Unraveling of Cooperation in Dynamic Collaboration

Suvi Vasama*

* Humboldt-Universität zu Berlin, Germany

SFB 649 ECONOMIC RISK BERLIN

Unraveling of Cooperation in Dynamic Collaboration[*]

Suvi Vasama

November 7, 2016

Abstract

We examine collaboration in a one-arm bandit problem in which the players' actions affect the distribution over future payoffs. The players need to exert costly effort both to enhance the value of a risky technology and to learn about its current state. Both product value and learning are public goods, which gives the players incentives to free-ride on each others' actions. This leads to an inefficiently low aggregate level of effort. When the players' actions affect the distribution over future payoffs, they eventually get trapped in the low action, causing an inefficient unraveling of the game. Moreover, the players' incentives to exert effort depend on the state that in turn depends on the aggregate effort. If the players start restricting effort when the belief decreases in expectation, the two effects play in the same direction. Higher effort encourages higher effort and vice versa. Unraveling leads to multiple symmetric Markov perfect equilibria.

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

JEL Classification: C73, D83, O31.

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 spent over three years and $50 million on an unsuccessful attempt to develop a new standard for an operating system for computers that would challenge Microsoft Windows.

Constant 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WAP Forum merged to a new Open Mobile Architecture to form a new initiative, Open Mobile Alliance (OMA), and has by now been replaced by new, more elaborate smart phone standards.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, the technology is a public good; the firms benefit each others efforts to facilitate the common standard.

The main result of the paper shows an inefficient unraveling of the cooperation in the dynamic collaboration relationship. The unraveling follows since the players fail to coordinate on the efficient action if they are not sufficiently confident about the viability of the technology. Due to the insufficiently high level of effort, the players quickly lose the market. Moreover, we show that the players' incentive to exert effort depend on their current assessment of the value of the technology, which again depends on the players' effort. Whenever the two effects play in the same direction, the game has multiple symmetric Markov Perfect equilibria. Otherwise, the symmetric Markov Perfect equilibrium is unique.

Specifically, the players engage in dynamic collaboration while facing uncertainty about the value of the technology that changes over time depending on the players' effort. At each point in time, each player chooses how much effort they exert in selling and developing the product. Effort is costly and the returns are uncertain and depend on the unobserved state of the world. The state changes stochastically over time and the probability distribution over future payoffs depends on the players' actions. When the players exert effort, they receive noisy signals about the state. Thus, effort has two consequences in the model: the players' receive more informative signals about the current state and they improve

---

[1]Such collaboration is legal in the US by the National Cooperative Research Act of 1984 and in the EU by Article 101(3) TFEU 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.

the probability distribution over future payoffs.

We analyze the noncooperative game in which each player individually decides how much effort to exert. The state of the world is the same across the players while each player carries the cost of effort individually. The players always have the option to wait and see if the other players' efforts resulted in success. In the noncooperative 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 game, we identify two important benchmark cases: the complete information problem and the social planner's solution when the current state is unobserved. With complete information, the planner always exerts effort if the current state is high. If the state becomes low, the planner has to decide if she should continue exerting effort. 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 exert full effort. Otherwise, the planner abandons the product as soon as its value becomes low. In absence of effort, the product value stays low and the planner never exerts effort again.

Next, we analyze the social planner's optimal decision if she does not observe the state, 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 strategy is a threshold policy that only conditions on the planner's belief about the state. If the planner becomes too pessimistic, she abandons the technology. Now if the planner does not exert effort, the technology 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 players individually decide how much effort to exert. The players observe both each others' payoffs and their effort levels, but face uncertainty about the current state. We solve for the symmetric Markov Perfect equilibria of the noncooperative 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 effort regimes, depending on the players' belief about the current state. Firstly, if the players are very optimistic, effort has positive instantaneous value and the players exert every effort to extract the surplus. 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 non-cooperative solution from the social planner's solution.

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 strategic learning literature. The encouragement effect follows since the value of information is positive. The players 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 value of the technology and the information are public goods, and reduces the players' 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 if they abstain from effort; and (b) on the impact on their efforts on the state. The first effect alleviates free-riding: a sufficiently high level of 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 impact on the state. Secondly, restricting effort leads to an inefficiently early unraveling in the nonooperative game. The aggregate effort is not high enough to sustain the product value, but the players gradually lose the market. Thirdly, the players' incentives to exert effort depend on the current state that again depends on the aggregate effort. Thus, a higher effort encourages other players to exert higher effort. We show that this leads to multiple symmetric Markov equilibria of the noncooperative game.

Our model builds on the two-armed bandit problem that has been exhaustively studied in both economics and applied mathematics literature.[23] The strategic learning game is well explored for the pure experimentation case in which exerting effort allows for the players to learn about the value of a risky arm. 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 by allowing the value of the arm to depend on the players' effort. We find that in the noncooperative equilibrium, the players fail to sustain a sufficiently high level of effort, leading to an inefficient unraveling of the game. The unraveling 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 game. Moreover, the symmetric Markov Perfect Equilibrium is always unique while our game has

---

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

[3]Bergemann and Välimäki (2008) provide a comprehensive review of earlier literature in applied mathematics and economics.

multiple equilibria. In the pure experimentation problem, the players' incentives to pull the arm depend on the state, but the state is independent of the force at which the players pull. As a consequence, the noncooperative game has a unique symmetric Markov Perfect Equilibrium. 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.[4]

Bandit problems that involve evolving arms are sometimes called restless bandits in the literature. Fryer and Harms (2013) 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 framework borrows from related models that consider an autonomously evolving risky arm, in particular, from Keller and Rady (1999, 2003) who examine the problem of selling in the market of stochastically evolving demand. Khromenkova (2016) analyzes of the case of strategic experimentation with autonomously evolving restless bandits and Poisson distributed payoffs. Our model distinguishes from those by letting the value of the risky arm to depend on the players' action.

This paper is among the first ones to consider strategic learning in a restless bandit framework with endogenously evolving arms. Safronov (2014) develops a two-armed bandit model with competing players in which the player eventually becomes experienced in pulling an arm, which increases the payoff from that particular arm. If experienced, the player eventually gets trapped with a less valuable arm, instead of developing a potentially more valuable, but yet unexplored arm. 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.

---

[4]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.

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 learning about its value. The player's profit depends on the unobserved product value and on exogenous circumstances. Moreover, the value is stochastic and may change over time. The change is unobserved and depends on the players' actions.

Each player is facing a capacity constraint; without loss of generality we normalize the maximal effort that he can exert to 1. Let $\alpha_{t,i} \in [0,1]$ denote the intensity at which the player i develops the product and $\alpha_t \equiv (\alpha_{t,1}, ..., \alpha_{t,N})^\top$ a vector summarizing all players' efforts at time t.[5]

The product value at time t is

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

where $s_t \in \{0,1\}$ is the 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

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

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

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

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

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

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

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

---

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

The probability, that the product has a high value on the market, increases in the players' aggregate effort α_t. Moreover, to keep the model as simple as possible, we assume that if *all* players abstain from effort, i.e. set $\alpha_{t,i} = 0$ for all i, the low state becomes more and more likely over time.[6] 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 effort, he receives a certain flow payoff that we normalize to 0. If the player decides to exert effort, 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 effort and $\sigma > 0$ the volatility of the player's profit.

We assume that $\mu > c_i$ for all i such that the effort is valuable in the high state. 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 state. 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 accuracy of information increases with the aggregate effort 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 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 state $s_t = 1$ 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 $E[s_t|\mathscr{F}^{\alpha}] = \phi_t$.

