Theoretical Economics 20 (2025), 93–130 1555-7561/20250093

Gradual learning from incremental actions

Tuomas Laiho
Department of Economics, Aalto University School of Business and Ministry of Finance

Pauli Murto
Department of Economics, Aalto University School of Business and Helsinki Graduate School of Economics

Julia Salmi
Department of Finance and Economics, Hanken School of Economics and Helsinki Graduate School of Economics

We introduce a collective experimentation problem where a continuum of agents choose the timing of irreversible actions under uncertainty and where public feedback from the actions arrives gradually over time. The leading application is the adoption of new technologies. 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 ignore the future actions of other agents, which lets us recast the complicated two-dimensional problem as a series of one-dimensional problems.

Keywords. Social learning, experimentation, optimal stopping, technology adoption.

JEL classification. C61, C73, D82, D83.

1. Introduction

Innovation adoption decisions have long-run consequences that can be observed only gradually over time. Consider an individual firm contemplating whether to invest in a novel production technology or an individual consumer contemplating whether to pur-

Tuomas Laiho: tuomas.s.laiho@gmail.com
Pauli Murto: pauli.murto@aalto.fi
Julia Salmi: julia.salmi@hanken.fi
We thank Aislinn Bohren, Martin Cripps, Rahul Deb, Nils Christian Framstad, Dino Gerardi, Yingni Guo, Daniel Hauser, Johannes Hörner, 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. We are grateful to the referees for their helpful suggestions. This project has received financial support from ERC Grant 683031 through PI Bård Harstad (Laiho), Academy of Finland (Murto), Emil Aaltonen Foundation, OP Group Research Foundation, Yrjö Jahnsson Foundation, and ERC Grant 682417 through PI Vasiliki Skreta (Salmi).

© 2025 The Authors. Licensed under the Creative Commons Attribution-NonCommercial License 4.0. Available at https://econtheory.org. https://doi.org/10.3982/TE5452

chase a novel durable good. Each new adopter begins to observe how well the technology functions in different situations and whether technical problems emerge over time. The experiences of the adopters spill over to potential future adopters through private communications, social media, or platforms that collect reliability statistics. An individual who has not yet made an adoption decision utilizes such information in deciding whether and when to adopt the new technology herself. Since past adopters continue to produce information over time, the current efficacy of learning is increasing in the number of past adopters.

In this paper, we analyze how this kind of endogenous gradual learning shapes the socially optimal path of incremental actions and we contrast this to a situation where the actions are decentralized. The key feature of our model is that each individual action has a long-run impact on the flow of information that is publicly observed.

In the model, a continuum of small agents decide whether and when to take an irreversible action (e.g., adopt the new innovation). We refer throughout the paper to the action as a decision to "stop." 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 that other agents observe 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.[1]

Our main question is how the incremental path of stopping decisions—the adoption path—is determined on the one hand in a decentralized equilibrium where agents optimize individually, and on the other hand, when a central planner coordinates the actions to maximize the expected welfare summed over the agents. The contribution of this paper is twofold. First, we develop a novel methodological approach with suitable solution techniques. Second, we analyze economic trade-offs that arise in the combination of gradual learning and irreversible decisions. Gradual learning creates a new trade-off for the socially optimal expansions: *the information generation effect* calls for aggressive expansion to improve information for future decisions and *the option value effect* calls for cautious expansion to have better information for the current decisions. Previous literature has studied environments where these two effects arise separately; we focus on the informational trade-off between them.[2] A social planner trades off these two effects dynamically over time, but individual agents internalize only the option value effect, and thus the decentralized equilibrium suffers from informational free-riding.

We approach experimentation under gradual learning by modeling the cumulative path of individual actions as a stock process, which controls the speed of learning. The micro-foundation for our specification is that each agent who has stopped produces a persistent stream of i.i.d. signals conditional on the true state. In continuous time,

[1]In reality individual adoption decisions are often costly to reverse, if not entirely irreversible, which is enough to add persistence in the social learning. We look at the extreme case of such persistence—adoption decisions that are infinitely costly to reverse—to understand clearly the theoretical implications of irreversibility and gradual learning.

[2]This informational trade-off would not arise in a model where learning is instantaneous rather than gradual because there would then be no scope for improving information by delaying adoption. Papers studying instantaneous learning include Bonatti (2011), Che and Hörner (2017), Frick and Ishii (2024), and Laiho and Salmi (2024).

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 that is increasing in 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 current stock and the current belief about the state affect the optimal decisions. 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 the optimal stopping decisions of agents who ignore that the other agents stop in the future, i.e., they take the stock as fixed on the current level. This property turns the two-dimensional problem into a series of one-dimensional stopping problems. Our proof for the equivalence between the two stopping problems builds on the fact that 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. Optimization under the assumption that the stock remains fixed does not work because the value of information depends on the expected future actions. We cannot solve the optimal policy in closed form, but we derive a nonlinear differential equation that determines the policy and allows us to characterize it. 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 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 decrease it for high beliefs due to the informational trade-off between information generation and the option value of waiting.

Solving the social planner's problem is an important part of our contribution. The problem is of independent interest also because it can be interpreted as the canonical problem of a single decision maker choosing how to expand a capital stock over time under uncertainty. The key difference to existing literature in this area (see, e.g., Dixit and Pindyck (1994)) is that in our model the uncertainty is resolved endogenously. This is precisely what causes the disparity between the social optimum and the decentralized equilibrium in 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. The option value effect does not arise in these papers because actions are reversible, and hence, learning always increases the level of optimal quantities relative to the no-learning benchmark. Strulovici (2010) shows that also collective decision making by voting has the effect of reducing experimentation below the socially optimal level.

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. In contrast to models with exogenous uncertainty, the social planner's solution diverges from the decentralized equilibrium.

A few papers investigate social learning with irreversible actions, which bears similarities with informational free-riding in our decentralized solution. Frick and Ishii (2024) 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 unlike in our model with gradual learning. In equilibrium, on the other hand, free-riding on the information generated by others creates an option value effect for individual agents, resulting in an 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 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 (2024) analyze monopoly pricing in a similar setup. Both in Che and Hörner (2017) and in Laiho and Salmi (2024), the presence of a social planner or a monopolist induces information generation effect. The crucial difference from the present paper is that there is no option value effect since the principle gets more information only by attracting new consumers.

Our assumption that learning is gradual implies that past actions matter for the current information flow. Two contemporaneous papers share this feature with us, although their models and key trade-offs are otherwise different from ours. Liski and Salanié (2023) 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, the present paper is related to the literature on innovation adoption. Traditionally, the theoretical literature has focused on explaining adoption patterns through

noninformational (Mansfield (1961), Farrell and Saloner (1986), and Jovanovic and Lach (1989)) or purely exogenous channels (Jensen (1982)). The few exceptions are Frick and Ishii (2024) (discussed above), Young (2009), and Wolitzky (2018). In Young (2009) and Wolitzky (2018), adopters are myopic, and hence, the option value effect does not arise. There is vast empirical evidence for social learning in innovation adoption, including Foster and Rosenzweig (1995), Duflo and Saez (2003), Munshi (2004), Bandiera and Rasul (2006), and Conley and Udry (2010).[3] The present paper contributes to this literature by proposing a tractable model that matches the key characteristic in the studied real-life settings: gradual learning from others' outcomes.

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: $v_H(\theta) \geq 0 > v_L(\theta)$.[4] 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 either $\omega = H$ or $\omega = L$ (or both).[5] 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. The payoff of not stopping (i.e., stopping at $t = \infty$) is zero. We normalize $v_H(\underline{\theta}) = 0$ so that type $\underline{\theta}$ is indifferent between stopping and never stopping if he is sure that $\omega = H$; types lower than $\underline{\theta}$ would be redundant since they would never want to stop.

