PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA
MAGISTER EN ECONOMIA

TESIS DE GRADO
MAGISTER EN ECONOMIA

Ochoa, Abett de la Torre, Fernando

Noviembre, 2017

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
INSTITUTO DE ECONOMIA
MAGISTER EN ECONOMIA

Coordinating Experimentation

Fernando Ochoa Abett de la Torre

Comisión

Nicolás Figueroa y Rodrigo Harrison

Santiago, Noviembre de 2017

Coordinating experimentation*

Fernando Ochoa A.†
Pontificia Universidad Católica de Chile

November 29, 2017

Abstract

We consider a simple two-agent model of strategic experimentation with payoff externalities and unobservable actions. Each agent decides whether to invest effort (or resources) into an independent risky project, if he invests and succeeds his payoff is greater if the other agent succeeds too. We characterize the unique symmetric Markov Perfect Equilibria of the game. The dynamic structure of the problem allows agents to coordinate their experimentation in such a way that they provide a strictly more effort than they would have provided in the absence of the externality. Nevertheless, they provide too little effort in a dynamically inefficient way relative to the first best. The magnitude of efficiency losses is critically linked to the common prior about projects' quality. If agents are too pessimistic they do not provide any effort even though it is socially optimal to do so.

*Thesis written as a Master student at the Microeconomics thesis seminar, Department of Economics, Pontificia Universidad Católica de Chile. I thank to my advisors Rodrigo Harrison and Nicolás Figueroa for their guidance. I am specially grateful to Nicolás Figueroa for unconditional advice and support. I would also like to thank Caio Machado, Juan Pablo Montero, Sofía Moroni, and Juan Pablo Xandri for helpful comments and discussions. Finally, I thank the ICSI Institute (Milenio P05-004F and Basal FBO-16) for financial support. Any error or omission is my sole responsibility.
†Email: fjochoa@uc.cl

1 Introduction

There are many industries where the profitability of a firm's innovation depends on other firms' innovations, and vice versa. In several markets new products' value increases if there are other new complementary products. For example, a new and more powerful computer hardware would be more valued by customers if new and more demanding PC software is created, but that potential software could be used only if there exist a more powerful hardware in the future. Also, decisions on how much and for how long to invest to create a new product should consider potential innovations of intermediate goods producers that might lower their production costs, as well as the investment in R&D to reduce production costs should consider that new products might boost intermediate goods demand in the future. For example, while many tech-companies were investing in R&D to reduce the size of lithium batteries for smartphones and other devices, the mining industry was investing in R&D to reduce lithium's production costs.[1]

The consideration of possible future innovations of other firms in the R&D decisions of a firm can be understood as an externality on the experimentation payoffs. This means that the success of a firm's R&D increments net payoff of other firm's R&D. If firms invest more when there are externalities than when there are none, then they are *coordinating their experimentation*. The necessary conditions that lead to the coordination of experimentation could be crucial to understand the success of some innovation based industries as well as the failure (or non-existence) of others.

This work provides a simple model of experimentation to understand how externalities on experimentation payoffs affect agents' behavior. We model experimentation as a variation of the continuous-time two-armed exponential bandit model of Keller et al. (2005). In particular, we follow the main variations used by Bonatti and Hörner (2011).

We consider the strategic behavior of experimenting agents when the externalities on the payoffs are exogenous and actions are unobservable. An example of this is when firms have to decide whether to dedicate their resources to selling their usual products or investing in R&D to try to develop a new product that will have a greater market value (or lower production cost) if other firms' R&D is also successful. Our goal is to determine how payoff externalities changes the dynamics of investment provision. We model this situation as one where two agents have to decide whether to allocate resources to experiment on a risky project (risky arm) of unknown quality or into a safe one (safe arm) that pays a fixed amount each period. We model the safe arm as the experimentation cost of opportunity. The risky arm can be bad and yield zero profits or good and yield a

[1] Although the motivation of this work is theoretical, it is worth mentioning that identification and measurement of R&D spillovers has been an active line of empirical research in the last decades. See for example Griliches (1992), Malerba et al. (2013), Steurs (1995), and Park (2004).

profit. Each agent's risky arm quality is independent from the others and it can be learnt only by experimentation. The strategic component arises because if the risky arm is good the profit that it yields is higher if the other agent were successful too.

Most of the recent strategic experimentation literature has focused on the problem of free riding in several settings such as teams, public goods, or experimental consumption (Bonatti and Hörner, 2011; Georgiadis, 2015; Heidhues et al., 2015). To the best of our knowledge, this is the first work that brings the coordination problem to the strategic experimentation literature. We hope that the insights provided by our simple model may be the first steps in order to understand the problem of coordination in dynamic strategic experimentation.

We show that the mix of the learning process with the dynamic structure of the problem allow agents to coordinate their experimentation. In particular, we characterize the unique symmetric Markov Perfect Bayesian equilibrium of the game. There are two forces that shape the equilibrium. On one hand, experimentation is more profitable *today* since the expected payoff is greater because of the externality, so agents are willing to provide more effort for more time. On the other hand, in order to save experimentation costs agents have incentives to stop their experimentation to *wait and see* if the other agent has success, and then experiment again. The evolution of beliefs about projects' quality is what changes the relative importance of both forces over time.

Although agents coordinate their experimentation in equilibrium, there is under-provision of effort. If the common prior over projects' quality is high enough, agents start exerting maximal effort for some time. As agents becomes more pessimistic about projects's quality effort starts to be provided in a dynamically inefficient way (i.e. they provide interior levels of effort). Finally, when agents are pessimistic enough, they stop experimentation. Numerical exercises shows that agents stop experimentation too early relative to the first best. This implies that the common prior over projects' quality is key for the level of inefficiency. For intermediate priors, agents provide too little effort (interior levels) for a while. For low enough priors, agents do not exert any effort despite it is socially optimal to exert maximal effort.

2 Related literature and contribution

This work fits into the literature of experimentation with exponential bandits. The experimentation literature studies decisions of an agent that faces a trade-off between *exploitation* of a known production technology or *exploration* of a new one in dynamic decision problems with learning.[2] The strategic experimentation literature (Bolton and Harris, 1999) has extended the single agent problem to a many-agent setting where an

[2]See Bergemann and Välimäki (2008) for a good survey.

agent can learn from the experimentation of others, this implies that information is a public good, and therefore a free-rider problem arises. The two-armed exponential bandit model version of experimentation (Keller et al., 2005) is a way to model the aforementioned problems that is broadly used because of its tractability. Our model is a variation of Keller et al. (2005) very similar to that used by Bonatti and Hörner (2011). The main contribution of this work is to study the strategic behavior of agents when there are externalities on the experimentation payoffs, and this implies to bring the coordination problem, rather than the usual free-riding, to the experimentation literature. Up to our knowledge this work is the first that studies the coordination problem in a dynamic experimentation setting.[3]

Strategic experimentation models has been used to model the problem of experimentation in several settings, such as exogenously and self-organized teams (Bonatti and Hörner, 2011; Georgiadis, 2015) or experimental consumption of goods (Heidhues et al., 2015).[4] This literature aims to answer questions such as how much experimentation can be expected in equilibrium, how fast it is done, optimal team size, and provision dynamics. A particular assumption broadly used by this literature is that the quality (good or bad) of agents' risky arms are perfectly correlated (i.e. every agent face the same risky arm). One of the main reasons why the strategic experimentation literature share that assumption is because equilibrium with imperfectly correlated risky arms is still an open question in the literature.[5] Our simple two-agent model departure from the literature because, without overcoming the aforementioned limitation, allows us study a new setting where each agent faces a different project whose quality is independent from other agent's project. The externalities on the payoffs of the risky arm is what keep the strategic component in our model of experimentation despite the non-correlation of the risky arms.

Our main result show that there exist a unique symmetric Markov Perfect Equilibria. This result is surprising because coordination games usually have multiplicity of equilibria. Therefore, our work suggest that the dynamic structure and the learning component of strategic experimentation games could "refine" the equilibria without the need to impose an extra refinement criteria. We also find some differences between the problem of coordination of experimentation and free-riding on experimentation (Bonatti and Hörner,

[3]In order to avoid possible confusions is worth to mention that the common learning in repeated games literature usually refers to experimentation coordination (See for example Wiseman, 2005) . Nevertheless, in that games there is an uncertain state of the world that can be learnt only if agents coordinate their actions. In our model coordination is necessary only to keep experimentation profitable, not to learn about projects quality. To be even more precise, there are payoff externalities instead of information externalities, and each project has an independent state chosen by nature (its quality).

[4]The insights of this literature has been incorporated by contract theory. In general terms, they study how to provide optimal incentives for teams that has to experiment. See for example, Moroni (2015), Shan (2017), Georgiadis (2015).

[5]Up to our knowledge the only exception is Klein and Rady (2011). They characterize the Markov perfect equilibrium for two agents that face negatively correlated bandits (they extend the results for three players).

2011). First, efficiency losses are critically linked to initial priors about project's quality: for low enough priors agents provide zero effort despite it is socially efficient to do so. Second, agents' patience plays a non trivial role as intermediate levels of patience improves welfare.

This work is also related to a vast R&D literature that studies the effect of R&D spillovers on firms' investment decisions. Nevertheless, this literature usually focus on the effects that intra-industry competition (or negative spillovers) changes the way that inter-industry spillovers affects decisions in static settings (Steurs, 1995; Bessen and Maskin, 2009), externalities in our model can be understood as inter-industy spillovers.[6] Chen (2012) analyze innovation frequency (i.e. timing of release) in an industry with two producers of perfect complementary goods, he shows that coordination failures arises in the timing of releases (too early or too late). Our model departs from this literature since we do not assume that goods are perfect complements or that innovation is sequential, that is an agent can be successful and receive a payoff from his innovation despite the results of the other. Also, we can explicitly study investment dynamics.

The closest related work in R&D spillovers literature is the work by Salish (2015). She studies a dynamic model of R&D investment à la Keller et al. (2005) with inter- and intra-sectoral R&D spillovers where the profitability of an invention in one sector (dependent) increases if there is an innovation in another sector (independent), but not vice versa. She found that if the independent firm learns faster than the dependent there is no effect of spillovers but if it learns slower the dependent firm will invest more relative to a setting without spillovers. Our model differs from Salish (2015) since we do not assume independence of one sector, so we have a game and the equilibrium is not a direct implication of Keller et al. (2005).[7]

3 A static example

Two symmetric agents face an static experimentation problem with payoff externalities. Each agent has an independent project that can be good or bad. Let $\bar{p} \in (0,1)$ be the common prior probability that the project is good.

Each agent has to decide how much effort $u_i \in [0,1]$ to exert. The cost of effort is $c(u_i) = cu_i$, where $c > 0$. If the project is good and agent exert effort u_i, a breakthrough arrives with probability u_i and receives a payoff $\pi > c$. If the project is bad a breakthrough never arrives and the project yields no payoff. If both agents have a breakthrough, each

[6]By static setting we refer to that the investment decisions are essentially static, despite the model is dynamic.

[7]The assumption of independence of Salish (2015) rules out the possibility of a game since the independent firm decisions are (obviously) independent from the dependent one. Their model is a particular case of ours.

agent receive an extra payoff $\gamma > 0$, which is the externality.

Agent i solves the following problem,

$$\max_{u_i \in [0,1]} (\bar{p}u_i)(\bar{p}u_j)(\pi + \gamma) + (\bar{p}u_i)((1 - \bar{p}) + \bar{p}(1 - u_j))\pi - cu_i$$
$$\Leftrightarrow \max_{u_i \in [0,1]} \left[\bar{p}^2 u_j(\pi + \gamma) + (\bar{p} - \bar{p}^2 u_j)\pi - c\right] u_i,$$

which clearly has a corner solution that depends on the value of parameters and a given strategy u_j. Note that if $u_j = 0$, agent i optimally decides to exert $u_i = 1$ iff

$$\bar{p} \geq \frac{c}{\pi}.$$

Also, if $u_j = 1$ agent decides $u_i = 0$ iff

$$\bar{p}^2\gamma + \bar{p}\pi - c < 0$$
$$\Leftrightarrow \bar{p} < \frac{-\pi + \sqrt{4\gamma c + \pi^2}}{2\gamma}.$$

Therefore, agents always experiment for $\bar{p} \in [c/\pi, 1)$ and never experiment for $\bar{p} \in (0, \frac{-\pi+\sqrt{4\gamma c+\pi^2}}{2\gamma})$. But for $\bar{p} \in [\frac{-\pi+\sqrt{4\gamma c+\pi^2}}{2\gamma}), c/\pi)$ the game has two Nash Equilibria in pure strategies, both agents experiment or no one does.

This is a simple example of a coordination game (or more generally, a supermodular game). There are well known tools to refine this kind of games to obtain uniqueness, one of the most popular is the global game approach. Broadly speaking, this tools implies to select an information structure to deal with the high order beliefs (Van Zandt and Vives, 2007). The goal of this work is to determine if the mix of the dynamic structure and learning lead to a unique equilibria of this game without the necessity of a refinement tool.