---

[6]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.

3 Beliefs

This section presents a heuristic derivation for the evolution of the players' beliefs that they assign to the high state at time t; the formal derivation is delegated to the Appendix. The effect of the players' efforts on the state as well as the fact that the state 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 state is high in the future. Second, it allows the players to learn if their past efforts resulted in success.

We derive the filtering equation that governs the players' belief that they assign to the high state. 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 state 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 state is low, it may become high if the players keep on exerting effort. The players assign the probability $1 - \phi_t$ to the low state 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 effort, they anticipate that they gradually lose the market. The parameter λ tells how quickly this happens. δ_i is the long term mean of the belief conditional on the player i exerting full effort.

Next, if the players exert effort, they receive informative signals about the state. 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 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 that we normalized to 0. 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 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

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

with $\alpha_t^{1/2} = (\alpha_{t,1}^{1/2},...,\alpha_{t,N}^{1/2})^{\top}$ and $dZ_t^{\alpha} = (dZ_{t,1}^{\alpha},...,dZ_{t,N}^{\alpha})^{\top}$ denoting vectors summarizing all players' 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 state satisfies the following filtering equation*[7]

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

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

Proof. See Appendix. $\qquad\square$

If $\lambda > 0$ such that the state 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 efforts. The next corollary confirms the observation

Corollary 1. *Fix a strategy profile α_t. The player's common belief for the high state at time t 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 $e^{-\lambda t}$. $\qquad\square$

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 $\sum_i \delta_i = \gamma$. The model includes the following special cases

---

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

- If $\lambda = 0$, the state is fixed over time. In particular, if the players keep on exerting effort, 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 $\lambda \to \infty$, with any positive effort the state becomes i.i.d. Then there is no learning. One can show that if $\delta_i \mu > c_i$, the noncooperative game has a unique Markov perfect equilibrium in which all players exert full effort in perpetuity. Moreover, if $\gamma \mu < c_i$, effort is never optimal for the player i.

4 Social Planner's Solution

Before analyzing the noncooperative game, we solve for the social planner's problem when she decides about the optimal effort. In particular, we derive the optimal strategy in two benchmark cases: observable and unobservable state. The social planner's solution coincides with the players' optimal effort decision if they could coordinate on effort.

4.1 Complete Information Solution

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

With complete information, the planner always exerts full effort if the state is high. If the state is low, the optimal strategy depends on the switching rate λ and the long term gains $\gamma \mu$ relative to the cost c_i. If $\gamma \mu - c_i \geq 0$, effort is has positive value in the long run. Then if λ is very high, the state is very likely to recover if it becomes low and it is always optimal to exert full effort.

Otherwise, if the long term gains are not high enough compared to the cost of effort, it is optimal to bandon the technology if the state becomes low. From (2) it follows that if the state is low, and the planner does not exert effort, the state stays low. If the planner loses the market, she never exerts positive effort again.

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 state is observed by the players; i.e. $\phi \in \{0, 1\}$. Then if the following condition holds*

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

the social planner exerts full effort independent of the state. The expected payoff at time 0 *is*

$$\bar{v}_i(\phi) = \frac{\phi(\mu - c_i)}{r + \lambda(1 - \gamma)}. \tag{10}$$

Vice versa, if (9) does not hold, the planner stops exerting effort if the state becomes low. The expected time 0 *payoff is*

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

Proof. See Appendix. □

4.2 Incomplete Information Solution

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

Recall from the previous section that if the long-term gain of effort is very high, it is optimal to exert full effort, regardless of the current state. When the state is observed, this is the case if the long term gains exceed the myopic cost; i.e. if the condition (9) holds. Otherwise, it is optimal to never exert effort if the state becomes low. A similar logic applies for the planner's optimal decision when the state is unobserved. However, since the planner does not observe the true state, she conditions her strategy on her belief. We focus here on the analysis in the case in which the planner stops exerting effort if she becomes too pessimistic.

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

Formally, the planner's optimal decision can be derived using standard dynamic programming tools. 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 long term mean of the efforts and the costs are identical across the players 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

$$rv_i(\phi) = \sup_{\alpha_i(\phi)} \left\{ \alpha_i(\phi)(\mu\phi - c) + \lambda(\gamma\alpha(\phi_i) - \phi)v_i'(\phi) + \alpha_i(\phi)\frac{N}{2}\frac{\mu^2}{\sigma^2}\phi^2(1 - \phi)^2 v_i''(\phi) \right\} \tag{12}$$

11

for each i. From (12) we can immediately see that the social planner's objective is linear in $\alpha_i(\phi)$. We will show below that 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 exerting effort, she anticipates that she gradually loses the market.[8] Therefore, it is never optimal to exert effort again.

Now suppose that $\lambda > 0$ such that the state evolves over time. Then with bad enough luck, the social planner eventually stops exerting effort even if the state is high. Then over long enough time horizon, the planner loses 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 state is low. Then the social planner keeps on exerting effort, 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.

We next derive some key properties of the planner's value function $v_i(\phi)$. Since $\alpha_i(\phi) = 0$ is always feasible, the value function is nonnegative. Moreover, we 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 decision. The result is summarized in the following lemma

Lemma 1. *Social planner's value function $v_i(\phi)$ is nonnegative, 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 a 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, if $\alpha_i(\phi) = 0$, the belief drifts down according to (8). Therefore, if the planner stops exerting effort, she stops exerting effort in perpetuity. Since $\alpha_i(\phi) = 0$ is feasible, we must have $v_i(\phi) \geq 0$. Moreover, if $\alpha_i(\phi) = 0$ in perpetuity, $v_i'(\phi) = 0$. Since $v_i''(\phi) \geq 0$ by the first part of the proof, we must have that $v_i'(\phi) \geq 0$. $\qquad\square$

---

[8] Formally, as we can see from (8), the planner's belief drifts downwards.

We show that the social planner's optimal decision has an isomorphic interpretation as an optimal stopping problem. In particular, we first show that there is a cutoff belief $\phi^* \geq 0$ above which the planner always exerts full effort and below which she abandons the risky technology. Formally,

Lemma 2. *For all i, there exists a unique cutoff belief ϕ^* such that $\alpha_i(\phi) = 1$ for all $\phi \geq \phi^* \geq 0$ and $\alpha_i(\phi) = 0$ otherwise. ϕ^* is an absorbing boundary.*

Proof. Let ϕ^* denote the highest belief at which $\alpha_i(\phi) = 0$ and recall from the proof of Lemma 5 that if the planner stops exerting effort, she never exerts positive effort again. Thus, $\alpha_i(\phi) = 0$ for all $\phi \leq \phi^*$ and ϕ^* is an absorbing boundary.

Next, notice that (12) implies that $\alpha_i(\phi) \in [0, 1]$ is optimal only if

$$c - \mu\phi = \lambda\gamma v_i'(\phi) + \frac{N}{2}\frac{\mu^2}{\sigma^2}\phi^2(1-\phi)^2 v_i''(\phi), \tag{13}$$

otherwise, either $\alpha_i(\phi) = 1$ or $\alpha_i(\phi) = 0$. Substituting (13) into (12) implies that

$$rv_i(\phi) = -\lambda\phi v'(\phi) = 0,$$

where the second equality follows since $v_i(\phi) \geq 0$ and $v_i'(\phi) \geq 0$ by Lemma 5. Then (13) implies that $\alpha_i \in [0, 1]$ is optimal at most at the single point ϕ^*. $\qquad\square$

Notice that Lemmas 1 and 2 imply that the social planner always exerts effort longer than is myopically optimal. Let $\phi^M \equiv c/\mu$ denote the cut-off belief at which a myopic player stops.

