Strategic Exploration for Innovation*

Shangen Li†

November 13, 2021

Abstract

We analyze a game of technology development where players allocate resources between exploration, which continuously expands the public domain of available technologies, and exploitation, which yields a flow payoff by adopting the explored technologies. The qualities of the technologies are correlated and initially unknown, and this uncertainty is fully resolved once the technologies are explored. We consider Markov perfect equilibria with the quality difference between the best available technology and the latest technology under development as the state variable. In all such equilibria, free-rider problem is mitigated by an encouragement effect. In the unique symmetric equilibrium, the encouragement effect is dominant if the players are sufficiently patient and the prospect of development is not too poor. Pareto improvements over the symmetric equilibrium can be achieved by asymmetric equilibria where players take turns performing exploration.

Keywords: Strategic experimentation, Encouragement effect, Innovation, Multi-armed bandit

JEL classification: C73; D83; O3.

*I am grateful to Marlon Azinovic, Steven Callander, Martin Cripps, Christian Ewerhart, David Hémous, Johannes Hörner, Bruno Jullien, Vijay Krishna, Alessandro Lizzeri, Margaret Meyer, Nick Netzer, Sven Rady, Anne-Katrin Roesler, Jakub Steiner, Bruno Strulovici, Leeat Yariv, Haoyuan Zeng, and seminar participants at University of Zürich and Swiss Theory Day 2021 for helpful comments. I am especially indebted to Marek Pycia, Armin Schmutzler, and Aleksei Smirnov for advice and encouragement. All errors are mine.

†Department of Economics, University of Zürich. E-mail address: `work@shangen.li`

1 Introduction

Without innovation, we would still be living in caves and hunting with stones. Though people celebrate disruptive inventions such as penicillin and the Internet, innovations are usually accomplished through hundreds and thousands of minor improvements. These incremental improvements are achieved from iterative experimentation through small changes, and can have a huge cumulative impact in the long run.[1] Incremental experimentation of this kind is particularly pertinent in today's data-driven world, in which experiments can be conducted more swiftly and inexpensively than ever before thanks to the fast-growing fields of data science and artificial intelligence. Tech giants such as Microsoft and Amazon run thousands of experiments every day to optimize their products, processes, customer experiences, and business models.[2]

However, even for these companies with digital roots that can run experiments at virtually zero cost, valuable resources still need to be allocated to the design and implementation of the experiments, as well as to the collection and analysis of the data they produce. Moreover, because of its uncertain nature, experimentation does not guarantee innovation. As a result, the trade-off between creating value through innovation——to explore——and capturing value through operations——to exploit——is behind all innovation processes in organizations. On top of such complexities, incentives for experimentation are susceptible to free-riding opportunities when successes and failures are publicly observed. Such a strategic effect inevitably creates inefficiency of learning, having a lasting impact on technological progress. Examples of strategic exploration certainly go beyond intracompany R&D activities. Research joint ventures formed by multiple firms, open-source software development, consumer search, and even academic researchers working on a joint paper are also virtually involved in strategic exploration.

This paper analyzes a game of strategic exploration in which a finite number of forward-looking players jointly search for innovative technologies. Given a set of publicly available technologies, at each point in time, each player decides simultaneously how to allocate a perfectly divisible unit of resource between exploration, which expands the set of feasible technologies at a rate proportional to the resource allocated, and exploitation, which yields a flow payoff from the adoption of one of the available technologies. This flow payoff is generated according to an (initially unknown) realized path of Brownian motion, which maps technologies to their qualities. Not all technologies are readily available to the public, and only the qualities associated with the

[1] Garcia-Macia, Hsieh, and Klenow (2019) estimate that improvements to already existing products accounted for about 77 percent of economic growth in the United States between 2003 and 2013.

[2] See Thomke (2020).

explored technologies are known. We focus on the scenario in which players' actions and outcomes are publicly observed. Hence, the explored technologies are pure public goods, and there are perfect knowledge spillovers between players.

One of the major contributions of this paper is to extend the classic game of strategic experimentation to a setting that features a continuum of correlated arms with the possibility of discovering new arms. When modeling technologies as arms, correlation between arms and discoveries of new arms are key features of technology development and innovation. However, a fixed set of independent arms is assumed in most literature on strategic experimentation because the analysis is substantially complicated by correlation. We demonstrate that under a specific form of correlation, such complexities can be circumvented by simply focusing on incremental experimentation.

We first consider the efficient benchmark in which the players work jointly to maximize the average payoff. The solution to this cooperative problem takes a simple cutoff form: All players allocate resources exclusively to exploration if and only if the quality difference between the best available technology and the latest outcome of the experiment is within a time-invariant cutoff. Players in a larger team can attain higher payoffs by being more tolerant of unsuccessful outcomes. As team size grows, because of the growing total learning capacity, exploration extends to arbitrarily pessimistic states in the limit, and the welfare under complete information can be achieved asymptotically.

In the strategic problem, we restrict attention to Markov perfect equilibria (MPE) with the difference between the qualities of the best available technology and the latest technology under development as the state variable. Despite the considerable disparities between our setting and the two-armed bandit models in the strategic experimentation literature, all MPE in our model exhibit a similar *encouragement effect*: The presence of other players encourages at least one of them to continue exploration at states where a single agent would already have given up. The players thus become more tolerant of failure in the hope that successful outcomes will bring the other players back to exploration in the future, promoting innovation and sharing its burden. In addition, exactly because of such an incentive, any player who never explores would strictly prefer the deviation that resumes the exploration process as a volunteer at the state where all exploration stops. Therefore, no player always free-rides in equilibrium in spite of the fact that exploration per se never generates any payoffs.

We further show that there is no MPE in which all players use simple cutoff strategies as in the cooperative solution. A symmetric equilibrium thus necessarily requires the players to choose an interior allocation of their resources at some states. We establish the existence and uniqueness of the symmetric equilibrium, and provide closed-form representations for the equilibrium strategies and the associated payoff functions.

Depending on the parameters, the symmetric equilibria can be classified into two types as follows. In the first type of equilibrium, all players allocate resources exclusively to exploration when the latest outcome is sufficiently close to the highest-known quality, exclusively to exploitation when it is sufficiently far away, and to both simultaneously when it is in between. In such a case, the resource constraints on exploration are binding at optimistic states, and therefore will be referred to as the *binding* case. In the second type of equilibrium, however, the resource constraints are never binding. An interior allocation is chosen by each player when the latest outcome is close to the highest-known quality, and all resources are allocated to exploitation if it is sufficiently far away. In this *non-binding* case, the incentive to free-ride is so strong that players even allocate a positive fraction of their resources to exploitation when the latest technology achieves the highest-known quality and thus innovation takes place instantaneously. Moreover, this free-rider effect completely freezes the encouragement effect: Additional players would not be able to encourage the players to explore at more pessimistic states and would bring down both the individual and the overall intensity of exploration.

We then show the following comparative statics: The symmetric equilibrium always belongs to the binding case for a small team and may or may not fall into the non-binding category irreversibly as team size grows. The non-binding case can be averted if and only if the players are patient enough and the prospect of development is sufficiently promising. In such a case, the players become increasingly tolerant of failure as team size grows and are willing to explore at arbitrary pessimistic states in the limit, sharing the same features as in the cooperative solution. Otherwise, free-riding incentives prevail as the team becomes large: The resource constraints eventually fail to be binding, and the tolerance for failure among the players stays the same for further increases in team size.

This result suggests that the rate of technological progress and long-run outcomes hinge on the strength of the encouragement effect. When the players are sufficiently patient, the promising prospect of the development is able to strengthen the encouragement effect to the extent that the free-rider effect is dominated in a large team, leading to technological trajectories resembling the ones in the cooperative solution.

Lastly, we investigate whether asymmetric equilibria can improve welfare and long-run outcomes over the symmetric equilibrium. We construct a class of asymmetric MPE in which the players take turns performing exploration at pessimistic states, so that each player achieves a higher payoff than in the symmetric equilibrium. It turns out that these asymmetric MPE are the best MPE in two-player games in terms of average payoffs. Unlike in the symmetric equilibrium, the players in these asymmetric MPE become increasingly tolerant of failure as team size grows, independent of the other parameters.

The intuition for this result is that when alternation is allowed, the burden of keeping the exploration process active can be shared among more players in larger teams, and thus the players are willing to explore at more pessimistic states. As a consequence, the long-run outcomes approach the first best as team size grows, irrespective of the patience of the players or the prospect of the development. Nevertheless, similar to the symmetric equilibrium with non-binding resource constraints, in these asymmetric equilibria innovations might arrive at a much slower rate than in the cooperative solution because of the low proportion of the overall resource allocated to exploration. As a result, the welfare loss might still be significant in large teams because of the strong free-rider effect.

1.1 Related Literature

This paper contributes to the literature on strategic experimentation, started by the Brownian model in Bolton and Harris (1999), and then followed by the exponential model in Keller, Rady, and Cripps (2005) and the Poisson model in Keller and Rady (2010) and Keller and Rady (2015). In all of these models, players face identical two-armed bandit machines, which consist of a risky arm with an unknown quality and a safe arm with a known quality. At each point in time, each player decides simultaneously how to split one unit of a perfectly divisible resource between these arms, so that learning occurs gradually by observing other players' actions and outcomes. These models differ in the assumptions on the probability distribution of the flow payoffs that each type of the arm generates. By contrast, players in our model face a continuum of correlated arms. Local learning—learning the quality of a particular arm—occurs instantaneously. However, as the set of arms is unbounded, global learning—learning the qualities of all arms—occurs gradually.

More broadly, this paper contributes to the literature of dynamic public good provision games. Admati and Perry (1991), Marx and Matthews (2000), Yildirim (2006), and Georgiadis (2014) study voluntary contributions to a joint project in dynamic settings. While the public good in these papers is the progress toward the completion of a project, the public good in our model is the knowledge—the feasible technologies—built over time, which can be exploited once developed.

The encouragement effect was first identified by Bolton and Harris (1999) in the symmetric equilibrium in their Brownian model. This effect is then established by Keller and Rady (2010) for all MPE in the Poisson model with inconclusive good news, and by Keller and Rady (2015) for the symmetric MPE in the Poisson model with bad news. Due to the absence of technological advancements, the encouragement effects in all these papers are not strong enough to dominate the free-rider effect. We are not only able to show the presence of encouragement effect in all MPE in our model, but also (1) express the stopping threshold in the unique symmetric MPE in closed form, and (2) perform comparative statics analysis of the stopping threshold in both the symmetric MPE and a class of asymmetric MPE to further study the strength of the encouragement effect. We find that the prospect of innovation and technological advancements allows the encouragement effect to overcome the free-rider effect to some extent.

Callander (2011) proposes modeling the correlation between arms by Brownian path and studies experimentation conducted by a sequence of myopic agents. Garfagnini and Strulovici (2016) extend this model to a setting with overlapping generations, in which short-lived yet forward-looking players search on a Brownian path for arms with higher qualities. They mainly focus on the search patterns and the long-run dynamics, and identify the stagnation of search and the emergence of a technological standard in finite time. By contrast, a byproduct of our analysis suggests that their findings might depend on the search cost or the underlying correlation structure. When the qualities of arms contribute exponentially to the payoffs, which is common in modeling technological progress in the macroeconomic growth literature, stagnation can be avoided even under a negative drift of the Brownian path. Our main assumption on the parameters (Assumption 1) provides the condition for stagnation in the cooperative problem.

