Dynamics of Collaboration: Cooperation and Retardation*

Suvi Vasama†

July 11, 2016

Abstract

We propose a new model of strategic experimentation in which the players' action affects the distribution over future payoffs. The players need to exert costly effort both to develop a risky technology and to learn about its value. Both product development are learning are public goods, which gives the players incentives to free-ride on each others' development efforts. Free-riding leads to an inefficiently low aggregate level of development effort. When the players' actions affect the distribution over future payoffs, this causes the firms eventually retarding the innovation, which leads to an inefficiently short lifetime of the product when compared to the efficient benchmark. Moreover, we find that the game exhibits multiple symmetric Markov perfect equilibria.

Keywords: Stochastic games, strategic experimentation, Bayesian learning, restless bandits, innovation

JEL Classification: C73, D83, O31, O33.

*I thank Eduardo Faingold, Johannes Hörner, Chiara Margaria, Larry Samuelson, Roland Strausz and Juuso Välimäki for many helpful comments and suggestions. Please visit https://sites.google.com/site/suvivasama/research for the most recent version of the paper.

†Department of Economics, Humboldt University of Berlin, suvi.vasama@hu-berlin.de

Contents

1 Introduction **3**

2 Setting **7**

3 Beliefs **9**

4 Social Planner's Solution **11**
4.1 Complete Information Solution 12
4.2 Incomplete Information Solution 13

5 Noncooperative Solution **15**
5.1 Markov Perfect Equilibria 16
5.2 Symmetric Equilibria 19

6 Poisson Bandits **25**
6.1 Beliefs 25
6.2 Social Planner's Solution 26
6.3 Noncooperative Solution 27

7 Discussion **28**

8 Appendix **30**

References **36**

1 Introduction

In a market for technological innovations, firms often engage in collaboration to facilitate development of a common standard or to create a new business entity.[1] In 1997, the cell phone companies Nokia, Motorola, and Ericsson started collaboration in creating a common standard for a wireless telecommunication format. The Wireless Application Protocol (WAP) soon became a widely used by mobile operators, equipment producers and software developers. However, such collaboration is related with great uncertainty and not every attempt to collaborate is successful. Starting from 1994, IBM, Apple and Hewlett-Packard spend three years and $50 million developing a new standard for an operating system for computers that would challenge Microsoft Windows.

Constant development effort is needed to both facilitate the standard and to learn if it is viable. Even if the firms manage to establish a new standard, it may lose its position and become obsolete. Indeed, in 2002, WAP Forum merged to a new Open Mobile Architecture to form a new initiative, Open Mobile Alliance (OMA), and has by now been replaced by new, more elaborate smart phone standards. Finally, development is a public good; the firms benefit each others development efforts that facilitate the common standard.

This paper introduces a new approach of modeling dynamics of collaboration in markets for risky technologies. The firms exert costly effort to establish a common standard. Success is uncertain and constant effort is needed both to facilitate the standard and to learn from its viability in the market. The public good feature of the problem leads to free riding on other firms' development efforts whenever the firms are not sufficiently confident of the viability of the standard. The aggregate development effort is not high enough to facilitate the standard, which leads to an inefficiently short lifetime compared to the first-best optimal benchmark.

We model the firms' problem of developing and facilitating a standard or a new business entity as a strategic learning game in which the players' actions affect the distribution of future payoffs. Thus, the players need to exert costly effort both to increase the value of the standard and to learn about its viability. We solve for the symmetric Markov perfect equilibria of the product development game and show it exhibits multiple equilibria. Multiple equilibria arise because product development is a coordination game: exerting higher effort is optimal only if other players also exert higher effort.

To be more concrete, we model the product development game as a strategic learning game in which the economic environment changes over time. At each point in time, each player chooses how much effort they exert in selling and de-

[1] Such collaboration is legal in the US by the National Cooperative Research Act of 1984 if the collaborative efforts do not have anticompetitive effects. See Schilling (2005) for further examples and analysis of circumstances which such a collaboration is beneficial for the firms.

veloping the product. Effort is costly and the returns are uncertain and depend on the unobserved value of the product. The value changes stochastically over time and the probability distribution over future payoffs depends on the players' actions. When the players develop the product, they receive noisy signals about its value. Thus, effort has two consequences in the model: the players' receive more informative signals about the current state and they improve the probability distribution over future payoffs.

We analyze the noncooperative product development game in which each player individually decides how much effort to exert. The value of the product is the same across the players while each player carries the cost of effort individually. The firms always have the option to wait and see if the other firms' development efforts were successful. In the noncooperative product development game both information acquisition and the value of the product resemble a public good problem. The players have an incentive to free-ride on each others' actions.

Before solving the noncooperative product development game, we identify two important benchmark cases: the complete information investment problem and the social planner's solution when the value of the product unobserved. With complete information, the planner always develops the product if its current value is high. If the product value becomes low, the planner has to decide if she still keeps on developing it. The optimal decision depends on how likely the planner is to win back the market if she lost it. If the probability of success is high enough, it is always optimal to exerts full effort. Otherwise, the planner abandons the product as soon as its value becomes low. In absence of product development, the product value stays low and the planner never exerts effort again.

Next, we analyze the social planner's optimal investment decision if she does not observe the product value, but learns about it by observing her payoffs. Again, the optimal policy depends on how likely the planner is to win back the market if she lost it. If the probability of success is not high enough, the planner eventually abandons the technology if she becomes too pessimistic about its value. The optimal investment strategy is a threshold policy that only conditions on the planner's belief about the product value. If the planner becomes too pessimistic, she abandons the technology. Now if the firms do not develop the product, is more and more likely to become obsolete. Once the technology is abandoned, it is never optimal to exert effort again.

The main part of the paper analyzes the version of the model in which the firms individually decide about the optimal effort strategy. The firms observe both each others' payoffs and their effort levels, but face uncertainty about the product value. We solve for the symmetric Markov Perfect equilibrium of the noncooperative product development game. The players choose their strategies as functions of their beliefs only. The restriction is applied for tractability and allows us to use standard dynamic programming tools to solve for the equilibrium.

As is standard in strategic learning games, we identify three qualitatively different investment regimes, depending on the players' belief about the product value. Firstly, if the players are very optimistic about the product value, the expected profit is positive and the firms exert every effort to develop the product. Secondly, if the players are sufficiently pessimistic, it is never optimal to develop the product. Thirdly, when beliefs are intermediate, the players optimally restrict their effort. The third investment regime arises because of free-riding and distinguishes the noncooperative solution from the social planner's optimal product development decision.

At the noncooperative optimum, the players' incentives to exert effort are driven by two important forces: free-riding and encouragement effect. Both effects are familiar from the classical strategic learning literature. The encouragement effect follows since option value of information is positive. The firms benefit by learning from each others' signals: by exerting higher effort they encourage other players to exert higher effort. The free-riding effect arises because both the firms' development efforts and the information about the product value are public goods, and reduces the firms' effort at the noncooperative equilibrium.

Letting the players' action affect the distribution over future payoffs adds several new features in the strategic learning game. Firstly, he players' incentives to free-ride on each others' actions depend on (a) how likely they are to lose the market; and (b) on the impact on their development efforts on the product value. The first effect alleviates free-riding; effort is needed for the players not to lose the market. The second effect makes free-riding more attractive: the players have an incentive to free-ride not only on each others' information acquisition but also on their investment on the project value. Secondly, restricting effort leads to an inefficiently short lifetime of a risky innovation. The aggregate development effort is not high enough to sustain the product value, but the players gradually lose the market. Thirdly, since the value of the product depends on the aggregate effort, the players play a coordination game. We show that this leads to multiple symmetric Markov equilibria of the noncooperative product development game.

Most of the analysis focuses on the case in which the players receive signals about the state that are disturbed by Brownian noise. We briefly review an alternative model in which the firm receives lump sum profits that arrive following a Poisson process if the value of the product is high and show that there is one crucial difference in model dynamics: the encouragement effect is absent. In the noncooperative product development game, the firms never develop the product longer than a single player would.

Our model builds on a classical two-armed bandit problem that has been exhaustively studied in both economics and applied mathematics literature.[2] The seminal contributions include Karatzas (1984) for the single player experimentation problem with Brownian bandits and Presman (1991) for Poisson bandits. Bergemann and Välimäki (2008) provide a comprehensive review of earlier literature in applied mathematics and economics.

The strategic learning game is well explored for the case in which the value of the arm remains the same over time. Our model builds on Bolton and Harris (1999) in which the payoff of the risky arm follows a Brownian motion with an unknown drift and a known volatility. We extend the framework to allow for the drift to evolve over time. In particular, the value of the risky arm depends on the players' strategy. We show that allowing for the risky arm to evolve over time results in the players eventually getting trapped in the bad state and retarding the risky arm. The retardation leads to an inefficiently short lifetime of the risky arm, and cannot occur if the value of the arm is fixed. In contrast, Bolton and Harris show that experimentation goes on indefinitely in the noncooperative experimentation game. Moreover, the symmetric Markov Perfect Equilibrium is always unique while our game eventually has multiple equilibria. Keller, Rady, and Cripps (2005); Keller and Rady (2010, 2015) analyze related models in which the risky arm generates lump sum payoffs that arrive at Poisson rate. For traditional bandits, Poisson case turns out to be much more tractable, and the players' value functions can be solved in closed form in most cases. In contrast, when the risky arm is evolving finding closed form solutions seems unlikely.[3]

Bandit problems that involve evolving arms are sometimes called restless bandits in the literature. Fryer and Harms study a model in which the value of the risky arm increases deterministically in the player's investments. Following bad news, the player eventually gets trapped in investing in the safe arm instead of developing the more valuable risky arm. The results are very different from our framework in which predictions are self-fulfilling. The players eventually abandon the good risky arm which then becomes bad with probability one over a long enough time horizon. Also, good luck may encourage a player to invest in it such that the arm eventually becomes good. Our model has also benefited from Keller and Rady (1999, 2003) who examine the problem of a monopoly that faces stochastically changing demand conditions.

As far as I know, this paper is among the first ones to consider strategic experimentation in a restless bandit framework. A notable exception is Safronov (2014) who develops a multi-armed bandit model with competing players. By investing

[2]In particular, exerting effort to develop the product corresponds to investing in the risky arm while abstaining from product development corresponds to investing in the safe arm.

[3]The classical strategic bandit problem has been extended in many important directions to answer relevant questions that remain yet unexplored in our framework. We do not attempt to review the literature here; Hörner and Skrzypacz (2016) provides an excellent review.

in a risky arm, the player eventually becomes experienced in it, which increases the payoff from the arm. If experienced, the player eventually gets trapped with a less valuable arm, instead of developing a potentially more valuable, but yet unexplored arm. However, the model exhibits no gradual learning, but players eventually learn the value of the arm after receiving an instantaneous lump-sum. The model simplifies tremendously and the players' value functions can be solved in closed form. Our model builds on a classical bandit, which, unfortunately, comes with the cost of losing on tractability.