Corollary 2. *The social planner keeps on exerting effort 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 $v_i'(\phi^*) = 0$ by Lemmas 1 and 2, 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. $\qquad\square$

The result holds for an arbitrary number of players, including $N = 1$. A single player always keeps on exerting effort beyond the myopic belief. We will show below that the same result holds in the noncooperative game since the aggregate effort can never cease as long as it is optimal for a single, isolated player to exert effort.

We can now write the social planner's problem as a standard optimal stopping problem. The result is summarized in the following proposition

Proposition 3 (Social Planner's Solution). *For all $\phi \geq \phi^*$, the planner's value function is the unique solution of the ordinary differential equation*

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

with the boundary conditions $rv_i(1) = \mu - c - \lambda(1-\gamma)v_i'(1)$, $v_i(\phi^*) = 0$ and $v_i'(\phi^*) = 0$.

Proof. See Appendix. □

(a) $\lambda = 0.5$, $\gamma = 1$; $\phi^* = 0.25$.　　　**(b)** $\lambda = 0.01$, $\gamma = 0.5$; $\phi^* = 0.47$.
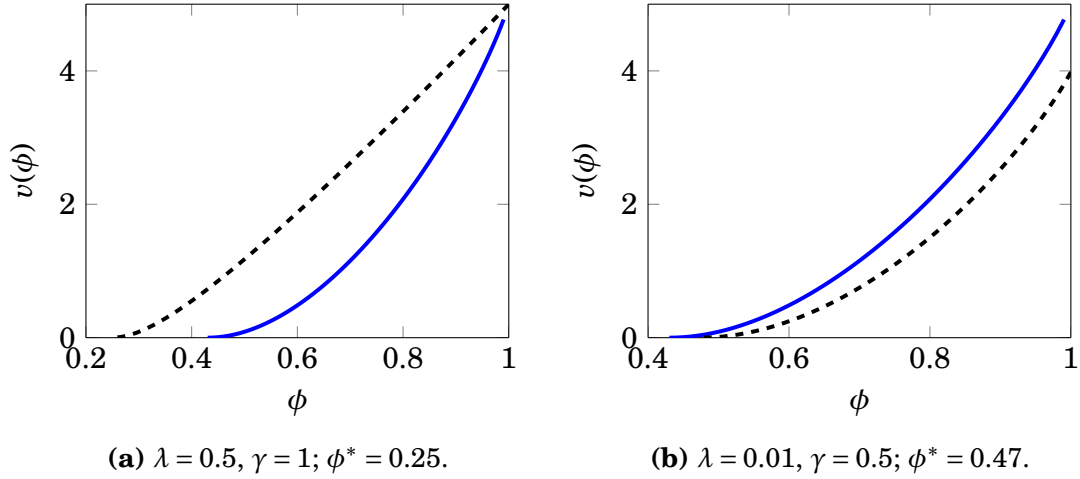
Figure 1: The effect of varying λ and γ on the social planner's optimal policy. The solid curve describes the pure experimentation problem for which $\phi^* = 0.43$; the dashed line describes the problem with an endogenous state. The parameter values are $\mu = 5$, $c = 4.5$, $\sigma = 5$, $r = 0.1$ and $N = 2$.

Unfortunately, explicit solutions are only available for the special case $\lambda = 0$ that is carefully analyzed in Bolton and Harris (1999). Figure 1 examines numerically how the changing state changes the optimal policy in comparison to the pure experimentation problem. Figure 1a describes the model in which $\gamma = 1$ such that the good state is absorbing. Then increasing λ increases the probability that the planner is able get to the good state by exerting effort. Consequently, the planner has a very strong incentive to exert effort also at low beliefs. As we will see below, this turns out to be a very strong prediction that holds also in the noncooperative game. Increasing λ always increases the players' incentives to exert effort.

Figure 1b demonstrates the effect of lowering the low term mean γ of beliefs conditional on effort. We can see that the planner's value decreases if γ is lower, but that the incentives to exert effort at low beliefs are hardly affected. We will show below that this effect is very different in the noncooperative game. A lower

long-term mean strengthens the players' temptation to free-ride on each other impact on the state. This results in a lower level of aggregate effort which again results in an earlier unraveling of the noncooperative game.

5   Noncooperative Solution

The main section of the paper studies the noncooperative game in which each firm decides individually how much effort it wants to exert. The state is the same for all players and they benefit from each others' efforts. However, effort is eventually costly since the state may be low. While effort benefits all firms, it is only profitable if the state is high.

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' actions. This gives incentives to free-ride. Second, value of information is positive and exerting higher effort encourages other players to exert effort. The two effects are familiar from the classical experimentation literature. The free-riding effect is affected if effort affects the distribution over future payoffs. In particular, the players are more likely to lose the market if they abstain from effort. This alleviates free-riding. However, the players have an incentive to free-ride on each others' impact on the state. The latter effect strengthens the free-riding effect.

Next, the players' incentives to exert effort depend on the state which again depends on the players' actions. Thus, if the players exert higher effort, they also encourage other players to exert higher effort. We show that if the players start restricting effort above the long term mean, the two effects play in the same direction, which leads to multiple symmetric Markov Perfect equilibria of the noncooperative game. Notice that multiplicity does not arise in the pure experimentation problem. While the players' incentives to exert effort do depend on state, the state does not depend on the actions. As a consequence, the belief uniquely pins down the symmetric Markov Perfect equilibrium.[9]

Finally, we show that if the state depends on the players' actions, they eventually get trapped in low effort, leading to an inefficient unraveling of the game. The unraveling is a consequence of the players' free-riding together with the fact that a sufficiently high level of effort is needed to keep the market. In the noncooperative solution, the aggregate effort is too low, which quickly drives down the state. Such unraveling is inefficient and does not occur in the social planners' optimum. Moreover, it cannot occur in Bolton and Harris (1999) in which the

---

[9]Of course, learning games typically have asymmetric equilibria that are qualitatively different. See, for example, Keller et al. (2005).

state does not depend on the players' action. In contrast, learning slows down but 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 $\mathscr{F}_t^\alpha$.*

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. We 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 $\alpha_{-i}(\phi)$ chosen by the other players.*

The restriction on stationary Markovian strategies is made for tractability and 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 can be derived from his Hamilton-Jacobi-Bellman equation[10]

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

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

By comparing an individual player's decision problem (15) with the social planner's Hamilton-Jacobi-Bellman equation (12), we can immediately see that the players' 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 3 (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 it 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 evolving state on the players' optimal choice.

The term in the parenthesis on the second line of (16) describes the player i's opportunity cost of effort. The first term, $\mu\phi - c_i$, which is eventually negative, describes the player's flow profit of effort. The second term, $\lambda\delta_i v_i'(\phi)$ describes the player's benefit from the effect of his own effort on the state. The last term,

---

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

$1/2\mu^2/\sigma^2\phi^2(1-\phi)^2v_i''(\phi)$ is the shadow value of information from learning. We will show below that $v_i'(\phi) \geq 0$ and $v_i''(\phi) \geq 0$ such that both a higher state and learning have positive value for the player.