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.

[3]The findings in Foster and Rosenzweig (1995), Munshi (2004), and Bandiera and Rasul (2006) support the importance of the option value effect and informational free-riding; individuals with good prospects to learn from others are less likely to be early adopters.

[4]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.

[5]This assumption ensures that the stopping payoff is *strictly* increasing in type for any interior belief about the state.

Specifically, the public learns about the state by observing a Brownian diffusion:[6]

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

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 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).[7]

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 $\sqrt{q_t}/\sigma$ 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.

2.3 *Solution concepts*

We will consider two outcomes of our model, one where individual agents choose their stopping times in a decentralized manner and one where a social planner chooses the stopping times in a centralized manner. In both cases, 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.

In the decentralized solution, 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 $\mathcal{F}_t$. Type θ solves

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

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

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

[6]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. Note that this formulation gives rise to a bounded rate of learning even when all the agents have stopped, i.e. when $q_t = 1$.

[7]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.

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) $\tau^E(\theta)$ solves (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)

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

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

In Section 3.5, we recast the problem of finding the social optimum as a control problem for the stock process $\{q_t\}$. This control problem is of independent interest for various applications as a single-agent experimentation model.

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*

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

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:*

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

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

Lemma 2 states that any nonmonotone stopping profile can be improved by rearranging the agents to stop in descending order of type while keeping the evolution of q_t unchanged.

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 stopping profiles where the agents stop in a descending order by type. It follows that 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 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 $\tilde{q}$, an increasing *policy function* $\tilde{x} : [0, 1] \to [0, 1]$, which maps the stock to the cutoff belief.[8] 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 toward 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.

[8]We use the term *policy* to describe a stock process Q that can depend on the news process in an arbitrary way. The term *policy function* defined here refers to the characterization of a particular type of policy—a boundary policy—in the (q, x)-space.

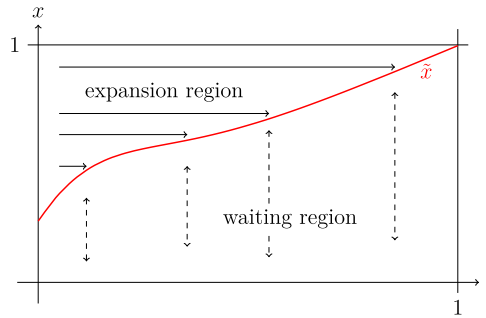

FIGURE 1. Dynamics in the waiting and expansion regions of the state space.

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: $xv_H(\theta) + (1-x)v_L(\theta) \geq 0$, or $x \geq -v_L(\theta)/(v_H(\theta) - v_L(\theta))$. Hence, the policy function associated with the no-learning benchmark is given by

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

where $v_\omega(q) := v_\omega(\theta(q))$.

Individually optimal 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 an equilibrium in the original problem is obtained by solving the problem with fixed stock separately for each individual type and tying these solutions together.[9] In effect this pins uniquely down a necessary condition for the threshold belief at which the "next" agents stops, given the current stock q_t, and hence, the procedure also establishes the uniqueness of the equilibrium. Intuitively, uniqueness arises because actions are strategic substitutes: an agent is less willing to stop if many other agents stop because then information arrives faster.

We elaborate here further the intuition for the equivalence between the problem with fixed stock and the original problem. 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 it is in the problem where stock is fixed. As this intuition suggests, the optimality of ignoring future changes in the stock is an equilibrium property and may well be violated against other (nonequilibrium) 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.[10] 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 a strictly increasing policy function x^E:*

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

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

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 the agent of type θ who assumes that the stock remains fixed at $q(\theta)$ forever. The term $\beta(q)$ reflects the cost of waiting for information. We have $\beta(q) > 1$ for all q, but $\lim_{q\to\infty} \beta(q) = 1$. The threshold $x^E(q(\theta))$ is decreasing in $\beta(q)$, which in turn is increasing in σ and r and decreasing in q.

[9]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 equilibrium behavior coincides with that of "myopic" investors who ignore the effect the future investments have on the price.

[10]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, Urgun, and Yariv (2023) for a similar equilibrium property in another context), but contrary to our setup there might be multiple equilibria.

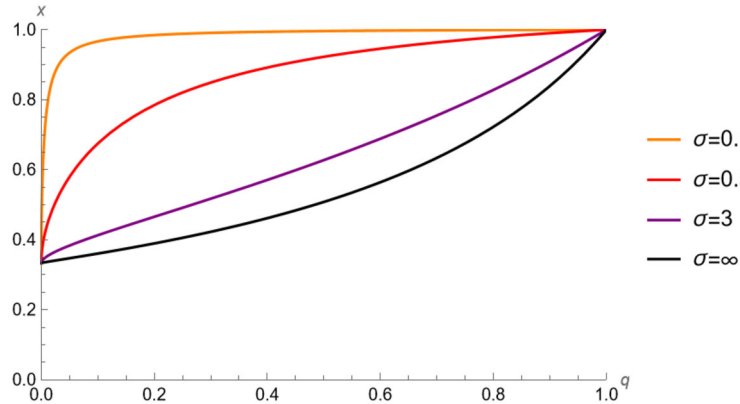

FIGURE 2. Equilibrium policy $x^E(q)$ for different σ when $v_H(q) = 1 - q$, $v_L(q) = -1/2$, and $r = 0.1$.

The decentralized equilibrium is a boundary policy: whenever the belief is about to cross the boundary $x^E(q)$, more agents stop. Notice that $x^E(q)$ 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 (by using that $\beta(q) > 1$):

COROLLARY 1. *The policy function in the decentralized equilibrium is strictly higher than the no-learning benchmark: $x^E(q) > x^{\text{myop}}(q)$ for all $\sigma \geq 0$ and all $q \geq 0$.*

Notice that the limit $\sigma \to 0$ corresponds to the case, where the state will be revealed immediately. In that limit, $x^E(q) \to 1$ for all $q > 1$. The agents who learn very quickly stop only once they are sure that the state is good.

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.

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 use notation $U(Q; x, q)$ to denote the expected total payoff of agents that have not yet stopped, given current state (x, q) and given a policy Q:

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

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

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

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

We derive here heuristically the solution to the HJB equation (the formal proof and the verification argument are in Appendix C). In principle, the optimal policy could consist of several waiting and expansion regions. We start by guessing that there is only one expansion and only one waiting region. Let $x^* : [0,1] \to [0,1]$ denote our candidate solution, which 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 also this property indeed holds.

We first pin down the functional form for the value function that solves the HJB equation (5) in the waiting region, i.e., below x^*. There it should be optimal to choose $q' = q$, and hence, (5) reduces to a differential equation:

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

This has a closed-form solution:[11]

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

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

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

The value function (6) captures the option value of the future actions for the planner in state (x, q).

