(q^*(x)) + (1-x)v_L'(q^*(x))\right). \end{aligned} \tag{31}$$

The second term on the right-hand side vanishes by the value-matching condition (21) and the third term vanishes by the smooth-pasting condition (22). Let us look at the last term. First, since (21) holds along the boundary $(x, q^*(x))$, we can totally differentiate it with respect to x to get:

$$\begin{aligned} 0 &= V_{qx}(x,q^*(x)) + V_{qq}(x,q^*(x))(q^*)'(x) + v_H(q^*(x)) - v_L(q^*(x)) \\ &\quad + [xv_H'(q^*(x)) + (1-x)v_L'(q^*(x))](q^*)'(x). \end{aligned}$$

Applying (22), several terms disappear and this reduces to

$$V_{qq}(x,q^*(x)) + xv_H'(q^*(x)) + (1-x)v_L'(q^*(x)) = 0.$$

The last term in (31) vanishes as well, and it follows that $V_{xx}(x,q) = V_{xx}(x,q^*(x))$. Since this holds for any $q < q^*(x)$, it immediately implies that $V_{xxq}(x,q) = 0$. $\square$

Proof of Lemma 9. This is by direct computation. Recall that the value function for $q \geq q^*(x)$ is $V(x,q) = B(q)\Phi(x,q)$ and hence

$$V_{xxq} = B_q(q)\Phi_{xx}(x,q) + B(q)\Phi_{xxq}(x,q).$$

Plugging in the expressions for $B_q(q)$ and $B(q)$ from (16) and (17), multiplying by r, simplifying, and evaluating at $q = q^*(x)$ gives:

$$\begin{aligned} rV_{xxq}\left(x, q^*\left(x\right)\right) &= \frac{\left(\beta\left(q^*\left(x\right)\right) - 1\right)^2 + x\left(1-x\right)}{x^2\left(1-x\right)^2} v_L\left(q^*\left(x\right)\right) \\ &\quad - \frac{\beta\left(q^*\left(x\right)\right)\left(\beta\left(q^*\left(x\right)\right) - 1\right)}{x^2\left(1-x\right)^2}\left(v_H\left(q^*\left(x\right)\right) - v_L\left(q^*\left(x\right)\right)\right). \end{aligned}$$

Noting that $\beta\left(q^*\left(x\right)\right) > 1$, $v_L\left(q^*\left(x\right)\right) < 0$, and $v_H\left(q^*\left(x\right)\right) - v_L\left(q^*\left(x\right)\right) > 0$, it follows that $V_{xxq}\left(x, q^*\left(x\right)\right) < 0$.

□

C.3 Proof of Proposition 3

Proof. First, recall that $x^*(0) < x^E(0) = x^{stat}(0)$ by the proof of Proposition 2. Using this together with the continuity of the policy functions we find 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 $x^*_q(1) < x^{stat}_q(1)$. The first part is immediate. For the second part, use Lemma 6 and the uniqueness of the solution to get $x^*_q(1) = x^E_q(1)$. Now it is enough to show that the derivative of the equilibrium is smaller than of the myopic 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 myopic and optimal solutions meet at $q = 1$ but the optimal solution reaches the point above the myopic 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 $\underline{x}$. □

C.4 Proof of Proposition 4

Proof. Part (a): We show the result by contradiction. By using the solution from Proposition 2 and the value function derived in its proof, we show that $q^* = 0$ cannot maximize the HJB equation (5) in the limit as $\sigma \to 0$ unless $\sqrt{q^*_\sigma(x)}/\sigma \to \infty$. If $q^*(x)$ goes to any other value than 0, the claim immediately follows.

By taking the first order condition from (5), we get

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

The first order condition is necessarily strictly positive at $q^* = 0$ in the limit as $\sigma \to 0$ once we show that $V_{xx}(x,q) > 0$ and $V_{xxq}(x,q)$ is finite.