The last term in the parenthesis of the first line of (16) describes the externality that the other players' effort has on the player i's value. The first term, $\lambda\delta_{-i}v_i'(\phi)$, describes the externality that results from the other players' efforts improving the state. The second term, $1/2\mu^2/\sigma^2\phi^2(1-\phi)^2v_i''(\phi)$, is the shadow value of information provided by the other players. Both effects are positive such that the other players' efforts 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 implies that the players 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 effort on the other players' value, but has an incentive to free-ride on the 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' impact on the state, not only on their learning. In particular, recall that when the players' action affects the state, the value deteriorates if the players' abstain from effort. The second term in (16), $-\lambda\phi v_i'(\phi)$, describes how quickly the players lose the market if they do not exert effort. We will show below that this alleviates free-riding and increases the players' effort at the noncooperative equilibrium.

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 again assume that each player's effort has the same effect on the state and that the players face the same 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 state, exerting full effort is a dominant strategy. Similarly, if the players are very pessimistic,

18

abandoning the technology is optimal. However, at intermediate beliefs, the players have an incentive to free-ride on each others' actions. At equilibrium, they do not exert full effort. Individually, each player is just indifferent between her strategies.

The player's optimal effort 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 $c - \phi\mu < \lambda\gamma/Nv_i'(\phi) + \mu^2\phi^2(1-\phi)^2v_i''(\phi)/(2\sigma^2)$, it is optimal to set $\alpha_i(\phi) = 1$.*

2. *If $c - \phi\mu = \lambda\gamma/Nv_i'(\phi) + \mu^2\phi^2(1-\phi)^2v_i''(\phi)/(2\sigma^2)$, the players choose the interior level of effort*

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

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 exert effort are stronger when the exogenous switching rate λ is higher. If $\lambda > 0$, constant effort is needed for the players' not to lose the market. Higher λ implies that abstaining effort results in the players losing the market more quickly. 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. Substituting for $\alpha_i(\phi) = \alpha_j(\phi)$, we find that

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

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

Now, let $\bar{\phi}$ denote the highest belief at which $\alpha_i(\phi) < 1$ and $\bar{\bar{\phi}}$ the highest belief at which $\alpha_i(\phi) = 0$. We will show below that the two beliefs are cutoff beliefs at which the players switch between the regimes. First if the players become

pessimistic enough, such that their belief falls below $\bar{\phi}$, they start to restrict effort. Next, if they become even more pessimistic, such that their belief falls below $\bar{\bar{\phi}}$, they stop exerting effort. In absence of effort, losing the market becomes more and more likely.; the belief drifts down deterministically according to (8). The players never exert effort again.

Before writing down the main result that characterizes the equilibria, we first derive some key properties of the equilibria. Recall from Corollary 2 that the social planner keeps on exerting effort 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 exert effort longer than a myopic player would. The argument follows by contradiction. Suppose that none of the players would exert effort at the myopic cutoff. 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 following lemma summarizes the result

Lemma 4. *In the N-player noncooperative game, $\bar{\bar{\phi}} \leq \phi^M$.*

Proof. 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 $\hat{\phi}^1$ denote the corresponding threshold belief at which the single player abandons the technology.

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

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

Proof. Suppose first that $\phi \geq \phi^M$. At $\phi = 1$, the boundary condition $rv(1) = \alpha_i(1)(\mu - c_i) - \lambda(1 - \gamma/N\alpha_i(\phi) - N/(N-1)\alpha_j)v_i'(1)$, $i \neq j$, implies that $\alpha_i(1) = 1$ is optimal. At the neighborhood of $\phi = 1$, $v_i'(\phi) > 0$ and $v_i''(\phi) \geq 0$ 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 $v_i''(\phi) \ge 0$. Such a point exists since $v_i''(\phi)$ is continuous and $v_i''(\bar{\phi}) \ge 0$. Let $[\phi^1, \phi^2]$, with $\phi^1 < \phi^2$ denote an interval on which $v_i''(\phi) \ge 0$. Then $v_i'(\phi^1) \le v_i'(\phi^2)$. (18) together with $c - \mu\phi^1 > c - \mu\phi^2$ and $v_i''(\phi^1)$ implies that

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

a contradiction. Thus, we must have $v_i''(\phi) \ge 0$. $\qquad\square$

Lemma 5 implies that the players' value increases with other players' effort. This demonstrates the encouragement effect in the model with an endogenous state. Indeed, we can conclude that

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

Now it follows from Lemma 5 that the noncooperative solution is characterized by three regimes that depend on the belief and can be separated by two threshold beliefs: the threshold $\bar{\phi}$ at which the players start restricting their effort and the threshold $\bar{\bar{\phi}}$ at which the players stop exerting effort. Between $\bar{\phi}$ and $\bar{\bar{\phi}}$ the players gradually decrease their effort. The next lemma confirms the observation

Lemma 6. *There exist two cutoff beliefs $\bar{\phi}$ and $\bar{\bar{\phi}}$ such that (i) $\alpha_i(\phi) < 1$ if and only if $\phi < \bar{\phi}$; (ii) $\alpha_i(\phi) = 0$ if and only if $\phi < \bar{\bar{\phi}}$; (iii) $\bar{\bar{\phi}}$ is an absorbing boundary.*

Proof. Recall that $\bar{\bar{\phi}}$ denotes the highest ϕ such that $\alpha_i(\phi) = 0$. The argument that $\bar{\bar{\phi}}$ is an absorbing boundary follows along the same lines than in the social planner's problem in Lemma 2. Since the bad state becomes more and more likely over time in absence of effort, the players never start to exert effort again.

Next, notice that we can write the player i's strategy as

$$\alpha_i(\phi) = \max\left\{\min\left\{\frac{rv_i(\phi) + \lambda v_i'(\phi)}{(N-1)(c-\mu\phi)}, 1\right\}, 0\right\} \tag{19}$$

by Proposition 4. Then (19) together with Lemma 5 implies that $\alpha_i(\phi)$ is nondecreasing in belief. Let $\bar{\phi}$ denote the highest belief at which $\alpha_i(\phi) = 1$. The result follows. $\qquad\square$

Finally, by Corollary 3, the players' incentives to exert effort depend on the other players' effort. Now if the state is evolving, the players' incentives to exert effort depend on the state transition which again depends on the players' effort. We show that if $\bar{\phi} > \gamma$, such that the players start restricting effort above the long term mean,[11] these two effects play in same direction. Then the game has multiple symmetric Markov Perfect equilibria.

Similarly to the social planner's problem, the solution to the noncooperative 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 Equilibria). *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 $\phi > \bar{\phi}$, $\alpha_i(\phi) = 1$ for all $i \in \{1,...,N\}$ and the player i's payoff is the unique solution of the following ordinary differential equation*

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

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

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

$$c - \mu\phi = \frac{\lambda\gamma}{N}v_i'(\phi) + \frac{\mu^2}{2\sigma^2}\phi^2(1-\phi)^2 v_i''(\phi). \qquad (21)$$

   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 multiple symmetric Markov perfect equilibria if $\bar{\phi} > \gamma$; otherwise the symmetric Markov perfect equilibrium is unique.

Proof. See Appendix. $\qquad\square$

---

[11]Notice that this implies that unraveling starts immediately at $\bar{\phi}$.

(21) is an indifference condition and describes the players' value in the region of beliefs at which they choose an interior effort level. (20) describes the players' value on the region on which they exert full effort. By comparing the value functions in the social planner's problem (12) and the noncooperative game (15), we immediately see that they coincide. Free-riding is the only source of inefficiency that restricts the players' efforts at equilibrium.