4 Model

There are two agents engaged in different and independent projects, each project may be good ($\theta_i = G$) or bad ($\theta_i = B$). Agent i's project is good with a probability $\bar{p} \in (0, 1)$ which is assumed to be public information.

Time is continuous in $t \in [0, \infty)$. Each agent continuously chooses a level of costly effort. Agent i's instantaneous cost of exerting effort $u_i \in [0, 1]$ is $c_i(u_i) = cu_i$.

If the project is good and agent i exerts effort u_i at time t, a breakthrough occurs according to a Poisson process with intensity $f(u_i) = \lambda u_i$, for some $\lambda > 0$. If the project is bad a breakthrough never arrives.

For each player the game ends if a breakthrough occurs on his project, which can be interpreted as the successful completion of the project.[8] A successful project is worth a (expected) net present value of π to the agent if the breakthrough occurs at some $t < \infty$. While a breakthrough does not occur agents reaps no benefits from their project. Nevertheless, conditional on having a breakthrough, each agent payoff depends on the other agent's risky arm quality. Agent i is said to be successful iff he has experienced a breakthrough at some $t < \infty$. A successful agent j generates an (expected) externality γ on the payoff of agent i's project. All agents discount the future at a common rate $r > 0$.

If agent i exert effort $(u_{i,t})_{t\geq 0}$ and agent j was successful at $t' < t$, and a breakthrough arrives at some $t < \infty$ the average discounted payoff to agent i is

$$e^{-rt}(\pi + \gamma) - c\int_0^t e^{-rs}u_{i,s}ds.$$

On the other hand, if agent j is not successful yet, the average discounted payoff of a breakthrough to agent i is

$$e^{-rt}(\pi + \mathbb{E}[\gamma|p_{j,t}, \{u_j\}_{\tau\geq t}]) - c\int_0^t e^{-rs}u_{i,s}ds,$$

where $\mathbb{E}[\gamma|p_{j,t}, (u_j)_{\tau\geq t}]$ is the expected discounted externality. This is the discounted value of the externality integrated over the probability density of a breakthrough on j's risky arm, which depends on the belief at t that j's arm is good $p_{j,t}$, and on the expected effort $(u_j)_{\tau\geq t}$ that agent j will exert thereafter.

Breakthroughs are public information, meaning that all agents observe the time at which a breakthrough occurs, and who obtained it.

Our objective is to identify the symmetric Markov Perfect Bayesian Equilibrium (MPBE) in pure strategies of this game. That is, each agent i chooses two measurable functions $u_i : \mathbb{R}_+ \to [0, 1]$ (pure strategy) to maximize his expected payoff. First, a function conditional on no breakthrough having occurred. Second, they choose another measurable function that maximizes his payoff if the other agent has a breakthrough at some $t \in [0, \infty)$.

In absence of a breakthrough, each agent *learns* about the quality of his project, this is captured in the evolution of his belief that his project is good.[9] By Bayes' rule,

[8] As Moroni (2015) notes, an alternative interpretation is that after a breakthrough there is a number of tasks with known feasibility to be done.

[9] Since the arrival of a breakthrough is a Poisson process, given an effort profile $u_{i,t}$, the probability that a breakthrough has not occurred between τ and t if the project is good is $exp\{-\lambda \int_\tau^t u_{i,t}ds\}$.

$$p_{i,t} = \frac{\bar{p}e^{-\lambda\int_0^t u_{i,\tau}d\tau}}{\bar{p}e^{-\lambda\int_0^t u_{i,\tau}d\tau} + (1-\bar{p})} \tag{1}$$

and the evolution of the belief in an interval of dt can be obtained from

$$p_{i,t+dt} = \frac{p_{i,t}e^{-\lambda u_{i,t}dt}}{p_{i,t}e^{-\lambda u_{i,t}dt} + (1-p_{i,t})}.$$

By manipulating the latter expression algebraically and taking limits ($dt \to 0$), we can describe the evolution of the belief by the familiar differential equation

$$\dot{p}_{i,t} = \frac{dp_{i,t}}{dt} = -p_{i,t}(1-p_{i,t})\lambda u_{i,t}. \tag{2}$$

Every time that an agent exert effort and a breakthrough does not arrives he becomes more pessimistic about project quality. Clearly, on the equilibrium path each agent learns about both projects.

5 Analysis of the problem

In this section we solve the autarky problem to fix some basic ideas from experimentation literature. Then we characterize agent's problem when there are payoff externalities.

5.1 Autarky

The autarky problem is the simplest problem of experimentation with exponential bandits, only one agent faces the "exploitation" vs "exploration" trade off. In particular, this is equivalent to Keller et al. (2005) with one myopic agent. Obviating the individual subscript, the agent's problem is

$$\Pi^A(\bar{p}) = \max_{u_t} \int_0^\infty [p_t\lambda\pi - c]u_t e^{-\int_0^t(p_s\lambda u_s + r)ds}dt,$$

subject to

$$\dot{p}_t = -p_t(1-p_t)\lambda u_t \wedge p_0 = \bar{p}.$$

Notice that the instantaneous probability assigned by the agent to a breakthrough occurring is $p_t \lambda u_t$, and the probability that no breakthrough has occurred by time $t \geq 0$ is $e^{-\int_0^t p_s \lambda u_s ds}$, then $p_t \lambda u_t e^{-\int_0^t p_s \lambda u_s ds}$ is the probability density that agent i obtains a breakthrough at time t.[10] Since the integrand is positive as long as $p_t \lambda \pi \geq c$, its clear that if that condition is satisfied is optimal to set $u_t = 1$, and $u_t = 0$ otherwise. Note that by (1):

$$p_{i,t} = \frac{\bar{p}}{\bar{p} + (1-\bar{p})e^{\int_0^t \lambda u_t ds}},$$

so we can rewrite the condition,

$$\begin{aligned} u_t = 1 \Leftrightarrow & \quad p_t \lambda \pi \geq c \\ \Leftrightarrow & \quad \tfrac{\bar{p}}{\bar{p}+(1-\bar{p})e^{\lambda t}} \geq \tfrac{c}{\lambda \pi} \\ \Leftrightarrow & \quad \tfrac{\bar{p}}{(1-\bar{p})} \left(\tfrac{\lambda \pi_i}{c} - 1 \right) \geq e^{\lambda t} \\ \Leftrightarrow & \quad t \leq T^A := \tfrac{ln\left(\frac{\lambda \pi - c}{c}\right) - ln\left(\frac{1-\bar{p}}{\bar{p}}\right)}{\lambda}. \end{aligned}$$

Lemma 1 summarizes the result.

Lemma 1 *In the autarky problem the agent's optimal experimentation rule is*

$$u_t^A = \begin{cases} 1 & if\, t \leq T^A, \\ 0 & if\, t > T^A, \end{cases}$$

where,

$$T^A := \lambda^{-1} \left[ln \left(\frac{\lambda \pi - c}{c} \right) - ln \left(\frac{1-\bar{p}}{\bar{p}} \right) \right] \tag{3}$$

This is a standard result in experimentation literature, qualitatively similar to the planner solution in Bonatti and Hörner (2011) and Moroni (2015).[11] Note that T^A is only

[10] One of the properties that makes exponential bandits a powerful tool to model experimentation is that they provide this tractable form to the stochastic process of breakthrough arrival.

[11] This is because both models consist on a team working on the same project, so the first best is equivalent to our autarky model. To be more precise, Moroni (2015) has a team that works on several milestones of a big project, but just in one milestone at a time.

well defined when $\lambda\bar{p}\pi > c$. To make exposition more clear, we avoid to prove that some sub-games are well defined by assuming the following:[12]

Assumption 1 *It is always profitable for any agent to experiment for at least some arbitrary small amount of time:* $\lambda\bar{p}_i\pi_i > c, \forall i$.

We are going to refer to

$$p^a = c/\lambda\pi, \tag{4}$$

as the threshold belief at which an agent stops experimentation in autarky. Also, we refer to effort at time t as the instantaneous individual effort, to total effort at t as the sum of individual effort at that time, and to aggregate effort at t as the integral of of total effort over all time up to t.

5.2 Agent's problem after a breakthrough

In order to characterize the problem we need to solve the subgames after a breakthrough. W.l.g. we are going to write the problem from the perspective of agent i. We start by considering the problem of agent i when agent j had a breakthrough at some arbitrary time t,

$$\Pi^S := \max_{u_{i,\tau}} \int_t^\infty [p_{i,\tau}\lambda(\pi+\gamma) - c]u_{i,\tau}e^{-\int_t^\tau (p_{i,s}\lambda u_{i,s}+r)ds}d\tau \tag{5}$$

$$\text{s.t.} \quad \dot{p}_{i,\tau} = -p_{i,\tau}(1-p_{i,\tau})\lambda u_{i,\tau}.$$

Note that the problem faced by agent i is equivalent to the autarky problem starting at time t with initial belief $p_{i,t}$. Therefore, Lemma 1 implies the following

Corollary 1 *If agent j has a breakthrough at some arbitrary t, agent i optimally chooses the effort profile*

$$u^S_{i,\tau} = \begin{cases} 1 & \text{if } \tau \leq T^S + t, \\ 0 & \text{if } \tau > T^S + t, \end{cases}$$

where,

[12]Our main results hold for priors $\bar{p}$ lower than p^a. Therefore, Assumption 1 is not necessary, it is only made for exposition purposes.

$$T^S := \lambda^{-1}\left[ln\left(\frac{\lambda(\pi+\gamma)-c}{c}\right) - ln\left(\frac{1-p_{i,t}}{p_{i,t}}\right)\right]. \tag{6}$$

As before, T^S is not well defined for every set of parameters $\{\lambda, p_{i,t}, \pi, \gamma, c\}$. Nevertheless, is always well defined for some t at which agents were experimenting. The following Lemma formalizes this statement.

Lemma 2 T^S *is always well defined as an outcome of a symmetric sub-game where players were experimenting.*

Proof If in any symmetric equilibrium agent j has a breakthrough at some t then it was rational for him to experiment at t. The symmetry implies that it was also rational for agent i to experiment at t. Therefore, T^S is well defined because at $t + dt$ (after j's breakthrough) agent i's risky arm payoff is strictly higher than at t. ∎

Note that Π^S depends only on $p_{i,t}$ (see Appendix A):

$$\Pi^S(p_{i,t}) = c\left\{\frac{p_{i,t}}{\lambda+r}\left[\beta - \beta^{-\frac{r}{\lambda}}\left(\frac{p_{i,t}}{1-p_{i,t}}\right)^{-(1+\frac{r}{\lambda})}\right] - \frac{(1-p_{i,t})}{r}\left[1-\left(\beta\frac{p_{i,t}}{1-p_{i,t}}\right)^{-\frac{r}{\lambda}}\right]\right\},$$

where $\beta := (\lambda(\pi+\gamma)-c)/c$.

5.3 Agent's problem before a breakthrough

When no agent has been successful yet, a breakthrough for agent i at some time t implies an instant payoff of π plus the expected discounted externality. The later is the discounted externality that he will receive given that agent j will face the problem $\Pi^S(p_{j,t})$:

$$\mathbb{E}[\gamma|p_{j,t}, \{u_j\}_{\tau\geq t}] := \Pi^{DE}(p_{j,t}) = \gamma\int_t^\infty p_{j,\tau}\lambda u_{j,\tau}e^{-\int_t^\tau(p_{j,s}\lambda u_{j,s}+r)ds}d\tau. \tag{7}$$

On the equilibrium path, agent i knows that j will follow the same investment rule when he face the problem of maximizing $\Pi^S(p_{j,t})$, that is $u_{j,t} = 1$ iff $T_j^S + t \geq \tau$. The expected discounted externality can be rewrite as (see Appendix A)

$$\Pi^{DE}(p_{j,t}) = p_{j,t}\bar{\gamma}\left[1-\left(\beta\frac{p_{j,t}}{1-p_{j,t}}\right)^{-(1+\frac{r}{\lambda})}\right]$$

where $\bar{\gamma} = \gamma/(1 + r/\lambda)$.

Now we can write the problem of agent i when j had no success yet[13]

$$\Pi_i^E = \max_{u_{i,t}} \int_0^\infty \left\{ \left[p_{i,t}\lambda \left(\pi + \Pi^{DE}(p_{j,t}) \right) - c \right] u_{i,t} + p_{j,t}\lambda u_{j,t} \Pi^S(p_{i,t}) \right\} e^{-\int_0^t (p_{i,s}\lambda u_{i,s} + p_{j,s}\lambda u_{j,s} + r) ds} dt \quad (8)$$

subject to

$$\dot{p}_{m,t} = -p_{m,t}(1 - p_{m,t})\lambda u_{m,t} dt \wedge p_{m,0} = \bar{p},$$

for $m = i, j$. To understand Π_i^E note that at every t there are three possibilities; agent i has a breakthrough, agent j has a breakthrough, or no one has breakthrough (see footnote 13). With instantaneous probability $p_{i,t}\lambda$ agent i has a breakthrough on his project receiving an expected payoff composed by the payoff of the project π plus the discounted expected externality $\Pi^{DE}(p_{j,t})$. On the other hand, with instantaneous probability $p_{j,t}\lambda u_{j,t}$ agent j has a breakthrough and agent i faces the problem $\Pi^S(p_{i,t})$. If no one has a breakthrough the agent reaps no benefits. The probability that no one had a breakthrough up to time t is $\exp\{-\int_0^t (p_{i,s}\lambda u_{i,s} + p_{j,s}\lambda u_{j,s}) ds\}$.