The next step is to find functions B and x^* that make sure that the right side of the HJB equation is maximized everywhere. To do this, we first derive heuristically two additional conditions that will pin down a candidate for B and x^*, and prove the optimality of the candidate afterwards. To understand these conditions, imagine the social planner solving the problem in small successive steps, where each step consists of choosing the optimal time to add the next increment dq to the current stock q. The

[11]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 $x \to 1$.

planner's value function $V(x, q)$ encompasses the values of options to all future stock increments. At the time of adding dq, the planner obtains direct payoff increment $(x(q)v_H(q) - (1 - x(q))v_L(q))\,dq$, but at the same time foregoes the option to add that increment at some later moment thus inducing a change $V_q(x, q)\,dq$ in the value function. Requiring these to be in balance at the moment of hitting the threshold $x^*(q)$ gives a condition analogous to the value-matching condition in the literature of optimal stopping: $V_q(x^*(q), q) + x^*(q)v_H(q) + (1 - x^*(q))v_L(q) = 0$. This is an accounting equation that would have to hold irrespective of whether stock increment is undertaken at the optimal time instant or not, so we need a second condition to guarantee the *optimality* of the timing. Analogous to the smooth-pasting optimality condition in the literature of optimal stopping, we require that derivatives with respect to x of the two terms match at the threshold $x^*(q)$: $V_{qx}(x^*(q), q) + v_H(q) - v_L(q) = 0$.[12] Using equation (6), we can write these two conditions as

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

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

Since we have derived conditions (7)–(8) heuristically, we proceed in the spirit of guess-and-verify: we use them to pin down a candidate policy, but we will not rely on them when we prove the optimality of the candidate.

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

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

where

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

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 $x^*(q) \le x^E(q)$ for all $q \in [0, 1]$ (proof of Lemma 5 in Appendix C.2). This solution is our candidate for social optimum. We then take the following steps in Appendix C. First, we verify that, together with the value function in (6), the candidate solves the HJB equation. We then 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. We have the following.

[12] See Dixit and Pindyck (1994), Chapter 11.1A or Pindyck (1988) for a more formal justification of analogous conditions in a model with an exogenous stochastic process.

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*). In decentralized equilibrium, the agents only account for the former effect, so it is the *information generation effect* that induces the social planner to adopt faster than the decentralized equilibrium.

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 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.4 for the proof).

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

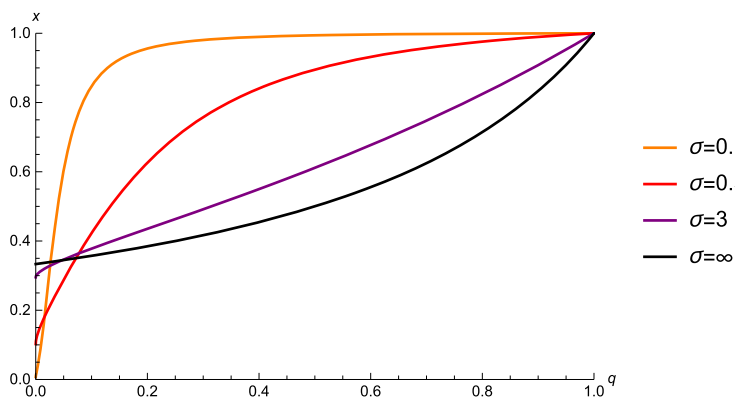

Figure 3. Socially optimal policy $x^*(q)$ for different σ when $v_H(q) = 1 - q$, $v_L(q) = -1/2$, and $r = 0.1$.

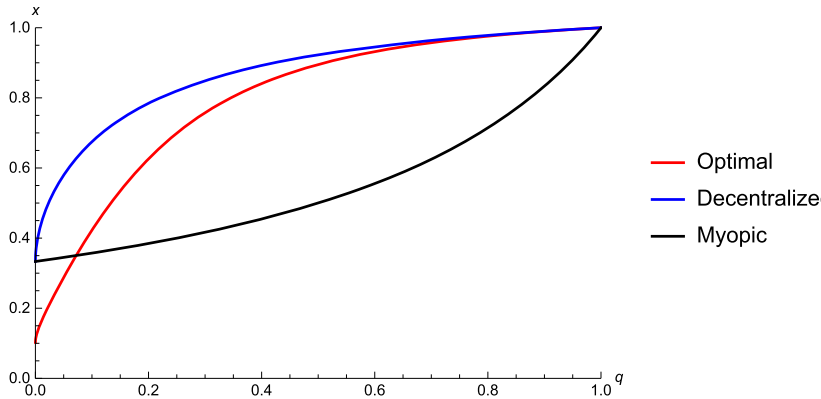

FIGURE 4. Different policies when $v_H(q) = 1 - q$, $v_L(q) = -1/2$, $\sigma = 0.5$, and $r = 0.1$.

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.

PROPOSITION 4. *(a) The socially optimal signal-to-noise ratio explodes as noise vanishes: $\sqrt{q^*_\sigma(x)}/\sigma \to \infty$ as $\sigma \to 0$ for all $x \in (0, 1)$. (b) The signal-to-noise ratio in decentralized equilibrium stays bounded as noise vanishes: $\sqrt{q^E_\sigma(x)}/\sigma \to a(x)$ as $\sigma \to 0$ where $a(x) = 0$ for all $x \le x^{\text{stat}}(0)$ and $a(x) \in (0, \infty)$ for all $x \in (x^{\text{stat}}(0), 1)$.*

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. In Appendix C.5, we prove Proposition 4 and derive the functional form for $a(x)$.

3.6 *Adoption path and long-run distribution of the stock*

Our model generates an S-shaped adoption path. We do not have a closed-form solution for the expected stock at a given time, but it is straightforward to generate one by simulation. Figure 5 shows the simulated average stock both in the decentralized equilibrium and in the social optimum, conditional on $\omega = H$. The adoption is first slow but then gets faster due to faster learning. Eventually, adoption slows as the marginal agent's valuation gets lower and the option value effect increases.

One can compute explicitly the probability distribution of the stock in the long-run for any boundary policy, including the decentralized equilibrium and social optimum.

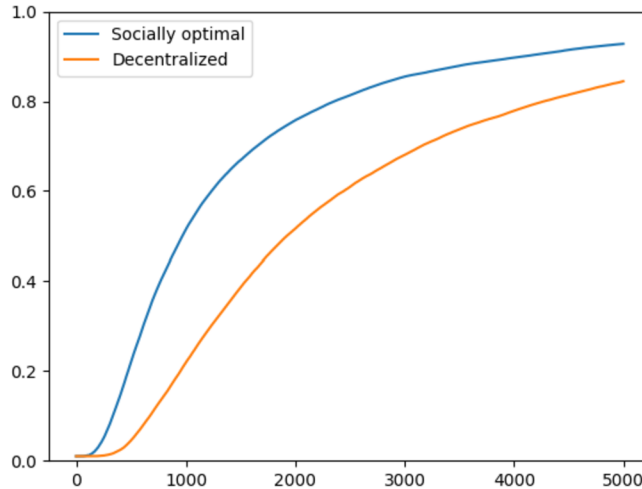

FIGURE 5. Expected adoption paths conditional on $\omega = H$. Parameters: $v_H(q) = 1 - q$, $v_L(q) = -1/2$, $\sigma = 0.5$, $r = 0.12$, $x_0 = 0.15$, $q_0 = 0.01$.

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 initial stock satisfies $q_{0^+} := \max(q_0, \widetilde{q}(x_0)) > 0$.*[13] *The probability distribution of the long-run stock is given by*

$$\Pr(q_\infty \le q|\omega = L) = \begin{cases} 0 & \text{if } q < q_{0^+}, \\ \dfrac{\widetilde{x}(q) - x_0}{\widetilde{x}(q)(1 - x_0)} & \text{if } q_{0^+} \le q < \widetilde{q}(1), \\ 1 & \text{if } q \ge \widetilde{q}(1), \end{cases}$$

$$\Pr(q_\infty \le q|\omega = H) = \begin{cases} 0 & \text{if } q < \widetilde{q}(1), \\ 1 & \text{if } q \ge \widetilde{q}(1). \end{cases}$$

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.*

[13] 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. Concluding remarks

Modeling gradual arrival of endogenous information 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.

Our model of gradual learning has also technical appeal as a tool for applied work. As demonstrated in this paper, the decentralized equilibrium can be solved in closed form. We believe that the solution method can be extended to richer environments, such as models where the stock controls a generic state process or where the actions of other players affect the profitability of stopping directly through payoff externalities.