2 Setting

We examine a game with N players in continuous time. Each player can decide how much effort to put in developing a product and to learn about its value. The player's profit depends on the unobserved value of the product and on exogenous circumstances. Moreover, the value of the product is stochastic and may change over time. The change depends on the players' development efforts.

Each firm is facing a capacity constraint; without loss of generality we normalize the maximal effort that it can exert to 1. Let $\alpha_{t,i} \in [0,1]$ denote the intensity of the player i's development efforts and $\alpha_t \equiv (\alpha_{t,1},...,\alpha_{t,N})^\top$ a vector summarizing all players' development efforts at time t.[4]

The product value at time t is

$$\mu_t = s_t \mu, \tag{1}$$

where $s_t \in \{0,1\}$ is unobserved state of the world. In particular, if $s_t = 1$, the product value is high and if $s_t = 0$, the product value is low. The state of the world s_t eventually changes over time following an unobserved stochastic process. In particular, the firms need to constantly exert effort to develop the product to facilitate its value in the market. The product value follows a continuous-time Markov chain with the transition probabilities

$$\Pr[s_z = i, \forall z \in [t, t+dt] | s_t = j] = \begin{cases} e^{-\lambda_s dt} & \text{if} \quad i = j, \\ 1 - e^{-\lambda_s dt} & \text{if} \quad i \neq j, \end{cases} \tag{2}$$

with $s \in \{0,1\}$.

The product value evolves at rate λ_s that depends on both on the development intensities α_t and on an exogenous parameter $\lambda \in [0,\infty)$. It is formally defined as

$$\lambda_s = \begin{cases} \lambda(1 - \delta^\top \alpha_t) & \text{if} \quad s = 1, \\ \lambda \delta^\top \alpha_t & \text{if} \quad s = 0, \end{cases} \tag{3}$$

[4] We let $A^\top$ denote the transpose of the matrix A.

with $\delta = (\delta_1, ..., \delta_N)$, and $\delta_i > 0$ for all i describes the effect of the player i's effort. In particular, if the product value is high at time t, the probability that it is still high at time $t + \Delta t$, is

$$\Pr[s_{t+\Delta t} = 1 | s_t = 1, \alpha_t] = 1 - (1 - \delta^\top \alpha_t)\Delta t + o(\Delta t),$$

and if it is low at time t, the probability that it is high at time $t + \Delta t$, is

$$\Pr[s_{t+\Delta t} = 1 | s_t = 0, \alpha_t] = \delta^\top \alpha_t \Delta t + o(\Delta t).$$

The probability, that the product has a high value on the market, increases in the players' total development intensity α_t. Moreover, to keep the model as simple as possible, we assume that if *all* players abstain from development, i.e. set $\alpha_{t,i} = 0$ for all i, the low state becomes more and more likely over time.[5] The terms of order $(\Delta t)^2$ and higher are collected in the term $o(\Delta t)$ and can be neglected.

If the player abstains from developing the product, he receives a certain flow payoff that we normalize 0. If the player decides to exert effort to develop the product he earns a stochastic payoff

$$d\pi_{t,i} = \alpha_{t,i}(\mu_t - c_i)dt + \sigma \alpha_{t,i}^{1/2} dZ_{t,i}, \tag{4}$$

where μ_t is the product value at time t as defined in (1) , $c_i > 0$, $i \in \{1, ..., N\}$, is the cost of developing the product and $\sigma > 0$ the volatility of the firm's profit.

We assume that $\delta_i \mu > c_i$ for all i such that the long term mean from development efforts is always positive. Moreover, $\sum_{i=1}^{N} \delta_i = \gamma$, where $\gamma \in (0,1]$ is a constant independent of the total number of players N. Notice that if $\gamma = 1$, the high state $s_t = 1$ is absorbing, conditional on the players' exerting full effort.

The players only observe (4) and not the exact decomposition of the profits. Thus, profit is a noisy signal about the product value. The player faces two sources of uncertainty: the uncertainty about the product value μ_t and an exogenous shock that is driven by the Brownian motion Z_i. The player only observes the flow payoff $d\pi_{t,i}$ and not the exact composition between the terms in (4). Thus, the player receives noisy information about the product value. The accuracy of information increases with the total development intensity α_t. Thus, higher effort both makes the high value more likely and facilitates information acquisition.

Both the players' actions and the payoffs are common knowledge. The drift μ_t is the same across the players, but the Brownian motions Z_i are independent across them. When the players exert effort, they receive independent signals

[5] The assumptions are sufficient to guarantee that the efficient policy is a threshold policy that can be solved using standard dynamic programming tools. Fryer and Harms (2013) provide very general sufficient conditions for applicability of standard dynamic programming tools in a single player's experimentation problem in a related setting.

about the state. The more the players exert effort the more accurate is the aggregate information that they receive. Thus, the player learns not only from his own payoffs, but also by observing the payoffs of the other players. Moreover, higher effort by one player makes the high state more likely for all players.

The filtration $\mathscr{F}^{\alpha}$ keeps track of the public history. Let ϕ_t denote the prior probability, that the players assign to the high value $s_t = 1$, at time t. At any point of time t, each player chooses the optimal effort $\alpha_{t,i}$ to maximize the expected payoff

$$v_{0,i} = E^{\phi}\left[\int_0^{\infty} e^{-rt} d\pi_{t,i} dt\right] = E^{\phi}\left[\int_0^{\infty} e^{-rt}\alpha_{t,i}(\mu\phi_t - c_i)dt\right]. \tag{5}$$

The equality follows using law of iterated expectations and $E[s_t|\mathscr{F}^{\alpha}] = \phi_t$.

3 Beliefs

This section presents a heuristic derivation for the evolution of the players' beliefs that they assign to the product value; the formal derivation is delegated to the appendix. The effect of the players' development efforts on the product value as well as the fact that the value may change due to exogenous circumstances, make the players' belief updating nonstandard.

When the players exert effort, they both learn about the current state of the world and affect the probability distribution over the future states. Thus, effort has two effects: First, it increases the probability that the product value is high in the future. Second, it allows the players to learn if their past development efforts were successful.

We derive the filtering equation that governs the players' belief that they assign to the high value. The filtering equation consists of two terms. The first term summarizes the effect of the state transition and is anticipated by the players. Even if the current value is high, it may become low. At time t, the players believe that the value is high with probability ϕ_t, but becomes low with probability $\lambda(1-\delta^{\top}\alpha_t) = \lambda(1-\sum_{i=1}^{N}\delta_i\alpha_{t,i})$ by (2). This effect decreases the players' belief by $-\lambda(1-\delta^{\top}\alpha_t)\phi_t dt$.

Moreover, even if the product value is low, it may become high if the firms keep on exerting effort. The players assign the probability $1-\phi_t$ to the low value but a probability $\lambda\delta^{\top}\alpha_t$ to the transition to the high state. The latter effect increases the players' belief by $\lambda\delta^{\top}\alpha_t(1-\phi_t)dt$. By summing up the effects, we can write the drift of the filtering equation as

$$\lambda(\delta^{\top}\alpha_t - \phi_t)dt. \tag{6}$$

In particular, the players' belief increases in effort. If the players abstain from developing the product, they anticipate that they gradually lose the market. The effect is stronger if δ_i is higher such that the product value is more sensitive to the firms' development efforts. Moreover, if λ is higher, the state is more likely to change for exogenous reasons.

Next, if the players exert effort, they receive informative signals about the product value. Based on their observations, they update their beliefs according to Bayes rule. Since the players learn from Brownian signals, this part of the filtering equation follows a martingale. The players incorporate all available information in their belief and any deviation from their estimate of the true state comes as a surprise.

The variance of the belief depends on the model parameters. Learning is more accurate if the high value μ is high relative to the low one. Similarly, if σ is higher, the profit is more volatile and the signal are noisier. Similarly, learning is faster close to the diffuse belief $\phi_t = 1/2$ when the players are very uncertain about the state of the world and becomes slower as beliefs tend towards 0 or 1. Finally, the players learn more by exerting higher effort $\alpha_{t,i}$. By summarizing the effects, we find that learning affects the players' posterior by

$$\frac{\mu}{\sigma}\phi_t(1-\phi_t)(\alpha_t^{1/2})^\top dZ_t^\alpha \tag{7}$$

with $\alpha_t^{1/2} = (\alpha_{t,1}^{1/2}, \ldots, \alpha_{t,N}^{1/2})^\top$ and $dZ_t^\alpha = (dZ_{t,1}^\alpha, \ldots, dZ_{t,N}^\alpha)^\top$ denoting vectors summarizing all players' development efforts and the Brownian motions that disturb they signals.

By summing up the drift and volatility terms (6) and (7), we find the law of motion for the players' common belief. The result is summarized in the following proposition

Proposition 1 (Beliefs). *Fix a prior belief ϕ_t and a strategy α_t. The common posterior belief, that the players assign to the high value satisfies the following filtering equation*[6][7]

$$d\phi_t = \lambda(\delta^\top \alpha_t - \phi_t)dt + \frac{\mu}{\sigma}\phi_t(1-\phi_t)(\alpha_t^{1/2})^\top dZ_t^\alpha, \tag{8}$$

with $\alpha_t^{1/2} = (\alpha_{t,1}^{1/2}, \ldots, \alpha_{t,N}^{1/2})^\top$ *and* $dZ_t^\alpha = (dZ_{t,1}^\alpha, \ldots, dZ_{t,N}^\alpha)^\top$.

Proof. See Appendix. □

If $\lambda > 0$ such that the product value changes over time, the players are never completely certain about the current market conditions. Indeed, (8) suggests that the beliefs are mean-reverting. They have the tendency to return to their long-term mean which depends on the players' collective investment in the product development. The next corollary confirms the observation

[6] We let $\mathbf{1} = (1, \ldots, 1)^\top$ denote the vector of ones of appropriate length.

[7] Cf. Liptser and Shiryayev (1977) and Keller and Rady (1999).

Corollary 1. *Fix a strategy profile α_t. The player's common belief for the high valuation is*

$$\phi_t = \phi_0 e^{-\lambda t} + \lambda \int_0^t e^{-\lambda(t-s)} \delta^\top \alpha_s ds + \int_0^t e^{-\lambda(t-s)} \frac{\mu}{\sigma} \phi_s (1-\phi_s)(\alpha_s^{1/2})^\top dZ_s^\alpha$$

Proof. Define

$$y(\phi_t, t) = \phi_t e^{\lambda t}$$

By Itô's lemma

$$dy(\phi_t, t) = \lambda \phi_t e^{\lambda t} dt + e^{\lambda t} d\phi_t = e^{\lambda t} \lambda \delta^\top \alpha_t dt + e^{\lambda t} \frac{\mu}{\sigma} \phi_t (1-\phi_t)(\alpha_t^{1/2})^\top dZ_t^\alpha.$$

The result follows by substituting back, integrating from 0 to t and multiplying by $e^{-\lambda t}$. □

One can show that if $\lambda > 0$ and $\alpha_{t,i} = 1$ for all i and $t \geq 0$, i.e. the players always exert full effort, the long term mean of the belief is γ. The model includes the following special cases

- If $\lambda = 0$, the product value is fixed over time. In particular, if the players keep on updating the product, they learn the true mean as $t \to \infty$. This is the classical Brownian bandit problem that is comprehensively examined by Bolton and Harris (1999).
- If $\lambda \to \infty$, any development effort causes the product value becomes i.i.d. Then there is no learning. One can show that since $\delta_i \mu > c_i$, the noncooperative game has a unique Markov perfect equilibrium in which all players exert every effort to develop the product in perpetuity.

