Restless Strategic Experimentation*

(Preliminary. Please do not circulate.)

Daria Khromenkova†

April 2, 2016

Abstract

I study a game of strategic experimentation with two-armed bandits in which the state of the world is restless, "reboots" at exponentially distributed random times. Players observe neither the initial state of the world nor reboot times, but may learn about whether the current state is good via news that arrives at exponential times. Unlike in standard good-news models of strategic experimentation in which the state is rested, there are parameters for which the encouragement effect is present and players experiment beyond the single-player threshold. There also exists a range of parameters for which the free-riding effect is mute and the equilibrium is efficient.

Keywords: Strategic experimentation, two-armed exponential bandit, restless bandit.

JEL Classification: C73, D83.

*I am grateful to Johannes Hörner for his support and guidance. I am also thankful to Larry Samuelson and Gonzalo Cisternas for helpful comments and suggestions.

†CDSE, University of Mannheim, daria.khromenkova@gess.uni-mannheim.de.

1 Introduction

None of all standard models of strategic experimentation like in Bolton and Harris (1999), Keller, Rady, and Cripps (2005), and Keller and Rady (2015) allows for the underlying state to change over time, or to be *restless*. The restless state, however, is what we often observe. Consumers experimenting with firms have either an ordinary experience or particularly happy with new features of firms' products, that is, the underlying quality of these products is restless from their point of view. Viruses mutate over time, and so, even though a new drug succeeds at the moment, it does not mean that it will aways perform better than placebo.

The contribution of this paper is twofold. First, it is theoretically relevant, for it tests whether some of the standard results, in particular, those in Keller, Rady, and Cripps (2005), rely on the state not changing over time, or being rested. Second, it addresses economically relevant questions as is clear from the applications just mentioned.

I consider a good-news, two-armed exponential bandit model as in Keller, Rady, and Cripps (2005) with the exception that the state of the world (or the type of the risky arm) which is common to all players changes, or "reboots," over time. There is a finite number of players, each with a unit of perfectly divisible resource. Each player continuously decides how to split her resource between two arms, safe and risky, that is, how much to experiment with the risky arm. The safe arm has a known payoff, whereas payoffs from the risky arm depend on its initially unknown type, good or bad. If the type is good, the risky arm generates higher payoffs than the safe one, while it yields nothing if it is bad. Payoffs from the good risky arm arrive at Poisson times independently across players and are referred to as news. They are publicly observable and reveal perfectly that the (current) state is good. The state reboots at exponentially distributed times unobserved by players and becomes good with a certain reboot probability independent of the previous state.

Restlessness of the state leads to "non-standard" dynamics of resource allocation in the social planner's problem and symmetric equilibrium. Even if players start with allocating all resources to the safe arm, they may switch to the risky arm eventually, for the state may become good after the reboot. On the other hand, the possibility that the state becomes bad may result in experimentation ceasing even after news has arrived. These do not occur if the state is rested. With the rested state and in absence of news, players become only more pessimistic over time and experimentation either ceases and never resumes again or never even starts. If news does arrive, players learn that the state is good, the state does not change over time, and hence they allocate all resources to the risky arm thereafter.

The first main result is for the case with the reboot probability equal to one, that is, in which the state becomes good after the reboot. I show that, *like* in Keller, Rady, and Cripps (2005), there is no encouragement effect, in the sense that experimentation does not go beyond the single-player threshold. This is not surprising, since an arrival of news reveals

that the current state is good and it will be good thereafter even if it reboots, so there is no value from other players' experimentation after that.

The second main result is for the case with the reboot probability equal to zero, that is, in which the state becomes bad after the reboot. *Unlike* in Keller, Rady, and Cripps (2005), there is the encouragement effect. Indeed, players choose to experiment with the risky arm because they want benefit from its higher payoff if it is initially good. An arrival of news reveals that the initial state is good, but it becomes bad after the reboot. That is why experimentation by other players is still valuable, for it helps to learn whether the reboot has occurred.

The free-riding effect is present in both cases. That is, starting at some threshold belief and in absence of news, each player gradually decreases the fraction of her resource which she allocates to the risky arm and thus free-rides on other players' experimentation. Interestingly, however, if the state is good after the reboot, there exists a range of parameters for which the free-riding effect is mute and the symmetric equilibrium is efficient. As is intuitive, this the case when the state reboots relatively often or there are few players.

In general, that is, for intermediate reboot probabilities, there are parameters for which there exists or does not exist the encouragement effect and for which the symmetric equilibrium is efficient. Overall, the two special cases span all possible efficient and equilibrium dynamic resource allocation and make economics behind it clear.

The paper is organized as follows. I discuss related literature next. The model is described in Section 2. As already pointed out, the main results are clear from the two special cases with the reboot probability equal to one and zero. The benchmark social planner's problem and the unique symmetric equilibrium for these cases are analyzed in Sections 3 and 4. The analysis of the general case is in Appendix A. Section 5 concludes. All the proofs are gathered in Appendix B.

Related literature. The paper contributes to the literature on strategic experimentation and, as discussed, is an extension to Keller, Rady, and Cripps (2005) allowing the underlying state to change over time. Bolton and Harris (1999) is the founding paper of the strategic experimentation literature with news arriving according to a Brownian process rather than a Poisson process. Keller and Rady (2015) strudies a two-armed exponential bandit model, but with bad news rather than with good news, that is, with news arriving if the state is bad. Keller and Rady also examine the case of inconclusive bad news. The case of inconclusive good news is analyzed in Keller and Rady (2010).

Vasama (2016) builds on the model of Bolton and Harris (1999) and allows the drift to depend on players' experimentation with the risky arm and thus to evolve over time. Vasama also extends Keller, Rady, and Cripps (2005) and allows players' valuations to change over

time. Unlike in this paper, the safe arm is absorbing in both versions, and the model is not solved in closed form.

Fryer and Harms (2015) studies a general two-armed bandit model in which the expected return from the risky arm increases if the arm is chosen and decreases otherwise. Fryer and Harms show that the optimal strategy can be described by Gittins index (Gittins, 1979).

Board and Meyer-ter-Vehn (2013) considers a model of firm reputation in which the reputation is restless from the point of view of consumers and depends on the firm's investments in the quality. Board and Meyer-ter-Vehn (2014) analyzes an extension with the firm not observing the underlying quality either and being allowed to exit the market. Halac and Prat (forthcoming) studies a two-sided moral hazard model of managerial attention and agent effort which has similar dynamics.

Keller and Rady (1999) analyzes a model of a monopoly who faces changing demand curve. Keller and Rady (2003) considers an extension to a duopoly.

A few papers in the literature on operations research and engineering study restless bandits and mostly focus on the optimality of the Whittle index (Whittle, 1988), a generalization of the Gittins index. For example, Dusonchet and Hongler (2003), Faihe and Müller (1998), Glazebrook, Hodge, and Kirkbride (2011), Glazebrook, Kirkbride, and Ruiz-Hernandez (2006), Hodge and Glazebrook (2015), La Scala and Moran (2008), Le Ny, Dahleh, and Feron (2008), Liu and Zhao (2010), Niño Mora (2001), Niño Mora (2002), Slivkins and Upfal (2008), Veatch and Wein (1996), Weber and Weiss (1990), and Zhao, Krishnamachari, and Liu (2008).

2 Model

Players, actions, and states. Time $t \in [0, \infty)$ is continuous, and the horizon is infinite. There are $I \geq 1$ players and two alternatives, or arms: a safe arm S and a risky one R. Each player is endowed with one unit of perfectly divisible resource per unit of time and continuously chooses how to split it between the two arms.

If player i allocates the fraction $x_{i,t} \in [0, 1]$ of her resource to R over an interval $[t, t+\mathrm{d}t)$, and consequently $1-x_{i,t}$ to S, she receives $(1-x_{i,t})s\mathrm{d}t$ from S, where $s > 0$ and $i = 1, \ldots, I$. That is, the safe arm yields a known constant flow payoff which is proportional to the fraction of the resource allocated to it. The risky arm's payoff at time t depends on an unknown binary state at that moment in time $\omega_t \in \{0, 1\}$. Specifically, the risky arm yields a lump-sum payoff $h > 0$ at some point in the interval $[t, t + \mathrm{d}t)$ with probability $\lambda X_t \mathrm{d}t$, where $\lambda > 0$ and $X_t = \sum_{i=1}^{I} x_{i,t}$ is the aggregate resource allocation at time t, if $\omega_t = 1$ and with probability zero if $\omega_t = 0$. I say that the risky arm is good (resp., bad) if $\omega_t = 1$ (resp., $\omega_t = 0$) and refer to the lump-sums as news received by players. Conditional on the state,

arrival of news is independent across players. Overall, player i's payoff over the interval $[t, t + \mathrm{d}t)$ is equal to

$$[(1 - x_{i,t})s + x_{i,t}g\omega_t]\mathrm{d}t,$$

where $g \coloneqq \lambda h$. Player i is said to experiment at time t if $x_{i,t} > 0$. Parameters are such that $g > s > 0$.

The state is restless, that is, it changes, or "reboots," over time. Independently of the current state ω_t and actions taken by players, the state reboots at some point in the interval $[t, t + \mathrm{d}t)$ with probability $\phi\mathrm{d}t$, where $\phi > 0$, and then $\omega_{t+\mathrm{d}t} = 1$ with probability $p^0 \in [0, 1]$, referred to as a reboot probability. The Poisson processes of the state reboots and news arrivals are conditionally independent.

Information and learning. Actions and news are publicly observable. News is also conclusive, that is, players learn that $\omega_t = 1$ if a lump-sum occurs. On the other hand, the reboot times are unobservable to players. All the above is common knowledge.