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 fully reversible actions such as Bonatti (2011) are significantly different. A natural extension to our model would be to analyze what happens if stopping decisions are partially reversible. While we believe that many of the qualitative properties of our results stay the same, pursuing such an extension is beyond the scope of this paper.

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 i.i.d. 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 $\epsilon_t(\theta) \sim N(0, \sigma^2(\pi_H(\theta) - \pi_L(\theta))^2)$. The noise term is scaled so that every increment in q is equally informative. This assumption is not necessary: in an earlier working paper version, we have analyzed the case of heterogeneous informativeness and shown 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

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

for all stopping rules τ. Or equivalently,

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

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 $\mathcal{T}^{\text{mon}}$ 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 $\theta > \theta'$. □

Appendix 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:

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

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:

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

where $\tau(S_\theta) = \inf(t : (x_t, q_t) \in S_\theta)$ is the first hitting time of S_θ and $u_\theta(x) := xv_H(\theta) + (1 - x)v_L(\theta)$ 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 $(x_t, q_t)_{t\ge 0}$. Although 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

$$F_\theta(x, q) = A_\theta(q)\Phi(x, q) + B_\theta(q)\widetilde{\Phi}(x, q), \tag{10}$$

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

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

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)):[14]

$$\frac{\partial}{\partial q}\big[F_\theta(x,q)\big]_{x=x^E(q)} = 0. \tag{11}$$

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 3. *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*

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

The corresponding value function is

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

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

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

[14]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 $\frac{\partial}{\partial q}[F_\theta(x,q)]_{x=x^E(q)} \neq 0$, then the expected instantaneous rate of change in the value function, $\mathbb{E}[dF_\theta(x,q)]/dt$, would explode at the moment of hitting the boundary.

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 3, this simple strategy gives value

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

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

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

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 3 and, therefore, also $F_\theta(x, q) > u_\theta(x)$. Hence, stopping immediately cannot be optimal for θ. □

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 6, i.e.,

$$S_\theta = \big\{(x, q) : q \geq q(\theta), x \in \big[\widehat{x}_\theta(q), x^E(q)\big]\big\}. \tag{12}$$

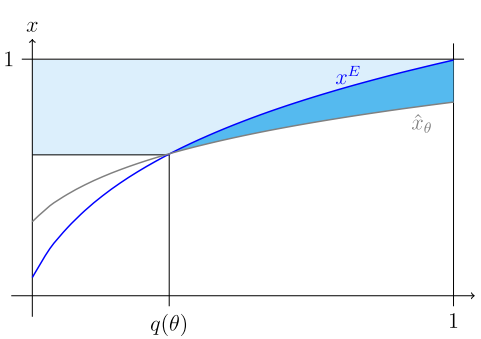

Figure 6. Optimal stopping for type θ.

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 4 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

$$\frac{\partial}{\partial x}\big[F_\theta(x, q)\big]_{x=x^E(q)} \geq \frac{\partial}{\partial x}\big[u_\theta(x)\big]_{x=x^E(q)}, \tag{13}$$

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 $(x^E(q), q) \notin S_\theta$, 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 (10) in the interval $(x', x^E(q))$ 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 $x \in (x', x^E(q))$, (13) 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 (10) 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, $\frac{\partial}{\partial x}[F_\theta(x, q)]_{x=x^E(q)} = A_\theta(q)\Phi_x(x^E(q), q)$. Our assumption $F_\theta(x^E(q), q) > u_\theta(x^E(q))$ is equivalent to

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

which further implies

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

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

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

which is the case if and only if $q \geq q(\theta)$. Noting that $\frac{\partial}{\partial x}[u_\theta(x)]_{x=x^E(q)} = v_H(\theta) - v_L(\theta)$, we may conclude that (13) holds in this case, too.

Given that (13) holds, the rate of change in $F_\theta(x, q)$ along the boundary is

$$\frac{d}{dq}F_\theta\big(x^E(q), q\big) = \frac{\partial}{\partial x}\big[F_\theta(x, q)\big]_{x=x^E(q)}\frac{d}{dq}x^E(q) + \frac{\partial}{\partial q}\big[F_\theta(x, q)\big]_{x=x^E(q)}$$

$$= \frac{\partial}{\partial x}\big[F_\theta(x,q)\big]_{x=x^E(q)} \frac{d}{dq} x^E(q)$$
$$\geq \frac{\partial}{\partial x}\big[u_\theta(x)\big]_{x=x^E(q)} \frac{d}{dq} x^E(q) = \frac{d}{dq} u_\theta\big(x^E(q)\big),$$

where the last term of the first line disappears by (11) and where the inequality follows from (13).

We have now shown that $F_\theta(x^E(q),q) > u_\theta(x^E(q))$ implies $\frac{d}{dq}F_\theta(x^E(q),q) \geq \frac{d}{dq}u_\theta(x^E(q))$. 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 6, i.e., $\{(x,q) : q \geq q(\theta), x \in [\widehat{x}_\theta(q), x^E(q)]\} \in S_\theta$, 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 $[\widehat{x}_\theta(q), x^E(q)]$.

We have now shown that the stopping rule defined in (12) 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 instantaneously to point $q_{0^+} := \{q : x^E(q) = x_0\}$. 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 6. 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 4, 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

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

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 3, 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 3. 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 $x^E(q) = x_t$.

Appendix C: Socially optimal policy

C.1 *Omitted calculations*

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

$$\Phi = \left(\frac{x}{1-x}\right)^{\beta(q)}(1-x), \qquad \Phi_q = \Phi\beta'(q)\ln\left(\frac{x}{1-x}\right),$$

$$\Phi_x = \Phi\frac{\big(\beta(q)-x\big)}{x(1-x)}, \qquad \Phi_{xx} = \Phi\frac{\beta(q)\big(\beta(q)-1\big)}{x^2(1-x)^2} = \Phi\frac{2r\sigma^2}{x^2(1-x)^2q},$$

$$\Phi_{xq} = \Phi\beta'(q)x^{-1}(1-x)^{-1}\left[1+\big(\beta(q)-x\big)\ln\left(\frac{x}{1-x}\right)\right],$$

$$\Phi_{xxq} = \Phi\frac{\beta'(q)}{x^2\big(1-x^2\big)}\left[\beta(q)+\big(\beta(q)-1\big)\left(1+\beta(q)\ln\left(\frac{x}{1-x}\right)\right)\right].$$

Deriving the differential equation We first show that the conditions (7) and (8) imply the differential equation in (9). Solving (7) and (8) for $B_q(q)$ and $B(q)$ yield

$$B_q(q) = A^1\big(x^*(q), q\big)x^*(q) + A^2\big(x^*(q), q\big), \tag{14}$$

$$B(q) = U^1\big(x^*(q), q\big)x^*(q) + U^2\big(x^*(q), q\big), \tag{15}$$

where

$$A^1(x, q) := \frac{-\Phi_{xq}(x, q)\big(v_H(q) - v_L(q)\big)}{\Phi(x, q)\Phi_{xq}(x, q) - \Phi_q(x, q)\Phi_x(x, q)},$$

$$A^2(x, q) := \frac{\Phi_{xq}(x, q)\big(-v_L(q)\big) + \Phi_q(x, q)\big(v_H(q) - v_L(q)\big)}{\Phi(x, q)\Phi_{xq}(x, q) - \Phi_q(x, q)\Phi_x(x, q)},$$

$$U^1(x,q) := \frac{\Phi_x(x,q)\big(v_H(q)-v_L(q)\big)}{\Phi(x,q)\Phi_{xq}(x,q)-\Phi_q(x,q)\Phi_x(x,q)},$$

$$U^2(x,q) := \frac{-\Phi_x(x,q)\big(-v_L(q)\big)-\Phi(x,q)\big(v_H(q)-v_L(q)\big)}{\Phi(x,q)\Phi_{xq}(x,q)-\Phi_q(x,q)\Phi_x(x,q)}.$$

Differentiating (15) with respect to q and using the chain rule gives

$$\begin{aligned} B_q(q) = {} & \big[U_x^1\big(x^*(q),q\big)x^{*\prime}(q)+U_q^1\big(x^*(q),q\big)\big]x^*(q)+U^1\big(x^*(q),q\big)x^{*\prime}(q) \\ & + U_x^2\big(x^*(q),q\big)x^{*\prime}(q)+U_q^2\big(x^*(q),q\big) \end{aligned} \tag{16}$$

Equating (14) and (16), solving for $x^{*\prime}(q)$, and simplifying yield the expression (9) in the text.