Both of their models focus on non-strategic environments, because in their discrete-time set-ups unexplored gaps between explored technologies create difficulty for further analysis in strategic environments. To overcome this challenge, we forgo the fine details of the learning dynamics for analytical tractability by imposing continuity on the experimentation process. This simplification can be interpreted as the qualities of the neighboring technology being revealed during experimentation, so that no unexplored territory remains between the explored technologies. Moreover, we impose a hard constraint on the intensity of exploration to capture the scenario that technologies far ahead of their time are infeasible to be explored today. These abstractions allow us to derive explicit expressions for the equilibrium payoffs and strategies, perform comparative statics analysis, and construct asymmetric equilibria.

From a modeling perspective, continuous exploration on a Brownian sample path appears in Wong (2021), Urgun and Yariv (2021) and Cetemen, Urgun, and Yariv (2021), however strategic interactions are absent from the first two papers. Cetemen, Urgun, and Yariv (2021) is the only paper we are aware of that studies collective exploration on a Brownian path in a strategic setting, as in this paper. They focus on the exit patterns during a search process conducted jointly by heterogeneous players. In their model, exploitation is only possible after an irreversible exit chosen endogenously by each player. The players in our model, however, are not faced with stopping problems and thus are free to choose between exploration and exploitation, or even both simultaneously, at all times.

2 The Exploration Game

Time $t \in [0, \infty)$ is continuous, and the discount rate is $r > 0$. There are $N \geq 1$ players, each endowed with one unit of perfectly divisible resource per unit of time. Each player has to simultaneously and independently allocate the resources between exploration, which expands the feasible technology domain, and exploitation, which allows her to adopt one of the explored technologies. The feasible technology domain, which is common to all players and contains all the explored technologies at time t, is modeled as an interval $[0, X_t]$ with $X_0 = 0$. If a player allocates the fraction $k_t \in [0, 1]$ to exploration over an interval of time $[t, t + \mathrm{d}t)$, the boundary X_t is pushed to the right by an amount of $k_t \, \mathrm{d}t$. With the fraction $1 - k_t$ allocated to exploitation, by adopting technology $x_t \in [0, X_t]$, the player receives a deterministic flow payoff $(1 - k_t) \exp(W(x_t)) \, \mathrm{d}t$, where $W(x_t)$ denotes the quality of the adopted technology.

The quality mapping $W : \mathbb{R}_+ \to \mathbb{R}$ is common to all players, but only the qualities of the feasible technologies in $[0, X_t]$ are known to each player at time t. It coincides with a realized path of a Brownian motion $W_+ \in C(\mathbb{R}_+, \mathbb{R})$ starting at $W_+(0) = y \in \mathbb{R}$, with drift $\theta \in \mathbb{R}$ and volatility normalized to $\sigma = \sqrt{2}$, except possibly for $W(0) = s \geq y$, which is interpreted as the quality of a status quo technology. At the outset of the game, all players know the parameters of the mapping but not the realized Brownian path W_+ on $\mathbb{R}_{++}$. Therefore, the process of exploration described above captures the dynamics of *research*—building knowledge on the qualities of technologies—and *development*—expanding the set of feasible technologies.

Given a player's actions $\{(k_t, x_t)\}_{t \geq 0}$, with $k_t \in [0, 1]$ and $x_t \in [0, X_t]$ measurable with respect to the information available at time t, her total expected discounted payoff, expressed in per-period unit, is

$$\mathsf{E}\left[\int_0^\infty re^{-rt}(1 - k_t)e^{W(x_t)} \, \mathrm{d}t\right].$$

Clearly, for any $\{k_t\}_{t \geq 0}$, each player maximizes her total expected discounted payoff by choosing $x_t = \arg\max_{x \in [0, X_t]} W(x)$, which is the best feasible technology, for all $t \geq 0$. So we can focus on such an exploitation strategy without loss and rewrite the above total payoff as

$$\mathsf{E}\left[\int_0^\infty re^{-rt}(1-k_t)e^{S_t}\,\mathrm{d}t\right],$$

where $S_t = \max_{x\in[0,X_t]} W(x)$.

Therefore, the environment above can be equivalently reformulated as follows. Players have prior beliefs represented by a filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t), \mathsf{P})$, where $\Omega = C(\mathbb{R}_+, \mathbb{R})$ is the space of Brownian paths, P is the law of standard Brownian motion $B = \{B_t\}_{t\geq 0}$, and $\mathscr{F}_t$ is the canonical filtration of B. Each player chooses her strategy from the space of admissible control processes $\mathcal{A}$, which consists of all processes $\{k_t\}_{t\geq 0}$ adapted to the filtration $(\mathscr{F}_t)_{t\geq 0}$ with $k_t \in [0,1]$. The public history of technology development is represented by a development process $\{Y_t\}_{t\geq 0}$, where $Y_t := W(X_t)$ denotes the quality of the technology being explored at time t. This process satisfies the stochastic differential equation

$$\mathrm{d}Y_t = \theta K_t\,\mathrm{d}t + \sqrt{2K_t}\,\mathrm{d}B_t\,, \quad Y_0 = y,$$

where $K_t = \sum_{1\leq n\leq N} k_{n,t}$ measures how much of the overall resources is allocated to exploration, and will be referred to as the *intensity of exploration* at time t.

Given a strategy profile $\boldsymbol{k} = \{(k_{1,t},\ldots,k_{N,t})\}_{t\geq 0}$, player n's total expected discounted payoff can be written as

$$\mathsf{E}_{as}\left[\int_0^\infty re^{-rt}(1-k_{n,t})e^{S_t}\,\mathrm{d}t\right],$$

where

$$S_t = \max\left(s, \max_{0\leq\tau\leq t} Y_\tau\right)$$

denotes the quality of the best explored technology at time t.

In addition, we use the term "gap", denoted by $A_t := S_t - Y_t \geq 0$, to refer to the quality difference between the best explored technology and the latest technology under development. Henceforth, we shall use a and s when referring to the state variables as opposed to the stochastic processes $\{A_t\}_{t\geq 0}$ and $\{S_t\}_{t\geq 0}$ (i.e., if $A_t = a$, then "the game is in state a at time t").

A Markov strategy $k_n : \mathbb{R}_+ \times \mathbb{R} \to [0,1]$ with (a,s) as the state variables specifies the action player n takes at time t to be $k_n(A_t, S_t)$. A Markov strategy is called *s-invariant* if it depends on (a,s) only through a. Thus an s-invariant Markov strategy $k_n : \mathbb{R}_+ \to [0,1]$ takes the gap a as the state variable. Finally, an s-invariant Markov strategy k_n is a *cutoff strategy* if there is a cutoff $\bar{a} \geq 0$ such that $k_n(a) = 1$ for all $a \in [0,\bar{a})$ and $k_n(a) = 0$ otherwise.

Given an s-invariant Markov strategy profile $\boldsymbol{k}$, player n's associated payoff at state

(a, s) can be written as $v_n(a, s|\boldsymbol{k}) = e^s v_n(a, 0|\boldsymbol{k})$.[3] It is thus convenient to write $v_n(a, s|\boldsymbol{k}) = e^s u_n(a|\boldsymbol{k})$, where $u_n(a|\boldsymbol{k}) = v_n(a, 0|\boldsymbol{k})$ stands for player n's payoff at state (a, s) normalized by e^s, the opportunity cost of exploration. We refer to $u_n : \mathbb{R}_+ \to \mathbb{R}$ as player n's *normalized payoff function*, or simply as *payoff function* when it is clear from the context.

2.1 Exploration under Complete Information

To study the value of information, here we consider an alternative setting under complete information. More specifically, how would N players allocate their resources cooperatively over time if the entire Brownian path W is publicly known at the outset of the game?

Formally, in this subsection we replace $\mathcal{F}_t$ with $\mathcal{F}$ for each $t \geq 0$ while maintaining the assumption that the feasible technology domain $[0, X_t]$ can only be expanded by continuously pushing forward the boundary at the rate of $\mathrm{d}X_t/\mathrm{d}t = K_t$. For a given Brownian path W, denote the average (ex-post) value under complete information by

$$\hat{v}(W) := \sup \int_0^\infty r e^{-rt}(1 - K_t/N)e^{S_t}\,\mathrm{d}t\,,$$

where $S_t = \max_{x \in [0, X_t]} W(x)$, and the supremum is taken over all measurable functions $t \mapsto K_t \in [0, N]$.

Lemma 1 (Complete-information Payoff)**.** *Denote the average (ex-ante) value under complete information at state* (a, s) *by* $\widehat{V}(a, s) := \mathsf{E}_{as}[\hat{v}(W)]$. *We have* $\widehat{V}(a, s) = e^s \widehat{U}(a)$, *where*

$$\widehat{U}(a) = \begin{cases} 1 + \exp(-\lambda a)/(\lambda - 1), & \text{if } \lambda > 1, \\ +\infty, & \text{otherwise,} \end{cases}$$

with $\lambda = r/N - \theta$.

When $\lambda > 1$, there almost surely exists a *first-best technology* $\hat{x}_N \in \mathbb{R}_+$ so that the value $\hat{v}(W) < +\infty$ is achieved by exploring with full intensity up to the point when $\hat{x}_N$ is developed, and thereafter exploiting $\hat{x}_N$.[4] On the contrary, if $\lambda \leq 1$, then with probability 1, the payoff can be improved indefinitely by delaying the exploitation, and thus $\hat{v}(W) = +\infty$ almost surely. As a consequence, the value under *incomplete* information becomes infinite as well. The assumption that $\lambda > 1$ can be equivalently written in the following way.

[3]See Lemma 5 in the Appendix.

[4]See the proof of Lemma 8 for the precise definition of $\hat{x}_N$.

Assumption 1. $N(1+\theta) < r$.

To ensure well-defined payoffs and deviations, we maintain Assumption 1 throughout the rest of the paper unless otherwise stated.

3 Joint Maximization of Average Payoffs

Suppose that $N \geq 1$ players work cooperatively to maximize the *average* expected payoff. Denote by $\mathcal{A}_N$ the space of all adapted processes $\{K_t\}_{t\geq 0}$ with $K_t \in [0, N]$. Formally, we are looking for the value function

$$v(a, s) = \sup_{K \in \mathcal{A}_N} v(a, s|K),$$

where

$$v(a, s|K) = \mathsf{E}_{as}\left[\int_0^\infty re^{-rt}(1 - K_t/N)e^{S_t}\,\mathrm{d}t\right]$$

is the average payoff function associated with the control process $K = \{K_t\}_{t\geq 0}$, and an optimal control $K^* \in \mathcal{A}_N$ such that $v(a, s) = v(a, s|K^*)$. Clearly, the structure of the problem allows us to focus on Markov strategies $K : \mathbb{R}_+ \times \mathbb{R} \to [0, N]$ with (a, s) as the state variables, so that the intensity of exploration at time t is specified by $K_t = K(A_t, S_t)$.