In our framework, constant effort is needed for the players not to lose the market. The assumption has several consequences. First, free-riding is alleviated. Second, if the players become too pessimistic about the state, it eventually becomes optimal to abandon the technology. If the players stop exerting effort, they gradually become more pessimistic. Then the players never exert positive effort again. Moreover, as we will show below, the players eventually get trapped in low effort, causing an inefficient unraveling of the equilibrium.
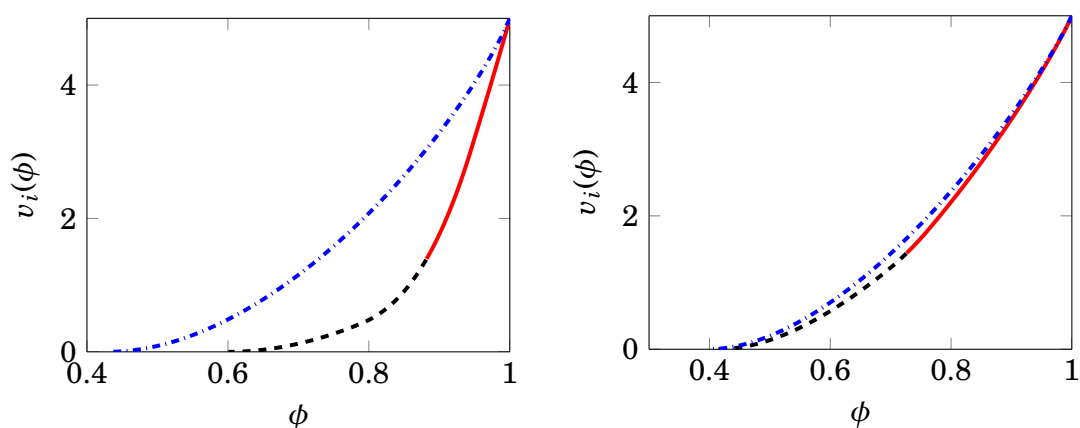
(a) Bolton and Harris (1999); $\phi^* = 0.43$ $\bar{\phi} =$ **(b)** $\lambda = 0.1$, $\gamma = 1$; $\phi^* = 0.40$ $\bar{\phi} = 0.73$, $\bar{\bar{\phi}} =$ 0.82, $\bar{\bar{\phi}} = 0.60$. 0.41.

Figure 2: Increasing λ alleviates free-riding. The dashed-dotted line describes the social planner's solution. Along the solid line the players exert full effort; along the dashed line the effort is restricted. The parameter values are $\mu = 5$, $c = 4.5$, $\sigma = 5$ $r = 0.1$ and $N = 2$.

Figures 2 and 3 examine the player's value function in noncooperative solution relative to the corresponding efficient benchmark. Figure 2 displays the numerical solution of Bolton and Harris (1999) on the left compared to a game with an endogenous state, $\lambda > 0$, but an absorbing good state, $\gamma = 1$. We can see that free-riding is alleviated and the players exert effort almost as long as a social planner would, already for a small increase in λ. The player become very eager to win the market, which increases their effort at equilibrium.

Figure 3 illustrates numerically the equilibrium unraveling result. For the

(a) $\lambda = 0.04$, $\gamma = 0.85$; $\bar{\phi} = 0.8$, $\bar{\bar{\phi}} = 0.53$.   **(b)** $\lambda = 0.01$, $\gamma = 0.5$; $\bar{\phi} = 0.85$, $\bar{\bar{\phi}} = 0.60$.
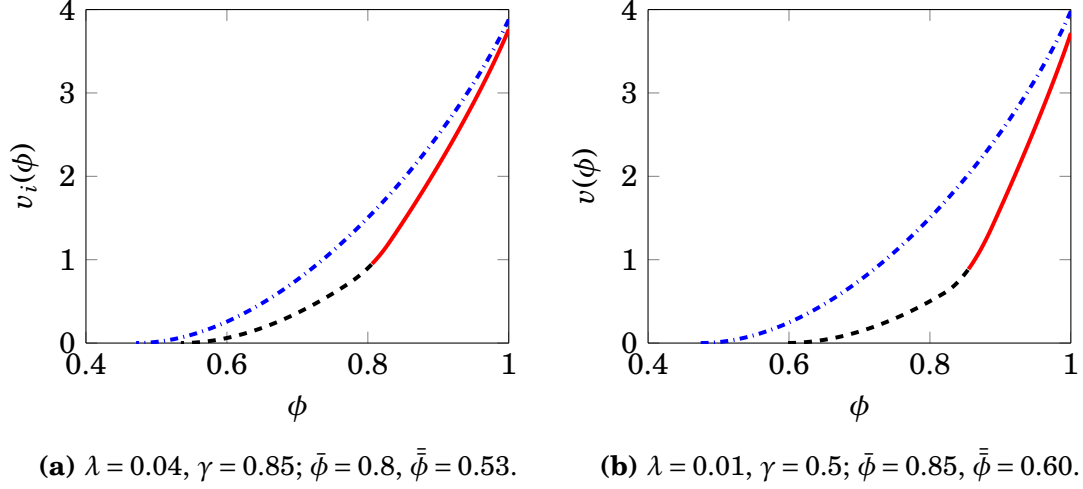
Figure 3: Unraveling. The dashed-dotted line describes the social planner's solution. Along the solid line the players exert full effort; along the dashed line the effort is restricted. The parameter values are $\mu = 5$, $c = 4.5$, $\sigma = 5$ $r = 0.1$ and $N = 2$. $\phi^* = 0.47$ in both cases.

left panel, the long term mean is relatively high conditional on effort, $\gamma = 0.85$; the exogenous switching rate is $\lambda = 0.04$. For the right panel, the long term mean is relatively low $\gamma = 0.5$; the switching rate is $\lambda = 0.01$. For both solutions, the social planner abandons the technology at the same belief. However, we can see that the noncooperative solution is relatively more inefficient in the right panel. In particular, $\Delta\bar{\phi} = 0.80 - 0.85 = -0.05$ while $\Delta\bar{\bar{\phi}} = 0.53 - 0.60 = -0.07$. Thus, there is an almost 30% higher decrease in the threshold $\bar{\bar{\phi}}$, at which the players abandon the technology, than in the threshold $\bar{\phi}$, at which the players start restricting effort. While higher switching rate λ alleviates free-riding, a lower long term mean γ leads to an earlier unraveling of the noncooperative game.

The rest of the chapter studies the unraveling result formally. Because of free-riding, the aggregate effort is not high enough, and the value deteriorates while learning slows down. Indeed, (8) implies that the belief drifts down if

$$\gamma \alpha_i(\phi_t) \leq \phi_t \tag{22}$$

It is immediate from (22) that unraveling is more likely if γ, i.e. the long term mean is lower conditional on effort.

Notice that such unraveling is not possible if $\lambda = 0$. Free-riding slows down the learning rate, but effort goes on indefinitely, see Bolton and Harris (1999). Moreover, the unraveling is inefficient and does not occur at the social planners' optimum. In contrast the social planner would exert every effort to improve the state.