4 Social Planner's Solution

Before analyzing the noncooperative product development game, we solve for the social planner's problem when she decides about the optimal level of development effort. In particular, we derive the optimal strategy in two important benchmark cases. First, we determine the full information solution when the product value is known by the planner. Then we derive the optimal strategy when the product value is unobserved.

With complete information, the planner always exerts full effort if the value of the product is high. If the value is low, the optimal strategy depends on the switching rate λ and the long term gains $\gamma\mu$ relative to the cost c. If λ and γ are very high, the state is very likely to recover if it becomes low and it is always optimal to exert full effort. The complete information solution allows us to identify conditions under which this is the case.

Secondly, we determine the social planner's problem when she does not observe the product value. She then conditions her strategy on her belief. Unlike with classical bandits, the social planner's value function cannot be solved in closed form. However, it offers an important benchmark for welfare considerations, and allows us to derive important insights that help as to make conclusions of the qualitative features when examining the noncooperative product development game.

4.1 Complete Information Solution

Before moving to the incomplete information problem, we first solve for the social planner's problem when the product value is observed. The planner's value function of the complete information solution can be solved in closed form.

The social planner always exerts full effort if the product value is high. If the planner loses the market, she has to decide if it optimal to exert effort to increase the probability of getting back to the high state. We show in the Appendix that it is optimal to exert effort in the bad state only if the following condition holds

$$\lambda(\gamma\mu - c_i) \geq c_i; \tag{9}$$

that is, if the myopic cost of effort c_i does not exceed the long term (net) gain of developing the product. We can immediately see from (9) that effort is more attractive in the low state if λ high. Then the planner is more likely to win back the market if she loses it. Effort is also more attractive if γ is high such that effort has a higher impact on the product value or if μ is high such that the gain of effort is high.

Vice versa, if (9) does not hold, the planner always stops developing the product if its value becomes low. From (2) it follows that if the product value is low, and the firms do not develop it, i.e. if $\alpha_t = 0$, the value stays low. The planner stops developing the product if she loses the market.

The complete information payoff can be solved in closed form and is summarized in the following proposition

Proposition 2 (Complete Information Solution)**.** *Suppose that the product value is observed by the players; i.e.* $\phi \in \{0,1\}$. *Then if the following condition holds*

$$\lambda(\gamma\mu - c_i) \geq c_i, \tag{10}$$

the social planner always exerts full effort. The expected payoff at time 0 *is*

$$\overline{v}_i(\phi) = \frac{\phi(\mu - c_i)}{r + \lambda(1-\gamma)}. \tag{11}$$

Vice versa, if (10) does not hold, the planner stops developing the product if the consumers get tired of it. The expected time 0 *payoff is*

$$\overline{v}_i(\phi) = \frac{\mu(\phi r + \lambda\gamma)}{r(r+\lambda)} - \frac{c_i}{r}. \tag{12}$$

Proof. See Appendix. □

4.2 Incomplete Information Solution

We next solve for the social planner's problem when she does not observe the product value, but learns from it by observing the payoffs. The social planner's solution provides both an appropriate benchmark for efficiency considerations and some useful insights that help to solve the noncooperative product development game.

Recall from the previous section that if the long-term gain from product development is very high, it is optimal to exert full effort, regardless of the current state. When the product value is known, this is the case if if the long term gains of product development exceed the myopic cost; i.e. if the condition (10) holds. Otherwise, it is optimal to stop development if the state becomes low. A similar logic applies for the planner's optimal investment strategy when the product value is unobserved. However, since the planner does not observe the true state, she conditions her strategy on her belief.

The social planner's optimal effort only depends on the belief that she assigns to the high product value. The belief summarizes both the long-term effect of the planner's development efforts and her information about the current value. The social planner's optimal decision is a bang-bang policy: she either exerts every effort to develop the product or entirely abstains from it. Effort is optimal if the belief is high enough, i.e. the planner is sufficiently optimistic about the product value.

Formally, the planner's optimal decision can be derived using standard dynamic programming methods. At any point of time, and for each $i \in \{1,...,N\}$, the social planner chooses $\alpha_i(\phi_t)$ such that it maximizes (5). We assume that both the effects of the firms' development efforts and the investment costs are identical across the firms such that $\delta_i = \delta_j$ and $c_i = c_j = c$ for all $i, j \in \{1,...,N\}$. Then by symmetry, $\alpha_i(\phi_t) = \sum_i \alpha_i(\phi_t)/N$.

Using symmetry together with $\sum_{i=1}^{N} \delta_i = \gamma$, we can write the social planner's Hamilton-Jacobi-Bellman equation at any point ϕ as

$$rv_i(\phi) = \sup_{\alpha_i(\phi)} \left\{ \alpha_i(\phi)\mu\phi - c + \lambda(\gamma\alpha(\phi_i) - \phi)v_i'(\phi) + \alpha_i(\phi)\frac{N}{2}\frac{\mu^2}{\sigma^2}\phi^2(1-\phi)^2 v_i''(\phi) \right\} \tag{13}$$

for each i. From (13) we can immediately see that the social planner's objective is linear in $\alpha_i(\phi)$. The optimal policy is characterized by two regimes: If the belief is high, it is optimal to exert full effort; i.e. $\alpha_i(\phi) = 1$. If the belief falls below a certain cutoff, it becomes optimal to set $\alpha_i(\phi) = 0$. By symmetry, the same effort strategy is optimal for all i. Moreover, if the planner stops developing the product, she anticipates that she gradually loses the market.[8] Therefore, it is never optimal to start developing the product again.

Now suppose that $\lambda > 0$ such that the product value may change over time. Then with bad enough luck, the social planner eventually stops developing the product even if its instantaneous value is high. Then over long enough time horizon, the firms lose the market with probability 1. Thus, the model entails self-fulfilling predictions. Of course, the converse is also possible: a sequence a good outcomes may occur even if the product value is low. Then the social planner keeps on developing the product, which eventually results in success over time. The resulting predictions are in contrast with Fryer and Harms (2013) in which the players eventually get trapped in abstaining from effort even if the product value is high.

The result is summarized in the following proposition

Proposition 3 (Social Planner's Solution)**.** *There exists a cut-off belief ϕ^* such that (i) if $\phi \geq \phi^*$, the social planner exerts every effort to develop the product, i.e. $\alpha_i(\phi) = 1$ for all $i \in \{1,...,N\}$. The firm i's value function is the unique solution of the ordinary differential equation*

$$rv_i(\phi) = \mu\phi - c + \lambda(\gamma - \phi)v_i'(\phi) + \frac{N}{2}\frac{\mu^2}{\sigma^2}\phi^2(1-\phi)^2 v_i''(\phi) \tag{14}$$

with the boundary conditions $rv_i(1) = \mu - c - \lambda(1-\gamma)v_i'(1)$, $v_i(\phi^*) = 0$ and $v_i'(\phi^*) = 0$. (ii) As soon as ϕ hits ϕ^*, it is optimal to set $\alpha_i(\phi_t) = 0$ in perpetuity.

Proof. See Appendix. □

Unfortunately, explicit solutions are only available for the special case $\lambda = 0$ that is carefully analyzed in Bolton and Harris (1999). However, we can derive some properties of the social planner's value function that are useful when examining the noncooperative product development game.

We first show that the social planner's value function is convex and increasing in her belief. Convexity reflects the fact that value of information is positive. The planner can only benefit from the resolution of uncertainty since it allows her to make a more efficient investment decision. The result is summarized in the following lemma

[8]Formally, as we can see from (8), the planner's belief drifts downwards. Notice that the decrease in belief is deterministic.

Lemma 1. *Social planner's value function $v_i(\phi)$ is nondecreasing and convex in ϕ.*

Proof. We first prove that the value function is convex. Notice that for arbitrary, fixed $\alpha_i(\phi)$, the social planner's objective (5) is linear in ϕ. Let $v_i^\alpha(\phi)$ denote the planner's value from such an investment strategy. Next, consider $\phi = \eta\phi^1 + (1-\eta)\phi^2$ with $\eta \in [0,1]$. Then

$$\begin{aligned} v_i^\alpha(\phi) =& \eta v_i^\alpha(\phi^1) + (1-\eta) v_i^\alpha(\phi^2) \\ \leq& \eta \sup v_i^\alpha(\phi^1) + (1-\eta) \sup v_i^\alpha(\phi^2) \\ =& \eta v_i(\phi^1) + (1-\eta) v_i(\phi^2). \end{aligned}$$

Taking the supremum on left hand side proves convexity. Next, $v_i'(\phi) \geq 0$ since $v_i'(\phi^*) = 0$ and $v_i'(\phi)$ is nondecreasing. □

Notice that Lemma 1 further implies that the social planner always experiments longer than is myopically optimal. Let $\phi^M \equiv c/\mu$ denote the cut-off belief at which a myopic player stops developing the product.

Corollary 2. *The social planner keeps on developing the product longer than myopically optimal; i.e. $\phi^* \leq \phi^M$.*

Proof. Since ϕ^* is the optimal threshold belief, at which the planner abandons the product and $v_i'(\phi^*) = 0$, we have

$$c - \mu\phi^* = \frac{N}{2}\frac{\mu^2}{\sigma^2}\phi^{*2}(1-\phi^{*2}) v_i''(\phi^*) \geq 0.$$

The last inequality follows by Lemma 1. The result follows by reorganizing. □

The result holds for an arbitrary number of players, including $N = 1$. A single player always keeps on developing the product beyond the myopic belief. We will show below that the same result holds in the noncooperative product development game since the firms never stops developing the product earlier than a single firm would.

5 Noncooperative Solution

The main section of the paper studies the noncooperative product development game. The analysis restricts the attention on a strict subset of equilibria, Markov perfect equilibria and, in particular, the symmetric ones. In a Markov perfect equilibrium, the player conditions her strategy only on the common belief ϕ_t. The focus on Markov perfect equilibria is a restriction. In particular, it does not allow the players to condition their strategies on the past play. The restriction is made for tractability and allows us to use standard dynamic programming tools to derive the equilibrium. The equilibrium concepts are adapted from Bolton and Harris (1999).

In the noncooperative product development game each firm decides individually how much effort it wants to exert to develop the product. The product value is the same for all firms and they benefit from each others' development efforts. However, product development is eventually costly since the product value may be low. While product development benefits all firms, exerting effort is only profitable if the state is high. Thus, product development is a public good.