According to the dynamic programming principle, we have

$$v(a, s) = \max_{K\in[0,N]} \left\{r\,(1 - K/N)\,e^s\,\mathrm{d}t + \mathsf{E}_{as}\left[e^{-r\mathrm{d}t}v(a + \mathrm{d}a\,,\, s + \mathrm{d}s)\right]\right\}.$$

First note that S_t can only change when $A_t = 0$, and thus $\mathrm{d}s = 0$ for all positive gaps. Hence, for each $a > 0$ at which $\partial^2 v/\partial a^2$ is continuous, the value function v satisfies the Hamilton-Jacobi-Bellman (HJB) equation

$$v(a, s) = \max_{K\in[0,N]} \left\{\left(1 - \frac{K}{N}\right)e^s + \frac{K}{r}\left(\frac{\partial^2 v(a, s)}{\partial a^2} - \theta\frac{\partial v(a, s)}{\partial a}\right)\right\}. \tag{1}$$

Assume, as will be verified, that the optimal strategy is s-invariant. Then by the homogeneity of the value function $v(a, s) = e^s v(a, 0) = e^s u(a)$, we can replace $v(a, s)$ with $e^s u(a)$, and divide both sides of equation (1) by the opportunity cost e^s, to obtain the normalized HJB equation

$$u(a) = 1 + \max_{K\in[0,N]} K\{\beta(a, u) - 1/N\}, \tag{2}$$

where $\beta(a, u) \coloneqq (u''(a) - \theta u'(a))/r$ is the ratio of the expected benefit of exploration

$(\partial^2 v/\partial a^2 - \theta\, \partial v/\partial a)/r$ to its opportunity cost e^s. It is then straightforward to see that the optimal action takes the following "bang-bang" form. If the shared opportunity cost of exploration, $1/N$, exceeds the full expected benefit, the optimal choice is $K(a) = 0$ (all agents choose exploitation exclusively), which gives value $u(a) = 1$. Otherwise, $K(a) = N$ is optimal (all agents choose exploration exclusively), and u satisfies the second-order ordinary differential equation (henceforth ODE),

$$\beta(a, u) = u(a)/N. \tag{3}$$

The optimal strategy could presumably depend on both a and s and hence might not be s-invariant. Indeed, both the benefit and the opportunity cost of exploration increase as innovation occurs. Nevertheless, due to our specific form of flow payoff where technologies contribute exponentially to the payoffs, the increased benefit of exploration exactly offsets the increased opportunity cost. As a result, the incentives for exploration for a fixed gap do not depend on the highest-known quality, which leads to an s-invariant optimal strategy.[5] This conjecture is confirmed by the following theorem.

Proposition 1 (Cooperative Solution)**.** *Suppose Assumption 1 holds. In the N-agent cooperative problem, there is a cutoff $a^* > 0$ given by*

$$a^* = \frac{1}{\gamma_2 - \gamma_1}\left(\ln\left(1 + \frac{1}{\gamma_2}\right) - \ln\left(1 + \frac{1}{\gamma_1}\right)\right)$$

with $\gamma_1 < \gamma_2$ being the roots of $\gamma(\gamma - \theta) = r/N$, such that it is optimal for all players to choose exploitation exclusively when the gap is above the cutoff a^* and it is optimal for all players to choose exploration exclusively when the gap is below the cutoff a^*.

The associated payoff at state (a, s) can be written as $V^*(a, s) = e^s U^*(a)$, where the normalized payoff function $U^* : \mathbb{R}_+ \to \mathbb{R}$ is given by

$$U^*(a) = \frac{1}{\gamma_2 - \gamma_1}\left(\gamma_2 e^{-\gamma_1(a^* - a)} - \gamma_1 e^{-\gamma_2(a^* - a)}\right)$$

when $a \in [0, a^*)$ and by $U^*(a) = 1$ otherwise.

If Assumption 1 is violated, then $U^*(a) = +\infty$ for all $a \geq 0$.

The cooperative solution is pinned down by the standard smooth pasting condition $u'(a^*) = 0$, and a *normal reflection* condition $(\partial v/\partial a + \partial v/\partial s)(0+, s) = 0$, which takes the form of $u(0) + u'(0+) = 0$ for s-invariant strategies.[6]

[5] This feature also appears in Urgun and Yariv (2021) under a setting of undiscounted linear preference with a flow cost convex in the learning rate.

[6] The normal reflection condition is not an optimality condition. It ensures that the infinitesimal change

Because of the lack of information on the qualities of technologies, the players might stop too early, giving up exploration before developing the first-best technology and thus ultimately adopting a suboptimal technology, or might stop too late, wasting too many resources for marginal improvement while the first-best technology has already been developed. The cooperative solution optimally balances these trade-offs between early and late stopping and therefore determines the *efficient* strategies under incomplete information.

Corollary 1 (Comparative Statics of the Cooperative Solution)**.** *The cooperate cutoff a^* is strictly increasing in N and strictly decreasing in r. For all $a \geq 0$, the cooperative payoff $U^*(a)$ is strictly below the complete-information payoff $\widehat{U}(a)$. For each $\theta \in \mathbb{R}$ and $r > 0$,*

- *if $\theta < -1$, then $|U_N^*(a) - \widehat{U}_N(a)| \to 0$ as $N \to +\infty$;*
- *if $\theta \geq -1$, then $U_N^* \to +\infty$ as $N \to r/(1+\theta) \in (1, +\infty]$.*[7]

In both cases $a_N^* \to +\infty$.

The stopping cutoff a^* represents the tolerance for failure among the players. The benefit of exploration decreases with r, and thus more patient players are willing to explore at more pessimistic states. Likewise, because the players work cooperatively, extra resources brought by additional players as team size increases enable a higher rate of exploration, so that for a fixed tolerance for failure, resources are wasted for a shorter period of time before the exploration fully stops, which in turn leads to greater tolerance.

Note that under the complete information setting in which players can observe the whole path, it is optimal for them to explore at arbitrarily large gaps as long as the gap would quickly shrink to zero if they explore a little bit longer. By contrast, under incomplete information, it is optimal to stop exploration whenever the gap reaches a^*. This informational disadvantage vanishes as either N or $1/r$ goes up because it originates from the suboptimal stopping decision in the cooperative solution due to the lack of information, rather than from the inefficient allocation of resources during exploration. As a result, the cooperative payoff converges to the complete-information payoff as the players become increasingly tolerant of failure.

of the payoff at a zero gap has a zero dS term, which is necessary for the continuation value process to be a martingale. See Peskir and Shiryaev (2006) for an introduction to the normal reflection condition in the context of optimal stopping problems, and the proof of Lemma 4 for more details.

[7]Here we allow N to be non-integral values for convenience. Also note that Assumption 1 is violated for $N \geq r/(1+\theta)$ if $\theta \geq -1$, in which case $U_N^* = +\infty$.

3.1 Long-run Outcomes

Consider an s-invariant strategy profile $\boldsymbol{k}$ such that the set of states at which the intensity of exploration is bounded away from zero takes the form of a half-open interval $[0, \bar{a})$. We denote by $\bar{x}(\bar{a}) := \lim_{t\to\infty} X_t = \int_0^\infty K_t \, \mathrm{d}t$ the *amount of exploration* under $\boldsymbol{k}$. In addition, we denote by $\bar{s}(\bar{a}) := \lim_{t\to\infty} S_t$ the *long-run technological standard*, which is defined as the quality of the best technology available as time approaches infinity. For a given Brownian path W, it is straightforward to see that both $\bar{s}(\bar{a})$ and $\bar{x}(\bar{a})$ depend only on the initial state (a, s) and the stopping threshold $\bar{a}$, and are independent of the intensity of exploration. Moreover, they are clearly nondecreasing in $\bar{a}$. We can explicitly express the prior belief on the distribution of $\bar{s}(\bar{a})$ as follows.

Lemma 2. *At state (a, s), for a strategy profile with stopping threshold $\bar{a}$, the long-run technological standard $\bar{s}(\bar{a})$ has the same distribution as* $\max\{s, M - a\}$*, where the random variable M has an exponential distribution with mean $(e^{\theta\bar{a}} - 1)/\theta$.*[8]

Recall from Section 2.1 that the first-best technology, denoted by $\hat{x}_N$, is the technology $x \geq 0$ that yields the highest payoff for N cooperative agents under complete information, taking the opportunity cost of its development into account. Denote by $q(\bar{a}) := \mathsf{P}_{as}(\hat{x}_N \in [0, \bar{x}(\bar{a})])$ the probability that the first-best technology will be explored in the long run under a strategy profile with stopping threshold $\bar{a}$.

Lemma 3. *At state (a, s), we have $q(\bar{a}) \to 1$ as $\bar{a} \to +\infty$.*

Therefore, the comparative statics in Corollary 1 imply that $\hat{x}_N$ will be developed under the cooperative solution with probability $q_N(a_N^*) \to 1$ as the team size grows.[9]

4 The Strategic Problem

From now on, we assume that there are $N > 1$ players acting noncooperatively. We study equilibria in the class of s-invariant Markov strategies, which are the Markov strategies with the gap as the state variable and will hereafter be referred to as *Markov strategies*. In this section, we provide characterization of best responses and the associated payoff functions, which help us to establish useful properties of the equilibria.

[8]For $\theta = 0$, we take $\lim_{\theta\to 0}(e^{\theta\bar{a}} - 1)/\theta = \bar{a}$ for the mean of M.

[9]Note that for a given W, the first-best technology $\hat{x}_N$ could depend on the team size N. However, the convergence in Lemma 3 does not require a constant sequence of parameters, provided that the parameters satisfy Assumption 1 along the sequence.

4.1 Best Responses and Equilibria

We denote by $\mathcal{K}$ the set of Markov strategies that are right-continuous and piecewise Lipschitz-continuous, and denote by $\mathcal{A}$ the space of admissible control processes as in Section 2.[10] A strategy $k_n^* \in \mathcal{K}$ for player n is a best response against her opponents' strategies $\boldsymbol{k}_{\neg n} = (k_1, \ldots, k_{n-1}, k_{n+1}, k_N) \in \mathcal{K}^{N-1}$ if

$$v_n(a, s | k_n^*, \boldsymbol{k}_{\neg n}) = \sup_{k_n \in \mathcal{A}} v_n(a, s | k_n, \boldsymbol{k}_{\neg n})$$

at each state $(a, s) \in \mathbb{R}_+ \times \mathbb{R}$. This turns out to be equivalent to

$$u_n(a | k_n^*, \boldsymbol{k}_{\neg n}) = \sup_{k_n \in \mathcal{K}} u_n(a | k_n, \boldsymbol{k}_{\neg n})$$

for each gap $a \geq 0$, with the normalized payoff function $u_n(a|\boldsymbol{k}) = e^{-s} v_n(a, s|\boldsymbol{k})$ defined as in Section 2. A Markov perfect equilibrium is a profile of Markov strategies that are mutually best responses.

Denote the intensity of exploration carried out by player n's opponents by $K_{\neg n}(a) = \sum_{l \neq n} k_l(a)$, and the benefit-cost ratio of exploration by $\beta(a, u_n)$ as in Section 3. The following lemma characterizes all MPE in the exploration game.