C.2 *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. The candidate is continuous and strictly increasing, and hence, defines a boundary policy. In part 2, we show that our candidate policy $x^*(q)$ satisfies the HJB equation (5). In part 3, we verify that the solution to the HJB equation solves the original problem.

Part 1: Solution to the initial value problem (9) We first establish some key properties of function g in (9).

Lemma 5. *For all (x,q) such that $q<1$ and $x \le x^E(q)$, function $g(x,q)$ in (9) is strictly positive, strictly increasing in x, and Lipschitz continuous. Furthermore, $g(x^E(q),q) > x^{E\prime}(q)$ for $q<1$, and $\lim_{q\to 1} g(x^E(q),q) = x^{E\prime}(1)$.*

The proof is by straightforward inspection of the properties of g in (9) and of $x^{E\prime}(q)$ in Proposition 1 and we relegate it to Appendix C.3.

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 $x(q_1)\le x^E(q_1)$ 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 $\lim_{q\to 1} g(x^E(q),q) = x^{E\prime}(1)$ 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 $g(x^E(q),q) > x^{E\prime}(q)$ 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 $x^*(q)\le x^E(q)$ for all $x\le q$. 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 $x \leq x^*(0)$. By construction of this policy, the value function of the planner following x^* is

$$V^*(x, q) = \begin{cases} \int_q^{q^*(x)} \big(xv_H(s) + (1-x)v_L(s)\big)\, ds + B\big(q^*(x)\big)\Phi\big(x, q^*(x)\big) & \text{for } q < q^*(x) \\ B(q)\Phi(x, q) & \text{for } q \geq q^*(x), \end{cases} \tag{17}$$

where $\Phi(x, q) = x^{\beta(q)}(1-x)^{1-\beta(q)}$ and $B(q)$ is given by (15), which simplifies to

$$B(q) = \frac{x^*(q)\big(\beta(q) - 1\big)v_H(q) + \big(1 - x^*(q)\big)\beta(q)v_L(q)}{\Phi\big(x^*(q), q\big)\beta'(q)}. \tag{18}$$

Recall also that we have derived $x^*(q)$ and $B(q)$ utilizing conditions (7)–(8), which must therefore hold. We rewrite them for convenience as

$$V_q^*\big(x^*(q), q\big) + x^*(q)v_H(q) + \big(1 - x^*(q)\big)v_L(q) = 0, \tag{19}$$

$$V_{qx}^*\big(x^*(q), q\big) + v_H(q) - v_L(q) = 0. \tag{20}$$

We next state three lemmas that state some further properties of the value function (17) that hold below, above, and at the boundary, respectively. The results follow from (17)–(20) by straightforward calculations, which are provided in Appendix C.3.

Lemma 6. *For all (x, q) with $q < q^*(x)$, we have*

$$V_{xx}^*(x, q) = V_{xx}^*\big(x, q^*(x)\big) = B\big(q^*(x)\big)\Phi_{xx}\big(x, q^*(x)\big).$$

Lemma 7. *For all (x, q) with $q > q^*(x)$, we have*

$$V_q^*(x, q) + xv_H(q) + (1-x)v_L(q) \leq 0.$$

Lemma 8. *For all $(x, q^*(x))$ with $q^*(x) > 0$, we have*

$$\frac{V^*\big(x, q^*(x)\big)}{q^*(x)} + xv_H\big(q^*(x)\big) + (1-x)v_L\big(q^*(x)\big) > 0.$$

Next, we show that the value function $V^*(x, q)$ in (17) satisfies

$$rV^*(x, q) = \max_{q' \geq q} \Pi\big(q'; q\big),$$

where

$$\Pi\big(q'; q\big) = r\int_q^{q'} \big(xv_H(s) + (1-x)v_L(s)\big) + \frac{1}{2}V_{xx}^*\big(x, q'\big)\frac{x^2(1-x)^2}{\sigma^2}q'. \tag{21}$$

Using Lemma 6 and noting that $\Phi_{xx}(x,q) = \Phi(x,q)\frac{2r\sigma^2}{x^2(1-x)^2 q}$, we can write

$$V^*_{xx}(x,q') = \begin{cases} B(q^*(x))\Phi_{xx}(x,q^*(x)) & \text{for } q' < q^*(x) \\ B(q')\Phi_{xx}(x,q') & \text{for } q' \geq q^*(x), \end{cases}$$

$$= \begin{cases} B(q^*(x))\Phi(x,q^*(x))\dfrac{2r\sigma^2}{x^2(1-x)^2 q^*(x)} & \text{for } q' < q^*(x) \\ B(q')\Phi(x,q')\dfrac{2r\sigma^2}{x^2(1-x)^2 q'} & \text{for } q' \geq q^*(x). \end{cases}$$

We can now rewrite (21) as

$$\Pi(q';q) := \begin{cases} r\displaystyle\int_q^{q'} \big(xv_H(s) + (1-x)v_L(s)\big) + rB(q^*(x))\Phi(x,q^*(x))\frac{q'}{q^*(x)} & \text{for } q' < q^*(x) \\ r\displaystyle\int_q^{q'} \big(xv_H(s) + (1-x)v_L(s)\big) + rB(q')\Phi(x,q') & \text{for } q' \geq q^*(x). \end{cases}$$

The function $\Pi(q';q)$ is continuous in q' and its derivative is

$$\frac{d\Pi(q';q)}{dq'} = \begin{cases} r\big(xv_H(q') + (1-x)v_L(q')\big) + \dfrac{rB(q^*(x))\Phi(x,q^*(x))}{q^*(x)} \\ r\big(xv_H(q') + (1-x)v_L(q')\big) + r\big(B_q(q')\Phi(x,q') + B(q')\Phi_q(x,q')\big) \end{cases}$$

$$= \begin{cases} r\left(\big(xv_H(q') + (1-x)v_L(q')\big) + \dfrac{V^*(x,q^*(x))}{q^*(x)}\right) & \text{for } q' < q^*(x) \\ r\big(\big(xv_H(q') + (1-x)v_L(q')\big) + V^*_q(x,q')\big) & \text{for } q' > q^*(x). \end{cases}$$

From Lemma 7, it follows that $\frac{d\Pi(q';q)}{dq'} \leq 0$ for $q' > q^*(x)$. Noting that $v_H(q)$ and $v_L(q)$ are decreasing in q, it follows from Lemma 8 that $\frac{d\Pi(q';q)}{dq'} > 0$ for $q' < q^*(x)$. Therefore, we have

$$\max_{q' \geq q} \Pi(q';q)$$

$$= \begin{cases} \Pi(q^*(x);q) & \text{for } q < q^*(x) \\ \Pi(q;q) & \text{for } q \geq q^*(x) \end{cases}$$

$$= \begin{cases} r\displaystyle\int_q^{q^*(x)} \big(xv_H(s) + (1-x)v_L(s)\big) + rB(q^*(x))\Phi(x,q^*(x)) & \text{for } q < q^*(x) \\ rB(q)\Phi(x,q) & \text{for } q \geq q^*(x). \end{cases}$$

$$= rV^*(x,q)$$

Notice as well that from Lemma 6 we have that the partial derivative V^*_{xx} is continuous. Thus, our candidate V^* satisfies the HJB equation (5).