6 Cooperative solution

The amount of effort that would be chosen cooperatively by agents (or by a planner) is the one that maximizes the sum of individual payoffs,

$$W(\bar{p}) := \sum_I \Pi^E = \int_0^\infty \left\{ \left[p_{i,t}\lambda \left(\pi + \Pi^{DE}(p_{j,t}) + \Pi^S(p_{j,t}) \right) - c \right] u_{i,t} \right.$$
$$\left. + \left[p_{j,t}\lambda \left(\pi + \Pi^{DE}(p_{j,t}) + \Pi^S(p_{j,t}) \right) - c \right] u_{j,t} \right\} e^{-\int_0^t (p_{i,s}\lambda u_{i,s} + p_{j,s}\lambda u_{j,s} + r) ds} dt,$$

subject to

$$\dot{p}_{m,t} = -p_{m,t}(1 - p_{m,t})\lambda u_{m,t} dt \wedge p_{m,0} = \bar{p},$$

for $m = i, j$. Replacing (5) and (7) in the objective function, we can rewrite the objective

[13] As is standard in the exponential bandits literature we ignore terms of $o(dt)$, this implies that the probability of both agents having a breakthrough in an interval dt is zero.

function as

$$W(\bar{p}) = \int_0^\infty \left\{ \left[p_{i,t}\lambda\left(\pi + \Pi^{SC}(p_{j,t})\right) - c \right] u_{i,t} \right.$$
$$\left. + \left[p_{j,t}\lambda\left(\pi + \Pi^{SC}(p_{i,t})\right) - c \right] u_{j,t} \right\} e^{-\int_0^t (p_{i,s}\lambda u_{i,s} + p_{j,s}\lambda u_{j,s} + r) ds} dt, \tag{9}$$

where,

$$\Pi^{SC}(p_{m,t}) := \max_{u_{m,\tau}} \int_t^\infty [p_{m,\tau}\lambda(\pi + 2\gamma) - c] u_{m,\tau} e^{-\int_t^\tau (p_{m,s}\lambda u_{m,s} + r) ds} d\tau$$

Note that this specification gives a straightforward interpretation. Cooperative agents consider that a breakthrough on any risky arm yield a payoff π plus Π^{SC}, the later is the (social) expected payoff of a breakthrough on the other arm. That is, they internalize that a second breakthrough implies twice the externality, which is clear since the only difference between Π^{SC} and Π^{S} is that the externality is multiplied by two. But they also consider the costs of experimentation, instead of Π^{DE} that only considers the expected externality. The following Proposition characterizes the cooperative solution, the proof is left to Appendix B.

Proposition 1 *The cooperative solution while no agent is successful is to set*

$$u_{i,t}^C = \begin{cases} 1 & \text{if } t \leq T^C, \\ 0 & \text{if } t > T^C, \end{cases}$$

where T^C (at which a belief p^c is reached) has no analytical solution but can be calculated from an implicit function. The cooperative solution implies maximal effort for more time than in autarky, $T^C > T^A$ for every set of parameters.

If there is a breakthrough in some arm at $t \leq T^C$. The cooperative solution for the unsuccessful agent is to set

$$u_\tau^{SC} = \begin{cases} 1 & \text{if } \tau \leq T^{SC} + t \\ 0 & \text{if } \tau > T^{SC} + t \end{cases}$$

where,

$$T^{SC} := \lambda^{-1} \left[ln\left(\frac{\lambda(\pi + 2\gamma) - c}{c} \right) - ln\left(\frac{1 - p_t}{p_t} \right) \right]$$

After a first breakthrough the cooperative solution implies maximal effort from the unsuccessful agent for more time than in the non-cooperative solution, $T^{SC} > T^S$ for every *set of parameters.*

This is not a surprising result because of the definition of an externality. In absence of a breakthrough agents would exert maximal effort for more time than in autarky since the expected value of their projects is greater because of the externality. Also, after a first breakthrough the unsuccessful agent would exert maximal effort for more time than in the non cooperative solution since he considers not only the increment in the expected value of his project because of the first breakthrough, but also the externality that a breakthrough on his arm will produce on the successful agent.

7 Non-cooperative equilibrium

In the symmetric non-cooperative equilibrium each agent decides an effort profile (function) in order to maximize (8). Since effort is not observed agents share a common belief only on the equilibrium path. This makes difficult to characterize the equilibrium for any arbitrary history since agent's best-response depends both on public and private beliefs. For this reason our proof relies on Pontryagin's principle where other agent's strategies can be treated as fixed, which simplifies the analysis.[14] The proof is presented in Appendix C.

In order to simplify the analysis we also assume that agents are sufficiently patient. Nevertheless, violation of this assumption does not changes the qualitative characteristics of the equilibrium for a large set of parameters.[15]

Assumption 2 *Agents are sufficiently patient. In particular,* $\lambda > r$.

The following proposition characterizes the unique symmetric equilibrium of the game. Despite the simplicity of our setting it is not possible to obtain analytical solutions to characterize the equilibrium. Therefore, we characterize the optimal path of our optimal control problem by an implicit function of $x^*_{i,t} = ln((1 - p^*_{i,t})/p^*_{i,t})$, and obtain numerical solutions for equilibrium effort.

Proposition 2 *There exist a unique symmetric equilibrium in which agents exert effort*

[14] In the strategic experimentation literature is common to use Dynamic Programming tools to characterize the equilibrium. Nevertheless, this tools are difficult to apply to settings in which public and private beliefs might differ because optimality equations are partial differential equations. Therefore, we follow the approach of Bonatti and Hörner (2011) to solve the equilibrium.

[15] This is because the effect of patience is of second order in the equilibrium. Codes are available at Appendix D to calculate the equilibria for every combination of parameters.

according to

$$u^*_{i,t} = \begin{cases} 1 & \text{if } p_t > \hat{p} \\ \dot{x}_{i,t}/\lambda & \text{if } p_t \leq \hat{p} \end{cases}$$

where $\hat{p}$ is the belief at which the upper bound on effort is not binding. That is agents exert maximal effort until their beliefs reach $\hat{p}$, and interior levels thereafter. The function $x^*_{i,t}$ is characterized by the solution of the following non-linear ODE

$$\dot{x}_t = \frac{r\left[e^{-2x_{i,t}}(\pi + \bar{\gamma} - \frac{c}{\lambda}) + e^{-x_{i,t}}\left(\pi - \frac{2c}{\lambda}\right) - e^{-(1-\frac{r}{\lambda})x_{i,t}}\beta^{-(1+\frac{r}{\lambda})}\bar{\gamma} - \frac{c}{\lambda}\right]}{-e^{-2x_{i,t}}\left(\pi + \bar{\gamma} - \frac{c}{\lambda} - \frac{c\beta}{\lambda+r}\right) - e^{-(1-\frac{r}{\lambda})x_{i,t}}\left(\bar{\gamma}\frac{r}{\lambda}\beta^{-(1+\frac{r}{\lambda})} + \frac{c\beta^{-\frac{r}{\lambda}}}{\lambda+r}\right) + e^{-x_{i,t}}\frac{c}{\lambda}} \tag{10}$$

where $\dot{x}_{i,t} = \lambda u_{i,t}$, with initial condition $x_{i,\hat{t}} = ln((1-\hat{p})/\hat{p})$.

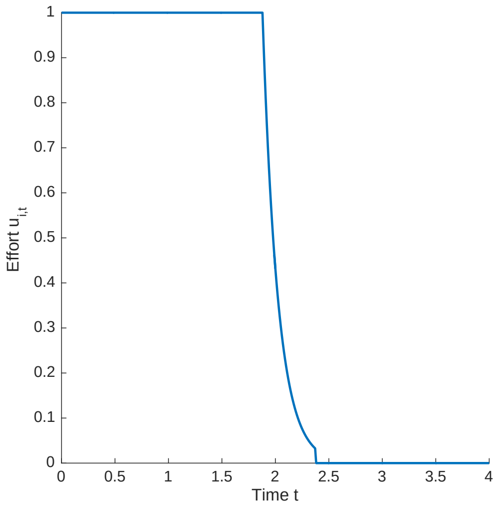
Effort is weakly decreasing, and strictly decreasing in the interior part of the equilibrium. A threshold belief $\underline{p} < \hat{p}$ could exist, if this threshold is reached at some $\underline{t}$ agents exert zero effort for all $t \geq \underline{t}$. If $\underline{p}$ does not exists or is never reached, agents exert positive effort forever.

A numerical example of equilibrium is presented in Figure 1.[16] The following statements are conclusions of numerical analyses of the equilibrium for several combinations of parameters.[17] All results can be replicated for any other set of parameters using codes available at Appendix D.

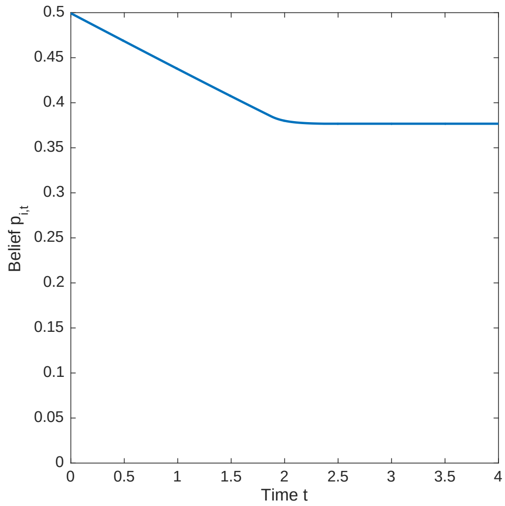
The symmetric equilibrium is qualitatively robust to any set of parameters. If agents have a high enough prior $\bar{p} > \hat{p}$, they exert maximal effort until some $\hat{t}$ at which $p_t = \hat{p}$. For $p \in (\underline{p}, \hat{p})$ they exert interior levels of effort until time $\underline{t}$ at which $p_t = \underline{p}$. For every $t > \underline{t}$, agents exert no effort. In the numerical exercises we have not found a set of parameters such that $\underline{p}$ does not exists. Clearly this not mean that $\underline{p}$ exists for every set of parameters.

[16]Parameters are normalized such that $\pi = 1$.

[17]The interior equilibrium effort was calculated by iterations of an integral equation (a transformation of the ODE into an integral equation). As can be noted in all figures there is a "kink" in the interior equilibrium effort. That is because a lot of computational precision is needed to calculate low levels of effort. The actual equilibrium involve a "fast" but continuous drop of effort until zero. Is possible to solve the ODE numerically, this allows us to obtain the equilibrium effort for more time but with a lot of noise. Codes are provided in Appendix D to replicate both ways, the results are qualitatively the same.



(a) Effort



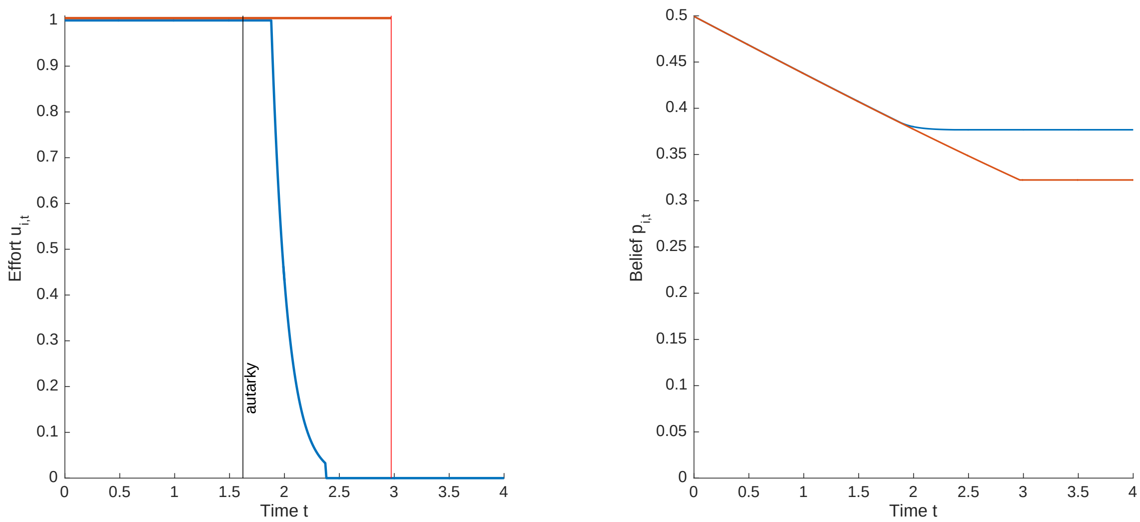
(b) Belief evolution

Figure 1: Non-cooperative equilibrium effort for parameters $(\pi, \gamma, c, \lambda, r, \bar{p}) = (1, 0.5, 0.1, 0.25, 0.2, 0.5)$

A comparison between non-cooperative and cooperative solution is presented in Figure 2. Agents provide maximal effort for more time than in autarky but for strictly less time than in the cooperative solution, $\hat{t} \in (T^A, T^C)$. For beliefs $p_t \in (\underline{p}, \hat{p})$ agents provide positive but inefficient levels of effort.[18] Finally, agents stop experimentation inefficiently early, $\underline{p} > p^c$.