Lemma 4 (Equilibrium Characterization)**.** *A strategy profile $\boldsymbol{k} = (k_1^*, \ldots, k_N^*) \in \mathcal{K}^N$ is a Markov perfect equilibrium with $u_n : \mathbb{R}_+ \to \mathbb{R}$ being the corresponding payoff function of player n, if and only if for each $n \in \{1, \ldots, N\}$, function u_n*

1. *is once continuously differentiable on $\mathbb{R}_{++}$;*

2. *is piecewise twice continuously differentiable on $\mathbb{R}_{++}$;*[11]

3. *satisfies the normal reflection condition*

$$u_n(0) + u_n'(0+) = 0; \tag{4}$$

4. *satisfies, at each continuity point of u_n'', the HJB equation*

$$u_n(a) = 1 + K_{\neg n}(a)\beta(a, u_n) + \max_{k_n \in [0,1]} k_n\{\beta(a, u_n) - 1\}, \tag{5}$$

[10] Piecewise Lipschitz-continuity means that $\mathbb{R}_+$ can be partitioned into a finite number of intervals such that the strategy is Lipschitz-continuous on each of them. This rules out the infinite-switching strategies considered in Section 6.2 of Keller, Rady, and Cripps (2005).

[11] This condition means that there is partition of $\mathbb{R}_{++}$ into a finite number of intervals such that u_n'' is continuous on the interior of each of them.

with $k_n^*(a)$ achieving the maximum on the right-hand side, i.e.,

$$k_n^*(a) \in \underset{k_n \in [0,1]}{\arg\max}\, k_n\{\beta(a, u_n) - 1\}.$$

These conditions are standard in optimal control problems. Condition 4 and the smooth pasting condition, which is implicitly stated in Condition 1, are the optimality conditions. The rest are properties for general payoff functions.

In any MPE, Lemma 4 provides the following characterization of best responses. If $\beta(a, u_n) < 1$, then $k_n^*(a) = 0$ is optimal and $u_n(a) = 1 + K_{\neg n}(a)\beta(a, u_n) < 1 + K_{\neg n}(a)$. If $\beta(a, u_n) = 1$, then the optimal $k_n^*(a)$ takes arbitrary values in $[0, 1]$ and $u_n(a) = 1 + K_{\neg n}(a)$. Finally, if $\beta(a, u_n) > 1$, then $k_n^*(a) = 1$ is optimal and $u_n(a) = (1 + K_{\neg n}(a))\beta(a, u_n) > 1 + K_{\neg n}(a)$. In short, player n's best response to a given intensity of exploration $K_{\neg n}$ by the others depends on whether u_n is greater than, equal to, or less than $1 + K_{\neg n}$.

On intervals where each k_n is continuous, HJB equation (5) gives rise to the ODE

$$u_n(a) = 1 - k_n(a) + K(a)\beta(a, u_n). \tag{6}$$

In particular, on the intervals where each k_n is constant, the ODE above admits the explicit solution

$$U(a) = 1 - k_n + C_1 e^{\tilde{\gamma}_1 a} + C_2 e^{\tilde{\gamma}_2 a}, \tag{7}$$

where $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ are the roots of the equation $\gamma(\gamma - \theta) = r/K$, and C_1, C_2 are constants to be determined.

Lastly, on intervals where an interior allocation is chosen, the ODE from the indifference condition $\beta(a, u_n) = 1$ has the general solution

$$U(a) = \begin{cases} C_1 + C_2 a + ra^2/2, & \text{if } \theta = 0, \\ C_1 + C_2 e^{\theta a} - ra/\theta, & \text{if } \theta \neq 0, \end{cases} \tag{8}$$

where C_1, C_2 are constants to be determined.

4.2 Properties of MPE

First note that in any MPE, the average payoff can never exceed the N-player cooperative payoff U_N^*, and no individual payoff can fall below the single-agent payoff U_1^*. The upper bound follows directly from the fact that the cooperative solution maximizes the average payoff. The lower bound U_1^* is guaranteed by playing the single-agent optimal strategy, as the players can only benefit from the exploration efforts exerted by others.

Second, all Markov perfect equilibria are inefficient. Along the efficient exploration path, the benefit of exploration tends to $1/N$ of its opportunity cost as the development process approaches the efficient stopping threshold. A self-interested player thus has an incentive to deviate to exploitation whenever the benefit of exploration drops below its full opportunity cost.

Also note that in any MPE, the set of states at which the intensity of exploration is positive must be an interval $[0, \bar{a})$ with $a_1^* \leq \bar{a} \leq a_N^*$. The bounds on the stopping threshold follow directly from the bounds on the average payoffs and imply that the long-run outcomes in any MPE cannot outperform the cooperative solution.

Corollary 2. *In any Markov equilibrium with stopping threshold $\bar{a}$, at any state (a, s) we have $\bar{s}(\bar{a}) \leq \bar{s}(a_N^*)$ almost surely, and $q(\bar{a}) \leq q(a_N^*)$.*

Moreover, the intensity of exploration must be bounded away from zero on any compact subset of $[0, \bar{a})$. If this were not the case, there would exist some gap $a < \bar{a}$ such that the development process $\{Y_t\}_{t \geq 0}$ starting from $Y_0 = s - a$ would never reach the best-known quality s because of diminishing intensity, and therefore allocating a positive fraction of resource to exploration at gap a is clearly not optimal for any player.

In the two-armed bandit models discussed in Section 1.1, the players always use the risky arm at beliefs higher than some myopic cutoff, above which the expected short-run payoff from the risky arm exceed the deterministic flow payoff from the safe arm. Because our model lacks such a myopic cutoff, it might seem reasonable to conjecture that some player n, with her payoff function u_n bounded from above by $1 + K_{\neg n}$, never explores, and thus free-rides the technologies developed by the other players. Such a conjecture is refuted by the following proposition.

Proposition 2 (No Player Always Free-rides)**.** *In any Markov perfect equilibrium, no player allocates resources exclusively to exploitation for all gaps.*

The intuition behind this result is that in equilibrium, the cost and benefit of exploration must be equalized at the stopping threshold for each player, whereas any player who never explores would find that this benefit outweighs the cost, manifested by a kink in her payoff function at the stopping threshold. She would then have a strict incentive to resume exploration immediately after the other players give up, hoping to reduce the gap to bring the development process back alive. Therefore, in equilibrium, every player must perform exploration at some states, respecting the smooth pasting condition (Condition 1 in Lemma 4).

This result shows that each player strictly benefits from the presence of the other players in equilibrium, which in turn encourages some of the players to explore at gaps larger than their single-agent cutoffs. Such an encouragement effect is exhibited in all MPE of the exploration game.

Proposition 3 (Encouragement Effect)**.** *In any Markov perfect equilibrium, at least one player explores at gaps above the single-agent cutoff* a_1^*.

With the same intuition as the encouragement effect, our last general result on Markov perfect equilibria concerns the nonexistence of equilibria where all players use cutoff strategies.

Proposition 4 (No MPE in Cutoff Strategies)**.** *In any Markov perfect equilibrium, at least one player uses a strategy that is not of the cutoff type.*

Next we look into Markov perfect equilibria in greater depth.

5 Symmetric Equilibrium

Our characterization of best responses and the nonexistence of MPE in cutoff strategies suggest that in any symmetric equilibrium, the players choose an interior allocation at some states. At these states of interior allocation, the benefit of exploration must be equal to the opportunity cost, and therefore the common payoff function solves the ODE $\beta(a, u) = 1$. As a consequence of equation (6), the payoff of each player at the states of interior allocation in the symmetric equilibrium must also satisfy $u = 1 + K_{\neg n} \leq N$. Therefore, whenever the common payoff exceeds N, each player allocates resources exclusively to exploration, and the payoff function satisfies the same ODE (3) as in the cooperative solution. However, it is worth pointing out that for some configurations of the parameters, because of the strength of free-riding incentives among the players, the common payoff could be below N for all gaps, and accordingly, the resource constraint of each player is not necessarily binding even when the gap is zero, in marked contrast to the cooperative solution.[12] Lastly, the common payoff satisfies $u = 1$ at the gaps for which the resource is exclusively allocated to exploitation.

The solutions to the corresponding ODEs provided in equations (7) and (8), together with the normal reflection condition (4) and the smoothness requirement on the

[12]Exploration at a zero gap immediately leads to innovation and thus has an "innovation benefit" that strictly dominates the benefit of exploitation. It might therefore seem puzzling why any player would not explore with full intensity at a zero gap. The reason is that, heuristically speaking, the innovation benefit can be achieved by *any* positive intensity of exploration at a zero gap. In the single-agent problem, the player can only reap such an innovation benefit by her own effort and thus would never stop exploration when the gap is zero. In the strategic problem, however, from each player's own perspective, the innovation benefit is secured by the others' effort whenever $K_{\neg n}(0) > 0$. In the extreme cases, it is even possible for some players to allocate all resources to exploitation at a zero gap. This happens in the asymmetric equilibria for some configurations of the parameters in the next section.

equilibrium payoff functions, uniquely pin down the strategies and the associated payoff functions in the symmetric equilibrium, which can be expressed in closed form as follows.

Proposition 5 (Symmetric Equilibrium)**.** *The N-player exploration game has a unique symmetric Markov perfect equilibrium with the gap as the state variable. There exists a stopping threshold $\tilde{a} \in (a_1^*, a_N^*)$ and a full-intensity threshold $a^\dagger \geq 0$ such that the fraction of resource $k^\dagger(a)$ that each player allocates to exploration at gap a is given by*

$$k^\dagger(a) = \begin{cases} 0, & \text{on } [\tilde{a}, +\infty), \\ \frac{r}{N-1} \int_0^{\tilde{a}-a} \phi_\theta(z)\,\mathrm{d}z \in (0,1), & \text{on } [a^\dagger, \tilde{a}), \\ 1, & \text{on } [0, a^\dagger) \text{ if } a^\dagger > 0, \end{cases} \tag{9}$$

with $\phi_\theta(z) := \frac{1-e^{-\theta z}}{\theta}$.[13]

The corresponding payoff function is the unique function $U^\dagger : \mathbb{R}_+ \to [1, +\infty)$ of class C^1 with the following properties: $U^\dagger(a) = 1$ on $[\tilde{a}, +\infty)$; $U^\dagger(a) = 1 + (N-1)k^\dagger(a) \in (0, N)$ and solves the ODE $\beta(a, u) = 1$ on $(a^\dagger, \tilde{a})$; if $a^\dagger > 0$, then $U^\dagger(a) > N$ and solves the ODE $\beta(a, u) = u/N$ on $(0, a^\dagger)$.

The closed-form expressions for the common payoff function $U^\dagger$ and the thresholds $a^\dagger$ and $\tilde{a}$ are provided in the Appendix.

As we have already pointed out, depending on the parameters, it is possible that $a^\dagger = 0$, in which case $k^\dagger(a) < 1$ for all $a > 0$, and hence will be referred to as the *non-binding case*. The opposite case, where $a^\dagger > 0$, will be referred to as the *binding case*. Figure 1 illustrates the symmetric equilibrium for these two cases.