We find several effects that drive the players' optimal decision at equilibrium. First, the players always have the possibility to delay their effort and benefit from each others' product development. This gives incentives to free-ride. Second, value of information is positive and exerting higher effort encourages other firms to develop the product. The two effects are familiar from the classical experimentation literature. If effort affects the distribution over future payoffs, the firms are more likely to lose the market if they do not develop the product. This alleviates free-riding. However, the players have an incentive to free-ride on each others' impact on the state. This strengthens the free-riding effect.

Finally, we show that if the product value is sensitive to the firms' development efforts, the firms eventually gets trapped in the bad state of the world. The entrapment is a consequence of the players' restricted level of effort together with the fact that a sufficiently high level of product development is need to keep the market. In the noncooperative solution, the aggregate development effort is too low, which gradually drives down the product value. The retardation of the innovation is inefficient and does not occur in the social planners' optimum. Moreover, it cannot occur in Bolton and Harris (1999) in which the value does not depend on the firms' action; in contrast, experimentation goes on indefinitely.

5.1 Markov Perfect Equilibria

We start the analysis by defining the players' strategies and the equilibrium concepts formally. In continuous time, a strategy is a stochastic process on which we impose appropriate measurability conditions

Definition 1 (Strategy)**.** *A strategy of player i is a stochastic process $\alpha_i = \{\alpha_{i,t} \in [0,1], 0 \le t < \infty\}$ progressively measurable with respect to the filtration $\mathcal{F}_t^{\alpha}$.*

We restrict the attention on a strict subset of possible strategies, stationary Markovian strategies. Each player chooses his strategy as a function of the belief ϕ_t only. Formally, we define

Definition 2 (Stationary Markovian Strategy)**.** *A strategy α_i is a stationary Markovian strategy if $\alpha_{t,i} = \alpha_i(\phi_t)$ for all $0 \le t < \infty$.*

At equilibrium, each player chooses his strategy such that it is a best response to the other players' strategies. In Markov perfect equilibria, all players play stationary Markovian strategies. At time t let

$$\alpha_{-i}(\phi_t) \equiv (\alpha_1(\phi_t), ..., \alpha_{i-1}(\phi_t), \alpha_{i+1}(\phi_t), ..., \alpha_N(\phi_t))$$

denote the vector summarizing a stationary Markovian strategy profile chosen by the players other than i. Then we can formally define the equilibrium as follows

Definition 3 (Markov Perfect Equilibrium)**.** *A Markov perfect equilibrium is a Nash equilibrium in which*

(i) *the players update their common belief according to the filtering equation (8), starting from the initial value $\phi_0 = \phi$;*

(ii) *each player i chooses a stationary Markovian strategy $\alpha_i(\phi)$ such that it is a best response to the stationary Markovian strategy profile $\alpha_{-i}(\phi)$.*

The restriction on stationary Markovian strategies allows us to use standard dynamic programming tools to derive the players' best response correspondences. Fix a stationary Markovian strategy profile $\alpha_{-i}(\phi_t)$ by the other players and consider the decision problem of player i. At any point, the player i's optimal effort level can be derived from his Hamilton-Jacobi-Bellman equation

$$rv_i(\phi) = \sup_{\alpha_i(\phi)\in[0,1]} \Bigg\{ \alpha_i(\phi)(\mu\phi - c_i) + \lambda(\delta_i\alpha_i(\phi) + \delta_{-i}^{\top}\alpha_{-i}(\phi) - \phi)v_i'(\phi) + \frac{1}{2}(\alpha_i(\phi) + \mathbf{1}^{\top}\alpha_{-i}(\phi))\frac{\mu^2}{\sigma^2}\phi^2(1-\phi)^2 v_i''(\phi)\Bigg\}, \quad (15)$$

where $\delta_{-i} = (\delta_1, ..., \delta_{i-1}, \delta_{i+1}, ..., \delta_N)$ summarizes the impact of the other players' development efforts on the product quality. Fixing the other players' strategies, we can derive the player i's best response from (15).

By comparing an individual firm's decision problem (15) with the social planner's Hamilton-Jacobi-Bellman equation (13), we can immediately see that the players' development efforts have an externality to the other players' value that a player does not fully internalize when choosing his strategy. We analyze the resulting free-riding problem in detail below.

Each player chooses his strategy to maximize his value in (15). Taking into account the strategies chosen by the other players, the optimal choice implies a best response by the player. The best response correspondence is summarized in the following lemma

Lemma 2 (Best Response). *A stationary Markovian strategy $(\alpha_i(\phi_t))_{t\geq 0}$ is a best response to the stationary Markovian strategy profile $(\alpha_{-i}(\phi_t))_{t\geq 0}$ with respect to the belief process $(\phi_t)_{t\geq 0}$ if and only if the effort strategy solves*

$$\alpha_i(\phi) \in \arg\sup_{\alpha\in[0,1]} \left\{ -\lambda\phi v_i'(\phi) + \alpha_{-i}^\top(\phi)\left(\lambda\delta_{-i}v_i'(\phi) + \frac{1}{2}\mathbf{1}\frac{\mu^2}{\sigma^2}\phi^2(1-\phi)^2 v_i''(\phi)\right) \right.$$
$$\left. +\alpha_i(\phi)\left[\phi\mu - c_i + \lambda\delta_i v_i'(\phi) + \frac{1}{2}\frac{\mu^2}{\sigma^2}\phi^2(1-\phi)^2 v_i''(\phi)\right]\right\}. \quad (16)$$

Proof. See Appendix. □

If $\lambda = 0$, (16) reduces to the player i's best response correspondence in Bolton and Harris (1999). If $\lambda > 0$, the player's best response correspondence contains additional terms that describe the effects of the players' development efforts and the exogenous state switching on the players' optimal choice.

The term in the parenthesis on the second line of (16) describes the player i's opportunity cost of developing the product. The first term, $\phi\mu - c_i$, which is eventually negative, describes the player's myopic net profit of product development. The second term, $\lambda\delta_i v_i'(\phi)$ describes the players' benefit from his own development efforts. The last term, $1/2\mu^2/\sigma^2\phi^2(1-\phi)^2 v_i''(\phi)$ is the shadow value of information that describes the gain of acquiring information about the product value. We will show below that $v_i'(\phi) \geq 0$ and $v_i''(\phi) \geq 0$ such that both the development efforts and learning have positive value for the player.

The last term in the parenthesis of the first line of (16) describes the effect of the other players' effort in the player i's value. The first term, $\lambda\delta_{-i}v_i'(\phi)$ describes the externality that the other players' development efforts have on the player i's value that results from the players' effort affecting the state. The second term, $1/2\mu^2/\sigma^2\phi^2(1-\phi)^2 v_i''(\phi)$, is the shadow value of information provided by the other players' effort. Both effects are positive such that the other players' effort have a positive externality on the player i's value.

We will show below that the player i's effort increases in his value $v_i(\phi)$. Now the other players' effort has a positive externality on the player i's value. This increases the players' effort at equilibrium and guarantees that the players are willing to exert effort longer than a single player would. Bolton and Harris (1999) first discovered the encouragement effect which is also present in our model.

However, the player does not fully internalize the effect of his development efforts on the other players' value, but has an incentive to free-ride on other players' actions. Also this effect was first discovered by Bolton and Harris (1999), and it is affected by the fact that the players have an incentive to free-ride on each others' development efforts, not only on their learning about the product value. However, we will argue below that the encouragement effect remains unaffected. The players benefit from each others' development efforts, but the gain comes too late, and does not benefit the players at the boundary.

The second term in (16), $-\lambda\phi v_i'(\phi)$, is new and describes the effect of the changing environment. If the consumers' preferences are more sensitive to the firms' development efforts, their valuation is harder to track. Obtaining information becomes more important if the preferences are more sensitive to the development efforts. We will show below that free-riding is alleviated.

5.2 Symmetric Equilibria

We next analyze the symmetric Markov perfect equilibria in more detail. In a symmetric equilibrium, each player faces the same decision problem and the equilibrium best responses coincide. The restriction allows us to make more detailed predictions about the properties of the equilibrium.

We analyze the equilibria in the game with $N \geq 2$ players. To facilitate symmetry, we assume that each player's development efforts have the same effect on the product value and that they face the same production cost. Thus, $\delta_i = \delta_j = \gamma/N$ and $c_i = c_j = c$ for all $i, j \in \{1, ..., N\}$.

Formally, a symmetric Markov perfect equilibrium (MPE) is defined as follows

Definition 4 (Symmetric MPE)**.** *A Markov perfect equilibrium is symmetric if* $\alpha_i(\phi_t) = \alpha_j(\phi_t)$ *for all* $t \geq 0$*, and all players i and j.*

In the symmetric Markov perfect equilibria, we can identify three qualitatively different effort regimes. If the players are very optimistic about the product value, exerting full effort is a dominant strategy. Similarly, if the players are very pessimistic, abandoning the product technology is optimal. However, at intermediate beliefs, the players have an incentive to free-ride on each others' development efforts. At equilibrium, they do not develop the product at the full scale. Individually, each player is just indifferent between her strategies.

The player's optimal investment strategy can be found by applying symmetry on (16) and is summarized in the following proposition

Proposition 4 (Effort Regimes)**.** *The symmetric Markov perfect equilibria are characterized by three effort regimes:*

1. *If* $c - \phi\mu < \lambda\gamma/N v_i'(\phi) + \mu^2\phi^2(1-\phi)^2 v_i''(\phi)/(2\sigma^2)$*, it is optimal to set* $\alpha_i(\phi) = 1$.
2. *If* $c - \phi\mu = \lambda\gamma/N v_i'(\phi) + \mu^2\phi^2(1-\phi)^2 v_i''(\phi)/(2\sigma^2)$*, the players choose an interior level of effort*

$$\alpha_i(\phi) = \frac{r v_i(\phi) + \lambda\phi v_i'(\phi)}{(N-1)(c - \phi\mu)}. \tag{17}$$

3. *If* $c-\phi\mu > \lambda\gamma/Nv_i'(\phi)+\mu^2\phi^2(1-\phi)^2v_i''(\phi)/(2\sigma^2)$*, it is optimal to set* $\alpha_i(\phi)=0$.

(17) suggests that the incentives to develop the product are stronger at the intermediate range of beliefs when the sensitivity λ of the product value for the product is higher. If $\lambda > 0$, the product needs constant development for the firms' not to lose the market. Higher λ implies a higher probability of the state changing. Then the risk of losing the market increases. This makes free-riding less attractive and increases the players' effort at equilibrium.

Proof. The Hamilton-Jacobi-Bellman equation (15) if linear in $\alpha_i(\phi)$. The optimal control is a bang-bang policy. If $c-\phi\mu < \lambda\gamma/Nv_i'(\phi)+\mu^2\phi^2(1-\phi)^2v_i''(\phi)/(2\sigma^2)$, it is optimal to set $\alpha_i(\phi)=1$. If $c-\phi\mu > \lambda\gamma/Nv_i'(\phi)+\mu^2\phi^2(1-\phi)^2v_i''(\phi)/(2\sigma^2)$, $\alpha_i(\phi)=0$ is optimal.