Next, a natural question arises if free-riding can lead to unraveling if the state 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 next proposition confirms the observations

Proposition 5 (Unraveling). *Consider a symmetric Markov perfect equilibrium of the noncooperative 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 game.*

- *Suppose that $\bar{\bar{\phi}} \leq \phi < \tilde{\phi}$. Then if*

$$(rv_i(\phi) + \lambda v_i'(\phi))\left(\frac{\gamma}{N} + \frac{\mu^2}{\lambda\sigma^2}\phi(1-\phi)^2\right) < \frac{N-1}{N}\phi(c - \mu\phi), \qquad (23)$$

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

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 $\tilde{\phi} > \bar{\bar{\phi}}$. Indeed, recall that the belief drifts down if (22) holds. Let $\tilde{\phi}$ denote the belief such that the 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 state 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[12]

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

at the noncooperative equilibrium. Reasoning along the same lines than in the first part of the proof, one can show that the drift of the true belief process is nonpositive if and only if (23) holds. $\square$

---

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

6   Discussion

We examine a new model of strategic experimentation in which the players' action affects the distribution over future payoffs. The players need to exert effort to improve the value of a product that is uncertain and unobserved by the players. Effort is needed both to enhance the product value and to learn about its viability in the market.

The main focus of the analysis is a version of the model in which each player individually decide how much effort to exert to develop and sell the product. We show that whenever the product value depends on the players' efforts, they eventually get trapped in the low action, which leads to an inefficient unraveling in the noncooperative equilibrium. For intermediate beliefs, the players have a strong incentive to free-ride on each others' actions, which results in them reducing their effort at the noncooperative equilibrium. The aggregate level of effort remains too low and the players gradually lose the market. The result is in sharp contrast with the results in pure experimentation problems, in which learning slows down, but experimentation goes on indefinitely.

Moreover, we show that if unraveling starts immediately when the players start restricting effort, the noncooperative game has multiple Markov Perfect equilibria. The players' incentives to exert effort depend on the belief that they assign for the state. However, now the state also depends on the players' efforts. Higher aggregate effort does not only lead to more accurate signals, but also to a higher value. This drives multiplicity of equilibria in our framework.

This paper only derives the equilibrium for a restricted class of processes and actions. In particular, we only consider the case in which the signals are distributed by Brownian noise and the low state becomes more and more likely if the players do not exert effort. The latter assumption implies that the players never start exerting effort again if they stop. An important line of future research seems to be allowing for a more general dependence of the state on the action, possibly in a framework with Poisson distributed profits. Also the Levy version of the bandits still remains unexplored in our framework.

7 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 players' 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 state and the case in which the state is unobserved.

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

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

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

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

$$d\xi_t = \frac{\mu}{\sigma}\xi_t(\alpha_t^{1/2})^\top dZ_t. \tag{24}$$

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

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 state eventually changes over time. The change depends on the players' 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).

At time t, the players assign a belief ϕ_{t^-} for the high state and a probability of $1 - \Delta N_t^1$ for the game actually remaining in the high state. Similarly, the players belief that the state 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 state.

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

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

The last equality follows since ϕ_t is continuous almost everywhere. $\qquad\square$

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 exerts effort and conditions under which she abandons the technology if the state becomes low.

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

Suppose first that the planner stops exerting effort if the state becomes low. Now as long as the state remains high, full effort is optimal, i.e. $\alpha_t = 1$. However, the state eventually becomes low, an event after which the planner stops exerting effort. The distribution of s_t follows (2) and is now observed. At $t = 0$, $s_0 = 1$ with probability ϕ. The expected profit is

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

---

[13]See Øksendal and Sulem (2005).

28

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

$$rv_i^1 = \mu - c_i + \lambda(1-\gamma)v_i^0 - \lambda(1-\gamma)v_i^1,$$
$$rv_i^0 = -c_i + \lambda\gamma v_i^1 - \lambda\gamma v_i^0.$$

The solution is

$$v_i^1 = \frac{(r+\lambda\gamma)\mu}{r(r+\lambda)} - \frac{c_i}{r}, \tag{25}$$

$$v_i^0 = \frac{\lambda\gamma\mu}{r(r+\lambda)} - \frac{c_i}{r}, \tag{26}$$

and the expected payoff is (11).

Finally, we derive conditions under which it is optimal to stop if the state becomes low. If the planner exerts no effort, she earns a continuation payoff 0. If she keeps on exerting effort, the payoff is (26). By comparing the payoffs, we find that it is optimal to keep on exerting effort in the low state only if (9) holds. $\quad\square$

The next step is to analyze the social planner's problem in the case in which she does not observe the state.

Proof of Proposition 3. 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 in a new domain in which standard theorems imply existence and uniqueness of the solution.

Since (12) is linear in $\alpha_i(\phi)$, the optimal policy satisfies

$$\alpha \begin{cases} = 1 & \text{if} \quad c - \mu\phi < \lambda\gamma v_i'(\phi) + N\mu^2\phi^2(1-\phi)^2 v_i''(\phi)/(2\sigma^2), \\ \in [0,1] & \text{if} \quad c - \mu\phi = \lambda\gamma v_i'(\phi) + N\mu^2\phi^2(1-\phi)^2 v_i''(\phi)/(2\sigma^2), \\ = 0 & \text{if} \quad c - \mu\phi > \lambda\gamma v_i'(\phi) + N\mu^2\phi^2(1-\phi)^2 v_i''(\phi)/(2\sigma^2). \end{cases} \tag{27}$$

We first verify optimality of the effort strategy (27). In particular, we show that the process

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

is a supermartingale for an arbitrary effort strategy, and a martingale if the strategy (27) is chosen.

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

$$e^{rt}dS_t = \alpha_i(\phi_t)(\mu\phi_t - c)dt - rv_i(\phi_t)dt + \lambda(\delta^\top\alpha(\phi_t) - \phi_t)dt$$

$$+ \mathbf{1}^\top\alpha(\phi_t)\frac{1}{2}\frac{\mu^2}{\sigma^2}\phi_t^2(1-\phi_t)^2 v_i''(\phi_t)dt + \frac{\mu}{\sigma}\phi_t(1-\phi_t)(\alpha(\phi_t)^{1/2})^\top dZ_t^\alpha.$$

29

From (12), we can see that S_t is a supermartingale and it is a martingale only if the effort policy (27) is chosen. 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.

Now it follows from Lemma 2 that $\alpha_i(\phi) = 1$ is optimal if $\phi \geq \phi^*$ and $\alpha_i(\phi) = 0$ otherwise. Since $v_i(\phi) = v_i'(\phi) = 0$ for all $\phi < \phi^*$, standard arguments imply that the value function has to satisfy the value matching condition $v_i(\phi^*) = 0$ and the smooth pasting condition $v_i'(\phi^*) = 0$. The optimal value function solves the optimal stopping problem in Proposition 3.

It remains to prove that the solution exists and is unique. We apply the change of variables $x = (1 - \phi)/(1 - \phi^*)$ to rewrite (12) as an ordinary differential equation with an unknown parameter ϕ^* on the domain $[0, 1]$

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

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

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

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

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

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. $\square$

The rest of the Appendix examines the noncooperative solution. We first verify that the strategies described in Lemma 3 form best responses when the players play stationary Markov strategies.

Proof of Lemma 3. We can conclude from (16) that the player i's optimal investment strategy satisfies

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

Optimality follows again 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

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

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 (21) as solving a system of ordinary differential equations on $x \in [0,1]$, with unknown parameters $\bar{\phi}$ and $\bar{\bar{\phi}}$

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

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

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 in x, $\bar\phi$ and $\bar{\bar\phi}$ on $x \in [0,1]$.