Players have a (common) prior belief $p_0 \in [0, 1]$ that the state is 1. As they share the same information, they have a (common) belief p_t that $\omega_t = 1$ at any time t. Given the aggregate resource allocation $X_t = \sum_{i=1}^{I} x_{i,t}$ and in absence of news, the belief p_t is determined by Bayes' rule:

$$p_{t+\mathrm{d}t} = p^0\phi\mathrm{d}t + (1 - \phi\mathrm{d}t)\frac{p_t(1 - \lambda X_t\mathrm{d}t)}{p_t(1 - \lambda X_t\mathrm{d}t) + (1 - p_t)} + o(\mathrm{d}t).$$

The first term $p^0\phi\mathrm{d}t$ reflects the possibility that the state reboots in $[t, t + \mathrm{d}t)$ and becomes 1 with probability p^0. The second term reflects players learning about the state if there is no news in $[t, t + \mathrm{d}t)$. Hence, the law of motion that governs the belief in absence of news is as follows:

$$\dot{p}_t = P(p_t) \coloneqq \phi(p^0 - p_t) - \lambda X_t p_t(1 - p_t). \tag{1}$$

In particular, even if $X_t = 0$, the belief does not stay the same and evolves with time as follows:

$$\dot{p}_t = \phi(p^0 - p_t).$$

If no resource continues to be allocated to R, the belief converges to p^0, the reboot probability.

If $X_t > 0$, the law of motion (1) can be rewritten as

$$\dot{p}_t = -\lambda X_t(p_t - \alpha_{X_t})(\beta_{X_t} - p_t),$$

where

$$\alpha_X \coloneqq \frac{1}{2X}\left(X + \psi - \sqrt{(X + \psi)^2 - 4X\psi p^0}\right), \tag{2}$$

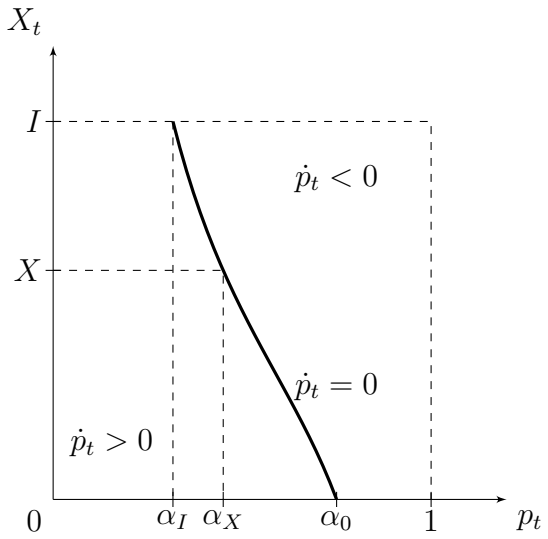

Figure 1: Evolution of the belief p_t if the aggregate resource allocation to the risky arm is $X_t = X$. Parameters: $(I, \psi, p^0) = (2, 1, 0.75)$.

$$\beta_X \coloneqq \frac{1}{2X}\left(X + \psi + \sqrt{(X+\psi)^2 - 4X\psi p^0}\right), \tag{3}$$

and $\psi \coloneqq \phi/\lambda$. In the special cases with the reboot probabilities $p^0 = 1$ and $p^0 = 0$, α_X and β_X take simpler forms and their dependence on X can be clearly seen. If $p^0 = 1$, $\alpha_X = \psi/X$ and $\beta_X = 1$ if $\psi/X \leq 1$, and $\alpha_X = 1$ and $\beta_X = 1 + \psi/X$ otherwise. If $p^0 = 0$, $\alpha_X = 0$ and $\beta_X = 1 + \psi/X$. Properties of α_X and β_X for $p^0 \in (0, 1)$ are captured by the next lemma.

Lemma 2.1. *For all $p^0 \in (0, 1)$ and $X > 0$, α_X and β_X satisfy the following properties:*

- *$\alpha_X \in [0, p^0]$ and $\beta_X \geq 1$;*
- *α_X is strictly decreasing with X;*
- *$\lim_{X \to 0} \alpha_X = p^0$.*

It follows that, if the same fraction X continues to be allocated to R and if no news arrives, the belief converges to α_X. The limit belief α_X is lower for larger X. It is convenient to define $\alpha_0 \coloneqq p^0$.

All in all, given $X_t = X$, the belief increases in the interval $[t, t + \mathrm{d}t)$, that is, $\dot{p}_t > 0$, if $p_t \in [0, \alpha_X)$, and decreases, that is, $\dot{p}_t < 0$, if $p_t \in (\alpha_X, 1]$. Figure 1 captures the belief dynamics in absence of news.

If news does arrive, the belief jumps to 1. However, unlike in Keller, Rady, and Cripps (2005) and except for the special case with $p^0 = 1$, the state may reboot and hence, even though news is conclusive, the belief does not stay at 1, but starts going down according to (1) in the next instant.[1]

[1] In the setting of Keller, Rady, and Cripps (2005), $\alpha_X = 0$ and $\beta_X = 1$ for any $X \in [0, I]$.

Payoffs. Players discount payoffs at a (common) rate $r > 0$. Given players' actions $x = \{x_t\}_{t\geq 0}$, where $x_t = (x_{1,t}, \ldots, x_{I,t})$ is measurable with respect to information available by time t, and the (common) prior belief p_0 that the state is 1, player i's expected normalized payoff is as follows:

$$\mathbf{E}\left[\int_0^\infty re^{-rt}[(1-x_{i,t})s + x_{i,t}g\omega_t]\mathrm{d}t \;\middle|\; p_0\right],$$

where the expectation is taken over $\{x_t\}_{t\geq 0}$ and $\{\omega_t\}_{t\geq 0}$.

At any time t, if player i allocates $x_{i,t}$ to R, her expected payoff over the interval $[t, t+\mathrm{d}t)$ is equal to

$$[(1-x_{i,t})s + x_{i,t}gp_t]\mathrm{d}t,$$

where p_t is the (common) belief that $\omega_t = 1$. By the Law of Iterated Expectations, her expected normalized payoff can be rewritten as follows:

$$\mathbf{E}\left[\int_0^\infty re^{-rt}[(1-x_{i,t})s + x_{i,t}gp_t]\mathrm{d}t \;\middle|\; p_0\right],$$

where the expectation is taken over $\{x_t\}_{t\geq 0}$ and $\{p_t\}_{t\geq 0}$.

Strategies and equilibrium. A pure strategy of player i is a map $x_{i,t}\colon [0,1] \to [0,1]$ such that, given the belief p_t and by allocating the fraction $x_{i,t}$ to R, she maximizes her expected normalized payoff

$$\mathbf{E}\left[\int_t^\infty re^{-r(\tau-t)}[(1-x_{i,\tau})s + x_{i,\tau}gp_\tau]\mathrm{d}\tau \;\middle|\; p_t\right],$$

where the expectation is taken over $\{x_\tau\}_{\tau\geq 0}$ and $\{p_\tau\}_{\tau\geq 0}$. I assume that players' strategies satisfy regularity conditions that ensure that (1) has a unique solution (see Klein and Rady, 2011). I look for Markov perfect equilibria in pure strategies with p as a state variable and in which players use symmetric strategies, that is, they allocate the same fraction of the resource to R given the belief.

3 Social Planner's Problem

As a benchmark, I consider the social planner's problem. The efficient resource allocation has the bang-bang property as in standard strategic experimentation literature, but its dynamics are richer. Similar to Board and Meyer-ter-Vehn (2013), the efficient threshold p_* can be one of two types:

– *convergent*: given the efficient resource allocation and in absence of news, the belief p stays at p_* upon reaching it, that is, $P(p_*) = 0$ (recall the law of motion (1));

– *permeable*: given the efficient resource allocation and in absence of news, the belief p goes through p_*, that is, either $P(p_*) > 0$ or $P(p_*) < 0$.

Whether the threshold is convergent or permeable depends on parameters and is associated with a different efficient dynamic resource allocation:

– *full experimentation*: experimentation always starts and never ceases afterwards, and all resources are allocated to R on path;

– *partial experimentation*: experimentation always starts and never ceases afterwards, but only a fraction of resources is allocated to R at the threshold p_* on path;

– *no experimentation*: experimentation never takes place or ceases eventually and never resumes again, and all resources are allocated to S on path.

All the three pattens of the efficient dynamic resource allocation are spanned by the special cases with the reboot probabilities $p^0 = 1$ and $p^0 = 0$. If the state is good after the reboot, that is, if $p^0 = 1$, the optimum exhibits either the full or partial experimentation property. If the state is bad after the reboot, that is, if $p^0 = 0$, the optimum always exhibits the no experimentation property. These are summarized by Propositions 3.1 and 3.2 and discussed in detail below.

The two special cases make economics behind each pattern clear. That is why I leave the intermediate case with $p^0 \in (0, 1)$, for which the efficient dynamic resource allocation exhibits the full, partial, or no experimentation property, for Appendix A.

3.1 Special Case with $p^0 = 1$

As already pointed out, if the state becomes good eventually, experimentation always starts and never ceases afterwards. If all resources are allocated to S, that is, if $X = 0$, the belief always drifts up and hence it reaches p_* and experimentation starts. If all resources are allocated to R, that is, if $X = I$, the belief reaches α_I and stays there is absence of news. (Recall that, if $p^0 = 1$, $\alpha_I = \psi/I$ if $\psi/I \leq 1$ and $\alpha_I = 1$ otherwise.) Hence, experimentation never ceases, but whether it is efficient to experiment fully, that is, whether $\alpha_I \geq p_*$, depends on whether the state reboots at a higher rate than news arrives and on the number of players, as emphasized in Proposition 3.1.

Proposition 3.1 (Efficient Allocation: $p^0 = 1$)**.** *The optimal strategy of the social planner*

is (essentially) unique. It is a cut-off strategy with

$$X(p) = \begin{cases} I & \text{if } p > p_*, \\ X_* & \text{if } p = p_*, \\ 0 & \text{if } p < p_*, \end{cases}$$

where X_* and p_* are as follow:

(i) if $\psi/I \geq s/g$, then $X_* \in [0,1]$ and $p_* = s/g$ (full experimentation);

(ii) if $\psi/I < s/g$, then $X_* = \psi/p_*$ and p_* is given by (4) (partial experimentation).