The result implies that efficiency losses are critically linked to the initial prior $\bar{p}$. If agents share low enough priors $\bar{p} \in (\underline{p}, \hat{p})$, they internalize (qualitatively) very little of the externality. Even more, there are some beliefs $\bar{p} \in (p^c, \underline{p})$ at which in the cooperative solution it is efficient to exert maximal effort but in the non-cooperative solution agents do not exert any effort.

[18]We refer to inefficient experimentation to interior levels of experimentation. Hypothetically, imagine that agents can commit to behave cooperatively but with a restriction over aggregate effort $U = \int_0^\infty (u_{i,t} + u_{j,t})dt$, which is less or equal than aggregate effort of the cooperative solution. In that case, agents will choose to exert maximal effort until they reach U.



(a) Effort

(b) Belief evolution

Figure 2: Comparison of cooperative (Red) solution and non-cooperative (Blue) equilibrium for parameters $(\pi, \gamma, c, \lambda, r, \bar{p}) = (1, 0.5, 0.1, 0.25, 0.2, 0.5)$

Why has this game a unique symmetric equilibrium despite having the *taste* of a usual coordination game with multiplicity of equilibria? Because of its dynamic structure. Let's use as an example the two usual extreme equilibria of a coordination game: both agents exert maximal effort until the autarky threshold belief p^a and zero thereafter, or both exert maximal effort until the cooperative threshold belief p^c and zero thereafter. If agent j follows the first strategy, at p^a is optimal for agent i to exert maximal effort because he knows that a breakthrough on his arm will make optimal for agent j to start experimenting again, which has an expected value $\Pi^{DE}(p_j^a)$ for agent i. We refer to this effect as incentives to experiment *today*. On the other hand, if agent j follows the later strategy agent i becomes more pessimistic about his and agent j's arm quality over time. Therefore, when agent i's beliefs reaches some threshold belief $\tilde{p}$ he considers optimal to *wait and see* if j has success. This is because the value of $\Pi^{DE}(\tilde{p}_j)$ is not enough to compensate the costs of experimentation, so he prefers to maintain his beliefs at $\tilde{p}_i$ and start experimenting again if only if j has a breakthrough. The combination of this forces is what makes agents to *coordinate their experimentation*, "refining" the equilibrium.

The learning component gives the non-cooperative equilibrium its shape by changing the incentives to experiment more *today* and to *wait and see* over time. For high enough beliefs p_t, the incentives to experiment *today* makes agents to exert maximal effort because they know that if they have a breakthrough the other agent will continue his experimentation. Nevertheless, as agents becomes more pessimistic about the quality of projects the weaker the incentives to experiment *today*, and the more tempting is to *wait and see*. The interior part of the equilibrium arises because both agents wants to *wait and see* but they need to provide incentives to the other agent to keep experimenting.

Finally, we are going to refer to off the equilibrium path behavior.[19] Despite there being infinite types of deviations, the only thing that is relevant for a deviated agent at time t is the aggregate effort that he exerted in $[0, t)$. This is because if aggregate effort of a deviated agent is lower (higher) than it would have been on the equilibrium path, he is more (less) optimistic about the quality of his arm than the other agent.[20] If the aggregate effort is lower, the deviated agent is more optimistic about the quality of his arm than the other. Therefore, if the deviated agent is more optimistic than the other agent, he will provide maximal effort until his private belief catches up with the other agent's belief. On the other hand, if the agent is more pessimistic he will provide zero effort until the other agent becomes as pessimistic as him about the quality of his arm.

8 Numerical exercises

In this section we study the sensibility of cooperative and non-cooperative equilibrium to changes in two key parameters; externality γ and patience r. Since we are interested in the effects of the externality, we are going to focus in the effort provided by agents with an initial prior $\bar{p} = p^a$. That allows us to study the changes in total and aggregate effort that would have not been provided in absence of externalities. Also, since after a first breakthrough the cooperative solution always implies more aggregate effort than the non-cooperative equilibrium (See Proposition 1), we are going to focus on effort provision before a first breakthrough to make exposition more clear. Finally, we define efficiency loss (EL) as

$$\text{EL} = \frac{\int_0^\infty (u_t^C - u_t^{NC})dt}{\int_0^\infty u_t^C dt},$$

where $u_t = u_{i,t} + u_{j,t}$, and the superscripts indicate cooperative (C) and non-cooperative (NC) solutions.

All codes to replicate results for any set of parameters are available in Appendix D.

8.1 Effect of externality

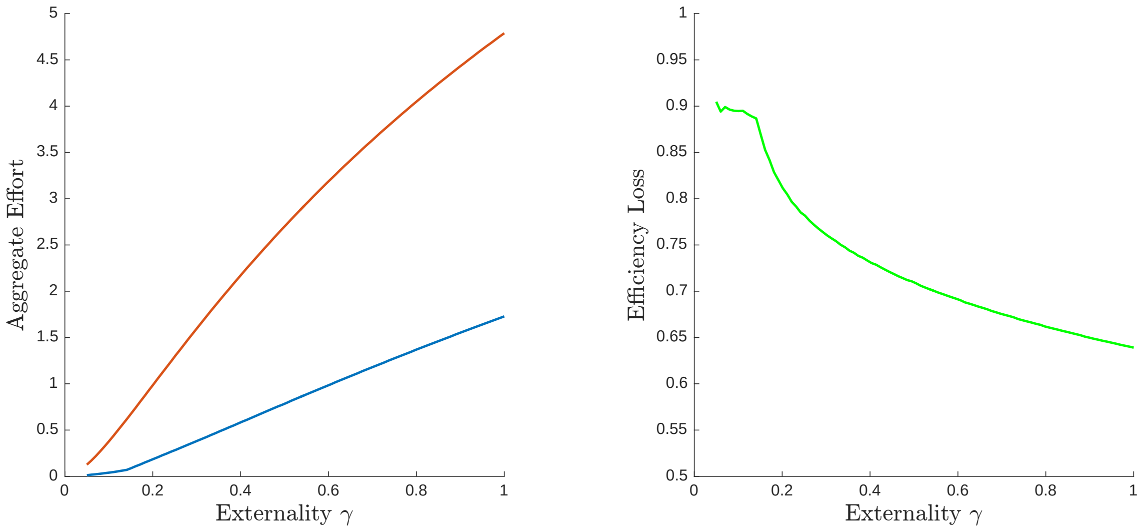
The externality γ has a direct effect on the cooperative solution, since $\partial \Pi^{SC}(p_t)/\partial \gamma > 0$ is easy to note that the integrand of (9) increases, and cooperative agents are willing to experiment for more time. Nevertheless, the effect on the non-cooperative

[19] This is almost direct from Bonatti and Hörner (2011).

[20] Clearly, for $t \in [0, \hat{t}]$ the only possible deviation is to provide less aggregate effort. Also, if $\underline{p}$ exists, a deviated agent that has provided too much effort until t with beliefs $p_t \leq \underline{p}$ will not provide any effort again.

equilibrium effort is not clear. On one hand, since $\partial\Pi^{DE}(p_t)/\partial\gamma > 0$ there is an incentive to provide maximal effort for more time. On the other hand, $\partial\Pi^{S}(p_t)/\partial\gamma > 0$ so there is also more incentives to *wait and see*.

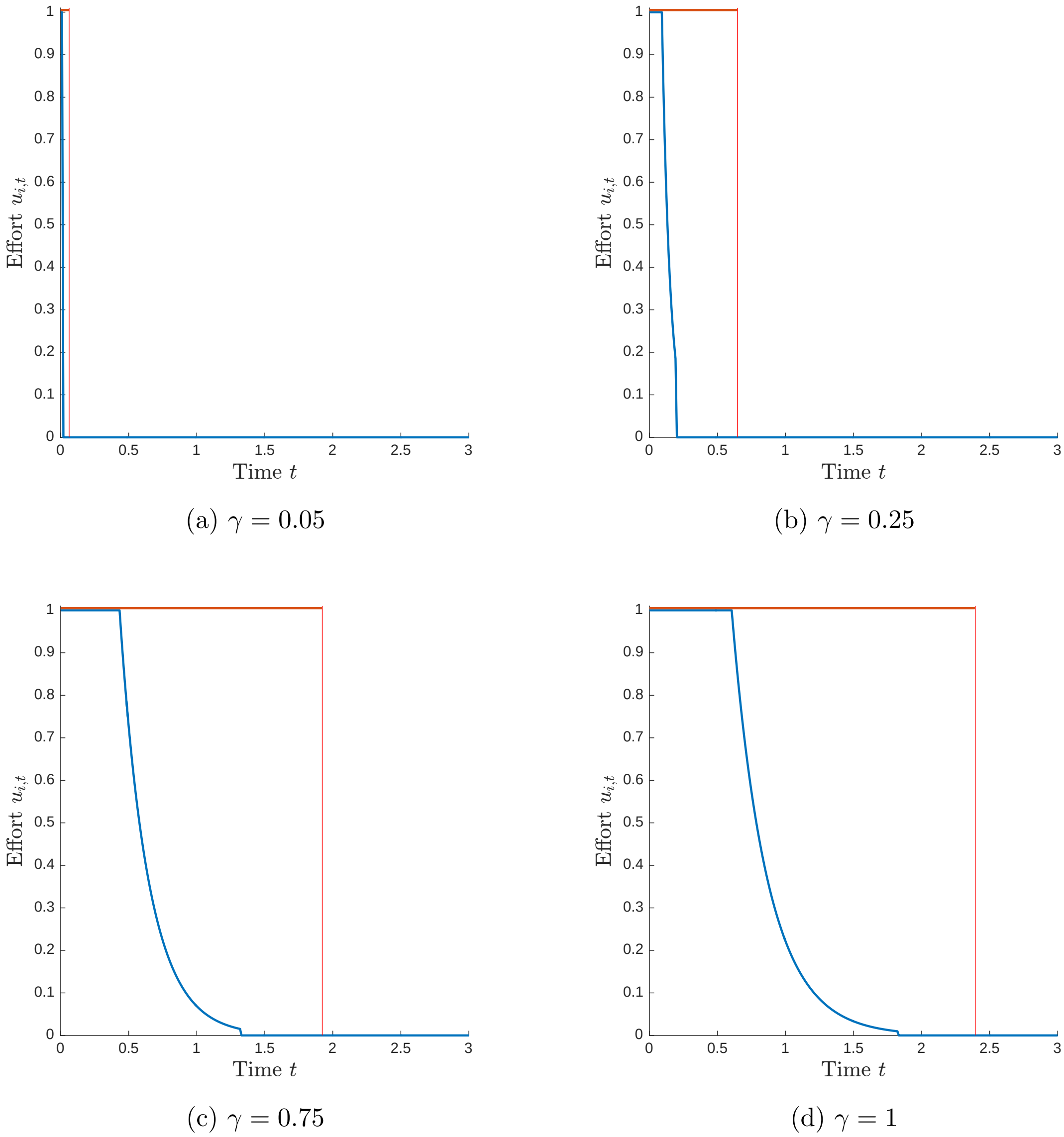
In Figure 3 we show changes in aggregate effort in response to changes in γ. As before, parameters are normalized such that $\pi = 1$. Therefore, γ must be interpreted as a proportion of π (i.e. $\gamma = 0.1$ implies that the externality represents a 10% of π).



(a) (b)

Figure 3: Sensibility of cooperative (red) and non-cooperative (blue) equilibrium to externality. Parameters $(\pi, c, \lambda, r, p^a) = (1, 0.1, 0.25, 0.2, 0.4)$

From Figure 3 is clear that both non-cooperative and cooperative aggregate effort are increasing in γ, while efficiency loss is almost decreasing in γ (for low levels of γ it is not necessarily decreasing). This mean that in the non cooperative equilibrium the incentives to experiment more *today* dominates the effect of *wait and see*. Also, as γ increases agents exert interior equilibrium effort for more time (see Figure 4). That is, despite that *wait and see* incentives makes agents to exert inefficient levels of effort, they want to keep experimentation because of higher payoffs.