Recall that the value function is $V(x,q) = B(q)\Phi(x,q)$ and its derivatives are then $V_{xx} = B(q)\Phi_{xx}$ and $V_{xxq} = B_q(q)\Phi_{xx} + B(q)\Phi_{xxq}$. By plugging in the values of Φ_{xx}, we get

$$V_{xx} = B(q)\beta(q)\Phi\frac{(\beta(q)-1)}{x^2(1-x)^2}.$$

We know that $B > 0$ for all $q < 1$ in the optimal solution and that $\Phi > 0$ for all $x \in (0,1)$. Then, $V_{xx} > 0$ whenever $\beta > 1$ which is true whenever the signal-to-noise ration is finite.

We can write V_{xxq} as

$$V_{xxq} = \frac{(\Phi_x\Phi_{xxq} - \Phi_{qx}\Phi_{xx})(xv_H + (1-x)v_L)}{\Phi\Phi_{qx} - \Phi_q\Phi_x} + \frac{(\Phi_q\Phi_{xx} - \Phi\Phi_{xxq})(v_H - v_L)}{\Phi\Phi_{qx} - \Phi_q\Phi_x}.$$

The first term equals $\frac{(\beta-x)^2+x(1-x)}{x^2(1-x)^2}(xv_H + (1-x)v_L)$ and the second term equals $-\frac{\beta+(\beta-1)ln\left(\frac{x}{1-x}\right)}{x(1-x)}(v_H - v_L)$. Both are finite for all $x \in (0,1)$.

Hence, we conclude that for the first order condition to be satisfied, we must have $\sqrt{q^*_\sigma(x)}/\sigma \to \infty$ as $\sigma \to 0$.

Part (b): We fix the belief to be $x \in (0,1)$. By rearranging the solution in Proposition 1, we get

$$\beta(q) = \frac{xv_H(q)}{xv_H(q) + (1-x)v_L(q)}.$$

We take the limit $\lim_{\sigma\to 0}\beta(q_\sigma^E(x)) = \frac{xv_H(0)}{xv_H(0)+(1-x)v_L(0)}$, which is strictly larger than 1 for all $x > x^{stat}(q)$ and hence further implying that $\lim_{\sigma\to 0}\sqrt{q_\sigma^E(x)}/\sigma < \infty$. More precisely, we get the limit of the signal-to-noise ratio as $a(x)$ satisfying $\frac{xv_H(0)}{xv_H(0)+(1-x)v_L(0)} = \frac{1}{2}\left(1+\sqrt{1+8ra(x)^{-2}}\right)$. □

C.5 Proof of Proposition 5

Take an arbitrary boundary policy $\widetilde{q}\,(x)$ with inverse $\widetilde{x}\,(q)$. The long-run stock, denoted q_∞, is equal to $\widetilde{q}\,(\overline{x})$, where $\overline{x} := \sup\left(x_t\,|t>0\right)$ is the long-run maximum value of the belief. Deriving the distribution of the long-run stock boils down to deriving the distribution of the maximum value of the belief. We do that utilizing the fact that the belief process x_t is a martingale with continuous path that eventually converges to truth.

Denote the initial belief by x_0, and consider some $x' \in (x_0, 1)$. Let $\tau\,(x') :=$ $\inf\left(t : x_t \geq x'\right)$ denote the time of reaching belief x' (with the convention $\tau_{x'} = \infty$ if x' is never reached). Since x_t has continuous path and will converge to either 0 or 1 (depending on true state), $x_{\tau(x')}$ is a random variable that takes value either x' or 0. By Doob's optional sampling theorem, we have

$$x_0 = \mathbb{E}\left(x_{\tau(x')}\right) = \Pr\left(\overline{x} \geq x'\right)\cdot x' + \Pr\left(\overline{x} < x'\right)\cdot 0,$$

from which we can solve

$$\Pr\left(\overline{x} \geq x'\right) = \frac{x_0}{x'}.$$

On the other hand, we can write

$$\Pr\left(\overline{x} \geq x'\right) = x_0 \Pr\left(\overline{x} \geq x'\,|\omega = H\right) + (1-x_0)\Pr\left(\overline{x} \geq x'\,|\omega = L\right)$$