It follows from Proposition 3.1 that, if the state reboots at a sufficiently higher rate than news arrives, that is, if $\psi/I \geq s/g$, where $\psi \coloneqq \phi/\lambda$ with ϕ the reboot rate and λ the news arrival rate, the efficient behavior is myopic, that is, $p_* = s/g$. If learning is slow compared to how fast the state reboots and becomes good, the opportunity to learn loses its value and it is efficient to allocate all resources to R only when the immediate payoff from the risky arm is higher.

All in all, myopic behavior takes place if $\psi/I \geq s/g$. Then $\psi/I \geq p_*$, that is, the belief stays in $[\psi/I, 1]$ and the optimum exhibits the full experimentation property. The threshold is permeable with $P(p_*) > 0$. Any $X_* \in [0, I]$ at p_* leads to the increase in the belief, so that $X = I$ is optimal in the next moment in time. Hence, any such X_* is optimal.

If there are many players or news arrives relatively fast, that is, if $s/g > \psi/I$, it becomes optimal to experiment for beliefs below the myopic one, that is, $p_* < s/g$. The efficient threshold is given by

$$p_* = \frac{-s(\psi_I - \mu_I) + \sqrt{\Delta_I}}{2(g(1+\mu_I) - s)}, \tag{4}$$

where $\psi_I \coloneqq \psi/I$, $\mu_I \coloneqq \mu/I$, $\mu \coloneqq r/\lambda$, and

$$\Delta_I \coloneqq s^2(\psi_I - \mu_I)^2 + 4s\psi_I(g(1+\mu_I) - s).$$

The more players are present, the larger is the range of beliefs for which it is efficient to allocate all resources to R. On the other hand, the higher is the rate of the state reboot, the higher the efficient threshold is. These are captured by the next two lemmata.

Lemma 3.1 (Number of Players: $p^0 = 1$)**.** *The efficient threshold given by (4) is strictly decreasing with the number of players I.*

Lemma 3.2 (Reboot Rate: $p^0 = 1$)**.** *The efficient threshold given by (4) is strictly increasing with the relative rate of the state reboot ψ.*

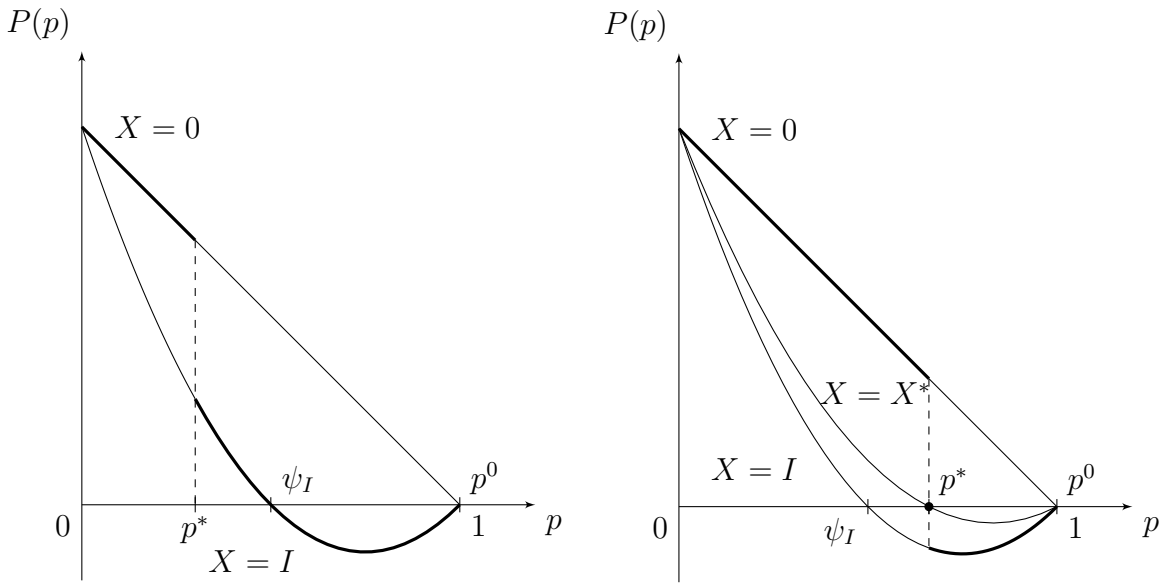

Figure 2: Evolution of the belief p in optimum that exhibits the full (left) and partial (right) experimentation properties. Parameters: $(I, \lambda, \phi, r, g, p^0) = (2, 1, 1, 1, 1, 1)$, $s = 0.3$ (left), and $s = 0.7$ (right).

The threshold is convergent, that is, $P(p_*) = 0$. There exists $X_* \in [0, I]$, namely, $X_* = \psi/p_*$, such that the belief stays at p_* if no news arrives. Any $X \in [0, I]$ different from X_* results in fluctuations of the belief around p_* in absence of news. Indeed,

- if $X > X_*$, the belief falls below p_*, which makes $X = 0$ optimal in the next moment in time, which in turn causes the belief to rise to p_* again;
- if $X < X_*$, the belief goes up above p_*, so that $X = I$ is optimal in the next instant, which pushes the belief down to p_* again.

As a result, experimentation never ceases, but only fraction X_* of resources is allocated to R at the threshold p_*.

Figure 2 shows how the belief evolves, in particular, whether it increases (resp., decreases) so that $P(p) > 0$ (resp., $P(p) < 0$), when the fraction $X = 0$ or $X = I$ of resources is allocated to R. The thick curves correspond to the efficient dynamic resource allocation.

3.2 Special Case with $p^0 = 0$

If the state becomes bad after the reboot, experimentation never takes place or ceases eventually and never resumes afterwards. Whatever the fraction of resources allocated to R is, that is, whatever $X \in [0, 1]$, the belief always drifts down in absence of news. If news arrives the belief jumps to 1, but starts drifting down again in the next instant. Hence, it reaches p_* eventually and experimentation ceases.

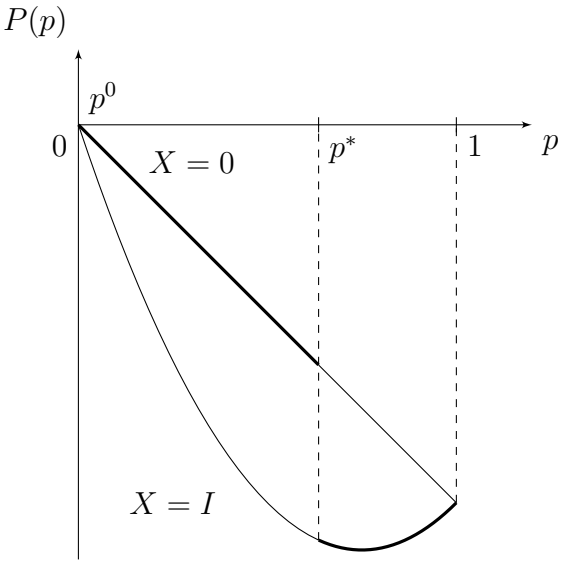

Figure 3: Evolution of the belief p in optimum that exhibits the no experimentation property. Parameters: $(I, \lambda, \phi, r, g, s, p^0) = (2, 1, 1, 1, 1, 0.7, 0)$.

Proposition 3.2 (Efficient Allocation: $p^0 = 0$)**.** *The optimal strategy of the social planner is (essentially) unique. It is a cut-off strategy with*

$$X(p) = \begin{cases} I & \text{if } p > p_*, \\ X_* & \text{if } p = p_*, \\ 0 & \text{if } p < p_*, \end{cases}$$

where $X_* \in [0, 1]$ *and* p_* *solves* (5) *(no experimentation).*

The opportunity to learn whether the initial state is good is valued, for there will not be a chance to benefit from the good state after the state reboot. The threshold is such that $p_* < s/g$ and it solves

$$s + \frac{\mu_I}{p_*}(s - gp_*) = g\frac{\mu_I}{\mu_I + \psi_I} + s\frac{1 + \psi_I - p_*}{\mu_I + \psi_I}\frac{\Phi_I(1)}{\Phi_I(p_*)}, \tag{5}$$

where

$$\Phi_I(p) \coloneqq (1 + \psi_I - p)\left(\frac{1 + \psi_I - p}{p}\right)^{\frac{\mu_I}{1+\psi_I}}. \tag{6}$$

The threshold is permeable with $P(p_*) < 0$. As mentioned, any $X_* \in [0, I]$ leads to the decrease in the belief in absence of news, so that $X = 0$ is optimal in the next instant. Hence, any such X_* is optimal. The thick curve in Figure 3 shows the efficient dynamic resource allocation.

4 Strategic Problem

Unlike in standard good-news models of strategic experimentation in which the state is rested, there are parameters for which the encouragement effect is present and players experiment beyond the single-player threshold. This is the case for relatively low reboot probabilities. There also exists a range of parameters for which free-riding effect is mute and the equilibrium is efficient. In particular, this is the case when the reboot probability is relative high and learning is slow compared to how fast the state reboots.

The (symmetric) equilibrium has either the bang-bang property or, in short, a gradual decrease in the level of experimentation. To be more precise in the latter case, there are two thresholds such that each player allocates the whole resource to R for beliefs above the upper threshold and the whole resource to S for the beliefs below the lower threshold, and she gradually decreases the fraction of resource allocated to R for beliefs between the two thresholds. The equilibrium with the bang-bang property has either the full or partial experimentation pattern of the dynamic resource allocation. The equilibrium with no bang-bang property has either partial or no experimentation pattern.

Again, all four cases of the equilibrium resource allocation are spanned by the special cases with the reboot probabilities $p^0 = 1$ and $p^0 = 0$, so I leave the intermediate case with $p^0 \in (0,1)$ for Appendix A. Similar to the social planner's problem, if $p^0 = 1$, the equilibrium exhibits either the full or partial experimentation property, while it exhibits the no experimentation property if $p^0 = 0$.

4.1 Special Case with $p^0 = 1$

If the state becomes good after the reboot, there is no encouragement effect, but there exists a range of parameters for which the symmetric equilibrium is efficient.