(a) $\gamma = 0.05$

(b) $\gamma = 0.25$

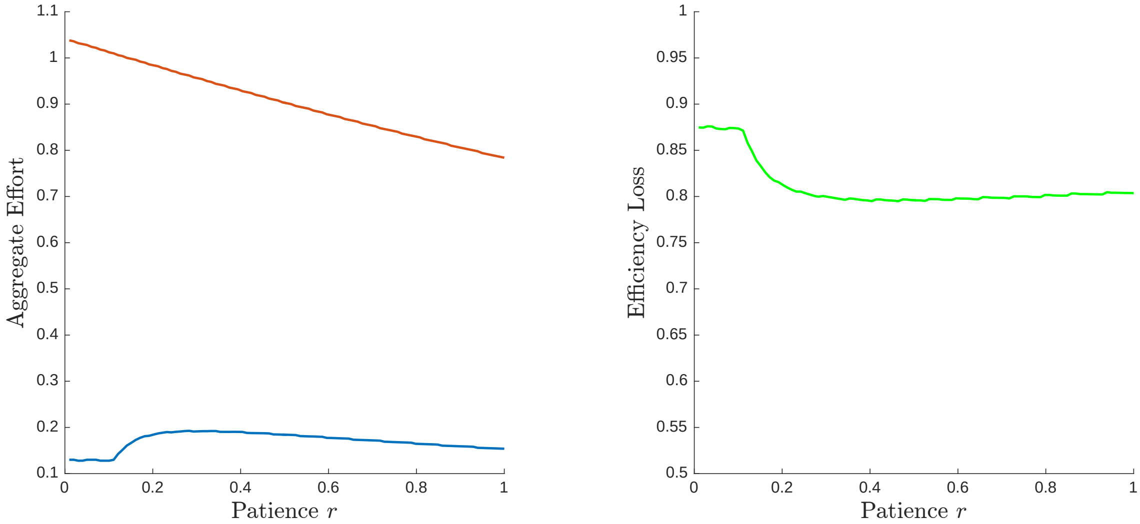
(c) $\gamma = 0.75$

(d) $\gamma = 1$

Figure 4: Examples of individual cooperative (red) and non-cooperative (blue) equilibrium effort for different levels of externality. Parameters $(\pi, c, \lambda, r, p^a) = (1, 0.1, 0.25, 0.2, 0.4)$

8.2 Effect of patience

The effect of patience r on the cooperative equilibrium is direct. The more impatient the agents are, the less valuable is a first breakthrough $\partial \Pi^{SC}(p_t)/\partial r < 0$, so they are willing to experiment for less time. As before, effect of patience on the non-cooperative equilibrium is not clear. On one hand, if agents are more impatient they value less the expected discounted externality $\partial \Pi^{DE}(p_t)/\partial r < 0$, which reduce incentives to experiment *today*. On the other hand, if agents are more impatient is less profitable to *wait and see* $\partial \Pi^{S}(p_t)/\partial r < 0$ since they want to have a breakthrough as soon as possible. In Figure 5 we show changes in aggregate effort in response to changes in r.[21]



(a) (b)

Figure 5: Sensibility of cooperative (red) and non-cooperative (blue) equilibrium to patience. Parameters $(\pi, \gamma, c, \lambda, p^a) = (1, 0.2, 0.1, 0.25, 0.4)$

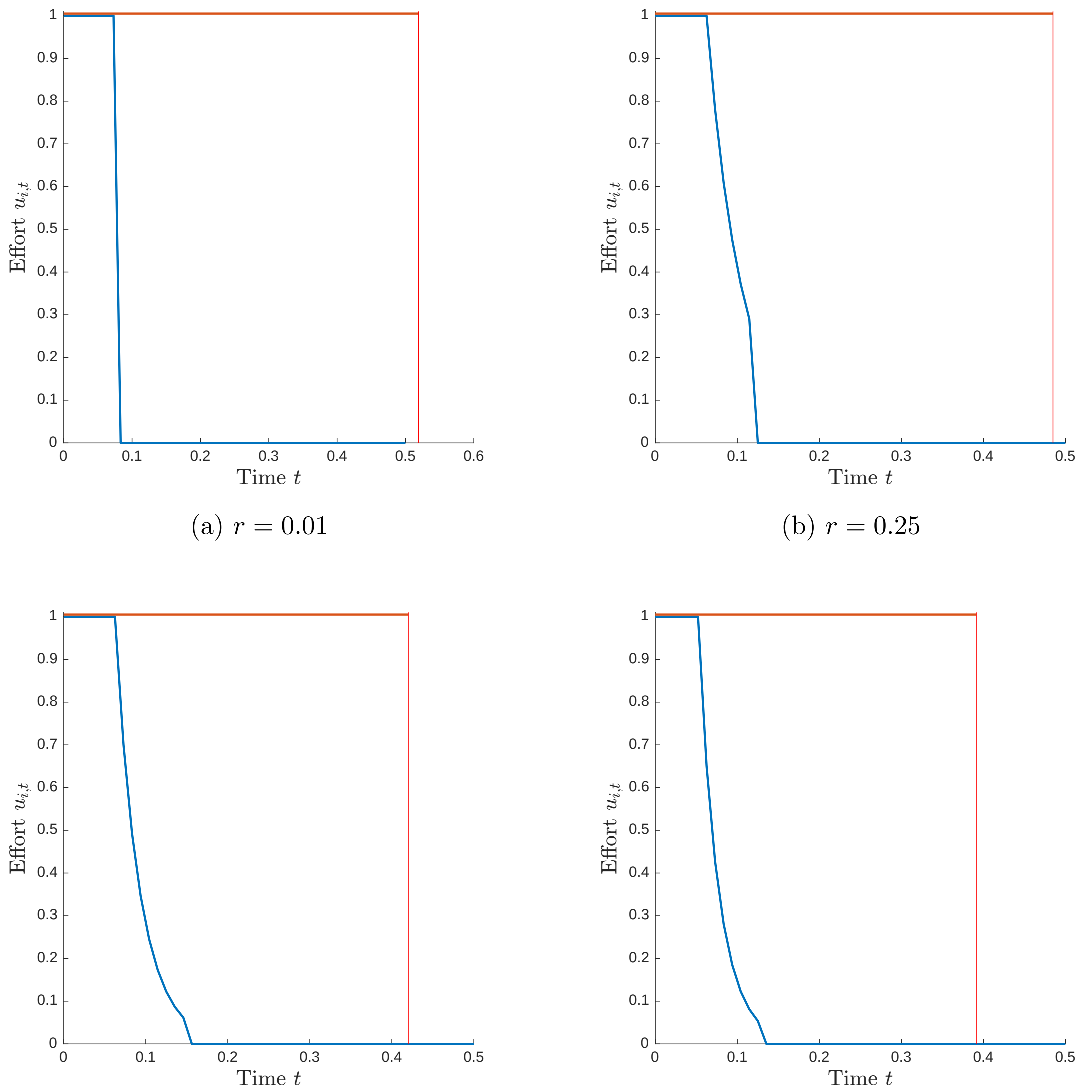
As discussed above, aggregate effort is decreasing in r in the cooperative solution. But the response of non-cooperative equilibrium aggregate effort is not trivial. Figure 6 show some numerical examples that makes easier to understand the following argument.

When agents are too patient $r \to 0$ (see panel (a) of Figure 6), the non-cooperative solution tends to a bang-bang solution, which makes aggregate effort very low. This is because *wait and see* incentives becomes too strong very soon, so once the expected discounted externality is low enough they just stop experimentation. Then, if agents are a little impatient (see panel (b)) they still value the discounted externality enough but *wait and see* incentives becomes weaker, so agents provide interior levels of effort for a while making aggregate effort to increase. Nevertheless, if agents are too impatient (see panel (c) and (d)) they have too little incentives to experiment *today* (Π^{DE} is too low) so the aggregate effort start to decrease again.

Note that efficiency losses have almost a U-shaped form. That means that an intermediate level of impatience makes the non-cooperative equilibrium more efficient. This goes against a "Folk theorem" intuition, which is is not surprising since a well known result in repeated games literature is that when we restrict to strong symmetric equilibria (in this case MPBE) it is not possible to obtain a Folk Theorem. This is because agents can not condition their strategies on any history, or in other words the set of possible punishments is very coarse (See Mailath and Samuelson, 2006, Ch. 9). Nevertheless, this

[21] Note that Assumption 2 is violated for some levels of r in this section. As commented above, violation of that assumption does not change the qualitative characteristics of the equilibrium.

result is still interesting because it departs from the literature. For example, Bonatti and Hörner (2011) who studies free-riding in teams in a context of experimentation, finds that aggregate effort is increasing in r. Our result could give some insights about the non trivial role of patience when we address the coordination problem in an experimentation setting.



(a) $r = 0.01$ (b) $r = 0.25$

(c) $r = 0.75$ (d) $r = 1$

Figure 6: Examples of individual cooperative (red) and non-cooperative (blue) equilibrium effort for different levels of patience. Parameters $(\pi, \gamma, c, \lambda, r, p^a) = (1, 0.2, 0.1, 0.25, 0.4)$

9 Discussion

Our simple model provide some insights about the problem of coordination in a dynamic strategic experimentation context. Nevertheless, in order to provide a simple model we made some assumptions that could be guiding our results. In this section we discuss how our results could vary if we relax some key assumptions.

Uniqueness and payoff functional form. We assumed that the independent payoff (π) and the externality (γ or Π^{DE}) are additively separable. This generates a dynamic effect that could be crucial to obtain a unique equilibrium. In absence of a breakthrough complementarity comes from the fact that if agent i has a breakthrough agent j will follow the investment rule described in Corollary 1, this can be understood as "ex-post" complementarities. At the same time, strategies are not perfect complements. The force that captures strategic substitutability is the *wait and see* incentive. Actually, the interior part of the equilibrium is shaped by the interaction of this two forces; agents want to *wait and see* while the other agent experiment (substitutability) but they need to provide some effort to keep the other agent experimenting (complementarities).

Therefore, uniqueness in our model could exist because complementarities are not too strong. Then, more general payoff functional forms could yield to multiplicity of equilibria. For example, imagine that the externalities come as synergies such that the payoff is $\pi \times \gamma$ if both are successful, and zero otherwise. This makes complementarities stronger because now agents receive a positive payoff iff both are successful. Nevertheless, *wait and see* incentives does not disappear. They might be even stronger because experimentation is more risky since receiving a positive payoff is less likely to happen, so the costs of experimentation are relatively higher. Is not clear if there exist a unique symmetric equilibrium with such payoff functions, but the insights provided by our model make us think that uniqueness depends on a balance between complementarities and substitutability forces. Furthermore, insights from coordination games literature suggest that the uniqueness in settings with stronger complementarities will be possible to obtain only for low enough γ_i (Bergemann and Morris, 2012). The general payoff functions that make possible to obtain a unique symmetric equilibrium is an interesting research agenda.

More agents. Our model is restricted to two agents since the non linear structure of sub-games' payoffs makes impossible to obtain analytical solutions for the non-cooperative equilibrium. Therefore, is very hard to characterize the sub-games for games with more than two agents.[22] Despite that is possible to make simulations for equilibrium with more agents, is very hard to obtain uniqueness or sufficiency results, which are the interesting results.

Is not clear what force becomes stronger with the inclusion of more agents. On one hand, the expected externality is higher so incentives to experiment *today* should be stronger. On the other hand, incentives to *wait and see* could become stronger too because each agent is not the only that incentivizes the other to experiment. Therefore, as $N \to \infty$ we could find more efficient or inefficient equilibria. To study under what conditions non-cooperative equilibria becomes more (or less) efficient as there are more

[22]Note that in a model with three agents, the non-cooperative equilibrium of our model is the sub-game that two unsuccessful agents play after a first breakthrough.

agents is a relevant question not only from a theoretical perspective but also from a policy perspective.

Commitment. We have focussed on two extreme solutions. A cooperative solution where agents have perfect commitment, and the non-cooperative equilibrium where agents can't commit to anything. Is interesting to study efficiency gains from intermediate levels of commitment. For example, to commit to deadlines or wages.

Following Bonatti and Hörner (2011) is possible that deadlines could make experimentation more efficient, by compromising to a deadline agents could exert the interior levels of effort in a more efficient way (i.e. provide maximal effort for less time). Also, commitment to wages is an interesting mechanism science is not direct from Bonatti and Hörner (2011). Note that a successful agent has willingness to pay to the other agent to exert effort for more time than T^S. Hypothetically, at T^S the unsuccessful agent could offer to the successful agent to exert more effort for a time interval $[T^S, T']$ in exchange for the total value of the expected discounted externality that is generated by that effort profile with initial prior $p_{j,T'}$. Since the successful agent is risk neutral, he would be indifferent between accepting the offer or rejecting it. This intuition suggest that if agents could commit to wages they could reach more efficient levels of experimentation.

More general structure of news arrival. The exponential bandit literature assumes that breakthroughs arrives according to a Poisson process. This assumprtion provide a tractable stochastic structure that usually make models more easy to solve. Nevertheless, this assumption is what makes the sub-game payoff non linear in our model. Since this is what puts heavy limitations to generalize or extend our model, alternative approaches should be considered.