and since we know that the belief converges to truth, we have

$$\Pr\left(\overline{x} \geq x'\,|\omega = H\right) = 1.$$

Using the equations above, we can then solve for:

$$\Pr\left(\overline{x} \geq x' \left| \omega = L\right.\right) = \frac{x_0\left(1-x'\right)}{x'\left(1-x_0\right)}$$

and so

$$\Pr\left(\overline{x} \leq x' \left| \omega = L\right.\right) = 1 - \Pr\left(\overline{x} \geq x' \left| \omega = L\right.\right) = \frac{x'-x_0}{x'\left(1-x_0\right)}.$$

Noting that

$$\Pr\left(q_\infty \leq q \left| \omega\right.\right) = \Pr\left(\overline{x} \leq \widetilde{x}\left(q\right) \left| \omega\right.\right),$$

we get the long-run distribution of the stock given in the proposition.

D Mechanism design

D.1 Proof of Proposition 6

Proof. Fix policy Q and posted price $P\left(x,q\right)$ as in Proposition 6. To show that Q is a decentralized equilibrium, we have to show that it is optimal for type θ to stop at $\tau\left(\theta\right) := \inf\left\{t : q_t \geq 1 - F\left(\theta\right)\right\}$. We call $\tau\left(\theta\right)$ the intended stopping time for type θ.

For an individual player this is a Markovian stopping problem with fixed policy Q and stopping payoff given by (10). From the structure of the problem it is clear that it is optimal for θ to stop at the first hitting time of *some* boundary point. To see this, note that whenever the state (x,q) is above the boundary, it moves immediately to the boundary point $(x,\widetilde{q}(x))$ and since we have defined $P\left(x,q\right) = P\left(x,\widetilde{q}\left(x\right)\right)$ above the boundary, any player is indifferent between stopping at (x,q) or $(x,\widetilde{q}(x))$ (or any point between those). Furthermore, we have defined $P\left(x,q\right)$ to be so large below the boundary that no player wants to stop there. Since every boundary point is an intended stopping point for some $\theta \in \left[\underline{\theta},\overline{\theta}\right]$, we only have to show that it is better for θ to stop at the intended stopping time $\tau\left(\theta\right)$ than at the intended stopping time for some other type $\tau\left(\widetilde{\theta}\right)$, $\widetilde{\theta} \neq \theta$.

Denote by $U\left(\theta,\widetilde{\theta}\right)$ the expected value at time zero for type θ who intends to stop at $\tau\left(\widetilde{\theta}\right)$, $\widetilde{\theta} \neq \theta$. To complete the proof, we must show that $U(\theta,\theta) \geq U(\theta,\widetilde{\theta})$ for all θ and for all $\widetilde{\theta}$.

We next proceed to derive the expression for $U(\theta,\widetilde{\theta})$. Denote by $P_0\left(\widetilde{\theta}\right)$ the expected discounted transfer payment that θ has to pay if she stops at the intended stopping time $\tau\left(\widetilde{\theta}\right)$:

$$P_0\left(\widetilde{\theta}\right) := \mathbb{E}\left[e^{-r\tau\left(\widetilde{\theta}\right)}P\left(x_{\tau\left(\widetilde{\theta}\right)}, q_{\tau\left(\widetilde{\theta}\right)}\right)\right].$$

Using (11), this can be written as:

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

With this, we get an expression for $U\left(\theta,\widetilde{\theta}\right)$:

$$U\left(\theta,\widetilde{\theta}\right) = \mathbb{E}\Big[e^{-r\tau\left(\widetilde{\theta}\right)}(x_{\tau\left(\widetilde{\theta}\right)}v_H(\theta) + (1 - x_{\tau\left(\widetilde{\theta}\right)})v_L(\theta))\Big] - P_0\left(\widetilde{\theta}\right).$$