The last step is to examine the uniqueness properties of the symmetric Markov Perfect equilibria. Recall that

$$y_1'(x,\bar\phi,\bar{\bar\phi}) = -(1-\bar\phi)v'(\phi;\bar\phi,\bar{\bar\phi}) \le 0 \quad \text{and} \quad y_2'(x,\bar\phi,\bar{\bar\phi}) = -(\bar\phi - \bar{\bar\phi})v'(\phi;\bar\phi,\bar{\bar\phi}) \le 0 \tag{42}$$

and

$$y_1''(x;\bar\phi,\bar{\bar\phi}) = (1-\bar\phi)^2 v''(\phi;\bar\phi,\bar{\bar\phi}) \ge 0 \quad \text{and} \quad y_2''(x,\bar\phi,\bar{\bar\phi}) = (\bar\phi - \bar{\bar\phi})^2 v''(\phi;\bar\phi,\bar{\bar\phi}) \ge 0 \tag{43}$$

by Lemma 5. In particular, we show that we can find two solutions $y(x;\bar\phi^1,\bar{\bar\phi}^1)$ and $y(x;\bar\phi^2,\bar{\bar\phi}^2)$ such that

$$\bar\phi^1 > \bar\phi^2, \quad \bar{\bar\phi}^1 > \bar{\bar\phi}^2 \quad \text{and} \quad \bar\phi^1 - \bar{\bar\phi}^1 < \bar\phi^2 - \bar{\bar\phi}^2 \tag{44}$$

that both satisfy the system (34)-(41). The last assumption states that the game unravels more quickly if the players start to restrict effort at a higher belief.

Notice that (38) and (39) imply that

$$y_2(0;\bar\phi^1,\bar{\bar\phi}^1) = y_2(1;\bar\phi^2,\bar{\bar\phi}^2) = 0 \quad \text{and} \quad y_2'(1;\bar\phi^1,\bar{\bar\phi}^1) = y_2'(1;\bar\phi^2,\bar{\bar\phi}^2) = 0.$$

Then (35) implies that at $x \in [0,1]$

$$-\frac{\lambda\gamma}{N}\left(\frac{y_2(x;\bar\phi^1,\bar{\bar\phi}^1)}{\bar\phi^1 - \bar{\bar\phi}^1} - \frac{y_2'(x;\bar\phi^2,\bar{\bar\phi}^2)}{\bar\phi^2 - \bar{\bar\phi}^2}\right)$$

$$+ \Phi_2\left(\left(\frac{\bar{\bar\phi}^1(1-\bar{\bar\phi}^1)}{\bar\phi^1 - \bar{\bar\phi}^1}\right)^2 y_2''(1;\bar\phi^1,\bar{\bar\phi}^1) - \left(\frac{\bar{\bar\phi}^2(1-\bar{\bar\phi}^2)}{\bar\phi^2 - \bar{\bar\phi}^2}\right)^2 y_2''(1;\bar\phi^2,\bar{\bar\phi}^2)\right)$$

$$= -\mu(1-x)(\bar\phi^1 - \bar\phi^2) - \mu x(\bar{\bar\phi}^1 - \bar{\bar\phi}^2) < 0. \tag{45}$$

Now since $\bar\phi^1 - \bar{\bar\phi}^2 < \bar\phi^2 - \bar{\bar\phi}^2$ it is immediate that we can find $\bar\phi^1$, $\bar\phi^2$, $\bar{\bar\phi}^1$, $\bar{\bar\phi}^2$, $y_2'(x;\bar\phi^1,\bar{\bar\phi}^1) \le y_2'(x;\bar\phi^2,\bar{\bar\phi}^2)$ and $y_2''(x;\bar\phi^1,\bar{\bar\phi}^1) \ge y_2''(x;\bar\phi^2,\bar{\bar\phi}^2)$ such that (45) is satisfied.[14] Notice that since $\bar\phi^1 - \bar\phi^2 < \bar{\bar\phi}^1 - \bar{\bar\phi}^2$, the right hand side of (45) increases as x decreases. Also, the left hand side increases since $y'(x;\cdot,\cdot)$ increases. This demonstrates that the earlier unraveling of the first equilibrium drives the difference between the thresholds.

At $x = 1$, (40) and (41) imply that $y_1(x;\bar\phi^1,\bar{\bar\phi}^1) \le y_1(x;\bar\phi^2,\bar{\bar\phi}^2)$ and $y_1'(x;\bar\phi^1,\bar{\bar\phi}^1) \le y_1'(x;\bar\phi^2,\bar{\bar\phi}^2)$. Moreover, (37) is satisfied since $\bar\phi^1/(1-\bar\phi^1) > \bar\phi^2/(1-\bar\phi^2)$, and

$$r(y(1;\bar\phi^1,\bar{\bar\phi}^1) - y(1;\bar\phi^2,\bar{\bar\phi}^2)) - \lambda\left(\frac{\bar\phi^1}{1-\bar\phi^1}y'(1;\bar\phi^1,\bar{\bar\phi}^1) - \frac{\bar\phi^2}{1-\bar\phi^2}y'(1;\bar\phi^2,\bar{\bar\phi}^2)\right)$$
$$= -(N-1)\mu(\bar\phi^1 - \bar\phi^2) < 0.$$

(34) is satisfied at any $x \in [0,1]$ if

$$r(y_1(x;\bar\phi^1,\bar{\bar\phi}^1) - y_1(x;\bar\phi^2,\bar{\bar\phi}^2)) = \mu x(\bar\phi^1 - \bar\phi^2)$$
$$+ \lambda\left(\frac{1-x-(\gamma-\bar\phi^1 x)}{1-\bar\phi^1}y_1'(x;\bar\phi^1,\bar{\bar\phi}^1) - \frac{1-x-(\gamma-\bar\phi^2 x)}{1-\bar\phi^2}y_1'(x;\bar\phi^2,\bar{\bar\phi}^2)\right)$$
$$+ \Phi_1((1-x(1-\bar\phi^1))^2 x^2 y_1(x;\bar\phi^1,\bar{\bar\phi}^1) - (1-x(1-\bar\phi^2))^2 x^2 y_1(x;\bar\phi^2,\bar{\bar\phi}^2)). \quad (46)$$

Clearly, (46) can only be satisfied if the second term in the right hand side is negative. This is the case if $\bar\phi^1 > \gamma$, that is, the players start restricting effort above the long term mean. In that case, unraveling starts at $\bar\phi$ and the game has multiple equilibria. If $\bar\phi^1 < \gamma$, the belief drifts up at the point, at which the players start restricting effort.

Now focus on the case in which $\bar\phi^1 > \gamma$. Then (36) implies that

$$r(y_1(0;\bar\phi^1,\bar{\bar\phi}^1) - y_1(0;\bar\phi^2,\bar{\bar\phi}^2)) = \lambda(1-\gamma)\left(\frac{y_1'(0;\bar\phi^1,\bar{\bar\phi}^1)}{1-\bar\phi^1} - \frac{y_1'(0;\bar\phi^2,\bar{\bar\phi}^2)}{1-\bar\phi^2}\right)$$

which is satisfied if

$$y(0;\bar\phi^1,\bar{\bar\phi}^1) < y(0;\bar\phi^2,\bar{\bar\phi}^2)$$

and

$$\frac{y_1'(0;\bar\phi^1,\bar{\bar\phi}^1)}{1-\bar\phi^1} < \frac{y_1'(0;\bar\phi^2,\bar{\bar\phi}^2)}{1-\bar\phi^2}.$$

Changing back the variables to recast the problem on the original domain, we have

$$v(1;\bar\phi^1,\bar{\bar\phi}^1) < v(1;\bar\phi^2,\bar{\bar\phi}^2) \ \text{ and } \ v'(1;\bar\phi^1,\bar{\bar\phi}^1) > v'(1;\bar\phi^2,\bar{\bar\phi}^2).$$