Külpmann (2015) studies the same setting of Bonatti and Hörner (2011) but without putting almost any restriction over news arrival structure. In particular, he characterize the "breakthrough effort distribution" by the hazard rate structure. That is, how much effort an agent has to spend for a certain chance of success, conditional on the effort that has been spent in the past. Exponential bandits implies a decreasing hazard rate, because as agents provide effort and no breakthrough arrives they become more pessimistic about project quality. Nevertheless, the hazard rate could be increasing if a certain amount of cumulative effort is needed for a breakthrough. For example, imagine that the cumulative effort needed for a breakthrough is uniformly distributed on some interval. Then as more effort was provided in the past more likely is that the agent reaches that threshold today.

Using the hazard rate instead of a particular distribution could make possible to obtain analytical solutions, which would make possible to relax some of the critical assumptions described in this section.

10 Concluding remarks

We considered a simple model of strategic experimentation with externalities on risky arm's payoffs and unobservable actions. We have shown that that the combination of a dynamic structure and learning lead to a unique symmetric MPBE, such that agents coordinate their experimentation.

In equilibrium agents coordinate in a way that they exert too little effort in a dynamically inefficient way, relative to the first best. The magnitude of efficiency losses is mainly defined by the common prior about projects' quality, the lower the prior the higher the efficiency losses. Even more, for low enough priors agents exert zero effort even if it is socially efficient to provide maximal effort. Also, efficiency losses are (almost) decreasing in the level of the externality, and are minimized for interior levels of patience.

The main message of this work is that the dynamic structure and the learning component in experimentation coordination games could lead to equilibrium uniqueness, making unnecessary to choose an equilibrium refinement criteria. Our model has many limitations, it is only a first step in order to understand the problem of experimentation coordination. Nevertheless, it provide several opportunities for future research. An interesting research agenda is to study if the main result holds for more general payoff functions or for an arbitrary number of agents.

References

Bergemann, D. and Morris, S. (2012). *Robust mechanism design: The role of private information and higher order beliefs*, volume 2. World Scientific.

Bergemann, D. and Välimäki, J. (2008). Bandit problems. In Durlauf, S. N. and Blume, L. E., editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, Basingstoke.

Bessen, J. and Maskin, E. (2009). Sequential innovation, patents, and imitation. *The RAND Journal of Economics*, 40(4):611–635.

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

Bonatti, A. and Hörner, J. (2011). Collaborating. *American Economic Review*, 101(2):632–663.

Chen, Y. (2012). Innovation frequency of durable complementary goods. *Journal of Mathematical Economics*, 48(6):407–421.

Georgiadis, G. (2015). Projects and team dynamics. *Review of Economic Studies*, 82(1):187.

Griliches, Z. (1992). The Search for R&D Spillovers. *The Scandinavian Journal of Economics*, pages S29–S47.

Heidhues, P., Rady, S., and Strack, P. (2015). Strategic experimentation with private payoffs. *Journal of Economic Theory*, 159:531–551.

Kamihigashi, T. (2001). Necessity of transversality conditions for infinite horizon problems. *Econometrica*, 69(4):995–1012.

Keller, G., Rady, S., and Cripps, M. (2005). Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68.

Klein, N. and Rady, S. (2011). Negatively correlated bandits. *Review of Economic Studies*, page rdq025.

Külpmann, P. (2015). Procrastination and projects.

Mailath, G. J. and Samuelson, L. (2006). *Repeated games and reputations: long-run relationships*. Oxford university press.

Malerba, F., Mancusi, M. L., and Montobbio, F. (2013). Innovation, international R&D spillovers and the sectoral heterogeneity of knowledge flows. *Review of World Economics*, 149(4):697–722.

Moroni, S. (2015). Experimentation in organizations. *Manuscript, Department of Economics, University of Pittsburgh.*

Park, J. (2004). International and intersectoral R&D spillovers in the OECD and East Asian economies. *Economic Inquiry*, 42(4):739–757.

Salish, M. (2015). Innovation, intersectoral R&D spillovers and intellectual property rights. Working paper.

Seierstad, A. and Sydsaeter, K. (1986). *Optimal control theory with economic applications.* Elsevier North-Holland, Inc.

Shan, Y. (2017). Optimal contracts for research agents. *The RAND Journal of Economics*, 48(1):94–124.

Steurs, G. (1995). Inter-industry R&D spillovers: what difference do they make? *International Journal of Industrial Organization*, 13(2):249–276.

Van Zandt, T. and Vives, X. (2007). Monotone equilibria in bayesian games of strategic complementarities. *Journal of Economic Theory*, 134(1):339–360.

Wiseman, T. (2005). A partial folk theorem for games with unknown payoff distributions. *Econometrica*, 73(2):629–645.

Appendix

A Calculating continuation payoffs

A.1 Expression for $\Pi^S(p_{j,t})$

We are going to derive an expression for

$$\Pi^S = \max_{u_{i,\tau}} \int_t^\infty [p_{i,\tau}\lambda(\pi+\gamma) - c] u_{i,\tau} e^{-\int_t^\tau (p_{i,s}\lambda u_{i,s} + r) ds} d\tau$$
$$\text{s.t. } \dot{p}_{i,\tau} = -p_{i,\tau}(1-p_{i,\tau})\lambda u_{i,\tau}$$

First, note that using $\dot{p}_\tau$, with a little abuse of notation

$$e^{-\int_\tau^{\tau'} p_s \lambda u_s ds} = e^{-\int_{p_\tau}^{p_{\tau'}} \frac{dp}{(1-p_s)}} = \frac{1-p_\tau}{1-p_{\tau'}}$$

using Corollary 1 we know that $u_\tau = 1$ until T^S and 0 otherwise. W.l.g. we can change the integration interval from $[t, \infty]$ to $[0, T^S]$, with $p_{i,0} = p_{i,t}$. We can rewrite the objective function as

$$c \int_0^{T^S} (p_{i,\tau}\lambda \frac{\pi+\gamma}{c} - 1) \frac{1-p_{i,t}}{1-p_{i,\tau}} e^{-r\tau} d\tau$$

Note that

$$\frac{1-p_{i,t}}{1-p_{i,\tau}} = \frac{1-p_{i,t}}{1-\frac{p_{i,t}}{p_{i,t}+(1-p_{i,t})e^{\lambda\tau}}} = p_{i,t}e^{-\lambda\tau} + (1-p_{i,t})$$

Then,

$$
\begin{aligned}
& c\int_0^{T^S}(p_{i,t}e^{-\lambda\tau}\lambda\tfrac{\pi+\gamma}{c} - p_{i,t}e^{-\lambda\tau} - (1-p_{i,t}))e^{-r\tau}d\tau \\
\Leftrightarrow\quad & c\int_0^{T^S}(p_{i,t}e^{-\lambda\tau}(\tfrac{\lambda(\pi+\gamma)}{c} - 1) - (1-p_{i,t}))e^{-r\tau}d\tau \\
\Leftrightarrow\quad & c\left\{p_{i,t}\left(\tfrac{\lambda(\pi+\gamma)-c}{c}\right)\left[-\tfrac{e^{-(\lambda+r)\tau}}{\lambda+r}\right]_0^{T^S} - (1-p_{i,t})\left[-\tfrac{e^{-r\tau}}{r}\right]_0^{T^S}\right\} \\
\Leftrightarrow\quad & c\left\{\tfrac{p_{i,t}}{\lambda+r}\left(\tfrac{\lambda(\pi+\gamma)-c}{c}\right)\left[1-e^{-(\lambda+r)T^S}\right] - \tfrac{(1-p_{i,t})}{r}\left[1-e^{-rT^S}\right]\right\}
\end{aligned}
$$

Note that

$$
e^{-aT^S} = \left(\frac{\lambda(\pi+\gamma)-c}{c}\frac{p_{i,t}}{1-p_{i,t}}\right)^{-a/\lambda}
$$

then,

$$
\begin{aligned}
c\Bigg\{ & \frac{p_{i,t}}{\lambda+r}\frac{\lambda(\pi+\gamma)-c}{c}\left[1-\left(\frac{\lambda(\pi+\gamma)-c}{c}\frac{p_{i,t}}{1-p_{i,t}}\right)^{-(1+\frac{r}{\lambda})}\right] \\
& - \frac{(1-p_{i,t})}{r}\left[1-\left(\frac{\lambda(\pi+\gamma)-c}{c}\frac{p_{i,t}}{1-p_{i,t}}\right)^{-\frac{r}{\lambda}}\right]\Bigg\}
\end{aligned}
$$

Define $\beta := \frac{\lambda(\pi+\gamma)-c}{c}$. Finally,

$$
\Pi^S(p_{i,t}) := c\left\{\frac{p_{i,t}}{\lambda+r}\left[\beta-\beta^{-\frac{r}{\lambda}}\left(\frac{p_{i,t}}{1-p_{i,t}}\right)^{-(1+\frac{r}{\lambda})}\right] - \frac{(1-p_{i,t})}{r}\left[1-\left(\beta\frac{p_{i,t}}{1-p_{i,t}}\right)^{-\frac{r}{\lambda}}\right]\right\}
$$

A.2 Expression for $\Pi^{DE}(p_{j,t})$

Now, we are going to find an expression for

$$
\Pi_i^{DE}(p_{j,t}) := \gamma\int_t^{\infty} p_{j,\tau}\lambda u_{j,\tau}e^{-\int_t^{\tau}(p_{j,s}\lambda u_{j,s}+r)ds}d\tau
$$

using the same techniques and notation as before

$$\begin{aligned} & \lambda\gamma \int_0^{T^S} p_{j,t} e^{-(\lambda+r)\tau} d\tau \\ \Leftrightarrow \quad & \tfrac{\lambda\gamma p_{j,t}}{\lambda+r}(1-e^{-(\lambda+r)T^S}) \\ \Leftrightarrow \quad & p_{j,t}\tfrac{\gamma}{1+\frac{r}{\lambda}}\left[1-\beta^{-(1+\frac{r}{\lambda})}\left(\tfrac{p_{j,t}}{1-p_{j,t}}\right)^{-(1+\frac{r}{\lambda})}\right] \end{aligned}$$

Define $\bar{\gamma} := \gamma/(1+r/\lambda)$. Finally,

$$\Pi_i^{DE}(p_{j,t}) := p_{j,t}\bar{\gamma}\left[1-\left(\beta\frac{p_{j,t}}{1-p_{j,t}}\right)^{-(1+\frac{r}{\lambda})}\right]$$

B Proof of Proposition 1

The cooperative problem is to maximize

$$W(\bar{p}) = \int_0^\infty \left\{ \left[p_{i,t}\lambda\left(\pi+\Pi^{SP}(p_{j,t})\right)-c\right]u_{i,t} + \left[p_{j,t}\lambda\left(\pi+\Pi^{SP}(p_{i,t})\right)-c\right]u_{j,t}\right\} e^{-\int_0^t(p_{i,s}\lambda u_{i,s}+p_{j,s}\lambda u_{j,s}+r)ds}dt$$

Subject to

$$\begin{aligned} \dot{p}_{m,t} &= -p_{m,t}(1-p_{m,t})\lambda u_{m,t}dt \\ p_{m,0} &= \bar{p} \end{aligned}$$

for $m = i, j$. Where,

$$\Pi^{SP}(p_{m,t}) := \max_{u_{m,\tau}} \int_t^\infty [p_{m,\tau}\lambda(\pi+2\gamma)-c]u_{m,\tau}e^{-\int_t^\tau(p_{m,s}\lambda u_{m,s}+r)ds}d\tau$$

is the expected value of experimentation in one of the projects after a breakthrough in the other project at some arbitrary t. Define $\omega := \frac{\lambda(\pi+2\gamma)-c}{c}$, from Lemma 1 is direct that (ignoring the individual subscript) the cooperative solution of Π^{SP} is to experiment according

$$u_\tau^{SP} = \begin{cases} 1 & \text{if } \tau \leq T^{SP} + t \\ 0 & \text{if } \tau > T^{SP} + t \end{cases}$$

Where,

$$T^{SP} := \lambda^{-1}\left[ln\left(\frac{\lambda(\pi+2\gamma)-c}{c}\right) - ln\left(\frac{1-p_t}{p_t}\right)\right]$$

is direct that $T^{SP} > T^S$ for any t, that is after a breakthrough the cooperative solution involves strictly more effort than the non-cooperative. With the same technics used in Appendix A, we can rewrite

$$\Pi^{SP}(p_t) := c\left\{\frac{p_t}{\lambda+r}\left[\omega - \omega^{-\frac{r}{\lambda}}\left(\frac{p_t}{1-p_t}\right)^{-(1+\frac{r}{\lambda})}\right] - \frac{(1-p_t)}{r}\left[1-\left(\omega\frac{p_t}{1-p_t}\right)^{-\frac{r}{\lambda}}\right]\right\}$$

Now, we can compute the cooperative solution in absence of a breakthrough. Since both agents are symmetric, the solution is to set $u_{i,t} = u_{j,t} = 1$ while

$$p_t\lambda\left(\pi + \Pi^{SP}(p_t)\right) - c \geq 0$$

Since $\dot{p} < 0$ and $\partial\Pi^{SP}/\partial p_t > 0$, $\exists T^C$ such that the equation is satisfied with equality and is optimal to stop experimenting in both projects. Is not possible to obtain an analytical solution for T^C but the computational solution is trivial.