Finally, consider ϕ such that $\lambda\gamma/Nv'(\phi)+\phi^2(1-\phi)^2v_i''(\phi)/(2\sigma^2)=c-\mu\phi$. Then (15) implies that $\alpha_i(\phi)\in[0,1]$ is optimal. We derive the equilibrium effort level that implies that each player is just indifferent between different effort levels. Substituting for $\alpha_i(\phi)=\alpha_j(\phi)$, we find that

$$\begin{aligned} rv_i(\phi) =&\alpha_i(\phi)(\phi\mu-c)+\lambda\left(\gamma\alpha_i(\phi)-\phi\right)v_i'(\phi)+\frac{N}{2}\alpha_i(\phi)\frac{\mu^2}{\sigma^2}\phi^2(1-\phi)^2v_i''(\phi)\\ =&\alpha_i(\phi)(\phi\mu-c)-\lambda\phi v_i'(\phi)+N\alpha_i(\phi)r(\rho-\phi\mu)\\ =&-\lambda\phi v_i'(\phi)+\alpha_i(\phi)(N-1)(c-\phi\mu). \end{aligned}$$

Rearranging yields the interior solution for the equilibrium effort level (17). □

We will show that we can identify two cutoff beliefs at which the players switch between the regimes. Let $\bar{\phi}$ denote the cutoff belief from which on the players choose an interior effort level and $\bar{\bar{\phi}}$ the cutoff belief at which the players stop developing the product. Notice that once the players abandon product development, they will never adopt it again since their belief drifts down deterministically according to (8). In absence of product development, losing the market becomes more and more likely.

Before writing down the main result that characterizes the equilibria, we next derive some key properties of the equilibria. Recall from Corollary 2 that the social planner keeps on developing the product longer than is myopically optimal. In particular, the result holds if there is only one player; i.e. $N=1$. We argue that also in the noncooperative game, the players develop the product longer than a myopic player would. The argument follows by contradiction. Suppose that none of the players would exert effort in a symmetric Markov perfect equilibrium as long as one player alone would. But then, once the players stop developing the product, exerting effort would still be profitable for each player alone. This is a profitable deviation and cannot be part of an equilibrium strategy.

The next lemma shows that the players always keep on developing the technology longer than a myopic player would

Lemma 3. *In the N-player noncooperative product development game,* $\bar{\bar{\phi}} \le \phi^M$.

Proof. We argue that the players never stop exerting effort if it is optimal for a single player to keep on developing the product. The proof follows by contradiction. The single player's decision problem is a special case of the social planner's problem with $N = 1$. Let ϕ^{*1} denote the corresponding threshold belief at which the single player abandons the technology.

Suppose that there exists a Markov perfect equilibrium of the noncooperative product development game with the strategy profile $\alpha^{MPE}(\phi)$ such that the players stop developing the product at some belief $\bar{\bar{\phi}} > \phi^{*1}$. Now for $\phi \in [\phi^{*1}, \bar{\bar{\phi}}]$ continuing selling the product is a profitable deviation for any single player. Thus, $\alpha^{MPE}(\phi)$ cannot be an equilibrium. It follows that $\bar{\bar{\phi}} \le \phi^{*1}$. The result follows since $\phi^{*1} \le \phi^M$ by Corollary 2. □

The next lemma shows that at the symmetric Markov perfect equilibrium, the value of information is positive; i.e. the value function $v_i(\phi)$ is convex. The players can only benefit from the resolution of uncertainty since this helps them to find the more efficient action. Moreover, we show that the value function increasing; the expected payoff is higher at higher beliefs.

Lemma 4. *At any symmetric Markov perfect equilibrium,* $v_i'(\phi) \ge 0$, $v_i''(\phi) \ge 0$ *and* $\lambda\gamma/N v_i'(\phi) + \mu^2\phi^2(1-\phi)^2 v_i''(\phi)/(2\sigma^2) \ge 0$.

Proof. Suppose first that $\phi \ge \phi^M$. At $\phi = 1$, the boundary condition $rv(1) = \alpha_i(1)(\mu - c_i) - \lambda(1-\gamma)$ that $\alpha_i(1) = 1$ is optimal. At the neighborhood of $\phi = 1$, $v_i'(\phi) > 0$ and $v_i''(\phi) \ge 0$ by (5), and therefore $\alpha_i(\phi) = 1$ by (17). The result follows by repeating the argument.

Next, suppose that $\phi \le \phi^M$. Again, if $\alpha_i(\phi) = 1$, the result follows from (5). If $\alpha_i(\phi) < 1$ is optimal, Proposition 4 implies that

$$\lambda\delta_i v_i'(\phi) + \frac{1}{2}\frac{\mu^2}{\sigma^2}\phi^2(1-\phi)^2 v_i''(\phi) = c_i - \phi\mu \ge 0, \tag{18}$$

with strict inequality if $\phi < \phi^M$. Therefore, we must have that either $v_i'(\phi) \ge 0$ or $v_i''(\phi) \ge 0$ or both. Notice that since $v_i'(\bar{\bar{\phi}}) = 0$, we must have both $v_i'(\phi) \ge 0$ and $v''(\phi) \ge 0$ at the neighborhood of $\bar{\bar{\phi}}$. Moreover, (5) implies that $v_i'(\phi) \ge 0$ and $v_i''(\phi) \ge 0$ at the neighborhood of $\bar{\phi}$. This implies that we would need to have $v_i''(\phi) < 0$ on a set of positive measure. We show that this cannot be the case.

Suppose that $v_i''(\phi) < 0$ on a set of positive measure and let ϕ^1 denote the largest point at which $v_i''(\phi) = 0$. Such a point exists since $v_i''(\phi)$ is continuous and

$v_i''(\bar{\phi}) \geq 0$. Let $[\phi^1, \phi^2]$, with $\phi^1 < \phi^2$ denote an interval on which $v_i''(\phi) \geq 0$. Then $v_i'(\phi^1) \leq v_i'(\phi^2)$. (18) together with $c - \mu\phi^1 > c - \mu\phi^2$ and $v_i''(\phi^1)$ implies that

$$\lambda\delta_i v_i'(\phi^1) > \lambda\delta_i v_i'(\phi^2) + \frac{1}{2}\frac{\mu^2}{\sigma^2}(\phi^2)^2(1-\phi^2)^2 v_i''(\phi^2) \geq \lambda\delta_i v_i'(\phi^2)$$

a contradiction. Thus, we must have $v_i''(\phi) \geq 0$. □

We argue that allowing the players' affect the distribution of future payoffs does not affect the encouragement effect. Formally, the value of information is always positive since $v_i''(\phi) \geq 0$ for all $\phi \geq \bar{\bar{\phi}}$.[9] The option value of investment is always positive with Brownian bandits.[10] However, $v_i'(\bar{\bar{\phi}}) = 0$ such that the value of additional development efforts is 0 at the boundary. The other players' development efforts eventually increase the value of the technology in the future, but it comes too late such that it will not benefit the players at the boundary.

Lemma 4 implies that the players' value increases with other players' investments. Indeed, we can conclude that

Corollary 3. *In a symmetric Markov perfect equilibrium, if* $\mathbf{1}^\top \alpha_{-i}(\phi) \geq \mathbf{1}^\top \tilde{\alpha}_{-i}(\phi)$, *then* $v_i(\phi) \geq \tilde{v}_i(\phi)$.

Similar to the social planner's problem, the solution to the noncooperative product development game has an isomorphic interpretation as a standard optimal stopping problem with which it is easier to deal. The symmetric Markov perfect equilibria are summarized in the following proposition

Theorem 1 (Symmetric Markov Perfect Equilibrium)**.** *There exists a symmetric Markov perfect equilibrium. Starting from a prior belief* ϕ_0, *the player i's payoff admits the following dynamics*

- *As long as* $\phi \geq \bar{\phi}$, $\alpha_i(\phi) = 1$ *for all* $i \in \{1, \ldots, N\}$ *and the player i's payoff is the unique solution of the following ordinary differential equation*

$$r v_i(\phi) = \phi\mu - c + \lambda(\gamma - \phi)v_i'(\phi) + \frac{N}{2}\frac{\mu^2}{\sigma^2}\phi^2(1-\phi)^2 v_i''(\phi) \tag{19}$$

  with the boundary conditions $(N-1)(c - \mu\bar{\phi}) = r v_i(\bar{\phi}) + \lambda\bar{\phi}v_i'(\bar{\phi})$ *and* $r v_i(1) = \mu - c - \lambda(1-\gamma)v_i'(1)$.

[9]In particular, one can show that $v_i''(\bar{\bar{\phi}}) > 0$. Otherwise, $\bar{\bar{\phi}} = \phi^M$, and $v_i'(\phi) = 0$ at the neighborhood of $\bar{\bar{\phi}}$, which would lead to a contradiction with (5).

[10]The same is not true for Poisson bandits for which the player's value is eventually concave if he does not experiment, cf. Keller et al. (2005).

- *When* $\bar{\bar{\phi}} \leq \phi < \bar{\phi}$, $\alpha_i(\phi) \in [0,1]$ *for all* i *and the player* i*'s payoff solves*

$$c - \mu\phi = \frac{\lambda\gamma}{N} v_i'(\phi) + \frac{\mu^2}{2\sigma^2}\phi^2(1-\phi)^2 v_i''(\phi). \tag{20}$$

The boundary conditions are $v_i(\bar{\bar{\phi}}) = 0$ *and* $v_i'(\bar{\bar{\phi}}) = 0$.

- *As soon as* ϕ *reaches* $\bar{\bar{\phi}}$, $\alpha_i(\phi) = 0$ *for all* i.

Moreover, the game has a unique symmetric Markov perfect equilibrium if $\lambda = 0$ *or* $\gamma = 1$ *and multiple symmetric Markov perfect equilibria if* $\lambda > 0$ *and* $\gamma < 1$.

Proof. See Appendix. □

(20) is an indifference condition and describes the players' value in the region of beliefs at which they choose an interior effort level. (19) describes the players' value on the region in which the exert full effort. By comparing the value functions that solve the average payoff in the social planner's problem (13) and the noncooperative game (15), when immediately see that they coincide. Free-riding is the only source of inefficiency that restricts the players' development efforts at equilibrium.

In our framework, the product needs constant development for the firms not to lose the market. The assumption has several consequences. First, free-riding is alleviated. Second, if the players become too pessimistic about the product value, it eventually becomes optimal to abandon the technology. If the firms stop developing the technology, they gradually become more pessimistic about the product value. Thus, if the technology is abandoned, the firms never exert positive effort again. Moreover, we will show below, the players eventually get trapped in the bad state and endogenously retard the technology.

Since the players' efforts "build up" the common state for the future, the product development game becomes a coordination game. We show in the Appendix that the game multiple symmetric Markov perfect equilibria. For intermediate beliefs, exerting higher effort is optimal at equilibrium if all players exert higher effort. Vice versa, it is optimal for the players to exert low effort if other players do so, too.