---

[14] Notice that $y_2(x;\bar\phi^1,\bar{\bar\phi}^1) \le y_2(x;\bar\phi^2,\bar{\bar\phi}^2)$ for all $x \in [0,1]$.
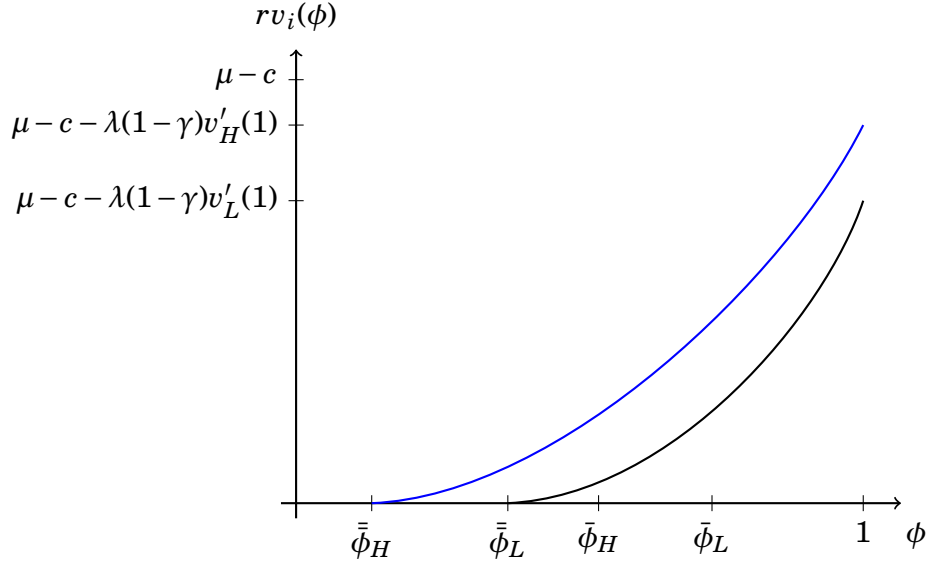
Figure 4: Multiple equilibria in the model with $\gamma < 1$ and $\lambda > 0$.

The players do not only have a lower value at the upper boundary $\phi = 1$, but the value also falls more rapidly. Since the players start restricting effort earlier, the game unravels earlier, which leads to them losing value more rapidly, even at high beliefs. Figure 4 describes the situation in a phase diagram. $\square$

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. Mimeo, Yale University and Stanford Graduate School of Business, 2016.

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.

Daria Khromenkova. Restless strategic experimentation. Mimeo, University of Mannheim, 2016.

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.

Mikhail Safronov. Obtaining experience in a multi-armed bandit. Mimeo, University of Northwestern, 2014.

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

SFB 649 Discussion Paper Series 2016

001 "Downside risk and stock returns: An empirical analysis of the long-run and short-run dynamics from the G-7 Countries" by Cathy Yi-Hsuan Chen, Thomas C. Chiang and Wolfgang Karl Härdle, January 2016.

002 "Uncertainty and Employment Dynamics in the Euro Area and the US" by Aleksei Netsunajev and Katharina Glass, January 2016.

003 "College Admissions with Entrance Exams: Centralized versus Decentralized" by Isa E. Hafalir, Rustamdjan Hakimov, Dorothea Kübler and Morimitsu Kurino, January 2016.

004 "Leveraged ETF options implied volatility paradox: a statistical study" by Wolfgang Karl Härdle, Sergey Nasekin and Zhiwu Hong, February 2016.

005 "The German Labor Market Miracle, 2003 -2015: An Assessment" by Michael C. Burda, February 2016.

006 "What Derives the Bond Portfolio Value-at-Risk: Information Roles of Macroeconomic and Financial Stress Factors" by Anthony H. Tu and Cathy Yi-Hsuan Chen, February 2016.

007 "Budget-neutral fiscal rules targeting inflation differentials" by Maren Brede, February 2016.

008 "Measuring the benefit from reducing income inequality in terms of GDP" by Simon Voigts, February 2016.

009 "Solving DSGE Portfolio Choice Models with Asymmetric Countries" by Grzegorz R. Dlugoszek, February 2016.

010 "No Role for the Hartz Reforms? Demand and Supply Factors in the German Labor Market, 1993-2014" by Michael C. Burda and Stefanie Seele, February 2016.

011 "Cognitive Load Increases Risk Aversion" by Holger Gerhardt, Guido P. Biele, Hauke R. Heekeren, and Harald Uhlig, March 2016.

012 "Neighborhood Effects in Wind Farm Performance: An Econometric Approach" by Matthias Ritter, Simone Pieralli and Martin Odening, March 2016.

013 "The importance of time-varying parameters in new Keynesian models with zero lower bound" by Julien Albertini and Hong Lan, March 2016.

014 "Aggregate Employment, Job Polarization and Inequalities: A Transatlantic Perspective" by Julien Albertini and Jean Olivier Hairault, March 2016.

015 "The Anchoring of Inflation Expectations in the Short and in the Long Run" by Dieter Nautz, Aleksei Netsunajev and Till Strohsal, March 2016.

016 "Irrational Exuberance and Herding in Financial Markets" by Christopher Boortz, March 2016.

017 "Calculating Joint Confidence Bands for Impulse Response Functions using Highest Density Regions" by Helmut Lütkepohl, Anna Staszewska-Bystrova and Peter Winker, March 2016.

018 "Factorisable Sparse Tail Event Curves with Expectiles" by Wolfgang K. Härdle, Chen Huang and Shih-Kang Chao, March 2016.

019 "International dynamics of inflation expectations" by Aleksei Netšunajev and Lars Winkelmann, May 2016.

020 "Academic Ranking Scales in Economics: Prediction and Imdputation" by Alona Zharova, Andrija Mihoci and Wolfgang Karl Härdle, May 2016.

SFB 649 Discussion Paper Series 2016

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

SFB 649 Discussion Paper Series 2016

For a complete list of Discussion Papers published by the SFB 649, please visit http://sfb649.wiwi.hu-berlin.de.

040 "Principal Component Analysis in an Asymmetric Norm" by Ngoc M. Tran, Petra Burdejová, Maria Osipenko and Wolfgang K. Härdle, October 2016.

041 "Forward Guidance under Disagreement - Evidence from the Fed's Dot Projections" by Gunda-Alexandra Detmers, October 2016.

042 "The Impact of a Negative Labor Demand Shock on Fertility - Evidence from the Fall of the Berlin Wall" by Hannah Liepmann, October 2016.

043 "Implications of Shadow Bank Regulation for Monetary Policy at the Zero Lower Bound" by Falk Mazelis, October 2016.

044 "Dynamic Contracting with Long-Term Consequences: Optimal CEO Compensation and Turnover" by Suvi Vasama, October 2016.

045 "Information Acquisition and Liquidity Dry-Ups" by Philipp Koenig and David Pothier, October 2016.

046 "Credit Rating Score Analysis" by Wolfgang Karl Härdle, Phoon Kok Fai and David Lee Kuo Chuen, November 2016.

047 "Time Varying Quantile Lasso" by Lenka Zbonakova, Wolfgang Karl Härdle, Phoon Kok Fai and Weining Wang, November 2016.

048 "Unraveling of Cooperation in Dynamic Collaboration" by Suvi Vasama, November 2016.