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^* be the candidate solution (17), which also solves 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 the generalized Itô's formula, we have[15]

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

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

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

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

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

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.

C.3 *Proofs of Lemmas 5, 7, 6, and 8*

Proof of Lemma 5. Taking the derivative of $g(x, q)$ with respect x gives

$$g_x(x, q) = -\big[\beta''(q)\big(x^2(1-2x)\big(\beta(q)-1\big)^3 v_H(q)^2 - 2(1-x)x\beta(q)\big(\beta(q)-1\big)$$

[15]To see that $V^* \in C^2$ check V_x^* at the boundary. The continuity of V_{xx}^* and V_{qq}^* follows from Lemma 6 and the continuity of V_q^* and V_{xq}^* are implied by conditions (7) and (8).

$$\times v_H(q)v_L(q)\big((1-2x)\beta(q)-x\big)+(1-x)^2(1-2x)\beta(q)^3v_L(q)^2\big)$$
$$+\beta'(q)(2x^2(2x-1)\big(\beta(q)-1\big)^2v_H(q)^2\beta'(q)+(1-x)^2\beta(q)^2v_L(q)$$
$$\times\big(2(1-2x)v_L(q)\beta'(q)-2x\big(\beta(q)-1\big)v_H'(q)-(1-2x)\beta(q)v_L'(q)\big)$$
$$+xv_H(q)\big(4(1-x)v_L(q)\beta'(q)\big((1-2x)\beta(q)^2+2x\beta(q)+x\big)$$
$$-x\big(\beta(q)-1\big)^2\big((1-2x)\big(\beta(q)-1\big)v_H'(q)+2(1-x)\beta(q)v_L'(q)\big)\big)\big)\big]$$
$$/\big[\big(x\big(\beta(q)-1\big)^2v_H(q)+(1-x)\big(\beta(q)\big)^2v_L(q)\big)^2\beta'(q)\big].$$

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

$$x^E(q)\big(\beta(q)-1\big)^2v_H(q)+\big(1-x^E(q)\big)\big(\beta(q)\big)^2v_L(q)<0, \tag{22}$$

for all $q\in[0,1)$. Notice that the left side is increasing in x, and hence, (22) implies the same inequality for all lower x. The condition (22) is equivalent with

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

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.[16] The first term is scaled by x, while 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 $x(1-x)$ give

$$\frac{\beta(q)v_L(q)\big(\beta'(q)\big(\beta(q)-1\big)v_H'(q)-\big(\big(\beta(q)-1\big)\beta''(q)-2\big(\beta'(q)\big)^2\big)v_H(q)\big)}{\beta(q)v_L(q)+\big(1-\beta(q)\big)v_H(q)}$$
$$+\frac{\big(1-\beta(q)\big)v_H(q)\big(\beta'(q)\beta(q)v_L'(q)-\big(\beta(q)\beta''(q)-2\big(\beta'(q)\big)^2\big)v_L(q)\big)}{\beta(q)v_L(q)+\big(1-\beta(q)\big)v_H(q)}.$$

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

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

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

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):

$$\begin{aligned} g\big(x^E(q), q\big) &= \left(\frac{-\beta(1-\beta)v_L v_H}{\big(\beta v_L + (1-\beta)v_H\big)^2} \Big/ \frac{\beta'\beta(1-\beta)v_L v_H}{\beta v_L + (1-\beta)v_H}\right) \\ &\quad \cdot \left(\frac{\beta'\beta(1-\beta)\big(v_L v_H' - v_L' v_H\big)}{\beta v_L + (1-\beta)v_H} + \frac{\beta v_L v_H\big(-2\beta'^2 + (\beta-1)\beta''\big)}{\beta v_L + (1-\beta)v_H}\right. \\ &\quad \left. + \frac{(\beta-1)v_L v_H\big(-2\beta'^2 + \beta\beta''\big)}{\beta v_L + (1-\beta)v_H}\right) \\ &= \frac{v_H\big(2v_L\beta' - (\beta-1)\beta v_L'\big) + (\beta-1)\beta v_L v_H'}{\big((\beta-1)v_H - \beta v_L\big)^2}. \end{aligned}$$

The derivative of the decentralized policy x^E is

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

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

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

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 6. Fixing some (x, q) such that $q < q^*(x)$, differentiating (17) twice with respect to x, and simplifying give

$$\begin{aligned} V_{xx}^*(x, q) &= V_{xx}^*\big(x, q^*(x)\big) + 2\big(q^*\big)'(x)\big(V_{xq}^*\big(x, q^*(x)\big) + v_H\big(q^*(x)\big) - v_L\big(q^*(x)\big)\big) \\ &\quad + \big(q^*\big)''(x)\big(V_q^*\big(x, q^*(x)\big) + x v_H\big(q^*(x)\big) + (1-x)v_L\big(q^*(x)\big)\big) \\ &\quad + \big(\big(q^*\big)'(x)\big)^2\big(V_{qq}^*\big(x, q^*(x)\big) + x v_H'\big(q^*(x)\big) + (1-x)v_L'\big(q^*(x)\big)\big). \end{aligned} \tag{23}$$

Noting that $q^*(x)$ is the inverse function of $x^*(q)$, the second term on the right-hand side vanishes by condition (20) and the third term vanishes by the condition (19). Let us look at the last term. First, since (19) holds along the boundary $(x, q^*(x))$, we can totally differentiate it with respect to x to get

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

Applying (20), several terms disappear and this reduces to

$$V^*_{qq}\big(x, q^*(x)\big) + xv'_H\big(q^*(x)\big) + (1-x)v'_L\big(q^*(x)\big) = 0.$$

The last term in (23) vanishes as well, and it follows that $V^*_{xx}(x, q) = V^*_{xx}(x, q^*(x))$. □

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

$$V^*_q(x, q) + xv_H(q) + (1-x)v_L(q) > 0. \tag{24}$$

We show that this leads to a contradiction by showing that (24) implies $V^*_{xq}(x, q) + v_H(q) - v_L(q) > 0$, which further implies that (24) 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 (19).

It remains to show that (24) implies $V^*_{xq}(x, q) + v_H(q) - v_L(q) > 0$. First, notice that $V^*_q(x, q) = B_q(q)\Phi(x, q) + B(q)\Phi_q(x, q)$, which then together with (24) 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:

$$\begin{aligned}
V^*_{xq} + v_H - v_L &= B_q\Phi_x + B\Phi_{xq} + v_H - v_L \\
&> -\frac{\Phi_q\Phi_x}{\Phi}B - \frac{\Phi_x}{\Phi}\big(xv_H + (1-x)v_L\big) \\
&\quad + B\Phi_{xq} + v_H - v_L \\
&= \Phi^{-1}\big[B(\Phi_{xq}\Phi - \Phi_q\Phi_x) + \Phi(v_H - v_L) - \Phi_x\big(xv_H + (1-x)v_L\big)\big].
\end{aligned} \tag{25}$$