Finally, we show that the players eventually get trapped in the bad state of the world. Because of free-riding, the aggregate product development is not high enough, and the value of the technology decreases even if the players exert effort. Free-riding leads to retardation in the noncooperative product development game. Notice that retardation is not possible if $\lambda = 0$. Free-riding slows down the learning rate, but the investment goes on indefinitely, see Bolton and Harris (1999). Moreover, the retardation is inefficient and does not occur at the social planners' optimum. In contrast the social planner would would exert every effort to develop the product. For low beliefs, the planner's belief always drifts up according to (8).

Next, a natural question arises if free-riding can lead to retardation if the product value is high at time t. The next proposition derives conditions under which this is possible. In particular, the market size is not allowed to be too large. If the market is very large, the players always obtain a large number of signals, even if they exert very low effort. Then information acquisition dominates, and causes the players' belief to drift towards the true state.

The result is summarized in the next proposition.

Proposition 5 (Retardation)**.** *Consider a symmetric Markov perfect equilibrium of the noncooperative product development game with $N < \infty$ players. Then*

- *There exists a belief $\tilde{\phi}$ such that if $\phi_t < \tilde{\phi}_t$, the players' common belief is a (strict) supermartingale in the noncooperative product development game.*
- *Suppose that $\bar{\bar{\phi}} \leq \phi < \tilde{\phi}$. Then if*

$$(rv_i(\phi) + \lambda v_i'(\phi))\left(\frac{\gamma}{N} + \frac{\mu^2}{\lambda\sigma^2}\phi(1-\phi)^2\right) < \frac{N-1}{N}\phi(c - \mu\phi), \tag{21}$$

the players' belief is a (strict) supermartingale $s_t = 1$.

Proof. Suppose that $\bar{\bar{\phi}} \leq \phi < \bar{\phi}$ such that the players do not exert full effort. We identify the cutoff belief $\tilde{\phi}$, below which the belief (8) drifts down, and argue that $\tilde{\phi} > \bar{\bar{\phi}}$. From (8) we can see that the drift of the belief process is nonpositive if

$$\gamma\alpha_i(\phi_t) \leq \phi_t$$

Let $\tilde{\phi}$ denote the belief such that the previous condition holds with equality. By substituting for (17) and reorganizing, we find that $\tilde{\phi}$ is implicitly defined in

$$v_i(\tilde{\phi}) + \lambda v_i'(\tilde{\phi}) = \frac{\tilde{\phi}}{\gamma}(c - \mu\tilde{\phi})(N-1).$$

Since $v_i(\bar{\bar{\phi}}) = v_i'(\bar{\bar{\phi}}) = 0$, and $0 < \bar{\bar{\phi}} < c/\mu = \phi^M$ for $N < \infty$, we find that $\tilde{\phi} > \bar{\bar{\phi}}$.

To the second point, suppose that $s_t = 1$ such that the product value is high at time t. One can show that from the point of view of someone, who observes the true state, the players' common belief evolves according to[11]

$$d\phi_t = \lambda(\gamma\alpha_i(\phi_t) - \phi_t)dt + N\alpha_i(\phi_t)\frac{\mu^2}{\sigma^2}\phi_t(1-\phi_t)^2 dt + \mu\phi_t(1-\phi_t)\alpha(\phi_t)dZ_t.$$

Reasoning along the same lines, one can show that the drift of the true belief process is nonpositive if and only if (21) holds. □

[11]See Keller and Rady (1999); Liptser and Shiryayev (1977).

6 Poisson Bandits

In this section, we shortly review an alternative version of the model with Poisson bandits. The model presented here generalizes the classical Poisson bandit problem of Keller et al. (2005) to allows for the players' action to affect the distribution over future payoffs. The transition process that drives the changes of the product value over time follows the continuous-time Markov chain (2) and admits a similar dynamics than in the Brownian model. The proofs follow by adjusting the arguments for the Brownian model for the Poisson bandits and are mostly omitted.

The product development game admits a slightly different dynamics with Poisson bandits. If the product value is high, the players eventually yield a lump-sum payoff at exponentially distributed stochastic times. News fully reveal the product value, but the state eventually changes over time. We will show that the model is very similar, only with one difference: there is no encouragement effect.[12]

We focus on the good news case: exerting effort eventually yields positive lump-sum payoffs at stochastic times. In particular, we assume that the lump-sum payoff occurs with probability $\mu\alpha_t s_t dt$, distributed independently of the state transition process (2). To economize on notation, we normalize the size of the lump-sum payoff to 1. If the players observe a lump-sum, they know that the product value is high at time t. However, the state may change over time, an effect that causes the players' belief to drift down instantaneously. Thus, when the product value evolves over time, the players are never completely certain about the true value of the product.

6.1 Beliefs

The analysis follows along the same lines than in the model with Brownian noise. We first derive the filtering equation that governs the players' common posterior belief about the product value at time t. If the players observe a lump-sum payoff, their belief jumps to one. However, if $\lambda > 0$, the state of the world may change over time. The belief instantaneously drifts downwards and the players are never completely certain about the true state.

In absence of lump-sum payoffs, the players become more pessimistic about the product value. Also, they anticipate the effect of their development efforts on the state and take it into account when forming their beliefs. The latter effect is analogous to the corresponding effect in the Brownian model.

[12]Alternative specifications would include both good and bad arms yielding positive lump-sums or the bad arm eventually generating losses. We do not review the alternatives here, Keller and Rady (2010, 2015) provide comprehensive analysis with classical bandits, and prove the presence of encouragement effect.

The filtering equation for the evolution of the players' common posterior belief is summarized in the following proposition

Proposition 6 (Beliefs with Poisson Profits)**.** *Conditional on receiving no lump-sum payoffs, the players' common posterior evolves according to*

$$d\phi_t = \lambda(\delta^\top \alpha_t - \phi_t)dt - \mu \mathbf{1}^\top \alpha_t \phi_t (1-\phi_t)dt. \tag{22}$$

Proof. Similar to the proof of Proposition 1. □

The first part of (22) describes the evolution of the belief that follows from the anticipated effect of the possible state transition. The second term describes the decrease in players' belief if no lump-sum payoff is observed.

6.2 Social Planner's Solution

We next analyze the social planner's problem with Poisson distributed profits. The planner does not observe the product value for the product but learns about it according to (22), based on the observations of her payoffs. In particular, if she observes a lump-sum payoff, she knows that the product value is high at time t. However, the state may change at any point of time. Thus, the planner is never completely certain about the product value.

Similar to the Brownian case, the social planner's optimal decision is a threshold policy. If the planner is sufficiently optimistic about the product value, she exerts every effort to develop it. After a long enough period of absence of lump-sum payoffs, the planner gradually becomes too pessimistic, and abandons the product technology. Subsequently, the belief drifts further away from the boundary and it is never optimal to start developing the product again.

The result is summarized in the following proposition

Proposition 7 (Social Planner's Solution with Poisson Profits)**.** *There exists a cut-off belief ϕ^* such that (i) if $\phi \geq \phi^*$, the social planner exerts every effort to develop the product; i.e. $\alpha_i(\phi) = 1$ for all $i \in \{1,...,N\}$. The firm i's value function is the unique solution of the ordinary differential equation*

$$rv_i(\phi) = \phi\mu - c + (\lambda(\gamma - \phi) - \mu N\phi(1-\phi))v_i'(\phi) + \mu N\phi(v(1) - v(\phi)) \tag{23}$$

with the boundary conditions $v_i(\phi^*) = 0$, $v_i'(\phi^*) = 0$ and $rv(1) = (\mu - c) - \lambda(1-\gamma)v'(1)$. (ii) As soon as ϕ hits ϕ^*, it is optimal to set $\alpha_i(\phi_t) = 0$ in perpetuity.

Proof. Similar to Proposition 3. □

Unfortunately, finding a closed form solution for the social planner value function seems unlikely, apart from the special case $\lambda = 0$ that is exhaustively analyzed in Keller et al. (2005). However, a comparison with (23) allows us to derive conclusions about the players' strategic behavior in the noncooperative problem. In particular, we will prove the absence of encouragement effect in the noncooperative product development game with Poisson distributed profits.

6.3 Noncooperative Solution

Next, we solve for the symmetric Markov perfect equilibria of the noncooperative product development game. Again, we restrict the analysis on symmetric equilibria. The resulting dynamics is very similar to the Brownian model.

Again, both the impact, that the action has on the state, and the information about the product value are public goods, which gives incentives to free-riding. Free-riding is alleviated if the product value is more likely to change over time. Only the encouragement effect is absent. News are fully revealing and two jumps are equally informative than one is. Further jumps have no instantaneous value. As a consequence, additional experimentation by one player does not encourage the other players' to experiment more.

As in the Brownian model, we can clearly identify three effort regimes. If players are very optimistic about the product value, exerting full effort is a dominant strategy. If no lump-sum payoffs occurred over a sufficiently long time horizon, the players become too pessimistic and abandon the technology. For intermediate beliefs, the players choose an interior level of effort.

The results are summarized in the following proposition

Proposition 8 (Effort Regimes with Poisson Profits)**.** *The symmetric Markov perfect equilibria are characterized by three effort regimes:*

1. *If* $c-\phi\mu < (\lambda\gamma/N+\mu\phi(1-\phi))v_i'(\phi)-\mu\phi(v(1)-v(\phi))$*, it is optimal to set* $\alpha_i(\phi) = 1$.
2. *If* $c-\phi\mu = (\lambda\gamma/N+\mu\phi(1-\phi))v_i'(\phi)-\mu\phi(v(1)-v(\phi))$*, the players choose an interior effort level*
$$\alpha_i(\phi) = \frac{rv_i(\phi)+\lambda\phi v_i'(\phi)}{(N-1)(c-\phi\mu)}. \tag{24}$$
3. *If* $c-\phi\mu > (\lambda\gamma/N+\mu\phi(1-\phi))v_i'(\phi)-\mu\phi(v(1)-v(\phi))$*, it is optimal to set* $\alpha_i(\phi) = 0$.

Proof. Similar to Proposition 4. □

We can see that the model with Poisson distributed profits is very similar to the Brownian model. In particular, we can see from (24) that the players again internalize the effect of their development efforts on the product value. If the switching rate λ is higher, the incentive to free-ride is less severe. Similarly, the incentive to free-ride is more severe if the number of players N or the opportunity cost of product development $c - \phi\mu$ is high.

We show that there is one crucial difference between the models: the simplest model with Poisson profits exhibits no encouragement effect. Obtaining one lump-sum payoff is sufficient for the player to be convinced about the current state of the world. Future lump-sum payoffs have no instantaneous value. The difference between Poisson and Brownian models is exhaustively explored in Keller et al. (2005) and continues to hold in our case. The next lemma extends the result to our case with product development

Lemma 5. *In the symmetric Markov perfect equilibrium of the noncooperative product development game, no player exerts positive effort beyond the single player optimal cutoff* ϕ^{*1}.