As the outcomes are publicly observed and newly developed technologies are freely available, the players have incentives to free-ride. Such a *free-rider effect* can be seen more clearly from the comparison between the benefit-cost ratio of exploration at the states of interior allocation in the symmetric MPE

$$\beta(a, U^\dagger) = 1,$$

and the one in the cooperative solution

$$\beta(a, U^*) = U^*(a)/N.$$

Exploration in equilibrium thus requires the benefit of exploration to cover the cost, whereas the efficient strategy entails exploration at states where the cost exceeds the

[13] We define $\phi_0(z) := \lim_{\theta \to 0} \phi_\theta(z) = z$.

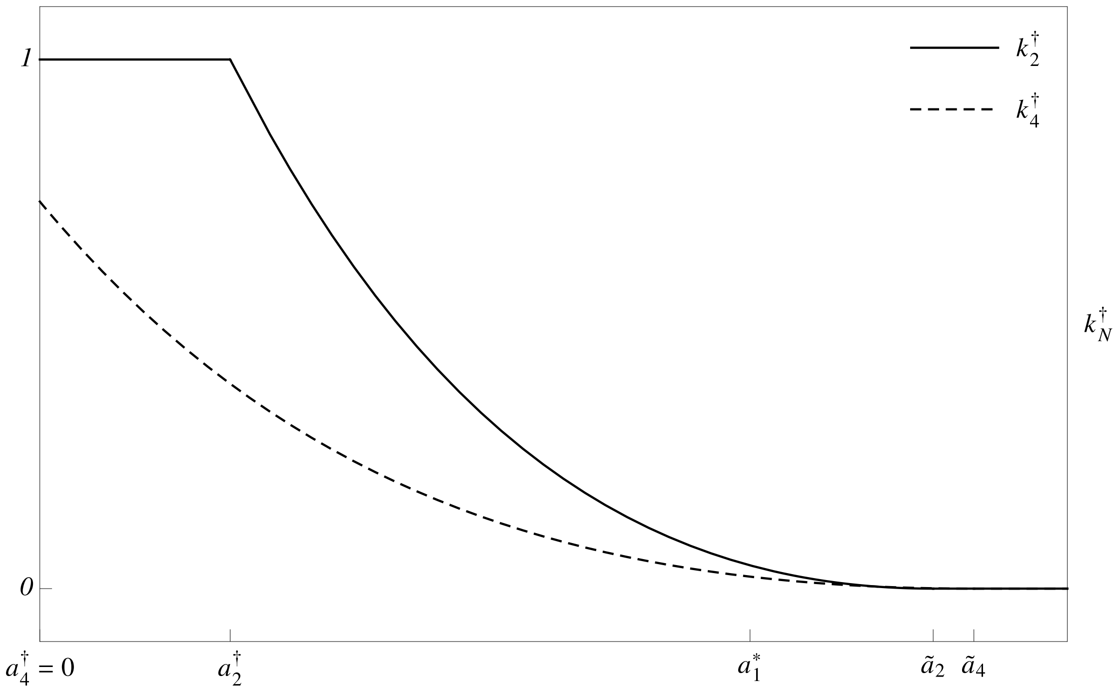


Figure 1. The symmetric equilibrium with binding resource constraints in a two-player game, and with non-binding constraints in a four-player game ($r = 1/2$, $\theta = -1$).

benefit, as $\beta(a, U^*) < 1$ whenever $U^*(a) < N$. Figure 2 illustrates the comparison between the common payoff function in the symmetric equilibrium and the cooperative solution in a two-player exploration game.

5.1 Comparative Statics

In this section we examine the comparative statics of the symmetric equilibrium with respect to the discount rate r and the number of players N.[14]

Corollary 3 (Effect of r)**.** *The stopping threshold $\tilde{a}_r$ is strictly decreasing in r, and the full-intensity threshold $a_r^\dagger$ is weakly decreasing in r. For any gap $a \geq 0$, the equilibrium strategy $k_r^\dagger(a)$ and the common payoff $U_r^\dagger(a)$ are weakly decreasing in r.*

As r decreases, the players becomes more patient and have greater incentives for exploration. Moreover, the increased exploration efforts of others encourage each player to raise their own effort further.

The common payoff is decreasing in r for two reasons. First, higher patience increases players' payoffs directly. Second, as in the cooperative solution, the increased

[14]The effect of discount rate r on the payoff U_r carries over to U_r/r. In other words, the following comparative statics with respect to r are not driven by the normalizing constant r in the flow payoff.

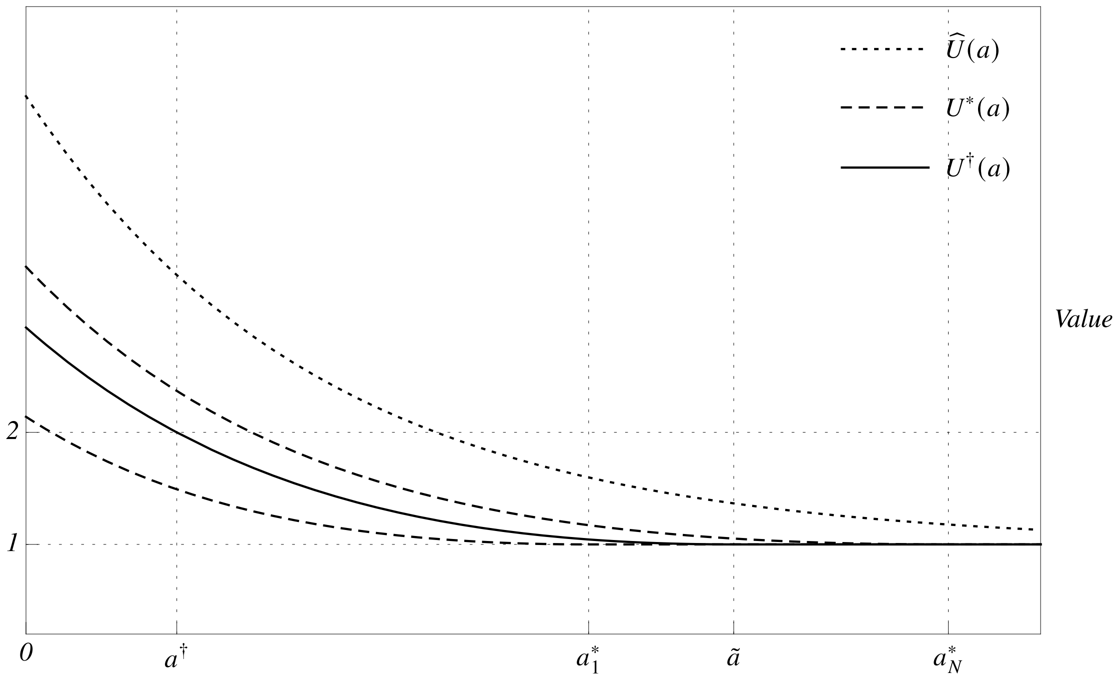


Figure 2. From top to bottom: complete-information payoff $\widehat{U}_N$, average payoff U_N^* in the cooperative solution, common payoff $U_N^\dagger$ in the symmetric equilibrium, and payoff in the single-agent optimum U_1^*. Parameter values: $r = 1/2$, $\theta = -1$, $N = 2$.

patience raises players' tolerance for failure $\tilde{a}_r$, and hence more advanced technologies will be developed and adopted in the long run.

Corollary 4 (Effect of N)**.** *On $\{ N \geq 1 \mid a_N^\dagger > 0 \}$, which is the range of N for which the players' resource constraints are binding in the symmetric equilibrium, the stopping threshold $\tilde{a}_N$ is strictly increasing in N and the common payoff function $U_N^\dagger$ is weakly increasing in N. Whereas on $\{ N \geq 1 \mid a_N^\dagger = 0 \}$, both $\tilde{a}_N$ and $U_N^\dagger$ are constant over N, and the equilibrium strategy $k_N^\dagger$ is weakly decreasing in N.*

In the binding case, notice that $k_N^\dagger(a)$ is not monotone in N because the full-intensity threshold $a_N^\dagger$ could be decreasing in N. This occurs when the free-rider effect outweighs the encouragement effect. On the one hand, extra encouragement brought by additional players raises the stopping threshold $\tilde{a}_N$. On the other hand, the increased free-riding incentives due to extra players tighten the resource binding requirement that $u > N$, which enlarges the region $(a_N^\dagger, \tilde{a}_N)$ of interior allocation. The total effect of increasing N on the intensity of exploration is determined by these two competing forces and hence is not monotone in N.

In the non-binding case, as N increases, each player adjusts their individual intensity of exploration downward, maintaining the same equilibrium payoff. This is a situation where the incentive to free-ride is so strong that it completely offsets further encouragement brought by additional players, and even the overall intensity of exploration $Nk^\dagger$ is decreasing in N for the gaps in $[0, \tilde{a}_N)$. Thus, free-riding slows down the development process considerably. In the worst scenario, the overall intensity when the gap is zero could even be lower than that in the single-agent problem.

Also notice that whenever the resource constraints are not binding, the full-intensity threshold $a_N^\dagger$ remains constant at zero for any further increase in N because $k_N^\dagger$ would be even lower. Therefore, if $a_N^\dagger$ ever hits zero as N goes up, the resource constraints in the symmetric equilibrium remain non-binding for any larger N.

As we have seen, depending on whether or not the resource constraints are binding in the symmetric equilibrium, a larger team size can have qualitatively different effects on the welfare and long-run outcomes. If the resource constraints are not binding in equilibrium, any extra resources brought by additional players translate entirely to free-riding, which results in a highly inefficient outcome in a large team in terms of average payoffs, the likelihood of developing the first-best technology, and the technological standard in the long run. The question then naturally arises of whether the resource constraints would ever fail to be binding in the symmetric equilibrium as N goes up. Or, conversely, would the encouragement effect eventually overcome the free-rider effect?