The first term can be simplified as

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

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

Now, (25) becomes

$$\begin{aligned}
&\frac{x^*\big(1-x^*\big)}{x(1-x)}\frac{\Phi}{\Phi^*\Phi^*}\big[\Phi^*_x\big(x^*v_H + \big(1-x^*\big)v_L\big) - \Phi^*(v_H - v_L)\big] \\
&\qquad - \frac{1}{\Phi}\big[\Phi_x\big(xv_H + (1-x)v_L\big) - \Phi(v_H - v_L)\big] \\
&\quad = \frac{1}{x(1-x)}\bigg(\frac{\Phi}{\Phi^*}\big((\beta-1)x^*v_H + \beta\big(1-x^*\big)v_L\big) - \big((\beta-1)xv_H + \beta(1-x)v_L\big)\bigg),
\end{aligned} \tag{26}$$

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

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

To conclude that (26) 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 $V_{xq}^* + v_H - v_L > 0$ and the proof is complete. □

Proof of Lemma 8. By definition of function $\Phi(x, q)$, the following holds for all $x > 0$, $q > 0$:

$$rB(q)\Phi(x, q) = \frac{1}{2}B(q)\Phi_{xx}(x, q)\frac{x^2(1-x)^2}{\sigma^2}q.$$

Differentiating w.r.t. q, the following holds as well:

$$\begin{aligned} &r\big(B_q(q)\Phi(x, q) + B(q)\Phi_q(x, q)\big) \\ &\quad = \frac{1}{2}B(q)\Phi_{xx}(x, q)\frac{x^2(1-x)^2}{\sigma^2} \\ &\qquad + \frac{1}{2}\big(B_q(q)\Phi_{xx}(x, q) + B(q)\Phi_{xxq}(x, q)\big)\frac{x^2(1-x)^2}{\sigma^2}q \\ &\quad = r\frac{B(q)\Phi(x, q)}{q} + \frac{1}{2}\big(B_q(q)\Phi_{xx}(x, q) + B(q)\Phi_{xxq}(x, q)\big)\frac{x^2(1-x)^2}{\sigma^2}q. \end{aligned}$$

In particular, this holds for any $q > 0$, $x = x^*(q)$:

$$\begin{aligned} &r\big(B_q(q)\Phi\big(x^*(q), q\big) + B(q)\Phi_q\big(x^*(q), q\big)\big) \\ &\quad = r\frac{B(q)\Phi\big(x^*(q), q\big)}{q} \\ &\qquad + \frac{1}{2}\big(B_q(q)\Phi_{xx}\big(x^*(q), q\big) + B(q)\Phi_{xxq}\big(x^*(q), q\big)\big)\frac{x^*(q)^2\big(1-x^*(q)\big)^2}{\sigma^2}q. \end{aligned} \tag{27}$$

From (19), we have

$$r\big(x^*(q)v_H(q) + \big(1-x^*(q)\big)v_L(q)\big) + r\big(B_q(q)\Phi\big(x^*(q), q\big) + B(q)\Phi_q\big(x^*(q), q\big)\big) = 0, \tag{28}$$

and so combining (27) and (28), we get

$$\begin{aligned} &r\big(x^*(q)v_H(q) + \big(1-x^*(q)\big)v_L(q)\big) + r\frac{B(q)\Phi\big(x^*(q), q\big)}{q} \\ &\quad + \frac{1}{2}\big(B_q(q)\Phi_{xx}\big(x^*(q), q\big) + B(q)\Phi_{xxq}\big(x^*(q), q\big)\big)\cdot\frac{x^*(q)^2\big(1-x^*(q)\big)^2}{\sigma^2}q = 0. \end{aligned} \tag{29}$$

Plugging in (14) and (15) for $B(q)$ and $B_q(q)$, we get by direct computation at $x = x^*(q)$:

$$B_q(q)\Phi_{xx}\big(x^*(q), q\big) + B(q)\Phi_{xxq}\big(x^*(q), q\big)$$
$$= \frac{x^*(q)\big(\beta(q)-1\big)^2 v_H(q) - \big(1-x^*(q)\big)\big(\beta(q)\big)^2 v_L(q)}{x^*(q)^2\big(1-x^*(q)\big)^2}. \tag{30}$$

Rearranging the equation that defines the policy function $x^E(q)$ of the decentralized equilibrium in Proposition 1, we have

$$x^E(q)\big(\beta(q)-1\big)v_H(q) - \big(1-x^E(q)\big)\beta(q)v_L(q) = 0.$$

We have shown in Part 1 of the Appendix C.2 that $x^*(q) < x^E(q)$. Noting that $\beta(q) > 1$, $v_H(q) > 0$, and $v_L(q) < 0$, it follows that

$$x^*(q)\big(\beta(q)-1\big)^2 v_H(q) - \big(1-x^*(q)\big)\big(\beta(q)\big)^2 v_L(q) < 0$$

and so it follows from (30) that

$$B_q(q)\Phi_{xx}\big(x^*(q), q\big) + B(q)\Phi_{xxq}\big(x^*(q), q\big) < 0. \tag{31}$$

Combining (29) and (31) give

$$r\big(x^*(q)v_H(q) + \big(1-x^*(q)\big)v_L(q)\big) + r\frac{B(q)\Phi\big(x^*(q), q\big)}{q} > 0,$$

which is equivalent to

$$xv_H\big(q^*(x)\big) + (1-x)v_L\big(q^*(x)\big) + \frac{B\big(q^*(x)\big)\Phi\big(x, q^*(x)\big)}{q^*(x)} > 0$$

for all x for which $q^*(x) > 0$. □

C.4 *Proof of Proposition 3*

Proof. First, recall that $x^*(0) < x^E(0) = x^{\text{stat}}(0)$ 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 $x^{\text{stat}}(q) > x^*(q)$ for all $q < \underline{q}$. As the policy functions are strictly increasing and continuous, the stocks $q^*(x)$ and $q^{\text{stat}}(x)$ are pinned down as the inverse of the policy functions for all $x \geq x^*(0)$ and $x \geq x^{\text{stat}}(0)$, respectively. In addition, $q^*(x) = 0$ for all $x \leq x^*(0)$ and $q^{\text{stat}}(x) = 0$ for all $x \leq x^{\text{stat}}(0)$, and hence, q^* and q^{stat} are continuous.

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

Next, we show the other direction by showing that $x^*(1) = x^{\text{stat}}(1) = 1$ and $x_q^*(1) < x_q^{\text{stat}}(1)$. The first part is immediate. For the second part, use Lemma 5 and the uniqueness of the solution to get $x_q^*(1) = x_q^E(1)$. Now it is enough to show that the derivative of

the equilibrium is smaller than of the myopic solution:

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

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 $x^*(q) > x^{\text{stat}}(q)$ 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.5 *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

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

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

$$\begin{aligned} V_{xxq} &= \frac{\big((\beta-x)^2 + x(1-x)\big)\big(x v_H + (1-x) v_L\big)}{\Phi\Phi_{xq} - \Phi_q\Phi_x} + \frac{(\Phi_q\Phi_{xx} - \Phi\Phi_{xxq})(v_H - v_L)}{\Phi\Phi_{xq} - \Phi_q\Phi_x} \\ &= \frac{(\beta-x)^2 + x(1-x)}{x^2(1-x)^2}\big(x v_H + (1-x) v_L\big) - \frac{\beta + (\beta-1)\ln\left(\dfrac{x}{1-x}\right)}{x(1-x)}(v_H - v_L). \end{aligned}$$

Both terms in this expression 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

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

We take the limit $\lim_{\sigma\to 0} \beta(q_\sigma^E(x)) = x v_H(0)/(x v_H(0) + (1-x) v_L(0))$, which is strictly larger than 1 for all $x > x^{\text{stat}}(q)$, and hence, further implying that $\lim_{\sigma\to 0}\sqrt{q_\sigma^E(x)}/\sigma <$

∞. More precisely, we get the limit of the signal-to-noise ratio as $a(x)$ satisfying $xv_H(0)/(xv_H(0)+(1-x)v_L(0)) = \frac{1}{2}(1+\sqrt{1+8ra(x)^{-2}})$. □

C.6 *Proof of Proposition 5*

Proof. 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 $\overline{x} := \sup(x_t | t > 0)$ 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 $\tau(x') := \inf(t : x_t \geq x')$ 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

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

from which we can solve $\Pr(\overline{x} \geq x') = \frac{x_0}{x'}$. On the other hand, we can write

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

Since the belief converges to truth, we also have $\Pr(\overline{x} \geq x' | \omega = H) = 1$.