Proof. Single player optimization problem solves the following Hamilton-Jacobi-Bellman equation

$$\begin{aligned} rv_i(\phi) &= \alpha_i(\phi)(\mu\phi - c) \\ &\quad + (\lambda(\alpha_i(\phi)\gamma - \phi) + \alpha_i(\phi)\mu\phi(1-\phi))v_i'(\phi) + \mu\phi\alpha_i(\phi)(v_i(1) - v_i(\phi)). \end{aligned}$$

Optimizing with respect to $\alpha_i(\phi)$ and using the boundary condition $v_i'(\phi^{*1}) = 0$, we find that

$$c - \phi^{*1}\mu = \mu\phi^{*1}(v_i(1) - v_i(\phi^{*1})).$$

Similarly, from Proposition 8 together with the boundary condition $v_i'(\bar{\bar{\phi}}) = 0$, we find that

$$c - \bar{\bar{\phi}}\mu = \mu\bar{\bar{\phi}}(v_i(1) - v_i(\bar{\bar{\phi}})).$$

If $v_i''(\bar{\bar{\phi}}) \neq 0$, we must have $\phi^{*1} = \bar{\bar{\phi}}$. Otherwise, if $v_i(\bar{\bar{\phi}}) = 0$, $v_i'(\phi) = 0$ and the result follows. □

7 Discussion

We examine a model of strategic experimentation in which the players' actions affect the probability distribution over future profits. The model finds a natural application in firms' developing a product under changing market conditions. The product value is uncertain and depends on the firms' development efforts. Effort is needed both to enhance the product value and to learn about its viability in the market.

We analyze the symmetric Markov perfect equilibria of the noncooperative product development game. We identify three qualitatively different effort regimes.

First, if the players are sufficiently optimistic about the product value, they always exert every effort to develop it. Second, if they become too pessimistic, they optimally abandon the technology. Third, at intermediate values of beliefs, the firms restrict their product development effort. The optimal effort strategy is such that the firms are just indifferent between the different effort levels at equilibrium.

We find that the players' incentives to exert effort are driven by two effects, free-riding and encouragement effect, that are similar to the classical experimentation literature. The firms' effort decision has an externality on other players' value that the player does not fully internalize. This gives incentives to free-ride on the other firms' actions. The free-riding effect is eventually stronger in our model since the players benefit, not only from each others' signals, but also on their impact on the common state. However, free-riding eventually results in the firms losing the market, an effect that alleviates free-riding.

The encouragement effect arises because other firms' payoffs provide valuable information about the product value. More information is valuable because it helps players to make a more efficient investment decision. This increases the option value of investment and encourages the players to keep on developing the product longer. We find that, while learning encourages effort, the fact that the players' actions affect the state does not.

Finally, we show that whenever the product value is sensitive to the firms' development efforts, the firms eventually get trapped in the bad state of the world. For intermediate beliefs, the firms have a strong incentive to free-ride on each others' actions. The aggregate level of development effort remains too low and the firms gradually lose the market. The firms retard the technology, which leads to an inefficiently short lifetime of the risky innovation.

8 Appendix

The appendix is organized as follows

1. We first derive the filtering equation governing the law of motion for the common beliefs. Since the firms' actions affect the distribution over future payoffs, belief updating is nonstandard.

2. We solve for the social planner's problem when she optimizes the profits of all N players simultaneously. We analyze both the case in which she observes the product value and the case in which the value is unobserved.

3. We solve for the symmetric Markov perfect equilibrium in which each player chooses his effort noncooperatively.

We start by deriving the filtering equation for the beliefs in Section 3. When the players exert effort, they both increase the probability of the good state in the future and learn from the success of past efforts. The filtering equation, that governs the evolution of beliefs, takes into account both effects.

Proof of Proposition 1. The payoffs in (4) are observationally equivalent to the signals

$$d\tilde{\pi}_{t,i} \equiv \frac{d\pi_{t,i}}{\alpha_{t,i}^{1/2}} = \alpha_{t,i}^{1/2}\mu\phi_t dt + \sigma dZ_{t,i}$$

for each player i, given the strategy $\alpha_{t,i}$. The players observe, not only their own signal, but also the other players' signals.

After observing the signals, the players update their common belief about the current state using Bayes rule,

$$\phi_{t^-} = \frac{\phi\xi_t}{\phi\xi_t + 1 - \phi},$$

where the likelihood ratio ξ_t evolves according to

$$d\xi_t = \frac{\mu}{\sigma}\xi_t(\alpha_t^{1/2})^\top dZ_t. \tag{25}$$

This uses Girsanov's theorem and the fact that the Brownian motions Z_i are independent across the players.

Notice that

$$\frac{\partial\phi_{t^-}}{\partial\xi_t} = \frac{p(1-p)}{(p\xi_t + 1 - p)^2} = \frac{\phi_{t^-}(1-\phi_{t^-})}{\xi_t},$$

$$\frac{\partial^2\phi_{t^-}}{\partial\xi_t^2} = -\frac{2p^2(1-p)}{(p\xi_t + (1-p))^3} = -\frac{2\phi_{t^-}^2(1-\phi_{t^-})}{\xi_t^2}.$$

By Itô's lemma, the players' posterior belief evolves according to

$$d\phi_{t^-} = (\alpha_t^{1/2})^\top \frac{\mu}{\sigma}\phi_{t^-}(1-\phi_{t^-})dZ_t^\alpha$$

prior to the state transition. The Brownian motions are related by $dZ_t^\alpha = dZ_t - \alpha_t^{1/2}\mu\phi_{t^-}dt$, i.e. the ith row of the vector dZ_t^α is $dZ_{t,i}^\alpha = dZ_{t,i} - \mu\phi_{t^-}\alpha_{t,i}^{1/2}dt$.

Next, the players anticipate that the product value eventually changes over time. The change depends on the players' development efforts. For $s \in \{0,1\}$, let $\Delta N_t^s = N_t^s - N_{s^-}^s$, where

$$\Delta N_t^s = \begin{cases} 1 & \text{if} \quad s_t \neq s_{t^-}, \\ 0 & \text{if} \quad s_t = s_{t^-}. \end{cases}$$

The process N_t^s follows the continuous-time Markov chain (2).

When forming their posterior belief for the high state at time t, the players assign a belief ϕ_{t^-} for the high valuation and a probability of $1-\Delta N_t^1$ for the remaining in the high state. Similarly, the players belief that the product value is low with probability $1-\phi_{t^-}$, but becomes high with probability ΔN_t^0. Thus, taking into account the possibility of the state transition, the players assign the belief

$$\phi_t = (1-\Delta N_t^1)\phi_{t^-} + (1-\phi_{t^-})\Delta N_t^0$$

to the high value.

Applying Itô's formula for jump diffusions,[13] we find that

$$\begin{aligned} d\phi_t = -\phi_{t^-}dN_t^1 + (1-\phi_{t^-})dN_t^0 + (1-(N_{t^-}^1 - N_{t^-}^1))d\phi_{t^-} - (N_{t-}^0 - N_{t^-}^0)d\phi_{t^-} \\ = \lambda(\delta^\top\alpha_t - \phi_t)dt + \frac{\mu}{\sigma}\phi_t(1-\phi_t)(\alpha_t^{1/2})^\top dZ_t^\alpha. \end{aligned}$$

The last equality follows since ϕ_t is continuous almost everywhere. □

We next derive the social planner's expected payoff when she observes the current state. In particular, we derive conditions under which the social planner always exert effort and conditions under which she abandons the product if the value becomes low.

Proof of Proposition 2. Let $\overline{v}_i(\phi)$ denote the complete information payoff from the investment i. We first derive the planner's expected payoff if she abandons the product as soon as its value becomes low. Then we derive the expected payoff when the planner always exerts effort and determine conditions under which effort is always optimal.

Consider first the situation in which the planner stops exerting effort if the product value becomes low. Of course, if the state is high, the planner exerts

[13] See Øksendal and Sulem (2005).

every effort to develop the product; i.e. $\alpha_t = 1$. However, the state eventually becomes low, an event after which the planner stops developing the product. The distribution of s_t follows (2) and is now observed. At $t = 0$, $s_0 = 1$ with probability ϕ. The expected profit is

$$\overline{v}_i(\phi) = \phi\left[\int_0^\infty e^{-rt} e^{-\lambda(1-\gamma)t}(\mu - c_i)dt\right] = \frac{\phi(\mu - c_i)}{r + \lambda(1-\gamma)}.$$

Next, suppose that the planner always exerts effort. Conditional on $s_t \in \{0,1\}$, the planner's payoff solves the system of equations

$$\begin{aligned} rv_i^1 &= \mu - c_i + \lambda(1-\gamma)v_i^0 - \lambda(1-\gamma)v_i^1, \\ rv_i^0 &= -c_i + \lambda\gamma v_i^1 - \lambda\gamma v_i^0. \end{aligned}$$

The solution is

$$v_i^1 = \frac{(r+\lambda\gamma)\mu}{r(r+\lambda)} - \frac{c_i}{r}, \tag{26}$$

$$v_i^0 = \frac{\lambda\gamma\mu}{r(r+\lambda)} - \frac{c_i}{r}, \tag{27}$$

and the expected payoff is (12).

Finally, we check if it is optimal to stop developing the product if its value becomes low. If the planner exerts no effort, she earns a continuation payoff 0. If she keeps on developing the product, the payoff is (27). By comparing the payoffs, we find that it is optimal to keep on developing the product if and only if (10) holds. □

The next step is to analyze the social planner's problem in the case in which she does not observe the product value. The proof follows in two steps: We first verify that the threshold policy suggested in Section 4 maximizes the social planner's expected payoff. Then we apply an appropriate change of variables to recast the problem on a new domain on which standard theorems imply existence and uniqueness of the solution.

Proof of Proposition 3. Since (13) is linear in $\alpha_i(\phi)$, the optimal policy satisfies

$$\alpha \begin{cases} = 1 & \text{if} \quad c - \phi\mu < \lambda\gamma v_i'(\phi) + N\mu^2\phi^2(1-\phi)^2 v_i''(\phi)/(2\sigma^2), \\ \in [0,1] & \text{if} \quad c - \phi\mu = \lambda\gamma v_i'(\phi) + N\mu^2\phi^2(1-\phi)^2 v_i''(\phi)/(2\sigma^2), \\ = 0 & \text{if} \quad c - \phi\mu > \lambda\gamma v_i'(\phi) + N\mu^2\phi^2(1-\phi)^2 v_i''(\phi)/(2\sigma^2). \end{cases}$$

If $\phi\mu - c + \lambda\gamma v_i'(\phi) + N\mu^2\phi^2(1-\phi)^2/(2\sigma^2)v_i''(\phi) \geq 0$, it is optimal to set $\alpha_i(\phi) = 1$. Otherwise, $\alpha_i(\phi) = 0$ is optimal. Therefore, it is optimal to choose either $\alpha_i(\phi) = 0$

or $\alpha_i(\phi) = 1$ almost everywhere. The planner's optimal investment strategy is a threshold policy.