Proposition 4.1 (Symmetric Equilibrium: $p^0 = 1$)**.** *There exists the (essentially) unique symmetric equilibrium such that*

$$x(p) = \begin{cases} 1 & \text{if } p > \bar{p}, \\ x_\dagger(p) & \text{if } p \in [\underline{p}, \bar{p}], \\ 0 & \text{if } p < \underline{p}, \end{cases}$$

where $x_\dagger$, $\underline{p}$, and $\bar{p}$ are as follows:

(i) if $\psi/I \geq s/g$, then $x_\dagger \in [0,1]$ and $p_\dagger = \underline{p} = \bar{p} = s/g$ (full experimentation);

(ii) if $\psi \geq s/g > \psi/I$, then $x_\dagger = \frac{\psi}{Ip_\dagger}$ and $p_\dagger = \underline{p} = \bar{p} = s/g$ (partial experimentation with bang-bang property);

(iii) if $s/g > \psi$, then $x_\dagger$ is given by (7), $\underline{p}$ is given by (8), and $\bar{p}$ solves (9) (partial experimentation with no bang-bang property).

Proposition 4.1 implies that, if the state reboots at a sufficiently higher rate than news arrives, that is, if $\psi \geq s/g$, the equilibrium behavior is myopic, that is, $\underline{p} = \bar{p} = s/g$. For $\psi/I \geq s/g$, such behavior is actually efficient, and hence the free-riding effect is mute. This is not the case for $\psi \geq s/g \geq \psi/I$, as it is efficient to experiment for beliefs below the myopic threshold, whereas players behave as if they were alone. In other words, there is no encouragement effect.

If the state reboots relatively slow, that is, if $s/g > \psi$, learning the initial state becomes valuable and players experiment beyond the myopic threshold. However, they free-ride on each others' experimentation, that is, as the belief falls from $\bar{p}$ to $\underline{p}$ in absence of news, they gradually decrease the fraction of their resource which they allocate to R. The equilibrium resource allocation for $p \in [\underline{p}, \bar{p}]$ is given by

$$x_\dagger(p) = \frac{s(\mu+\psi)\frac{p-\underline{p}}{\underline{p}} + s\psi\frac{(p-\underline{p})^2}{p\underline{p}^2} - s(\mu+\psi)(1-p)\ln\left(\frac{p}{1-p}\frac{1-\underline{p}}{\underline{p}}\right)}{(I-1)(s-gp)}, \tag{7}$$

where $\underline{p}$ is equal to

$$\underline{p} = \frac{-s(\psi-\mu)+\sqrt{\Delta}}{2(g(1+\mu)-s)} \tag{8}$$

with

$$\Delta := s^2(\psi-\mu)^2 + 4s\psi(g(1+\mu)-s),$$

and $\bar{p}$ solves

$$s(\mu+\psi)\frac{\bar{p}-\underline{p}}{\underline{p}} + s\psi\frac{(\bar{p}-\underline{p})^2}{\bar{p}\underline{p}^2} - s(\mu+\psi)(1-\bar{p})\ln\left(\frac{\bar{p}}{1-\bar{p}}\frac{1-\underline{p}}{\underline{p}}\right) = (I-1)(s-g\bar{p}). \tag{9}$$

Observe that the efficient threshold given by (4) for $I = 1$ coincides with $\underline{p}$, that is, the encouragement effect is absent for this range of parameters as well. This is due to the continuation payoff being independent of the number of players, for, if news arrives, the belief jumps to 1 and stays there thereafter, and thus there is no value from other players' experimentation.

4.2 Special Case with $p^0 = 0$

If the state becomes bad after the reboot, the encouragement effect is present.

Proposition 4.2 (Symmetric Equilibrium: $p^0 = 0$)**.** *There exists the (essentially) unique symmetric equilibrium such that*

$$x(p) = \begin{cases} 1 & \text{if } p > \bar{p}, \\ x_\dagger(p) & \text{if } p \in [\underline{p}, \bar{p}], \\ 0 & \text{if } p < \underline{p}, \end{cases}$$

where $x_\dagger$ is given by (12), whereas $\underline{p}$ and $\bar{p}$ solve (10) and (11) (no experimentation).[2]

The opportunity to learn the initial state is valued, that is, each player allocates the whole resource to R even for beliefs below s/g. Moreover, the encouragement effect is present, for $\underline{p}$ is below the single-player threshold pinned down by (5) for $I = 1$. All in all, the thresholds $\underline{p}$ and $\bar{p}$ solve

$$s(\mu+\psi)\frac{\bar{p}-\underline{p}}{\underline{p}} - s(\mu-(\mu+\psi)\bar{p})\ln\left(\frac{\bar{p}}{1-\bar{p}}\frac{1-\underline{p}}{\underline{p}}\right) = (I-1)(s-g\bar{p}) \tag{10}$$

and

$$\frac{I\mu_I(s-g\bar{p})}{\bar{p}}\frac{1+\mu_I+\psi_I}{\mu_I+\psi_I} = \left(s + \frac{I\mu_I(s-g\underline{p})}{\underline{p}} - g\frac{\mu_I}{\mu_I+\psi_I}\right)\left(1 + \frac{\mu_I-(\mu_I+\psi_I)\bar{p}}{(1+\psi_I-\bar{p})\bar{p}}\frac{\Phi_I(\bar{p})}{\Phi_I(1)}\right). \tag{11}$$

The equilibrium resource allocation for $p \in [\underline{p}, \bar{p}]$ is given by

$$x_\dagger(p) = \frac{s(\mu+\psi)\frac{p-\underline{p}}{\underline{p}} - s(\mu-(\mu+\psi)p)\ln\left(\frac{p}{1-p}\frac{1-\underline{p}}{\underline{p}}\right)}{(I-1)(s-gp)}, \tag{12}$$

that is, as the belief falls from $\bar{p}$ to $\underline{p}$ in absence of news, players gradually decrease the fraction of their resource allocated to R and thus free-ride on each others' experimentation.

5 Conclusion

The paper has highlighted that some results in standard strategic experimentation models, in particular, in Keller, Rady, and Cripps (2005), rely on the state being rested, not changing over time. The equilibrium property that experimentation stops at the single-player threshold, that is, that there is no encouragement effect, is a by-product not only of conclusive

[2]The system of equations (10) and (11) has two solutions such that $\underline{p} < \bar{p}$, but only one of them is admissible. The inadmissible solution is the one with $\bar{p} = \mu/(\mu+\psi)$.

good news, but also of the rested state. If the state is restless and is unlikely to be good after the reboot, players experiment longer than they would on their own. Furthermore, even though the interior resource allocation, that is, players free-riding on each others' experimentation, also takes place if the state is restless, there exists a range of parameters for which the free-riding effect is mute. This is the case when the state reboots relatively often and is likely to be good the reboot.

References

[1] Board, Simon, and Moritz Meyer-ter-Vehn (2013): "Reputation for Quality," *Econometrica*, **81**, 2381–2462.

[2] Board, Simon, and Moritz Meyer-ter-Vehn (2014): "A Reputational Theory of Firm Dynamics," working paper, UCLA.

[3] Bolton, Patrick, and Christopher Harris (1999): "Strategic Experimentation," *Econometrica*, **67**, 349–374.

[4] Dusonchet, Fabrice, and Max-Olivier Hongler (2003): "Continuous-Time Restless Bandit and Dynamic Scheduling for Make-To-Stock Production," *IEEE Transactions on Robotics and Automation*, **19**, 977–990.

[5] Faihe, Yassine, and Jean-Pierre Müller (1998): "Behaviors Coordination Using Restless Bandits Allocation Indexes," In *Proceedings of the Fifth International Conference on Simulation of Adaptive Behavior*, 159–164.

[6] Fryer, Roland G., and Philipp Harms (2015): "Two-Armed Restless Bandits with Imperfect Information: Stochastic Control and Indexability," working paper, Harvard University, and ETH Zurich.

[7] Gittins, John (1979): "Bandit Processes and Dynamic Allocation Indices," *Journal of Royal Statistical Society. Series B (Methodological)*, **41**, 148–177.

[8] Glazebrook, Kevin D., David J. Hodge, and Christopher Kirkbride (2011): "General Notion of Indexability for Queueing Control and Asset Management," *Annals of Applied Probability*, **21**, 876–907.

[9] Glazebrook, Kevin D., Christopher Kirkbride, and Diego Ruiz-Hernandez (2006): "Spinning Plates and Squad Systems: Policies for Bi-Directional Restless Bandits," *Advances in Applied Probability*, **38**, 95–115.

[10] Halac, Marina, and Andrea Prat: "Managerial Attention and Worker Performance," *American Economic Review*, forthcoming.

[11] Hodge, David J., and Kevin D. Glazebrook (2015): "On the Asymptotic Optimality of Greedy Index Heuristics for Multi-Action Restless Bandits," *Advances in Applied Probability*, **47**, 652–667.

[12] Keller, Godfrey, and Sven Rady (1999): "Optimal Experimentation in a Changing Environment," *Review of Economic Studies*, **66**, 475–507.

[13] Keller, Godfrey, and Sven Rady (2003): "Price Dispersion and Learning in a Dynamic Differentiated-Goods Duopoly," *RAND Journal of Economics*, **34**, 138–165.

[14] Keller, Godfrey, and Sven Rady (2010): "Strategic Experimentation with Poisson Bandits," *Theoretical Economics*, **5**, 275–311.

[15] Keller, Godfrey, and Sven Rady (2015): "Breakdowns," *Theoretical Economics*, **10**, 175–202.

[16] Keller, Godfrey, Sven Rady, and Martin Cripps (2005): "Strategic Experimentation with Exponential Bandits," *Econometrica*, **73**, 39–68.

[17] Klein, Nicolas, and Sven Rady (2011): "Negatively Correlated Bandits," *Review of Economic Studies*, **78**, 693–732.

[18] La Scala, Barbara F., and Bill Moran (2006): "Optimal Target Tracking with Restless Bandits," *Digital Signal Processing*, **16**, 479–487.

[19] Le Ny, Jerome, Munther Dahleh, and Eric Feron (2008): "Multi-UAV Dynamic Routing with Partial Observations using Restless Bandit Allocation Indices," In *2008 American Control Conference*.

[20] Liu, Keqin, and Qing Zhao (2010): "Indexability of Restless Bandit Problems and Optimality of Whittle Index for Dynamic Multichannel Access," *IEEE Transactions on Information Theory*, **56**, 5547–5567.