Finally, since $\Pi^{SP} > 0$ it follows that $T^C > T^A$. ∎

C Proof of Proposition 2

This proof relies on the Pontryagin's principle and was inspired by the proof of Theorem 1 of Bonatti and Hörner (2011).

C.1 Preliminaries

Since agent's objective function (8) is complex we are going to simplify the problem before solving it. First, note that we can rewrite $\exp\{-\int_0^t(p_{i,s}\lambda u_{i,s} + p_{j,s}\lambda u_{j,s} + r)ds\} = \frac{(1-\bar{p})^2}{(1-p_{i,t})(1-p_{j,t})}e^{-rt}$ (See Appendix A).

On the other hand, since $\dot{p}_t = -p_t(1-p_t)\lambda u_t$, we have $p_t \lambda u_t = -\dot{p}_t/(1-p_t)$ and $u_t = -\dot{p}_t/p_t(1-p_t)\lambda$. We can rewrite (8) as

$$\int_0^\infty \left\{ \frac{-\dot{p}_{i,t}}{(1-p_{i,t})}(\pi + \Pi^{DE}(p_{j,t})) + \frac{\dot{p}_{i,t}}{p_{i,t}(1-p_{i,t})}\frac{c}{\lambda} + p_{j,t}\lambda u_{j,t}\Pi^S(p_{i,t}) \right\} \frac{(1-\bar{p})^2}{(1-p_{i,t})(1-p_{j,t})} e^{-rt} dt$$

subject to

$$\dot{p}_{m,t} = -p_{m,t}(1-p_{m,t})\lambda u_{m,t} dt$$
$$p_{m,0} = \bar{p}$$

for $m \in \{i,j\}$. Now we are going to work with some terms of the objective function in order to make the optimal control problem easier, we are going to ignore the constants. By integrating by parts the following results:

$$(1-\bar{p})^2 \int_0^\infty \frac{-\dot{p}_{i,t}}{(1-p_{i,t})^2} \frac{\pi}{(1-p_{j,t})} e^{-rt} dt$$

consider,

$$dv = \frac{-\dot{p}_{i,t}}{(1-p_{i,t})^2} \Leftrightarrow \qquad v = -\tfrac{1}{(1-p_{i,t})}$$
$$u = \frac{\pi}{(1-p_{j,t})} e^{-rt} \Leftrightarrow \quad du = -\pi \left[\tfrac{r}{(1-p_{j,t})} + \lambda u_{j,t} \tfrac{p_{j,t}}{(1-p_{j,t})} \right] e^{-rt}$$

then,

$$(1-\bar{p})^2 \int_0^\infty \frac{-\dot{p}_{i,t}}{(1-p_{i,t})^2} \frac{\pi}{(1-p_{j,t})} e^{-rt} dt = C - (1-\bar{p})^2 \int_0^\infty \frac{\pi}{(1-p_{i,t})} \left[\frac{r}{(1-p_{j,t})} + \lambda u_{j,t} \frac{p_{j,t}}{(1-p_{j,t})} \right] e^{-rt}$$

Now we are going to work with the second term

$$(1-\bar{p})^2 \int_0^\infty \frac{-\dot{p}_{i,t}}{(1-p_{i,t})^2} \frac{\Pi^{DE}(p_{j,t})}{(1-p_{j,t})} e^{-rt} dt =$$
$$(1-\bar{p})^2 \int_0^\infty \frac{-\dot{p}_{i,t}}{(1-p_{i,t})^2} \frac{p_{j,t}}{(1-p_{j,t})} \bar{\gamma} \left[1 - \left(\beta \frac{p_{j,t}}{1-p_{j,t}} \right)^{-(1+r/\lambda)} \right] e^{-rt} dt$$

Take,

$$dv = \frac{-\dot{p}_{i,t}}{(1-p_{i,t})^2} \Leftrightarrow \qquad v = -\tfrac{1}{(1-p_{i,t})}$$

$$u = \frac{p_{j,t}}{(1-p_{j,t})}\bar{\gamma}\left[1-\left(\beta\frac{p_{j,t}}{1-p_{j,t}}\right)^{-(1+r/\lambda)}\right]e^{-rt} \Leftrightarrow \quad du = -e^{rt}\bar{\gamma}\left[\tfrac{p_{j,t}}{1-p_{j,t}}(\lambda u_{j,t}+r)+\beta^{-(1+\frac{r}{\lambda})}\left(\tfrac{p_{j,t}}{1-p_{j,t}}\right)^{-\frac{r}{\lambda}}\tfrac{r}{\lambda}(\lambda u_{j,t}-\lambda)\right]$$

Then

$$(1-\bar{p})^2\int_0^\infty \frac{-\dot{p}_{i,t}}{(1-p_{i,t})^2}\frac{\Pi^{DE}(p_{j,t})}{(1-p_{j,t})}e^{-rt}dt =$$

$$C-(1-\bar{p})^2\int_0^\infty \frac{\bar{\gamma}}{(1-p_{i,t})}\left[\frac{p_{j,t}}{1-p_{j,t}}(\lambda u_{j,t}+r)+\beta^{-(1+\frac{r}{\lambda})}\left(\frac{p_{j,t}}{1-p_{j,t}}\right)^{-\frac{r}{\lambda}}\frac{r}{\lambda}(\lambda u_{j,t}-\lambda)\right]e^{-rt}dt$$

Now, we focus on the term

$$(1-\bar{p})^2\int_0^\infty \frac{\dot{p}_{i,t}}{p_{i,t}(1-p_{i,t})^2}\frac{c}{\lambda(1-p_{j,t})}e^{-rt}dt$$

consider

$$dv = \frac{\dot{p}_{i,t}}{p_{i,t}(1-p_{i,t})^2} \Leftrightarrow \qquad v = \tfrac{1}{(1-p_{i,t})} + ln(\tfrac{p_{i,t}}{1-p_{i,t}})$$

$$u = \frac{c}{\lambda(1-p_{j,t})}e^{-rt} \Leftrightarrow \quad du = -\tfrac{c}{\lambda}\left[\tfrac{r}{1-p_{j,t}}+\tfrac{p_{j,t}}{1-p_{j,t}}\lambda u_{j,t}\right]e^{-rt}$$

then,

$$(1-\bar{p})^2\int_0^\infty \frac{\dot{p}_{i,t}}{p_{i,t}(1-p_{i,t})^2}\frac{c}{\lambda(1-p_{j,t})}e^{-rt}dt =$$

$$C+(1-\bar{p})^2\int_0^\infty \frac{c}{\lambda}\left[\frac{1}{(1-p_{i,t})}+ln(\frac{p_{i,t}}{1-p_{i,t}})\right]\left[\frac{r}{1-p_{j,t}}+\frac{p_{j,t}}{1-p_{j,t}}\lambda u_{j,t}\right]e^{-rt}$$

Now we can rewrite the objective function (ignoring constants)

$$\int_0^\infty \Bigg\{ -\frac{\pi}{(1-p_{i,t})}\left[\frac{r}{(1-p_{j,t})}+\lambda u_{j,t}\frac{p_{j,t}}{(1-p_{j,t})}\right]$$
$$-\frac{\bar{\gamma}}{(1-p_{i,t})}\left[\frac{p_{j,t}}{1-p_{j,t}}(\lambda u_{j,t}+r)+\beta^{-(1+\frac{r}{\lambda})}\left(\frac{p_{j,t}}{1-p_{j,t}}\right)^{-\frac{r}{\lambda}}\frac{r}{\lambda}(\lambda u_{j,t}-\lambda)\right]$$
$$+\frac{c}{\lambda}\left[\frac{1}{(1-p_{i,t})}+ln(\frac{p_{i,t}}{1-p_{i,t}})\right]\left[\frac{r}{1-p_{j,t}}+\frac{p_{j,t}}{1-p_{j,t}}\lambda u_{j,t}\right]$$
$$+\frac{p_{j,t}}{1-p_{j,t}}\lambda u_{j,t}c\left[\frac{p_{i,t}}{1-p_{i,t}}\frac{\beta}{\lambda+r}+\left(\beta\frac{p_{i,t}}{1-p_{i,t}}\right)^{-r/\lambda}\left(\frac{1}{r}-\frac{1}{\lambda+r}\right)-\frac{1}{r}\right]\Bigg\}e^{-rt}dt$$

we make the further change of variables $x_{m,t}=ln((1-p_{m,t})/p_{m,t})$ and rearrange some terms. Agent i's problem

$$\int_0^\infty \Bigg\{ -(1+e^{-x_{i,t}})\left[\left(r(1+e^{-x_{j,t}})+e^{-x_{j,t}}\lambda u_{j,t}\right)\left(\pi-\frac{c}{\lambda}\right)+\bar{\gamma}\left(e^{-x_{j,t}}(\lambda u_{j,t}+r)+e^{\frac{r}{\lambda}}\beta^{-(1+\frac{r}{\lambda})}\frac{r}{\lambda}(\lambda u_{j,t}-\lambda)\right)\right]$$
$$-\frac{c}{\lambda}x_{i,t}\left(r(1+e^{-x_{j,t}})+e^{-x_{j,t}}\lambda u_{j,t}\right)+e^{-x_{j,t}}\lambda u_{j,t}c\left[e^{-x_{i,t}}\frac{\beta}{\lambda+r}+e^{\frac{r}{\lambda}x_{i,t}}\beta^{-\frac{r}{\lambda}}\left(\frac{\lambda}{r(\lambda+r)}\right)-\frac{1}{r}\right]\Bigg\}e^{-rt}dt$$

Subject to

$$\dot{x}_{m,t}=\lambda u_{m,t};x_{m,o}=ln((1-\bar{p})/\bar{p})$$

given an arbitrary function $u_{j,t}$. The hamiltonian of this problem is

$$H=\eta_0\Bigg\{ -(1+e^{-x_{i,t}})\left[\left(r(1+e^{-x_{j,t}})+e^{-x_{j,t}}\lambda u_{j,t}\right)\left(\pi-\frac{c}{\lambda}\right)+\bar{\gamma}\left(e^{-x_{j,t}}(\lambda u_{j,t}+r)+e^{\frac{r}{\lambda}}\beta^{-(1+\frac{r}{\lambda})}\frac{r}{\lambda}(\lambda u_{j,t}-\lambda)\right)\right]$$
$$-\frac{c}{\lambda}x_{i,t}\left(r(1+e^{-x_{j,t}})+e^{-x_{j,t}}\lambda u_{j,t}\right)+e^{-x_{j,t}}\lambda u_{j,t}c\left[e^{-x_{i,t}}\frac{\beta}{\lambda+r}+e^{\frac{r}{\lambda}x_{i,t}}\beta^{-\frac{r}{\lambda}}\left(\frac{\lambda}{r(\lambda+r)}\right)-\frac{1}{r}\right]\Bigg\}e^{-rt}+\hat{\eta}_{i,t}\lambda u_{i,t}$$

w.l.g. we are going to work with the current value hamiltonian (See Seierstad and Sydsaeter, 1986, Ch.2.4, Exercise 6).[23] Define $\eta_{i,t}=e^{rt}\hat{\eta}_{i,t}$. Then,

$$\begin{aligned}H^c=&\eta_0\Bigg\{ -(1+e^{-x_{i,t}})\Big[\left(r(1+e^{-x_{j,t}})+e^{-x_{j,t}}\lambda u_{j,t}\right)\left(\pi-\frac{c}{\lambda}\right)\\&+\bar{\gamma}\left(e^{-x_{j,t}}(\lambda u_{j,t}+r)+e^{\frac{r}{\lambda}}\beta^{-(1+\frac{r}{\lambda})}\frac{r}{\lambda}(\lambda u_{j,t}-\lambda)\right)\Big]-\frac{c}{\lambda}x_{i,t}\left(r(1+e^{-x_{j,t}})+e^{-x_{j,t}}\lambda u_{j,t}\right)\\&+e^{-x_{j,t}}\lambda u_{j,t}c\left[e^{-x_{i,t}}\frac{\beta}{\lambda+r}+e^{\frac{r}{\lambda}x_{i,t}}\beta^{-\frac{r}{\lambda}}\left(\frac{\lambda}{r(\lambda+r)}\right)-\frac{1}{r}\right]\Bigg\}+\eta_{i,t}\lambda u_{i,t}\end{aligned} \tag{11}$$

Note that the change of variables has two advantages. First, it makes easier to work with

[23] All references to Seierstad and Sydsaeter (1986), Ch. 2, apply to finite horizon problems. Since all extensions to infinite horizon are straightforward and standard in the literature we are going to omit them.

the hamiltonian because the control variable $u_{i,t}$ is no longer in the objective function, and the state variable is no longer in the evolution of the state variable $\dot{x}_{i,t} = \lambda u_{i,t}$. On the other hand, note that the evolution of the state variable can be interpreted as the marginal effect of instant effort in the function of beliefs. This gives an straightforward interpretation for many of the optimality conditions that will arise later.