Moreover, notice that if the planner exerts no effort, she gradually becomes more pessimistic about the product value. In particular, if the technology is abandoned, the belief, that planner assigns to the high valuation, drifts further away from the boundary according to (8). Therefore, if it is optimal to abandon the technology, it will never be adopted again. By substituting for the optimal strategy into the Hamilton-Jacobi-Bellman equation (13), we can see that the value function $v_i(\phi)$ has to satisfy the value matching condition, $v_i(\phi^*) = 0$, and the smooth pasting condition $v_i'(\phi^*) = 0$.

The result now follows now from a standard verification argument. We show that the process

$$S_t = \int_0^t e^{-rs} \alpha_i(\phi_s)(\phi_s \mu - c) ds + e^{-rt} v_i(\phi_t) \tag{28}$$

is a supermartingale for an arbitrary investment policy, and a martingale if the optimal policy is chosen.

Applying Itô's lemma with (8) on (28), we find that

$$\begin{aligned} e^{rt} dS_t = {} & \alpha_i(\phi_t)(\phi_t \mu - c) dt - r v_i(\phi_t) dt + \lambda(\delta^\top \alpha(\phi_t) - \phi_t) dt \\ & + \mathbf{1}^\top \alpha(\phi_t) \frac{1}{2} \frac{\mu^2}{\sigma^2} \phi_t^2 (1 - \phi_t)^2 v_i''(\phi_t) dt + \frac{\mu}{\sigma} \phi_t (1 - \phi_t)(\alpha(\phi_t)^{1/2})^\top dZ_t^\alpha . \end{aligned}$$

From (13), we can see that S_t is a supermartingale and it is a martingale if $\alpha_i(\phi_t) = 1$ if $\phi_t \geq \phi^*$ and $\alpha_i(\phi_t) = 0$ if $\phi_t < \phi^*$.

The firm value at $t = 0$ satisfies

$$E\left[\int_0^\infty e^{-rt} \alpha_i(\phi_t)(\phi_t \mu - c) dt\right] = E[S_\infty] \leq S_0 = v_i(\phi_0),$$

with equality if and only if the optimal policy is chosen.

It remains to prove that the solution exists and is unique. We apply the change of variables $x = (1-\phi)/(1-\phi^*)$ to rewrite (13) as on $x \in [0,1]$

$$\begin{aligned} r y(x;\phi^*) = {} & \mu(1 - x(1-\phi^*)) - c \\ & - \frac{\lambda(\gamma - 1 + x(1-\phi^*))}{1-\phi^*} y'(x;\phi^*) + \Phi(1 - x(1-\phi^*))^2 x^2 y''(x;\phi^*), \end{aligned} \tag{29}$$

with $\Phi = N\mu^2/(2\sigma^2)$ and the boundary conditions

$$r y(0;\phi^*) - \frac{\lambda(1-\gamma)}{1-\phi^*} y'(0;\phi^*) - (\mu - c) = 0, \tag{30}$$

$$y(1;\phi^*) = 0, \tag{31}$$

$$y'(1;\phi^*) = 0. \tag{32}$$

and an unknown parameter ϕ^*. Since $\phi^* \leq \phi^M < 1$, (29) satisfies the standard Lipschitz and growth conditions in x and ϕ^*. Then by standard arguments, the solution of (29)-(32) exists and is unique. □

We next verify that the strategies described in Lemma 2 form best responses when the players play stationary Markov strategies.

Proof of Lemma 2. We can conclude from (16) that the player i's optimal investment strategy satisfies

$$\alpha \begin{cases} = 1 & \text{if} \quad c - \phi\mu < \lambda\delta_i v_i'(\phi) + \mu^2\phi^2(1-\phi)^2 v_i''(\phi)/(2\sigma^2), \\ \in [0,1] & \text{if} \quad c - \phi\mu = \lambda\delta_i v_i'(\phi) + \mu^2\phi^2(1-\phi)^2 v_i''(\phi)/(2\sigma^2), \\ = 0 & \text{if} \quad c - \phi\mu > \lambda\delta_i v_i'(\phi) + \mu^2\phi^2(1-\phi)^2 v_i''(\phi)/(2\sigma^2). \end{cases}$$

The result follows now from a standard verification argument. We show that the process

$$V_t = \int_0^t e^{-rt}\alpha(\phi_s)(\phi_s\mu - c)ds + e^{-rt}v_i(\phi_t) \tag{33}$$

is a supermartingale for an arbitrary control and a martingale if the optimal control is chosen.

Applying Itô's lemma with (8) on (33), we find that

$$\begin{aligned} e^{rt}dV_t =& \alpha_i(\phi_t)(\phi_t\mu - c)dt - rv_i(\phi_t) \\ &+ \lambda(\delta_i(\alpha_i(\phi_t) + \mathbf{1}^\top\alpha_{-i}(\phi_t)) - \phi_t)v_i'(\phi_t)dt \\ &+ \frac{1}{2}(\alpha_i(\phi_t) + \mathbf{1}^\top\alpha_{-i}(\phi_t))\frac{\mu^2}{\sigma^2}\phi^2(1-\phi)^2 v_i''(\phi_t)dt \\ &+ \frac{\mu}{\sigma}\phi_t(1-\phi_t)v_i'(\phi_t)(\alpha^{1/2}(\phi_t)dZ_{t,i} + \alpha_{-i}^{1/2}(\phi_t)dZ_{t,-i}^{\alpha}). \end{aligned}$$

We can now see from (15) that V_t is a supermartingale. It is a martingale only if the optimal control is applied.

The firm value at time 0 satisfies

$$E\left[\int_0^\infty e^{-rt}\alpha_i(\phi_t)(\phi_t\mu - c)dt\right] = E[V_\infty] \leq V_0 = v_i(\phi_0)$$

with equality if and only if the optimal control is chosen. □

We next proof existence of solution and examine the uniqueness properties of the symmetric Markov perfect equilibria.

Proof of Theorem 1. We apply the change of variables $x = (1-\phi)/(1-\bar{\phi})$ on $(\bar{\phi}, 1]$ and $x = (\bar{\phi}-\phi)/(\bar{\phi}-\bar{\bar{\phi}})$ on $(\bar{\bar{\phi}}, \bar{\phi}]$ to recast (15) and (20) as solving a system of ordinary differential equations on $x \in [0,1]$

$$
\begin{aligned}
ry_1(x;\bar{\phi},\bar{\bar{\phi}}) = \mu(1-x(1-\bar{\phi})) - c - \frac{\lambda(\gamma-1+x(1-\bar{\phi}))}{(1-\bar{\phi})} y_1'(x;\bar{\phi},\bar{\bar{\phi}}) \\
+\Phi_1(1-x(1-\bar{\phi}))^2 x^2 y_1''(x;\bar{\phi},\bar{\bar{\phi}}),
\end{aligned} \tag{34}
$$

$$
\begin{aligned}
-\frac{\lambda\gamma}{N(\bar{\phi}-\bar{\bar{\phi}})} y_2'(x;\bar{\phi},\bar{\bar{\phi}}) + \frac{\Phi_2(\bar{\phi}-x(\bar{\phi}-\bar{\bar{\phi}}))^2(1-\bar{\phi}+x(\bar{\phi}-\bar{\bar{\phi}}))^2}{(\bar{\phi}-\bar{\bar{\phi}})^2} y_2''(x;\bar{\phi},\bar{\bar{\phi}}) \\
= c - \mu(\bar{\phi}-x(\bar{\phi}-\bar{\bar{\phi}})),
\end{aligned} \tag{35}
$$

where $\Phi_1 = N/2\mu^2/\sigma^2$ and $\Phi_2 = \mu^2/(2\sigma^2)$; with the boundary conditions

$$ry_1(0;\bar{\phi},\bar{\bar{\phi}}) - \frac{\lambda(1-\gamma)}{1-\bar{\phi}} y_1'(0;\bar{\phi},\bar{\bar{\phi}}) - (\mu - c) = 0, \tag{36}$$

$$ry_1(1;\bar{\phi},\bar{\bar{\phi}}) - \frac{\lambda\bar{\phi}}{1-\bar{\phi}} y_1'(1;\bar{\phi},\bar{\bar{\phi}}) - (N-1)(c-\mu\bar{\phi}) = 0, \tag{37}$$

$$y_2'(1;\bar{\phi},\bar{\bar{\phi}}) = 0, \tag{38}$$

$$y_2(1;\bar{\phi},\bar{\bar{\phi}}) = 0, \tag{39}$$

$$-\frac{y_1'(1;\bar{\phi},\bar{\bar{\phi}})}{1-\bar{\phi}} + \frac{y_2'(0;\bar{\phi},\bar{\bar{\phi}})}{\bar{\phi}-\bar{\bar{\phi}}} = 0, \tag{40}$$

$$y_1(1;\bar{\phi},\bar{\bar{\phi}}) - y_2(0;\bar{\phi},\bar{\bar{\phi}}) = 0. \tag{41}$$

Since $0 < \bar{\bar{\phi}} < \bar{\phi} < 1$, the system (34)-(35) satisfies Lipschitz and growth conditions on $x \in [0,1]$.

If $\lambda = 0$ or γ, the system (34)-(41) has standard mixed boundary conditions[14] and admits a unique solution. If $\lambda > 0$ and $\gamma < 1$, the boundary conditions (36), (37) and (40) depend on $\bar{\phi}$. This is a system with periodic boundary conditions that has infinitely many solutions. □

[14] The boundary conditions of a system of ordinary equations are called mixed if they contain both boundary values and their derivatives.

References

Dirk Bergemann and Juuso Välimäki. Bandit problems. In Steven N. Durlauf and Lawrence E. Blume, editors, *The New Palgrave Dictionary of Economics*. Macmillan Press, 2nd edition, 2008.

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

Roland G Fryer and Philipp Harms. Two-armed restless bandits with imperfect information: Stochastic control and indexability. Technical report, National Bureau of Economic Research, 2013.

Johannes Hörner and Andrzej Skrzypacz. Learning, experimentation and information design. 2016.

Ioannis Karatzas. Gittins indices in the dynamic allocation problem for diffusion processes. *The Annals of Probability*, pages 173–192, 1984.

Godfrey Keller and Sven Rady. Optimal experimentation in a changing environment. *The Review of Economic Studies*, 66(3):475–507, 1999.

Godfrey Keller and Sven Rady. Price dispersion and learning in a dynamic differentiated-goods duopoly. *RAND Journal of Economics*, pages 138–165, 2003.

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

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

Godfrey Keller, Sven Rady, and Martin Cripps. Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68, 2005.

Robert S Liptser and AN Shiryayev. *Statistics of Random Processes: Vol.: 1.: General Theory*. Springer-Verlag, 1977.

Bernt Karsten Øksendal and Agnes Sulem. *Applied stochastic control of jump diffusions*, volume 498. Springer, 2005.

Ernst L Presman. Poisson version of the two-armed bandit problem with discounting. *Theory of Probability & Its Applications*, 35(2):307–317, 1991.

Mikhail Safronov. Obtaining experience in a multi-armed bandit. 2014.

Melissa A Schilling. *Strategic management of technological innovation*. Tata McGraw-Hill Education, 2005.