[21] Niño-Mora, José (2001): "Restless Bandits, Partial Conservation Laws and Indexability," *Advances in Applied Probability*, **33**, 76–98.

[22] Niño-Mora, José (2002): "Dynamic Allocation Indices for Restless Projects and Queueing Admission Control: a Polyhedral Approach," *Mathematical Programming*, **93**, 361–413.

[23] Seierstad, Atle, and Knut Sydsæter (1987): *Optimal Control Theory with Economic Applications*. Amsterdam: North-Holland.

[24] Slivkins, Aleksandrs, and Eli Upfal (2008): "Adapting to a Changing Environment: the Brownian Restless Bandits," In *Procedings of the 21st Annual Conference on Learning Theory*.

[25] Vasama, Suvi (2016): "Dynamics of Innovation: Cooperation and Retardation," working paper, Humboldt University of Berlin.

[26] Veatch, Micheal H., and Lawrence M. Wein (1996): "Scheduling a Make-To-Stock Queue: Index Policies and Hedging Points," *Operations Research*, **44**, 634–647.

[27] Weber, Richard R., and Gideon Weiss (1990); "On an Index Polity for Restless Bandits," *Journal of Applied Probability*, **27**, 637–648.

[28] Whittle, Peter (1988): "Restless Bandits: Activity Allocation in a Changing World," *Journal of Applied Probability*, **25**, 287–298.

[29] Zhao, Qing, Bhaskar Krishnamachari, and Keqin Liu (2008): "On Myopic Sensing for Multi-Channel Opportunistic Access: Structure, Optimality, and Performance," *IEEE Transactions on Wireless Communications*, **7**, 5431–5440.

A General Results

This appendix presents results in general, that is, for any reboot probability $p^0 \in [0,1]$. As can be easily checked propositions in Sections 3 and 4 are special cases of propositions below.

A.1 Social Planner's Problem

Proposition A.1 (Efficient Allocation: $p^0 \in [0,1]$)**.** *The optimal strategy of the social planner is (essentially) unique. It is a cut-off strategy with*

$$X(p) = \begin{cases} I & \text{if } p > p_*, \\ X_* & \text{if } p = p_*, \\ 0 & \text{if } p < p_*, \end{cases}$$

where X_* and p_* are as follows:

(i) if $\alpha_I \geq s/g$, then $X_* \in [0,1]$ and $p_* = s/g$ (full experimentation);

(ii) if neither $\alpha_I \geq s/g$ *nor condition* (13) *holds,* $X_* = \frac{\psi(p^0-p_*)}{p_*(1-p_*)}$ *and* p_* *solves* (16) *(partial experimentation);*

(iii) if condition (13) *holds, then* $X_* \in [0,1]$ *and* p_* *solves* (15) *(no experimentation).*

The optimum exhibits the no experimentation property (region (iii) in Figure 4, left) when

$$s(\mu+\psi)(\mu+Ip^0) - g[\mu^2+\mu(I+\psi)+I\psi p^0]p^0 \geq sI^2p^0(1-p^0)\frac{\Phi_I(1)}{\Phi_I(p^0)}, \tag{13}$$

where[3]

$$\Phi_I(p) \coloneqq (\beta_I - p)\left(\frac{\beta_I - p}{p-\alpha_I}\right)^{\frac{\mu}{I}+\alpha_I}_{\beta_I-\alpha_I}. \tag{14}$$

The threshold p_* is such that $p_* > p^0$ and is pinned down by the following equation

$$s(\mu+\psi)(\mu+Ip_*) - g[\mu^2+\mu(I+\psi)+I\psi p^0]p_* = sI[Ip_*(1-p_*)-\psi(p^0-p_*)]\frac{\Phi_I(1)}{\Phi_I(p_*)}. \tag{15}$$

The optimum exhibits a partial experimentation property for parameters that violate both $\alpha_I \geq s/g$ and (13) above (region (ii) in Figure 4, left). The threshold p_* is such that $p_* \in [\alpha_I, p^0]$ pinned down by the following equation

$$\begin{aligned} s\left((\mu+\psi)(\mu+Ip_*)p_* + \frac{\psi(p^0-p_*)[\mu(1-p_*)+\psi(p^0-p_*)]}{1-p_*}\right) \\ - g[\mu^2+\mu(I+\psi)+I\psi p^0]p_*^2 \\ = s\frac{[Ip_*(1-p_*)-\psi(p^0-p_*)]^2}{1-p_*}\frac{\Phi_I(1)}{\Phi_I(p_*)}. \end{aligned} \tag{16}$$

A.2 Strategic Problem

Proposition A.2 (Symmetric Equilibrium: $p^0 \in [0,1]$)**.** *The symmetric equilibrium is such that*

$$x(p) = \begin{cases} 1 & \text{if } p > \bar{p}, \\ x_\dagger(p) & \text{if } p \in [\underline{p}, \bar{p}], \\ 0 & \text{if } p < \underline{p}, \end{cases}$$

where $x_\dagger$*,* $\underline{p}$*, and* $\bar{p}$ *are as follows:*

(i) if $\alpha_I \geq s/g$*, then* $x_\dagger \in [0,1]$ *and* $\underline{p} = \bar{p} = s/g$ *(full experimentation);*

(ii) if $\alpha_1 \geq s/g > \alpha_I$*, then* $x_\dagger(p_\dagger) = \frac{\psi(p^0-p_\dagger)}{Ip_\dagger(1-p_\dagger)}$ *and* $p_\dagger = \underline{p} = \bar{p} = s/g$ *(partial experimentation with bang-bang property);*

[3]Note that (6) is a special case of (14) for $\alpha_I = 0$ and $\beta_I = 1+\psi_I$.

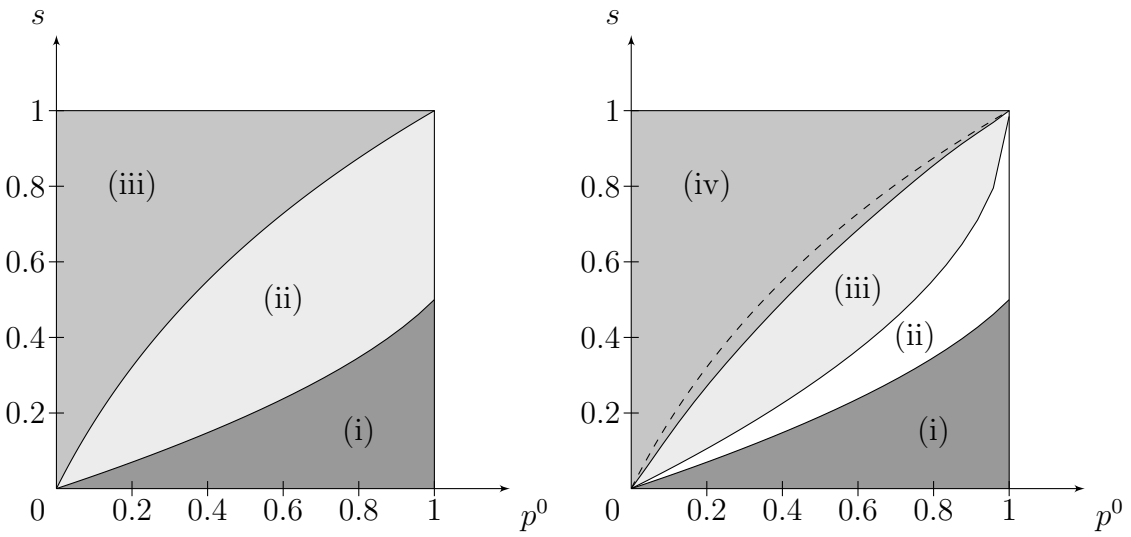

Figure 4: Parameter ranges for different patterns of the dynamic resource allocation in the social planner's problem (left) and the symmetric equilibrium (right). Parameters: $(I, \lambda, \phi, r, g) = (2, 1, 1, 1, 1)$.

(iii) if neither $\alpha_1 \geq s/g$ *nor the condition (to be found) holds,* $x_\dagger$ *is given by* (20) *and* $\underline{p}$ *and* $\bar{p}$ *solve* (21) *and* (22) *(partial experimentation with no bang-bang property);*

(iv) if the condition (to be found) holds, then $x_\dagger(p)$ *is given by* (17) *and* $\underline{p}$ *and* $\bar{p}$ *solve* (18) *and* (19) *(no experimentation).*[4]

The equilibrium exhibits a no experimentation property when the condition (to be found and is approximated numerically by the curve that separates regions (iii) and (iv) in Figure 4, right) holds. (The dashed line in Figure 4, right, corresponds to the curve that separates regions (ii) and (iii) in Figure 4, left.) The equilibrium resource allocation for $p \in [\underline{p}, \bar{p}]$ is as follows:

$$x_\dagger(p) = \frac{s\mu\frac{p-\underline{p}}{p\underline{p}} - s[\mu(1-p)+\psi(p^0-p)]\left(\frac{p-\underline{p}}{p\underline{p}} + \ln\left(\frac{p}{1-p}\frac{1-\underline{p}}{\underline{p}}\right)\right)}{(I-1)(s-gp)}, \tag{17}$$

that is, as the belief falls from $\bar{p}$ to $\underline{p}$ in absence of news, players gradually decrease the fraction of their resource allocated to R. The upper and lower thresholds, $\bar{p}$ and $\underline{p}$, are such that $\bar{p} > \underline{p} > p^0$ and solve the system of the following two equations:

$$\begin{aligned} &s(\mu+\psi)\frac{\mu+\underline{p}}{\underline{p}} - g[\mu^2+\mu(1+\psi)+\psi p^0] - \frac{s-g\bar{p}}{\bar{p}}[\mu^2+\mu(I+\psi)+I\psi p^0] \\ &\quad = \left(s(\mu+\psi)\frac{\mu+\underline{p}}{\underline{p}} - g[\mu^2+\mu(1+\psi)+\psi p^0]\right)\frac{\mu(1-\bar{p})+\psi(p^0-\bar{p})}{\psi(p^0-\bar{p})-I\bar{p}(1-\bar{p})}\frac{\Phi_I(\bar{p})}{\Phi_I(1)} \end{aligned} \tag{18}$$