To investigate this question, we now examine the effect on the symmetric equilibrium as N increases toward infinity, while keeping the other parameters fixed. For ease of exposition, we allow $N \geq 1$ to take non-integral values and drop Assumption 1 for the rest of this section.[15]

Corollary 5 (Asymptotic Effect of N)**.** *If $\theta > -1$ and $r < \hat{r} := \theta^2/(\theta - \ln(1+\theta))$, then we have $U_N^\dagger \to +\infty$, $a_N^\dagger \to +\infty$, and $q_N(\tilde{a}_N) \to 1$ as $N \to r/(1+\theta) \in (1, +\infty)$; if instead $\theta \leq -1$ or $r \geq \hat{r}$, we have $a_N^\dagger = 0$ for sufficiently large N, and we have $\lim U_N^\dagger(a) < \lim U_N^*(a)$ for each $a \geq 0$, $\tilde{a}_N$ is bounded, and $q_N(\tilde{a}_N)$ is bounded away from 1 as $N \to +\infty$.*

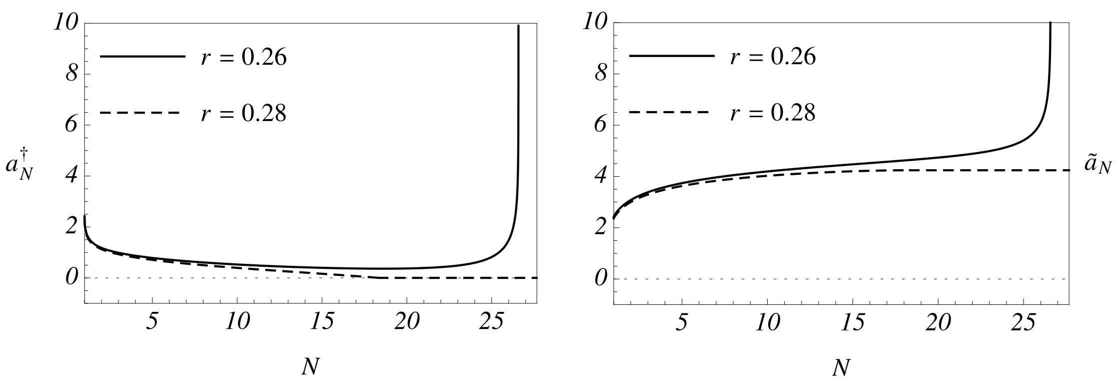
Corollary 5 highlights the key role of the prospect of innovation in the strength of the encouragement effect, which has been largely missing in literature of strategic experimentation. The absence of technological advancements is partly responsible for the prevalence of free-riding in two-armed bandit models. Our result suggests that allowing for technological advancements is crucial for understanding the incentives for experimentation and technology development in dynamic environments. Next we discuss this result for $\theta < -1$ and $\theta \geq -1$ separately.

When $\theta < -1$, recall that the cooperative payoff converges to the finite limit of

[15] Caution needs to be taken in the case of $\theta > -1$, in which Assumption 1 could be violated for large N, and the existence and uniqueness of the symmetric equilibrium is no longer guaranteed.

the complete-information payoff as $N \to +\infty$, with the cooperative cutoff $a^*_N \to +\infty$ and thus the long-run probability of developing the first-best technology $q_N(a^*_N) \to 1$. In contrast, the symmetric equilibrium always falls into the non-binding category for large N. After the full-intensity threshold $a^\dagger_N$ reaches zero, the encouragement effect is completely frozen by the free-rider effect: The stopping threshold $\tilde{a}_N$ remains constant for any further increase of N, so both the amount and the rate of exploration are highly inefficient for large N. Consequently, as $N \to +\infty$, the average payoff is bounded away from the cooperative payoff in the limit, and $q_N(\tilde{a}_N)$ is bounded away from 1.

On the other hand, when $\theta \geq -1$, recall that in addition to the divergence of a^*_N, the cooperative payoff tends to infinity as $N \to r/(1+\theta) \in (1, +\infty]$. Whether this is the case for the symmetric equilibrium now depends on the patience of the players, as illustrated in Figure 3. If the players are not patient enough (large r), the symmetric equilibrium always falls into the non-binding category for large N, same as the case of $\theta < -1$. However, if the players are sufficiently patient, the resource constraints are always binding in the symmetric equilibrium. Not only does the stopping threshold $\tilde{a}_N$ tend to infinity, but the full-intensity threshold $a^\dagger_N$ does as well. So the encouragement effect eventually overcomes the free-rider effect when the players are patient enough and the prospect of development is sufficiently promising. Both the amount and the rate of exploration resemble the cooperative solution for large N. As a consequence, $q_N(\tilde{a}_N) \to 1$ and the average equilibrium payoff tends to infinity as N goes up, bearing a resemblance to the cooperative payoff.



(a) Full-intensity threshold $a^\dagger_N$

(b) Stopping threshold $\tilde{a}_N$

Figure 3. Full-intensity thresholds and the stopping thresholds in the symmetric equilibrium ($\theta = -0.09$) for different discount rate r. If the players are sufficiently patient (solid curves), resources constraints are binding ($a^\dagger_N > 0$) for all N. Otherwise (dashed curve), resources constraints are not binding ($a^\dagger_N = 0$) for sufficiently large N.

Lastly, as yet another manifestation of the strength of the free-rider effect, in the case of $\theta > -1$, Corollary 5 asserts that symmetric equilibrium with finite payoff exists for all N provided that $r \geq \hat{r}$, standing in marked contrast to the cooperative solution, which fails to exist due to unbounded payoffs for large N.

6 Asymmetric Equilibria and Welfare Properties

Note that in the symmetric equilibrium, the intensity of exploration dwindles down to zero as the gap approaches the stopping threshold. As a result, the threshold is never reached and exploration never fully stops. This suggests that welfare can be improved if the players take turns between the roles of explorer and free-rider, keeping the intensity of exploration bounded away from zero until all exploration stops. In this section we investigate this possibility by constructing a class of asymmetric Markov perfect equilibria.

6.1 Construction of Asymmetric Equilibria

Our construction of asymmetric MPE is based on the idea of the asymmetric MPE proposed in Keller and Rady (2010). We let the players adopt the common actions in the same way as in the symmetric equilibrium whenever the resulting average payoff is high enough to induce an overall intensity of exploration greater than one, and let the players take turns exploring at more pessimistic states in order to maintain the overall intensity at one. Such alternation between the roles of explorer and free-rider leads to an overall intensity of exploration higher than in the symmetric equilibrium, yielding higher equilibrium payoffs.

Here we briefly address the two main steps in our construction. In the first step, we construct the average payoff function $\bar{u}$. We let $\bar{u}$ solve the same ODE $\beta(a, u) = \max\{u(a)/N, 1\}$ as the common payoff function in the symmetric equilibrium whenever $u > 2 - 1/N$, so that the corresponding overall intensity is $Nk(a) = N(u(a) - 1)/(N - 1) \geq 1$. Whenever $1 < u < 2 - 1/N$, we let $\bar{u}$ solve the ODE $u(a) = 1 - 1/N + \beta(a, u)$, which is the ODE for the average payoff function among N players associated with an overall intensity $K = 1$. The boundary conditions for the average payoff function $\bar{u}$, namely the smooth pasting condition at the stopping threshold and the normal reflection condition (4), are identical to the conditions in Lemma 4, simply because these conditions remain unchanged after taking the average. The unique solution of class $C^1(\mathbb{R}_{++})$ to the ODE above serves as the average payoff function, which also gives thresholds $a^\flat > a^\sharp \geq 0$ such that $\bar{u} = 1$ on $[a^\flat, +\infty)$, $1 < \bar{u} < 2 - 1/N$ on $(a^\sharp, a^\flat)$ and $\bar{u} > 2 - 1/N$ on $[0, a^\sharp)$. In the second step, equilibrium-compatible actions are assigned to each player. On $[0, a^\sharp)$, if it is nonempty, we let the players adopt the common action

$k_n(a) = \min\{(\bar{u}(a) - 1)/(N - 1), 1\}$ in the same way as in the symmetric equilibrium. On $[a^\sharp, a^\flat)$, players alternate between the roles of explorer and free-rider so as to keep the overall intensity at one. We first split $[a^\sharp, a^\flat]$ into subintervals in an arbitrary way and then meticulously choose the switch points of their actions, so that all individual payoff functions have the same values and derivatives as the average payoff function at the endpoints of these subintervals.[16] Lastly, our characterization of MPE in Lemma 4 confirms that the assigned action profile is compatible with equilibrium. We leave the method for choosing the switch points and further details to the Appendix.

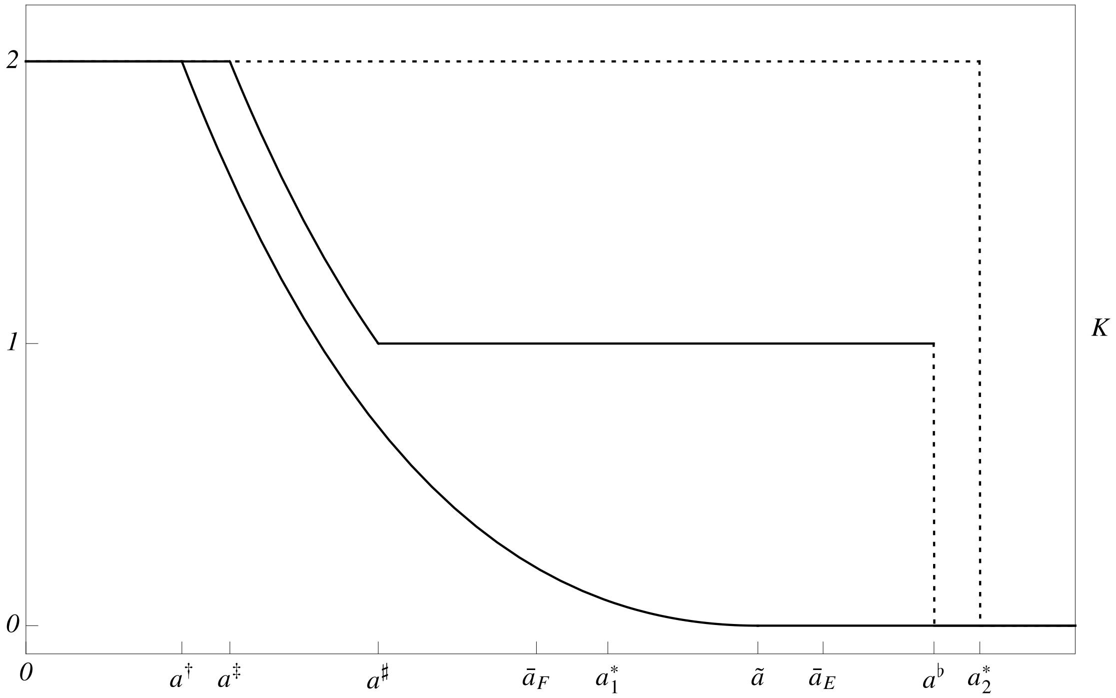


Figure 4. From top to bottom: intensity of exploration in the cooperative solution, the asymmetric equilibria, and the symmetric equilibrium ($r = 1/2$, $\theta = -1$, $N = 2$).

For $N = 2$, Figure 4 illustrates the intensity of exploration in the asymmetric MPE, compared with the symmetric equilibrium. We can see that the resource constraints are binding in the depicted equilibria, but it is worth bearing in mind that it may well be the other way around for different parameters. For example, if the the players are too impatient, the average payoff function could be bounded by $2-1/N$ from above, resulting in an intensity of exploration equal to 1 on the entire region $[0, a^\flat)$. The states $\bar{a}_F$ and $\bar{a}_E$ in the figure demarcate the switch points at which these two players swap roles when they take turns exploring on $[a^\sharp, a^\flat)$. The volunteer explores on $[a^\sharp, \bar{a}_F) \cup [\bar{a}_E, a^\flat)$,

[16]In fact, this technique can be used to construct asymmetric equilibria with strategies that take values in $\{0, 1\}$ only, which are referred to as simple equilibria in Keller, Rady, and Cripps (2005). See Proposition 10 in the Appendix. However, it is not clear whether such equilibria achieve higher average payoffs than the symmetric equilibrium.