Its partial derivative with respect to θ is

$$U_1\left(\theta,\widetilde{\theta}\right) = \mathbb{E}\Big[e^{-r\tau\left(\widetilde{\theta}\right)}(x_{\tau\left(\widetilde{\theta}\right)}v'_H(\theta) + (1 - x_{\tau\left(\widetilde{\theta}\right)})v'_L(\theta))\Big]. \tag{32}$$

The key property that we want to prove is that $U_1\left(\theta,\widetilde{\theta}\right)$ is increasing in $\widetilde{\theta}$. To do that, note that applying the law of iterated expectations, we can write $U_1\left(\theta,\widetilde{\theta}\right)$ in terms of the initial belief x_0 as

$$\begin{aligned} U_1\left(\theta,\widetilde{\theta}\right) &= x_0\mathbb{E}\left(e^{-r\tau\left(\widetilde{\theta}\right)}v'_H(\theta)\,|\omega = H\right) + (1 - x_0)\,\mathbb{E}\left(e^{-r\tau\left(\widetilde{\theta}\right)}v'_L(\theta)\,|\omega = L\right) \\ &= x_0v'_H(\theta)\,\mathbb{E}\left(e^{-r\tau\left(\widetilde{\theta}\right)}\,|\omega = H\right) + (1 - x_0)\,v'_L(\theta)\,\mathbb{E}\left(e^{-r\tau\left(\widetilde{\theta}\right)}\,|\omega = L\right). \end{aligned}$$

Since $v'_H(\theta) \geq 0$ and $v'_L(\theta) \geq 0$ (with at least one of the inequalities strict), both terms in the above expression are positive. Type $\widetilde{\theta}$ enters the expression only through the discounting terms $\mathbb{E}\left(e^{-r\tau\left(\widetilde{\theta}\right)}\,|\omega\right)$. Since $\theta'' > \theta'$ implies that $\tau(\theta'') := \inf\{t : q_t \geq 1 - F(\theta'')\} \leq \tau(\theta') := \inf\{t : q_t \geq 1 - F(\theta')\}$ with probability 1, it follows that $\mathbb{E}\left(e^{-r\tau\left(\widetilde{\theta}\right)}\,|\omega\right)$ is increasing in $\widetilde{\theta}$ irrespective of state ω, and hence $U_1\left(\theta,\widetilde{\theta}\right)$ is increasing in $\widetilde{\theta}$ as well.

We now utilize this property to complete the proof. Note first that we have:

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

Therefore, for arbitrary θ' and θ'', we have

$$U(\theta'',\theta'') - U(\theta',\theta') = \mathbb{E}\left[\int_{\theta'}^{\theta''} e^{-r\tau(s)}\Big(x_{\tau(s)}v'_H(s) + (1 - x_{\tau(s)})v'_L(s))\Big)ds\right]. \quad (33)$$

We can now write:

$$\begin{aligned}
U\left(\theta,\widetilde{\theta}\right) &= U\left(\widetilde{\theta},\widetilde{\theta}\right) + \int_{\widetilde{\theta}}^{\theta} U_1\left(s,\widetilde{\theta}\right) ds \\
&\leq U\left(\widetilde{\theta},\widetilde{\theta}\right) + \int_{\widetilde{\theta}}^{\theta} U_1\left(s,s\right) ds = \\
&= U\left(\widetilde{\theta},\widetilde{\theta}\right) + \mathbb{E}\left[\int_{\widetilde{\theta}}^{\theta} e^{-r\tau(s)}\Big(x_{\tau(s)}v'_H(s) + (1 - x_{\tau(s)})v'_L(s))\Big)ds\right] \\
&= U\left(\theta,\theta\right),
\end{aligned}$$

where the inequality uses the property that $U_1\left(\theta,\widetilde{\theta}\right)$ is increasing in $\widetilde{\theta}$, the second last equality uses (32), and the last equality uses (33).