[4]The system of equations (10) and (11) has two solutions such that $\underline{p} < \bar{p}$, but only one of them is admissible. The inadmissible solution is the one with $\bar{p} = (\mu + \psi p^0)/(\mu + \psi)$.

and

$$s\mu\frac{\bar{p}-\underline{p}}{\bar{p}\underline{p}}-s[\mu(1-\bar{p})+\psi(p^0-\bar{p})]\left(\frac{\bar{p}-\underline{p}}{\bar{p}\underline{p}}+\ln\left(\frac{\bar{p}}{1-\bar{p}}\frac{1-\underline{p}}{\underline{p}}\right)\right)=(I-1)(s-g\bar{p}). \quad (19)$$

If the equilibrium exhibits a partial experimentation property, the upper and lower thresholds are such that $\bar{p}\geq\alpha_I$ and $\underline{p}\leq p^0$. It is not bang-bang if $s/g>\psi$ and the condition that pins down equilibria with the no experimentation property is violated (region (iii) in Figure 4, right). The equilibrium resource allocation for $p\in[\underline{p},\bar{p}]$ is as follows:

$$x_\dagger(p)=\frac{s\mu\frac{p-\underline{p}}{p\underline{p}}+s\frac{\psi(p^0-\underline{p})(\bar{p}-\underline{p})}{(1-\underline{p})\underline{p}^2}-s[\mu(1-p)+\psi(p^0-p)]\left(\frac{p-\underline{p}}{p\underline{p}}+\ln\left(\frac{p}{1-p}\frac{1-\underline{p}}{\underline{p}}\right)\right)}{(I-1)(s-gp)}, \quad (20)$$

that is, as the belief falls from $\bar{p}$ to $\underline{p}$ in absence of news, players gradually decrease the fraction of their resource allocated to R. The upper and lower thresholds, $\bar{p}$ and $\underline{p}$, solve the system of the following two equations:

$$\begin{aligned}
&s(\mu+\psi)\frac{\mu+\underline{p}}{\underline{p}}+s\frac{\psi(p^0-\underline{p})[\mu(1-\underline{p})+\psi(p^0-\underline{p})]}{\underline{p}^2(1-\underline{p})}\\
&\qquad-g[\mu^2+\mu(1+\psi)+\psi p^0]-\frac{s-g\bar{p}}{\bar{p}}[\mu^2+\mu(I+\psi)+I\psi p^0]\\
&=\left(s(\mu+\psi)\frac{\mu+\underline{p}}{\underline{p}}+s\frac{\psi(p^0-\underline{p})[\mu(1-\underline{p})+\psi(p^0-\underline{p})]}{\underline{p}^2(1-\underline{p})}-g[\mu^2+\mu(1+\psi)+\psi p^0]\right)\\
&\qquad\times\frac{\mu(1-\bar{p})+\psi(p^0-\bar{p})}{\psi(p^0-\bar{p})-I\bar{p}(1-\bar{p})}\frac{\Phi_I(\bar{p})}{\Phi_I(1)} \quad (21)
\end{aligned}$$

and

$$\begin{aligned}
&s\mu\frac{\bar{p}-\underline{p}}{\bar{p}\underline{p}}+s\frac{\psi(p^0-\underline{p})(\bar{p}-\underline{p})}{\underline{p}^2(1-\underline{p})}-s[\mu(1-\bar{p})+\psi(p^0-\bar{p})]\left(\frac{\bar{p}-\underline{p}}{\bar{p}\underline{p}}+\ln\left(\frac{\bar{p}}{1-\bar{p}}\frac{1-\underline{p}}{\underline{p}}\right)\right)\\
&\qquad=(I-1)(s-g\bar{p}). \quad (22)
\end{aligned}$$

There exists $p_\dagger\in[\underline{p},\bar{p}]$ such that $\alpha_{Ix_\dagger}=p_\dagger$, where

$$x_\dagger(p_\dagger)=\frac{\psi(p^0-p_\dagger)}{Ip_\dagger(1-p_\dagger)}.$$

B Proofs

B.1 Proof of Lemma 2.1

Clearly, $\alpha_X \in (0, p^0)$ and $\beta_X > 1$. To show the dependence on X, take a derivative of α_X with respect to X:

$$\frac{\mathrm{d}\alpha_X}{\mathrm{d}X} = \frac{\psi}{2X^2\sqrt{(X+\psi)^2 - 4X\psi p^0}}\left(X + \psi - 2Xp^0 - \sqrt{(X+\psi)^2 - 4X\psi p^0}\right).$$

Observe that

$$\begin{gathered}\sqrt{(X+\psi)^2 - 4X\psi p^0} > X + \psi - 2Xp^0,\\ (X+\psi)^2 - 4X\psi p^0 > (X+\psi)^2 - 4Xp^0(X+\psi) + 4X^2(p^0)^2,\\ 1 > p_0.\end{gathered}$$

Hence, $\partial\alpha_X/\partial X < 0$. Finally, to show the limit, apply l'Hôpital rule.

B.2 Proof of Lemmata 3.1 and 3.2

For $s/g > \psi_I := \psi/I$, the efficient threshold p_* is given by (4). The derivatives of p_* with respect to I and ψ_I are as follows:

$$\frac{\mathrm{d}p_*}{\mathrm{d}I} = -s(g-s)\frac{\psi_I - \mu_I}{I(g(1+\mu_I) - s)\sqrt{\Delta}}\left(\frac{\psi_I}{\psi_I - \mu_I} - p_*\right) < 0$$

and

$$\frac{\mathrm{d}p_*}{\mathrm{d}\psi_I} = s\frac{1-p_*}{\sqrt{\Delta}} > 0.$$

The statements of the lemmata follow.

B.3 Proofs of Propositions 3.1, 3.2, and A.1

The social planner is concerned about which fraction $X = \{X_t\}_{t\geq 0}$, where $X_t \in [0, I]$, to allocate to R in each moment at time in order to maximize the sum of payoffs, or equivalently the average expected payoff. By the Principle of Optimality, the social planner's problem can be written as the Hamilton-Jacobi-Bellman (HJB) equation:

$$v(p) = \max_{X\in[0,I]}\left\{r\mathrm{d}t \cdot \left[\left(1 - \frac{X}{I}\right)s + \frac{X}{I}gp\right] + (1 - r\mathrm{d}t)\cdot \mathbf{E}[v(p + \mathrm{d}p) \mid p]\right\} + o(\mathrm{d}t),$$

where $v(p)$ denotes the average value function and X stands for the current fraction of resources allocated to R. To find $\mathbf{E}[v(p+\mathrm{d}p) \mid p]$, observe that

- with probability $\lambda X p \mathrm{d}t$, news arrives: the value function jumps to $v(1)$;
- with probability $p(1-\lambda X \mathrm{d}t) + (1-p) = 1 - \lambda X p \mathrm{d}t$, there is no news: assuming differentiability, the value function becomes

$$v(p) + v'(p)\mathrm{d}p = v(p) + \left[\phi(p^0 - p) - \lambda X p(1-p)\right] v'(p)\mathrm{d}t.$$

Therefore, the HJB equation for the social planner's problem takes the form:

$$v(p) = s + \frac{\psi}{\mu}(p^0 - p)v'(p) + \max_{X \in [0,I]} \left\{ X \left(b_v(p) - \frac{c(p)}{I} \right) \right\}, \tag{23}$$

with $\mu \coloneqq r/\lambda$, $\psi \coloneqq \phi/\lambda$,

$$b_v(p) \coloneqq \frac{1}{\mu} p[v(1) - v(p) - (1-p)v'(p)], \qquad \text{and} \qquad c(p) \coloneqq s - gp,$$

where $b_v(p)$ is the expected discounted benefit from the risky arm that captures the jump in the value function $v(1) - v(p)$ when good news arrives and value depreciation when there is no news, whereas $c(p)$ stands for the opportunity cost of playing risky.

Propositions 3.1, 3.2, and A.1 are proved using the verification argument.

B.3.1 Proof of Proposition 3.1

If $\psi/I \geq s/g$ and players adopt the stated strategy, the value function is equal to

$$v_*(p) = \begin{cases} g\frac{\mu p + \psi}{\mu + \psi} & \text{if } p > p_*, \\ s + g\frac{\psi(1-p_*)}{\mu+\psi}\left(\frac{1-p_*}{1-p}\right)^{\frac{\mu}{\psi}} & \text{if } p \leq p_*, \end{cases} \tag{24}$$

with $p_* = s/g$.

If $\psi/I < s/g$ and players adopt the stated strategy, the value function is equal to

$$v_*(p) = \begin{cases} g\frac{\mu_I p + \psi_I}{\mu_I + \psi_I} + C_I \Phi_I(p) & \text{if } p > p_*, \\ s + s\frac{\psi_I^2(1-p_*)}{(\mu_I + \psi_I)p_*^2}\left(\frac{1-p_*}{1-p}\right)^{\frac{\mu_I}{\psi_I}} & \text{if } p \leq p_*, \end{cases} \tag{25}$$

where

$$C_I \coloneqq \frac{\mu_I(s - gp_*)(p_* - \psi_I)}{(\mu_I + \psi_I)p_*\Phi_I(p_*)},$$

$\Phi_I(p)$ is given by (14) for $\alpha_I = \psi_I$ and $\beta_I = 1$, and p_* is given by (4).

It can be verified that, in each of the cases, $b_{v_*}(p) > c(p)/I$ for $p > p_*$ and $b_{v_*}(p) \leq c(p)/I$ for $p \leq p_*$. That is, $v_*(p)$ solves the HJB equation (23) and hence is the value function for the social planner's problem. For all p, the action prescribed by Proposition 3.1 achieves the maximum in the HJB equation, so this common strategy is optimal.