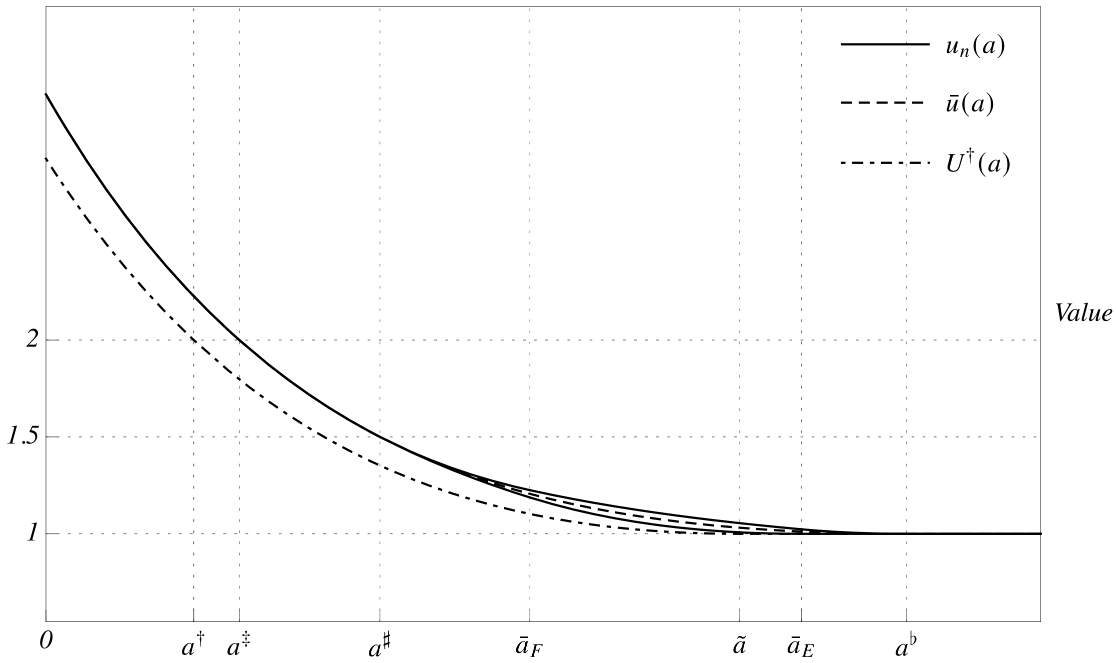


Figure 5. Average payoff and possible individual payoffs in the best two-player asymmetric equilibria, compared to the common payoff in the symmetric equilibrium ($r = 1/2$, $\theta = -1$, $N = 2$).

whereas the free-rider explores on $[\bar{a}_F, \bar{a}_E)$. These switch points are chosen in a way that ensures their individual payoff functions are of class $C^1(\mathbb{R}_{++})$ and coincide on $[0, a^{\sharp})$.

Figure 5 illustrates the associated average payoff function (dashed curve) and the individual payoff functions (solid curves) that can arise in the equilibria in a two-player game, compared with the common payoff function in the symmetric equilibrium (solid dotted curve). Note that the payoff function of the volunteer is strictly higher than that of the free-rider at the states immediately to the left of the stopping threshold $a^{\flat}$. In fact, the free-rider has a payoff equal to 1 on $[\bar{a}_E, a^{\flat})$. This observation stands in marked contrast to the models in Keller, Rady, and Cripps (2005) and Keller and Rady (2010), where the volunteer is worse off on this region. The intuition behind this feature is similar to that in Proposition 2. A kink has to be created at $a^{\flat}$ in the free-rider's payoff function in order to attain a higher payoff than the volunteer's at states immediately to the left of $a^{\flat}$, because the free-rider's ODE $u(a) = 1 + \beta(a, u)$ must be satisfied.[17] In such a case, the free-rider has a strict incentive to take over the role of volunteer to kickstart the development process at a larger gap. Therefore, in equilibrium, the volunteer must be compensated for acting as a lone explorer at pessimistic states by bearing relatively less burden at more optimistic states.

[17]Even though the free-rider has a payoff equal to 1 around $a^{\flat}$, she still benefits from free-riding on this region. This benefit, however, is offset by the relatively high burden of exploration effort she must bear in equilibrium at more optimistic states to reward the volunteer.

For arbitrary N, we have the following result.

Proposition 6 (Asymmetric MPE)**.** *The N-player exploration game admits Markov perfect equilibria with thresholds $0 \le a^{\ddagger} \le a^{\sharp} < a^{\flat} < a^{*}$, such that on $[0, a^{\sharp}]$, the players have a common payoff function; on $[0, a^{\ddagger}]$, all players choose exploration exclusively; on $(a^{\ddagger}, a^{\sharp})$, the players allocate a common interior fraction of the unit resource to exploration, and this fraction decreases in the gap; on $[a^{\sharp}, a^{\flat})$, the intensity of exploration equals 1 with players taking turns exploring on consecutive subintervals; on $[a^{\flat}, +\infty)$, all players choose exploitation exclusively. The intensity of exploration is continuous in the gap on $[0, a^{\flat})$. The average payoff function is strictly decreasing on $[0, a^{\flat}]$, once continuously differentiable on $\mathbb{R}_{++}$, and twice continuously differentiable on $\mathbb{R}_{++}$ except for the cutoff $a^{\flat}$. On $[0, a^{\flat})$, the average payoff is higher than in the symmetric MPE, and $a^{\flat}$ lies to the right of the threshold $\tilde{a}$ at which all exploration stops in that equilibrium.*

6.2 Welfare Results

For $N \ge 3$, further improvements can be easily achieved by letting the players take turns exploring, maintaining the intensity of exploration at K whenever $K < u < K + 1 - K/N$ for all $K \in \{1, \dots, N\}$, rather than for $K = 1$ only, as in the Proposition above. However, it is not clear whether such improvements achieve the highest welfare among all MPE of the N-player exploration game. For $N = 2$, the asymmetric equilibria of Proposition 6 are the best among all MPE.

Proposition 7 (Best MPE for $N = 2$)**.** *The average payoff in any Markov perfect equilibrium of the two-player exploration game cannot exceed the average payoff in the equilibria of Proposition 6.*

In the construction of the asymmetric MPE depicted in Figure 4, the interval $[a^{\sharp}, a^{\flat}]$ is not split into subintervals. This can be confirmed by the observation from Figure 5 that the players' payoff functions match values only at the endpoints of $[a^{\sharp}, a^{\flat}]$, not in the interior. Our construction allows an arbitrary partition on $[a^{\sharp}, a^{\flat}]$ during the splitting procedure, thus a trivial partition of $[a^{\sharp}, a^{\flat}]$, as in Figure 4, suffices. A finer partition, however, produces equilibria in which the players exchange roles more often, which allows them to share the burden of exploration more equally. Sufficiently frequent alternation of roles on $[a^{\sharp}, a^{\flat})$ guarantees each player a payoff close enough to the average payoff and thus yields a Pareto improvement over the symmetric equilibrium.[18]

[18]The payoffs of both players in the asymmetric MPE depicted in Figure 5 are higher than in the symmetric equilibrium on $[0, \bar{a}_E)$; however, this might not be the case in general when the trivial partition of $[a^{\sharp}, a^{\flat}]$ is used in the construction, as $\bar{a}_E$ could lie on the left of $\tilde{a}$.

Proposition 8 (Pareto Improvement over the Symmetric MPE)**.** *For any $\epsilon > 0$, the N-player exploration game admits Markov perfect equilibria as in Proposition 6 in which each player's payoff exceeds the symmetric equilibrium payoff on* $[0, a^\flat - \epsilon]$.

Recall that the symmetric equilibrium exhibits a limited encouragement effect when the players are too impatient. The reason is that the common payoff function on the interval of interior allocation must satisfy the ODE $\beta(a, U^\dagger) = 1$, which does not depend on N. As a result, the common payoff function in the symmetric equilibrium is constant over the team size when the resource constraints are not binding. By contrast, the average payoff always increases in the number of players in the asymmetric MPE in which the players take turns exploring at states immediately to the left of the stopping cutoff. This is because the burden of keeping the overall intensity at one at these states can be shared among more players in larger teams. As a result, the players would be able to exploit more often on average, which in turn encourages them to explore at more pessimistic states. Therefore, unlike the comparative statics of the symmetric equilibrium, the encouragement effect in the asymmetric equilibria is not frozen by the free-rider effect as N goes up, irrespective of the patience of the players and the prospect of the development.

Proposition 9. *The stopping cutoff $a^\flat_N$ in the asymmetric MPE of Proposition 6 goes to infinity as $N \to +\infty$.*[19]

Therefore, the amount of exploration, and thus the long-run outcomes, can be improved significantly over the symmetric equilibrium for large N by letting the players take turns exploring before the exploration fully stops. This positive result, unfortunately, does not extend as far to the welfare, as the rate of exploration might still be too low because of non-binding resource constraints, similar to the situation faced in the symmetric equilibrium. More precisely, when $\theta < -1$, the full-intensity threshold $a^\ddagger_N$ always hits 0 as $N \to +\infty$; when $\theta \geq 1$, whether this happens again depends on the patience of the players, just like in the symmetric equilibrium. It can be shown that the results regarding the average payoff and the full-intensity threshold in Corollary 5 extend to the asymmetric MPE of Proposition 6 with a larger $\hat{r}$.

7 Conclusion

This paper provides a novel and tractable framework for analyzing strategic interaction among a team of players conducting incremental experimentation. We show how welfare

[19]When $\theta > -1$, if the asymmetric MPE of Proposition 6 fail to exist due to unbounded payoffs for $N \geq r/(1+\theta)$, then by $N \to +\infty$ we actually mean $N \to r/(1+\theta)$.

and long-run outcomes depend on the strength of two competing forces—the free-rider effect and the encouragement effect.

Several limitations in our model are noteworthy. First, the scope of experimentation is restricted: Players do not have complete freedom to choose where to explore. Admittedly, this is indeed a simplifying assumption when our model is considered as one of the models of spatial experimentation as in Callander (2011) and Garfagnini and Strulovici (2016). Yet, incremental experimentation in our model can be alternatively viewed as an approximation of a sequential search process in which the search results follow a Markov process, with the search frequency being controlled by the players. In this sense, the restriction imposed on the scope of experimentation is no different from Markov assumption on the search results. Second, under the interpretation that the Brownian path is realized at the outset as in Callander (2011), no new direction of exploration exists around the feasible technologies explored in the past. Allowing for the opportunities to start exploration in new direction from past explored technologies would raise the prospect of the development and thus strengthen the encouragement effect. Third, in our model, all actions and outcomes are publicly observable. Hence, there are perfect knowledge spillovers between players, and the developed technologies are pure public goods. Extensions that allow for some degree of excludability of the technologies or friction in the diffusion of knowledge could be useful for applications in the industrial organization literature.

References

Admati, Anat R. and Motty Perry. "Joint Projects without Commitment". In: *The Review of Economic Studies* 58.2 (Apr. 1991), pp. 259–276.

Barbu, Viorel. "Existence and Uniqueness for the Cauchy Problem". In: *Differential Equations*. Cham: Springer International Publishing, 2016, pp. 29–77.