□

D.2 Proof of Lemma 3

Proof. Here, we show how to derive the virtual valuation representation for the designer's value. Suppose transfers follow an arbitrary policy, P_τ, adapted to $\mathcal{F}_t$. The designer's expected revenue can be written as

$$\begin{aligned}
&\int_{\underline{\theta}}^{\overline{\theta}} \mathbb{E}\Big[e^{-r\tau(\theta)}\Big(x_{\tau(\theta)}v_H(\theta) + (1 - x_{\tau(\theta)})v_L(\theta) - W(\theta, x_{\tau(\theta)})\Big)\Big]f(\theta)d\theta \\
&= \int_{\underline{\theta}}^{\overline{\theta}} \Big(\mathbb{E}\Big[e^{-r\tau(\theta)}\Big(x_{\tau(\theta)}v_H(\theta) + (1 - x_{\tau(\theta)})v_L(\theta)\Big)\Big]f(\theta)d\theta \\
&- \int_{\underline{\theta}}^{\overline{\theta}} \mathbb{E}\Big[\int_{\underline{\theta}}^{\theta(q)} e^{-r\tau(s)}\Big(x_{\tau(s)}v'_H(s) + (1 - x_{\tau(s)})v'_L(s)\Big)ds\Big]f(\theta)d\theta,
\end{aligned}$$

where we have used the envelope theorem for the agent's value $W(\theta, x)$. We can use Fubini's theorem to change the order of integration in the second term:

$$\begin{aligned}
&\int_{\underline{\theta}}^{\overline{\theta}} \mathbb{E}\Big[\int_{\underline{\theta}}^{\theta(q)} e^{-r\tau(s)}\Big(x_{\tau(s)}v'_H(s) + (1 - x_{\tau(s)})v'_L(s)\Big)ds\Big]f(\theta)d\theta \\
=&\mathbb{E}\Big[\int_{\underline{\theta}}^{\overline{\theta}}\int_{s}^{\overline{\theta}} e^{-r\tau(s)}\Big(x_{\tau(s)}v'_H(s) + (1 - x_{\tau(s)})v'_L(s)\Big)f(\theta)d\theta ds\Big] \\
=&\mathbb{E}\Big[\int_{\underline{\theta}}^{\overline{\theta}} e^{-r\tau(s)}\Big(x_{\tau(s)}v'_H(s) + (1 - x_{\tau(s)})v'_L(s)\Big)(1 - F(s))ds\Big].
\end{aligned}$$

The rest is simply to plug the above expression back into the designer's payoff and to write the integral over quantities rather than types (where we use $1 - F(\theta(q)) = q$). The profit maximizing designer's objective becomes

$$\mathbb{E}\Big[\int_0^1 e^{-r\tau(q)}(x_{\tau(q)}\phi_H(q) + (1 - x_{\tau(q)})\phi_L(q))dq\Big],$$

where $\tau(q)$ is the stopping time of the q highest type buyer and $\phi(q)$ is his virtual valuation:

$$\phi_\omega(q) := v_\omega(\theta(q)) - v'_\omega(\theta(q))\frac{1 - F(\theta(q))}{f(\theta(q))}.$$

□

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.

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.

Cetemen, Doruk, Can Urgun, and Leeat Yariv, "Collective Progress: Dynamics of Exit Waves," *Journal of Political Economy*, 2023, *accepted.*

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.

Décamps, Jean-Paul and Thomas Mariotti, "Investment timing and learning externalities," *Journal of Economic Theory*, 2004, *118* (1), 80–102.

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.

Laiho, Tuomas and Julia Salmi, "Coasian Dynamics and Endogenous Learning," *Working paper*, 2023.

_ , Pauli Murto, and Julia Salmi, "Timing of Irreversible Actions under Informational and Payoff Externalities," *Working paper*, 2023.

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.

Peskir, G., A. Shiryaev, and A.N. Shirayev, *Optimal Stopping and Free-Boundary Problems* number nid. 10. In 'Lectures in Mathematics. ETH Zürich.', SPRINGER VERLAG NY, 2006.

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.