B.3.2 Proof of Proposition 3.2

If players adopt the stated strategy, the value function is equal to

$$v_*(p) = \begin{cases} (v_*(1) + g\mu_I)\frac{p}{1+\mu_I+\psi_I} + C_I\Phi_I(p) & \text{if } p > p_*, \\ s & \text{if } p \leq p_*, \end{cases} \tag{26}$$

where

$$C_I \coloneqq \frac{\mu_I(s - gp_*)}{p_*\left(\Phi_I(1) + \frac{\mu_I - (\mu_I+\psi_I)p_*}{(1+\psi_I-p_*)p_*}\Phi_I(p_*)\right)}$$

and p_* solves (5). Equation (5) is a special case of equation (15), and thus it is suffices to argue the uniqueness of the solution in the general case with $p^0 \in [0,1]$, which is done in Appendix B.3.3.

It can be verified that $b_{v_*}(p) > c(p)/I$ for $p > p_*$ and $b_{v_*}(p) \leq c(p)/I$ for $p \leq p_*$. That is, $v_*(p)$ solves the HJB equation (23) and hence is the value function for the social planner's problem. For all p, the action prescribed by Proposition 3.2 achieves the maximum in the HJB equation, so this common strategy is optimal.

B.3.3 Proof of Proposition A.1

If $\alpha_I \geq s/g$ and players adopt the stated strategy, the value function is equal to[5]

$$v_*(p) = \begin{cases} g\frac{\mu p + \psi p^0}{\mu+\psi} & \text{if } p > p_*, \\ s + g\frac{\psi(p^0-p_*)}{\mu+\psi}\left(\frac{p^0-p_*}{p^0-p}\right)^{\frac{\mu}{\psi}} & \text{if } p \leq p_*, \end{cases} \tag{27}$$

with $p_* = s/g$.

If (13) holds and players adopt the stated strategy, the value function is equal to[6]

$$v_*(p) = \begin{cases} (v_*(1) + g\mu_I)\frac{\mu_I p + \psi_I p^0}{\mu_I^2 + \mu_I(1+\psi_I) + \psi_I p^0} + C_I\Phi_I(p) & \text{if } p > p_*, \\ s & \text{if } p \leq p_*, \end{cases} \tag{28}$$

[5]Note that (24) is a special case with $p^0 = 1$.
[6]Note that (26) is a special case with $p^0 = 0$.

where

$$C_I := \frac{\mu_I(s-gp_*)}{p_*\left(\Phi_I(1)+\frac{\mu_I(1-p_*)+\psi_I(p^0-p_*)}{p_*(1-\star)-\psi_I(p^0-p_*)}\Phi_I(p_*)\right)},$$

and p_* solves (15). The right-hand side of (15) is strictly increasing in p_* for all $p_* \in (\alpha_I, 1)$. Also, the right-hand side evaluated at $p_* = 1$ is equal to $Is\psi(1-p^0)$, whereas the left-hand side is as follows:

$$s(\mu+\psi)(\mu+I) - g(\mu^2+\mu(I+\psi)+I\psi p^0) < Is\psi(1-p^0)$$

because $g > s$. (The efficient threshold is below the myopic one, that is, $p_* < s/g$. Numerically, the left-hand side is below the right-hand side at $p_* = s/g$, but this is still to be shown analytically.) Therefore, there exists $p_* \geq p^0$ if and only if the left-hand side evaluated at $p_* = p^0$ is above the right-hand side, that is, if and only if the condition (13) holds.

If neither $\alpha_I \geq s/g$ nor (13) holds, and if players adopt the stated strategy, the value function is equal to[7]

$$v_*(p) = \begin{cases} (v_*(1)+g\mu_I)\frac{\mu_I p+\psi_I p^0}{\mu_I^2+\mu_I(1+\psi_I)+\psi_I p^0} + C_I\Phi_I(p) & \text{if } p > p_*, \\ s + s\frac{\psi_I^2(p^0-p_*)}{(\mu_I+\psi_I)p_*^2}\left(\frac{p^0-p_*}{p^0-p}\right)^{\frac{\mu_I}{\psi_I}} & \text{if } p \leq p_*, \end{cases} \tag{29}$$

where

$$C_I := \frac{\mu_I(s-gp_*)}{p_*\left(\Phi_I(1)+\frac{\mu_I(1-p_*)+\psi_I(p^0-p_*)}{p_*(1-\star)-\psi_I(p^0-p_*)}\Phi_I(p_*)\right)},$$

and p_* solves (16). The right-hand side of (16) is non-negative, strictly increasing, and convex in p_* for $p_* \geq \alpha_I$. Indeed, the first-order derivative with respect to p_* is as follows:

$$s\frac{[Ip_*(1-p_*)-\psi(p^0-p_*)][2I(1-p_*)^2+\mu(1-p_*)+\psi(2-p_*-p^0)]}{(1-p_*)^2}\frac{\Phi_I(1)}{\Phi_I(p_*)} > 0$$

for $p_* \geq \alpha_I$. The second-order derivative with respect to p_* takes the form:

$$\begin{aligned}\frac{s}{(1-p_*)^3}\Big((1-p_*)^2[2I^2(1-p_*)^2+I\mu+2I\mu(1-p_*)+\mu^2] \\ +\psi(1-p_*)[2I(1-p_*)^2+2I-I(1+p_*)p^0+\mu(1-p_*)+2\mu(1-p^0)] \\ +2\psi^2(1-p^0)^2\Big)\frac{\Phi_I(1)}{\Phi_I(p_*)} > 0.\end{aligned}$$

The left-hand side of (16) is not monotone in p_*, but it has at most one inflection point.

[7] Note that (25) is a special case with $p^0 = 1$.

Indeed, the third-order derivative with respect to p_* is as follows:

$$s\frac{6(1-p^0)^2\psi^2}{(1-p_*)^4} > 0.$$

The left-hand side is equal to $\psi p^0(\mu+\psi p^0) > 0$ at $p_* = 0$, and it goes to plus infinity as p_* goes to 1 (and thus it is convex for high enough p_*). Furthermore, the left-hand side evaluated at $p_* = s/g$ is equal to

$$-s\frac{\left[\mu\left(1-\frac{s}{g}\right)+\psi\left(p^0-\frac{s}{g}\right)\right]\left[I\frac{s}{g}\left(1-\frac{s}{g}\right)-\psi\left(p^0-\frac{s}{g}\right)\right]}{1-\frac{s}{g}} \leq 0$$

for $(\mu+\psi p^0)/(\mu+\psi) > s/g \geq \alpha_I$, whereas its derivative with respect to p_* at $p_* = s/g$ is as follows:

$$-s\frac{\left[\mu\left(1-\frac{s}{g}\right)+\psi\left(p^0-\frac{s}{g}\right)\right]\left[2I\left(1-\frac{s}{g}\right)^2+\mu\left(1-\frac{s}{g}\right)+\psi\left(2-\frac{s}{g}-p^0\right)\right]}{\left(1-\frac{s}{g}\right)^2} < 0.$$

Therefore, there exists p_* satisfying (16) and $p_* \in [\alpha_I, p^0]$ if (i) the left hand side of (16) is below its right-hand side at $p_* = p^0$ and (ii) it is above its right-hand side at $p_* = \alpha_I$. Trivially, (i) is satisfied if the condition (13) with the opposite inequality holds. The condition (ii) is equivalent to

$$\frac{s}{g} \geq \frac{[\mu^2+\mu(I+\psi)+I\psi p^0]\alpha_I^2}{(\mu+\psi)(\mu+I\alpha_I)\alpha_I+\frac{\psi(p^0-\alpha_I)[\mu(1-\alpha_I)+\psi(p^0-\alpha_I)]}{1-\alpha_I}} = \alpha_I.$$

It can be verified that, in each of the cases, $b_{v_*}(p) > c(p)/I$ for $p > p_*$ and $b_{v_*}(p) \leq c(p)/I$ for $p \leq p_*$. That is, $v_*(p)$ solves the HJB equation (23) and hence is the value function for the social planner's problem. For all p, the action prescribed by Proposition A.1 achieves the maximum in the HJB equation, so this common strategy is optimal.

B.4 Proofs of Propositions 4.1, 4.2, and A.2

Player i is concerned about how much of her resource $x_i \in [0,1]$ to allocate to the risky arm in order to maximize her expected utility $v_i(p)$, where $i = 1, \ldots, I$. Thus, she solves

$$v_i(p) = \max_{x_i \in [0,1]} \{r\mathrm{d}t \cdot [(1-x_i)s + x_i g p] + (1-r\mathrm{d}t) \cdot \mathbf{E}[v_i(p+\mathrm{d}p)]\} + o(\mathrm{d}t).$$

To find $\mathbf{E}[v_i(p+\mathrm{d}p)]$, observe that

– with probability $\lambda(X_{-i}+x_i)p\mathrm{d}t$, where $X_{-i}=\sum_{j\neq i}x_j$ is the aggregate of resources allocated to the risky arm by other players, good news arrives: the value function jumps to $v_i(1)$;

– with probability $p(1-\lambda(X_{-i}+x_i)\mathrm{d}t)+(1-p)=1-\lambda(X_{-i}+x_i)p\mathrm{d}t$, there is no news: assuming differentiability, the value function becomes

$$v_i(p)+v_i'(p)\mathrm{d}p=v_i(p)+\left[\phi(p^0-p)-\lambda(X_{-i}+x_i)p(1-p)\right]v_i'(p)\mathrm{d}t.$$

Therefore, the player i's problem can be rewritten as follows:

$$v_i(p)=s+\frac{\psi}{\mu}(p^0-p)v_i'(p)+X_{-i}b_{v_i}(p)+\max_{x_i\in[0,1]}\left\{x_i(b_{v_i}(p)-c(p))\right\}, \tag{30}$$

where

$$b_{v_i}(p)=\frac{1}{\mu}p[v_i(1)-v_i(p)-(1-p)v_i'(p)]\qquad\text{and}\qquad c(p)=s-gp$$

are the expected discounted benefit and the opportunity cost of playing risky.

Propositions 4.1, 4.2, and A.2 are proved using verification argument.