Using the equations above, we get $\Pr(\overline{x} \geq x' | \omega = L) = \frac{x_0(1-x')}{x'(1-x_0)}$ and so

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

Noting that $\Pr(q_\infty \leq q | \omega) = \Pr(\overline{x} \leq \widetilde{x}(q) | \omega)$, yields the long-run distribution of the stock in the proposition. □

References

Bandiera, Oriana and Imran Rasul (2006), "Social networks and technology adoption in northern Mozambique." *The Economic Journal*, 116, 869–902. [0097]

Bergemann, Dirk and Juuso Välimäki (1997), "Market diffusion with two-sided learning." *RAND Journal of Economics*, 28, 773–795. [0096, 0098, 0109]

Bergemann, Dirk and Juuso Välimäki (2000), "Experimentation in markets." *Review of Economic Studies*, 67, 213–234. [0096, 0098]

Bolton, Patrick and Christopher Harris (1999), "Strategic experimentation." *Econometrica*, 67, 349–374. [0095, 0098, 0102, 0112]

Bonatti, Alessandro (2011), "Menu pricing and learning." *American Economic Journal: Microeconomics*, 3, 124–163. [0094, 0096, 0098, 0109]

Cetemen, Doruk, Can Urgun, and Leeat Yariv (2023), "Collective progress: Dynamics of exit waves." *Journal of Political Economy*, 131, 2402–2450. [0102]

Chamley, Christophe and Douglas Gale (1994), "Information revelation and strategic delay in a model of investment." *Econometrica*, 62, 1065–1085. [0096]

Che, Yeon-Koo and Johannes Hörner (2017), "Recommender systems as mechanisms for social learning." *The Quarterly Journal of Economics*, 133, 871–925. [0094, 0096]

Conley, Timothy G. and Christopher R. Udry (2010), "Learning about a new technology: Pineapple in Ghana." *American Economic Review*, 100, 35–69. [0097]

Décamps, Jean-Paul and Thomas Mariotti (2004), "Investment timing and learning externalities." *Journal of Economic Theory*, 118, 80–102. [0102]

Dixit, Avinash (1989), "Entry and exit decisions under uncertainty." *Journal of Political Economy*, 97, 620–638. [0096]

Dixit, Avinash K. and Robert S. Pindyck (1994), *Investment Under Uncertainty*. Princeton University Press, Princeton. [0095, 0096, 0101, 0102, 0105, 0112]

Duflo, Esther and Emmanuel Saez (2003), "The role of information and social interactions in retirement plan decisions: Evidence from a randomized experiment." *The Quarterly Journal of Economics*, 118, 815–842. [0097]

Farrell, Joseph and Garth Saloner (1986), "Installed base and compatibility: Innovation, product preannouncements, and predation." *The American Economic Review*, 76, 940–955. [0097]

Fleming, Wendell H. and Halil Mete Soner (2006), *Controlled Markov Processes and Viscosity Solutions*, volume 25. Springer. [0120]

Foster, Andrew D. and Mark R. Rosenzweig (1995), "Learning by doing and learning from others: Human capital and technical change in agriculture." *Journal of Political Economy*, 103, 1176–1209. [0097]

Frick, Mira and Yuhta Ishii (2024), "Innovation adoption by forward-looking social learners." *Theoretical Economics*, 19, 1505–1541. [0094, 0096, 0097]

Gittins, John C. and David M. Jones (1974), "A dynamic allocation index for the sequential allocation of experiments." In *Progress in Statistics* (J. Gani, ed.), 241–266. [0095]

Grossman, Sanford J., Richard E. Kihlstrom, and Leonard J. Mirman (1977), "A Bayesian approach to the production of information and learning by doing." *The Review of Economic Studies*, 44, 533–547. [0095]

Jensen, Richard (1982), "Adoption and diffusion of an innovation of uncertain profitability." *Journal of Economic Theory*, 27, 182–193. [0097]

Jovanovic, Boyan and Saul Lach (1989), "Entry, exit, and diffusion with learning by doing." *The American Economic Review*, 79, 690–699. [0097]

Keller, Godfrey, Sven Rady, and Martin Cripps (2005), "Strategic experimentation with exponential bandits." *Econometrica*, 73, 39–68. [0095]

Keller, Godfrey and Sven Rady (2010), "Strategic experimentation with Poisson bandits." *Theoretical Economics*, 5, 275–311. [0096]

Laiho, Tuomas and Julia Salmi (2024), "Coasian dynamics and endogenous learning." Working paper. [0094, 0096]

Leahy, John V. (1993), "Investment in competitive equilibrium: The optimality of myopic behavior." *The Quarterly Journal of Economics*, 108, 1105–1133. [0102]

Liski, Matti and Francois Salanié (2023), "Catastrophes, delays, and learning." Toulouse School of Economics Working Papers No 1148. [0096]

Mansfield, Edwin (1961), "Technical change and the rate of imitation." *Econometrica*, 29, 741–766. [0097]

Martimort, David and Louise Guillouet (2020), "Precaution, information and time-inconsistency: On the value of the precautionary principle." CEPR Discussion Paper Series DP15266. [0096]

McDonald, Robert and Daniel Siegel (1986), "The value of waiting to invest." *The quarterly journal of economics*, 101, 707–727. [0096]

Moscarini, Giuseppe and Lones Smith (2001), "The optimal level of experimentation." *Econometrica*, 69, 1629–1644. [0098]

Munshi, Kaivan (2004), "Social learning in a heterogeneous population: Technology diffusion in the Indian green revolution." *Journal of Development Economics*, 73, 185–213. [0097]

Murto, Pauli and Juuso Välimäki (2011), "Learning and information aggregation in an exit game." *The Review of Economic Studies*, 78, 1426–1461. [0096]

Peskir, Goran and Albert Shiryaev (2006) In *Optimal Stopping and Free-Boundary Problems*, volume 10 of Lectures in Mathematics. ETH Zürich. Springer, Berlin. [0111, 0112]

Pindyck, Robert S. (1988), "Irreversible investment, capacity choice, and the value of the firm." *The American Economic Review*, 78, 969–985. [0096, 0105]

Prescott, Edward C. (1972), "The multi-period control problem under uncertainty." *Econometrica*, 40, 1043–1058. [0095]

Rob, Rafael (1991), "Learning and capacity expansion under demand uncertainty." *The Review of Economic Studies*, 58, 655–675. [0096]

Rothschild, Michael (1974), "A two-armed bandit theory of market pricing." *Journal of Economic Theory*, 9, 185–202. [0095]

Strulovici, Bruno (2010), "Learning while voting: Determinants of collective experimentation." *Econometrica*, 78, 933–971. [0096]

Wolitzky, Alexander (2018), "Learning from others' outcomes." *American Economic Review*, 108, 2763–2801. [0097]

Young, H. Peyton (2009), "Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning." *American Economic Review*, 99, 1899–1924. [0097]

Co-editor Bruno Strulovici handled this manuscript.

Manuscript received 6 October, 2022; final version accepted 4 August, 2024; available online 15 August, 2024.