C.2 Necessary conditions:

Is direct that that this is not an abnormal problem (See Seierstad and Sydsaeter, 1986, Ch.2.4, Note 5), so $\eta_0 = 1$. By Pontryagin's principle, there must exist a continuous function $\eta_{m,t}$ for all $m \in \{i, j\}$ such that:

(i) Maximum principle: For each $t \geq 0$, $u_{i,t}$ maximizes $H^c \Leftrightarrow \eta_{i,t}\lambda = 0$.

(ii) Evolution of co-state variable: The function $\eta_{i,t}$ satisfies

$$\dot{\eta}_{i,t} = r\eta_{i,t} - \frac{\partial H^c}{\partial x_{i,t}}$$

(iii) Transversality condition: If x^* is the optimal trajectory, $\lim_{t\to\infty} \eta_{i,t}(x_t^* - x_t) \leq 0$ for all feasible trajectories x_t (Kamihigashi, 2001).[24]

C.3 Candidate (interior) equilibrium:

From (i), since $\lambda > 0$ our candidate of equilibrium satisfies $\eta_{i,t} = 0$ for all $t \geq 0$. Then, condition (ii) becomes

$$-\frac{\partial H^c}{\partial x_{i,t}} = 0 \tag{12}$$

$$-\frac{\partial H^c}{\partial x_{i,t}} = -e^{-x_{i,t}} \left[\left(r(1 + e^{-x_{j,t}}) + e^{-x_{j,t}}\lambda u_{j,t} \right) \left(\pi - \frac{c}{\lambda} \right) + \bar{\gamma} \left(e^{-x_{j,t}}(\lambda u_{j,t} + r) + e^{\frac{r}{\lambda}x_{j,t}} \beta^{-(1+\frac{r}{\lambda})} \frac{r}{\lambda}(\lambda u_{j,t} - \lambda) \right) \right]$$
$$+ \frac{c}{\lambda} \left(r(1 + e^{-x_{j,t}}) + e^{-x_{j,t}}\lambda u_{j,t} \right) - e^{-x_{j,t}}\lambda u_{j,t} c \left[e^{\frac{r}{\lambda}x_{i,t}} \frac{\beta^{-\frac{r}{\lambda}}}{\lambda + r} - e^{-x_{i,t}} \frac{\beta}{\lambda + r} \right]$$

Since we are interested in the symmetric equilibrium, $u_{j,t} = u_{i,t}$ and $x_{i,t} = x_{j,t}$.

$$0 = -e^{-2x_{i,t}} \left[r \left(\pi + \bar{\gamma} - \frac{c}{\lambda} \right) + \lambda u_{i,t} \left(\pi + \bar{\gamma} - \frac{c}{\lambda} - \frac{c\beta}{\lambda + r} \right) \right] - e^{-x_{i,t}} \left[r(\pi - \frac{2c}{\lambda}) - \frac{c}{\lambda}\lambda u_{i,t} \right]$$
$$-e^{-(1-\frac{r}{\lambda})x_{i,t}} \left[\bar{\gamma}\beta^{-(1+\frac{r}{\lambda})} \frac{r}{\lambda}(\lambda u_{i,t} - \lambda) + \lambda u_{i,t} c \frac{\beta^{-\frac{r}{\lambda}}}{\lambda + r} \right] + \frac{cr}{\lambda}$$

[24] This is the same transversality condition used by Bonatti and Hörner (2011).

If the solution is interior the following must hold

$$\frac{\partial H^c(u_{i,t}, x_{i,t})}{\partial x_{i,t}}\Big|_{\substack{u=1\\t=0}} < 0 \tag{13}$$

the intuition is that the agent derives disutility from changing the state variable with full effort ($\dot{x}_{i,t} = \lambda$). Therefore, the upper bound of effort is not binding. That is both agents can reach a level of indifference between exerting effort in an interval $[0, dt)$ and $[dt, 2dt)$ with an interior level of effort. By continuity of the exponential function is easy to see that condition (13) can not be satisfied for a high enough $\bar{p}$ (low enough x_0), we address this issue later.

Now we can characterize the interior equilibrium. Since $\lambda u_{i,t} = \dot{x}_{i,t}$ the optimal belief function x_t^* is characterized by the following non-linear ODE

$$\dot{x}_t = \frac{r\left[e^{-2x_{i,t}}(\pi + \bar{\gamma} - \frac{c}{\lambda}) + e^{-x_{i,t}}\left(\pi - \frac{2c}{\lambda}\right) - e^{-(1-\frac{r}{\lambda})x_{i,t}}\beta^{-(1+\frac{r}{\lambda})}\bar{\gamma} - \frac{c}{\lambda}\right]}{-e^{-2x_{i,t}}\left(\pi + \bar{\gamma} - \frac{c}{\lambda} - \frac{c\beta}{\lambda+r}\right) - e^{-(1-\frac{r}{\lambda})x_{i,t}}\left(\bar{\gamma}\frac{r}{\lambda}\beta^{-(1+\frac{r}{\lambda})} + \frac{c\beta^{-\frac{r}{\lambda}}}{\lambda+r}\right) + e^{-x_{i,t}}\frac{c}{\lambda}} \tag{14}$$

this differential equation has not analytical solution. Therefore, we solve it by numerical methods (See Appendix D).

C.4 Corner solutions

As discussed before, condition (13) is not satisfied if $\bar{p}$ is high enough. In order to make this more clear consider

$$\begin{aligned}\frac{\partial H^c}{x_{i,t}} = e^{-2x_{i,t}}&\left[r\left(\pi + \bar{\gamma} - \frac{c}{\lambda}\right) + \lambda u_{i,t}\left(\pi + \bar{\gamma} - \frac{c}{\lambda} - \frac{c\beta}{\lambda + r}\right)\right] + e^{-x_{i,t}}\left[r(\pi - \frac{2c}{\lambda}) - \frac{c}{\lambda}\lambda u_{i,t}\right]\\ &+e^{-(1-\frac{r}{\lambda})x_{i,t}}\left[\bar{\gamma}\beta^{-(1+\frac{r}{\lambda})}\frac{r}{\lambda}(\lambda u_{i,t} - \lambda) + \lambda u_{i,t}c\frac{\beta^{-\frac{r}{\lambda}}}{\lambda + r}\right] - \frac{cr}{\lambda}\end{aligned} \tag{15}$$

$$\begin{aligned}\Rightarrow \frac{\partial H^c}{\partial x_{i,t}}\Big|_{\substack{u=1\\t=0}} = e^{-2x_{i,0}}&\left[r(\pi + \bar{\gamma} - \frac{c}{\lambda}) + \left(\frac{r(\lambda\pi - c)}{\lambda + r}\right)\right] + e^{-x_{i,0}}\left[r\left(\pi - \frac{2c}{\lambda}\right) - c\right]\\ &+e^{-(1-r/\lambda)x_{i,0}}\left[\frac{\lambda c}{\lambda + r}\left(\frac{c}{\lambda(\pi + \gamma) - c}\right)^{\frac{r}{\lambda}}\right] - \frac{cr}{\lambda}\end{aligned} \tag{16}$$

note that $x \in (-\infty, \infty)$, with $\lim_{p\to 1} x = -\infty$ and $\partial x/\partial p < 0$. Let's consider a $\bar{p} > 1/2$, which implies $x_{i,t} < 0$. Also, if Assumption 1 is satisfied the first and third RHS terms in brackets of (16) are positive. If Assumption 2 is satisfied the arguments of the three exponential functions are positive. This implies that RHS is strictly positive and an interior equilibrium is not possible.[25]

If $p < 1/2$ then $x > 0$, and the value of the RHS is not clear. Note that as x

[25] Assumption 1 does not guarantee that the RHS second term in brackets is positive, but even if its negative the effect is dominated by the others.

increases, the value of the exponential functions decreases. Then, since there is a cost $-cr/\lambda$, by continuity of the exponential function $\exists\hat{p}$ such that for every $p_t < \hat{p}$, (13) is not binding. Is not possible to obtain a closed form for $\hat{p}$, but is easy to find numerically using the implicit function (16).

Therefore, given $\{\pi, \gamma, c, l, r\}$ if $\bar{p}$ is high enough condition (13) dos not hold and we have a corner solution $u_{i,t} = 1$ (See Appendix D for codes that allows a numerical analysis). This implies that in the symmetric equilibrium, for $\bar{p} \geq \hat{p}$, agents exert maximal effort until a time $\hat{t}$ at which belief $\hat{p}$ is reached, and for every $p_t < \hat{p}$ they exert interior levels of effort that satisfies the non-linear ODE (14).

Finally, as in the interior equilibrium agents exert a strictly possitive amount of effort, is possible that a belief $\underline{p}$ is reached at which

$$\frac{\partial H^c(u_{j,t}, x_{i,t}, x_{j,t})}{\partial x_{i,t}}\Big|_{u\in(0,1]} < 0 \tag{17}$$

and the solution is to exert zero effort for all $t > \underline{t}$ (in the numerical analysis is not possible to find a set of parameters for which $\underline{p}$ does not exists, see Appendix D).

C.5 Uniqueness:

Despite we can not find a closed form for our optimal path candidate $x^*_{i,t}$. We can proof some characteristics of it that makes easier to proof uniqueness of our symmetric equilibrium candidate.

Effort is (weakly) decreasing in time: Assume for the sake of contradiction that for an arbitrary t, $u_{i,t} < u_{i,t+dt}$. If at t condition (13) does not hold the solution is $u_{i,t} = 1$ and by the upper bound of u, $u_{t+dt} > u_t$ is a contradiction. On the other hand, if condition (13) holds, then $\partial H^C/\partial x_{i,t} = 0$ for an interior $u_{i,t}$. But $\partial H^c/\partial x_{i,t}$ is decreasing in x, and $\dot{x}_{i,t} > 0$ implies $x_{i,t} < x_{i,t+dt}$. Therefore, if $u_{i,t+dt} \geq u_{i,t}$ then $\partial H^c/\partial x_{i,t+dt} < 0$, which is a contradiction. Note that for all $t > \hat{t}$ effort must be strictly decreasing.

Since effort is weakly decreasing, $\dot{x}^*_{i,t}$ is also weakly decreasing (strictly decreasing for all $t > \hat{t}$). Then, there are two possibilities: $x^*_{i,t}$ converges and agents exert effort forever, or $\exists\underline{p}$ and $x_{i,t}$ reaches a value less than infinity ($p_t > 0$) where agents stop experimentation.

We now use the trajectory $(\eta^*_{j,t}, \eta^*_{i,t}, x^*_{j,t}, x^*_{i,t})$ and the necessary conditions to eliminate other admissible trajectories. Note that the optimal trajectories are

$$\eta_t^* = \begin{cases} > 0 & \text{if } t \leq \hat{t} \\ = 0 & \text{if } t \in (\underline{t}, \hat{t}) \\ < 0 & \text{if } t \geq \underline{t} \end{cases}$$

Clearly, if $\underline{t}$ does not exists, $\eta_t = 0$ for all $t \geq \hat{t}$.

While $\frac{\partial H^c(u_{j,t}, x_{i,t}, x_{j,t})}{\partial x_{i,t}}\Big|_{u=1} \geq 0$ any strategy $u_{i,t} \neq 1$ violates necessary condition (i). Therefore, we need to prove that our symmetric equilibrium candidate is unique for $p_t \leq \hat{p}$.

- Consider paths that start with $\eta_{i,t} < 0$, for $t \geq \hat{t}$. From necessary condition (i), $u_{i,t} = 0$. Condition (ii) implies that, $\dot{\eta}_{i,t} < 0$. Clearly, $x_{i,t} = x_{i,\hat{t}}$ for all $t \geq \hat{t}$. Since in our reference path x^* is greater than $x_{\hat{t}}$ in the limit, and $\eta_{i,t} \to -\infty$. It follows that this path violates the transversality condition.
- Now consider paths that start with $\eta_{i,t} > 0$. By an analogous argument, $u_{i,t} = 1$ and $\dot{\eta}_{i,t} > 0$. Then, $x_{i,t} > x^*_{i,t}$ for all $t > \hat{t}$ and $\eta_{i,t} \to +\infty$. Therefore, this path violates the transversality condition.

Since we are restricting the analysis to symmetric equilibrium, the above arguments eliminate any other possible trajectory. In particular, any trajectory that differ from our candidate must start with a co-state variable > 0 or < 0, then the same arguments starting at t' eliminate the alternative candidate path. Therefore, our candidate is the only symmetric equilibrium of the game.

C.6 Sufficiency:

From (15) the maximized hamiltonian is concave in x_t, because if $\partial^2 H^c / \partial x_{i,t}^2 > 0$ then $\partial H^c / \partial x_{i,t} < 0$ for all $u \in (0, 1]$, but values of x_t that makes that possible are not reached in the symmetric equilibrium.

Therefore, from the Arrow sufficiency theorem necessary conditions are also sufficient (Seierstad and Sydsaeter, 1986, Ch. 3, Theorem 17). ∎

D MATLAB codes

All numerical exercises presented on this work can be replicated for any combination of parameters with the MATLAB codes available here. The *readme.pdf* document explains the function and how to use each *mfile*.