B.4.1 Proof of Proposition 4.1

If $\psi\geq s/g$ and players adopt the stated strategy, the value function is equal to

$$v_\dagger(p)=\begin{cases}g\frac{\mu p+\psi}{\mu+\psi} & \text{if } p>p_\dagger,\\ s+g\frac{\psi(1-p_\dagger)}{\mu+\psi}\left(\frac{1-p_\dagger}{1-p}\right)^{\frac{\mu}{\psi}} & \text{if } p\leq p_\dagger,\end{cases} \tag{31}$$

with $p_\dagger=\underline{p}=\bar{p}=s/g$.

If $s/g>\psi$ and players adopt the stated strategy, the value function is equal to

$$v_\dagger(p)=\begin{cases}g\frac{\mu p+\psi}{\mu+\psi}+C_I\Phi_I(p) & \text{if } p>\bar{p},\\ g\frac{(1+\mu)\psi}{\mu+\psi}+s\frac{\mu(p-\psi)}{(\mu+\psi)p}+(I-1)x_\dagger(p)\frac{\mu(s-gp)}{\mu+\psi} & \text{if } p\in[\underline{p},\bar{p}],\\ s+\frac{C_0}{(1-p)^{\frac{\mu}{\psi}}} & \text{if } p<\underline{p},\end{cases} \tag{32}$$

where

$$C_I:=\frac{I\mu_I(s-g\bar{p})(\bar{p}-\psi_I)}{(\mu_I+\psi_I)\bar{p}\Phi_I(\bar{p})},\qquad C_0:=\left(g-s-\frac{\mu(s-g\underline{p})}{\underline{p}}\right)\frac{\psi(1-\underline{p})^{\frac{\mu}{\psi}}}{\mu+\psi},$$

$x_\dagger$ is given by (7), $\underline{p}$ is given by (8), and $\bar{p}$ is pinned down by (9). The upper threshold $\bar{p}$

that solves (9) is unique. Gathering all terms in the left-hand side and taking a derivative with respect to $\bar{p}$ yield

$$s(\mu+\psi)\left(\frac{\bar{p}-\underline{p}}{\bar{p}\underline{p}}+\ln\left(\frac{\bar{p}}{1-\bar{p}}\frac{1-\underline{p}}{\underline{p}}\right)\right)+s\psi\frac{\bar{p}^2-\underline{p}^2}{\bar{p}^2\underline{p}^2}+(I-1)g>0.$$

Moreover, the left-hand side of (9) evaluated at $\bar{p}=\underline{p}$ is zero, whereas the right-hand side is equal to $(I-1)(s-g\underline{p})>0$. On the other hand, the left-hand side of (9) evaluated at $\bar{p}=1$ is equal to $s\mu(1-\underline{p})/\underline{p}+s\psi(1-\underline{p})/\underline{p}^2>0$, whereas the right-hand side is equal to $(I-1)(s-g)<0$. (The upper threshold is below the myopic one, that is, $\bar{p}<s/g$. Numerically, the left-hand side of (9) evaluated at $\bar{p}=s/g$ is above the right-hand side, but this is still to be shown analytically.)

It can be verified that, in each of the cases, $b_{v_\dagger}(p)>c(p)$ for $p>\bar{p}$, $b_{v_\dagger}(p)=c(p)$ for $p\in[\underline{p},\bar{p}]$, and $b_{v_\dagger}(p)<c(p)$ for $p<\underline{p}$. That is, $v_\dagger(p)$ solves the HJB equation (30) and hence is the value function in the symmetric equilibrium.

B.4.2 Proof of Proposition 4.2

If players adopt the stated strategy, the value function is equal to

$$v_\dagger(p)=\begin{cases}(v_\dagger(1)+g\mu_I)\frac{p}{1+\mu_I+\psi_I}+C_I\Phi_I(p) & \text{if } p>\bar{p},\\ -(v_\dagger(1)+\mu p)\frac{\psi p}{\mu-(\mu+\psi)p}+s\frac{\mu(1+\psi-p)}{\mu-(\mu+\psi)p}+(I-1)x_\dagger(p)\frac{\mu(1-p)(s-gp)}{\mu-(\mu+\psi)p} & \text{if } p\in[\underline{p},\bar{p}],\\ s & \text{if } p\le\underline{p},\end{cases}\tag{33}$$

where

$$C_I:=\frac{I\mu_I(s-g\bar{p})}{\bar{p}\left(\Phi_I(1)+\frac{\mu_I-(\mu_I+\psi_I)\bar{p}}{(1+\psi_I-\bar{p})\bar{p}}\Phi_I(\bar{p})\right)},$$

$x_\dagger$ is given by (12), and $\underline{p}$ and $\bar{p}$ solve the system of two equations (10) and (11). Numerically (and to be argued analytically), the system admits two solutions such that $\bar{p}>\underline{p}$, but one of them, namely, with $\bar{p}=\mu/(\mu+\psi)$, is inadmissible. The admissible solutions has the property $s/g>\bar{p}>\underline{p}$.

It can be verified that $b_{v_\dagger}(p)>c(p)$ for $p>\bar{p}$, $b_{v_\dagger}(p)=c(p)$ for $p\in[\underline{p},\bar{p}]$, and $b_{v_\dagger}(p)<c(p)$ for $p<\underline{p}$. That is, $v_\dagger(p)$ solves the HJB equation (30) and hence is the value function in the symmetric equilibrium.

B.4.3 Proof of Proposition A.2

If $\alpha_1 \geq s/g$ and players adopt the strategy specified, the value functions is equal to[8]

$$v_\dagger(p) = \begin{cases} g\frac{\mu p+\psi p^0}{\mu+\psi} & \text{if } p > p_\dagger, \\ s + g\frac{\psi(p^0-p_\dagger)}{\mu+\psi}\left(\frac{p^0-p_\dagger}{p^0-p}\right)^{\frac{\mu}{\psi}} & \text{if } p \leq p_\dagger, \end{cases} \tag{34}$$

with $p_\dagger = \underline{p} = \bar{p} = s/g$.

If neither $\alpha_1 \geq s/g$ nor the condition on parameters (that pins down the no experimentation case and which is to be found) holds, and if players adopt the strategy specified, the value function is equal to[9]

$$v_\dagger(p) = \begin{cases} (v_\dagger(1)+g\mu_I)\frac{\mu_I p+\psi_I p^0}{\mu_I^2+\mu_I(1+\psi_I)+\psi_I p^0} + C_I\Phi_I(p) & \text{if } p > \bar{p}, \\ (v_\dagger(1)+g\mu)\frac{\psi(p^0-p)}{\mu(1-p)+\psi(p^0-p)} + s\mu\frac{p(1-p)-\psi(p^0-p)}{p[\mu(1-p)+\psi(p^0-p)]} & \\ \qquad +(I-1)x_\dagger(p)\frac{\mu(1-p)(s-gp)}{\mu(1-p)+\psi(p^0-p)} & \text{if } p \in [\underline{p}, \bar{p}], \\ s + \frac{C_0}{(1-p)^{\frac{\mu}{\psi}}} & \text{if } p < \underline{p}, \end{cases} \tag{35}$$

where

$$C_I \coloneqq \frac{\mu(s-g\bar{p})}{\bar{p}\left(\Phi_I(1) + \frac{\mu(1-\bar{p})+\psi(p^0-\bar{p})}{I\bar{p}(1-\bar{p})-\psi(p^0-\bar{p})}\Phi_I(\bar{p})\right)},$$

$$C_0 \coloneqq \left(v_\dagger(1) - s - \frac{\mu(s-g\underline{p})}{\underline{p}}\right)\frac{\psi(p^0-\underline{p})^{\frac{\mu}{\psi}+1}}{\mu(1-\underline{p})+\psi(p^0-\underline{p})},$$

$x_\dagger$ is given by (20), and $\underline{p}$ and $\bar{p}$ solve the system of two equations (21) and (22). Numerically (and to be argued analytically), the system admits a unique solution such that $\bar{p} > \underline{p}$. This solution also has the property $\underline{p} < p^0$ and $s/g > \bar{p} > \alpha_I$.

If the condition (to be found) holds, and if players adopt the strategy specified, the value function is equal to[10]

$$v_\dagger(p) = \begin{cases} (v_\dagger(1)+g\mu_I)\frac{\mu_I p+\psi_I p^0}{\mu_I^2+\mu_I(1+\psi_I)+\psi_I p^0} + C_I\Phi_I(p) & \text{if } p > \bar{p}, \\ (v_\dagger(1)+g\mu)\frac{\psi(p^0-p)}{\mu(1-p)+\psi(p^0-p)} + s\mu\frac{p(1-p)-\psi(p^0-p)}{p[\mu(1-p)+\psi(p^0-p)]} & \\ \qquad +(I-1)x_\dagger(p)\frac{\mu(1-p)(s-gp)}{\mu(1-p)+\psi(p^0-p)} & \text{if } p \in [\underline{p}, \bar{p}], \\ s & \text{if } p < \underline{p}, \end{cases} \tag{36}$$

[8] Note that (31) is a special case with $p^0 = 1$.
[9] Note that (32) is a special case with $p^0 = 1$.
[10] Note that (33) is a special case with $p^0 = 0$.

where

$$C_I := \frac{\mu(s - g\bar{p})}{\bar{p}\left(\Phi_I(1) + \frac{\mu(1-\bar{p})+\psi(p^0-\bar{p})}{I\bar{p}(1-\bar{p})-\psi(p^0-\bar{p})}\Phi_I(\bar{p})\right)},$$

$x_\dagger$ is given by (20), and $\underline{p}$ and $\bar{p}$ solve the system of two equations (21) and (22). Numerically (and to be argued analytically), the system admits two solutions such that $\bar{p} > \underline{p}$, but one of them, namely, with $\bar{p} = (\mu + \psi p^0)/(\mu + \psi)$, is inadmissible. The admissible solutions has the property $s/g > \bar{p} > \underline{p}$.

It can be verified that, in each of the cases, $b_{v_\dagger}(p) > c(p)$ for $p > \bar{p}$, $b_{v_\dagger}(p) = c(p)$ for $p \in [\underline{p}, \bar{p}]$, and $b_{v_\dagger}(p) < c(p)$ for $p < \underline{p}$. That is, $v_\dagger(p)$ solves the HJB equation (30) and hence is the value function in the symmetric equilibrium.