Bolton, Patrick and Christopher Harris. "Strategic Experimentation". In: *Econometrica* 67.2 (1999), pp. 349–374.

Callander, Steven. "Searching and Learning by Trial and Error". In: *American Economic Review* 101.6 (Oct. 2011), pp. 2277–2308.

Cetemen, D., C. Urgun, and L. Yariv. *Collective Progress: Dynamics of Exit Waves*. Working Paper. 2021.

Chung, K. L. and R. J. Williams. "Generalized Ito Formula, Change of Time and Measure". In: *Introduction to Stochastic Integration*. New York, NY: Springer New York, 2014, pp. 183–215.

Fleming, W. and H. Soner. "Optimal Control of Markov Processes: Classical Solutions". In: *Controlled Markov Processes and Viscosity Solutions*. New York, NY: Springer New York, 2006, pp. 119–149.

Garcia-Macia, Daniel, Chang-Tai Hsieh, and Peter J. Klenow. "How Destructive Is Innovation?" In: *Econometrica* 87.5 (2019), pp. 1507–1541.

Garfagnini, Umberto and Bruno Strulovici. "Social Experimentation with Interdependent and Expanding Technologies". In: *The Review of Economic Studies* 83.4 (Feb. 2016), pp. 1579–1613.

Georgiadis, George. "Projects and Team Dynamics". In: *The Review of Economic Studies* 82.1 (Sept. 2014), pp. 187–218.

Keller, Godfrey and Sven Rady. "Strategic experimentation with Poisson bandits". In: *Theoretical Economics* 5.2 (2010), pp. 275–311.

— "Breakdowns". In: *Theoretical Economics* 10.1 (2015), pp. 175–202.

Keller, Godfrey, Sven Rady, and Martin Cripps. "Strategic Experimentation with Exponential Bandits". In: *Econometrica* 73.1 (2005), pp. 39–68.

Lehoczky, John P. "Formulas for Stopped Diffusion Processes with Stopping Times Based on the Maximum". In: *The Annals of Probability* 5.4 (1977), pp. 601–607.

Marx, Leslie M. and Steven A. Matthews. "Dynamic Voluntary Contribution to a Public Project". In: *The Review of Economic Studies* 67.2 (Apr. 2000), pp. 327–358.

Peskir, Goran and Albert Shiryaev. "Methods of solution". In: *Optimal Stopping and Free-Boundary Problems*. Basel: Birkhäuser Basel, 2006, pp. 143–241.

Rogers, L. C. G. and David Williams. "Introduction to Itô Calculus". In: *Diffusions, Markov Processes and Martingales*. 2nd ed. Vol. 2. Cambridge Mathematical Library. Cambridge University Press, 2000, pp. 1–109.

Shreve, Steven E. *Stochastic Calculus for Finance II*. Springer New York, 2004.

Strulovici, Bruno and Martin Szydlowski. "On the smoothness of value functions and the existence of optimal strategies in diffusion models". In: *Journal of Economic Theory* 159 (2015). Symposium Issue on Dynamic Contracts and Mechanism Design, pp. 1016–1055.

Taylor, Howard M. "A Stopped Brownian Motion Formula". In: *The Annals of Probability* 3.2 (1975), pp. 234–246.

Thomke, S. H. "Experimentation Works: The Surprising Power of Business Experiments". In: Harvard Business Review Press, 2020.

Urgun, C. and L. Yariv. *Retrospective Search: Exploration and Ambition on Uncharted Terrain*. Working Paper. Department of Economics, Princeton University, 2021.

Wong, Yu Fu. *Spatial Experimentation*. Working Paper. Department of Economics, Columbia University, 2021.

Yildirim, Huseyin. "Getting the Ball Rolling: Voluntary Contributions to a Large-Scale Public Project". In: *Journal of Public Economic Theory* 8.4 (2006), pp. 503–528.

Appendices

In this Appendix, we let $\theta := 2\mu/\sigma^2$ and $\rho := \sigma^2/(2r)$, where $\mu \in \mathbb{R}$ denotes the drift of the Brownian path W_+, without normalizing the volatility σ to $\sqrt{2}$. Consequently, we replace λ in Lemma 1 with $\lambda := 1/(N\rho) - \theta$. Moreover, we let $\delta := r/N$ for the sake of notational convenience. Assumption 1, that $\lambda > 1$, is maintained throughout this Appendix unless it is explicitly stated otherwise.

A Explicit Representation of the Symmetric MPE

Corollary 6. *The explicit representation for the normalized payoff function $U^\dagger$ in the unique symmetric equilibrium on $[a^\dagger, \tilde{a})$ is given by*

$$U^\dagger(a) = \begin{cases} 1 + \frac{1}{2\rho}(\tilde{a} - a)^2, & \text{if } \theta = 0, \\ 1 + \frac{1}{\rho\theta}\left(\tilde{a} - a + \frac{1}{\theta}\left(e^{-\theta(\tilde{a}-a)} - 1\right)\right), & \text{if } \theta \neq 0. \end{cases}$$

If $a^\dagger > 0$, $U^\dagger$ on $[0, a^\dagger)$ is given by

$$U^\dagger(a) = \frac{N}{\gamma_2 - \gamma_1}\left((\gamma_2 + \iota)e^{-\gamma_1(a^\dagger - a)} - (\gamma_1 + \iota)e^{-\gamma_2(a^\dagger - a)}\right),$$

with $\gamma_1 < \gamma_2$ being the roots of the equation $\gamma(\gamma - \theta) = \frac{1}{N\rho}$ and $\iota > 0$ being

$$\iota := \begin{cases} \frac{1}{N\rho\theta}\left(1 + W_0\left(-\exp(-1 - (N-1)\rho\theta^2)\right)\right), & \text{if } \theta > 0, \\ \sqrt{\frac{2}{N\rho}(1 - 1/N)}, & \text{if } \theta = 0, \\ \frac{1}{N\rho\theta}\left(1 + W_{-1}\left(-\exp(-1 - (N-1)\rho\theta^2)\right)\right), & \text{if } \theta < 0, \end{cases}$$

where W_0 and W_{-1} denote the two real branches of the Lambert W function.

If $\iota < 1$, then the equilibrium belongs to the binding case. The full-intensity threshold is given by

$$a^\dagger = \frac{1}{\gamma_2 - \gamma_1}\left(\ln\left(\frac{1 + \gamma_2}{\iota + \gamma_2}\right) - \ln\left(\frac{1 + \gamma_1}{\iota + \gamma_1}\right)\right),$$

and the stopping threshold is given by

$$\tilde{a} = a^{\dagger} + N\rho(\iota + \theta(1 - 1/N)).$$

If $\iota \geq 1$, then the equilibrium belongs to the non-binding case with the full-intensity threshold $a^{\dagger} = 0$, whereas the stopping threshold $\tilde{a}$ is given by

$$\tilde{a} = \begin{cases} \frac{1}{\theta}\left(1 + \theta - \theta^2\rho + W_0\left(-(1+\theta)e^{-1-\theta+\theta^2\rho}\right)\right), & \text{if } \theta < 0, \\ 1 - \sqrt{1 - 2\rho}, & \text{if } \theta = 0, \\ \frac{1}{\theta}\left(1 + \theta - \theta^2\rho + W_{-1}\left(-(1+\theta)e^{-1-\theta+\theta^2\rho}\right)\right), & \text{if } \theta > 0. \end{cases}$$

Proof. From Lemma 4 and explicit calculations. □

B Properties of Payoff Functions

Lemma 5 (Homogeneity)**.** *Player n's payoff function for an s-invariant Markov strategy profile $\boldsymbol{k} \in \mathcal{K}^N$ can be written as $v_n(a, s|\boldsymbol{k}) = e^s v_n(a, 0|\boldsymbol{k})$.*

Proof. At state (a, s), player n's payoff associated to $\boldsymbol{k}$ is given by

$$\begin{aligned} v_n(a, s|\boldsymbol{k}) &= \mathsf{E}\left[\int_0^{\infty} re^{-rt}(1 - k_n(A_t))e^{S_t}\,\mathrm{d}t \,\middle|\, A_0 = a, S_0 = s\right] \\ &= e^s\mathsf{E}\left[\int_0^{\infty} re^{-rt}(1 - k_n(A_t))e^{S_t - s}\,\mathrm{d}t \,\middle|\, A_0 = a, S_0 = s\right] \\ &= e^s\mathsf{E}\left[\int_0^{\infty} re^{-rt}(1 - k_n(A_t))e^{S_t}\,\mathrm{d}t \,\middle|\, A_0 = a, S_0 = 0\right] \\ &= e^s v_n(a, 0|\boldsymbol{k}), \end{aligned}$$

where the second-to-last equality comes from the Markov property of the diffusion process $\{A_t\}_{t\geq 0}$.

□

Lemma 6 (Normal Reflection Condition and Feynman-Kac Equation)**.** *Given an s-invariant Markov strategy profile, the associated payoff function of each player satisfies Feynman-Kac equation (6) at each state at which it is twice continuously differentiable. Moreover, if the intensity of exploration is bounded away from 0 in some neighborhood of the state $a = 0$, the payoff function also satisfies the normal reflection condition (4).*

Proof. A more general version of this lemma is provided in Lemma 7 below for Markov strategy profile $\boldsymbol{k}$ with state variables (a, s). The normal reflection condition $(\partial v_n/\partial a +$

$\partial v_n/\partial s)(0+, s) = 0$ takes form of $u_n(0) + u_n'(0+) = 0$ because of the homogeneity of the payoff functions for s-invariant strategies. □

Given Markov strategy profile $\boldsymbol{k}$ with state variables (a, s) in which the strategies are not necessarily best responses against each other, let

$$\mathsf{L} = \eta(a, s)\frac{\partial}{\partial a} + \frac{1}{2}\alpha^2(a, s)\frac{\partial^2}{\partial a^2}$$

with $\eta(a, s) = -\mu K(a, s)$ and $\alpha(a, s) = \sigma\sqrt{K(a, s)}$. Denote by $f(a, s) = r(1 - k_n(a, s))e^s$ the flow payoff that player n receives at state (a, s), and by $v(a, s) = v_n(a, s|\boldsymbol{k})$ her payoff at state (a, s).

Lemma 7 (Properties of Payoff Function)**.** *If $a \mapsto v(a, s)$ is twice continuously differentiable at a, then it satisfies the Feynman-Kac formula*

$$rv(a, s) = \mathsf{L}v(a, s) + f(a, s). \tag{10}$$

Moreover, if v is continuously differentiable on $\{a = 0\}$,[20] *and $K(a, s)$ is bounded away from 0 on some open set containing $\{a = 0\}$, then v satisfies the* normal reflection *condition*

$$\frac{\partial v(0+, s)}{\partial a} + \frac{\partial v(0+, s)}{\partial s} = 0. \tag{11}$$

Proof. Given strategy profile $\boldsymbol{k}$, consider player n's continuation value at time T, that is,

$$v(A_T, S_T) = \mathsf{E}\left[\int_T^\infty e^{-r(t-T)} f(A_t, S_t)\,\mathrm{d}t \middle| \mathcal{F}_T\right],$$

and her total discounted payoff at time 0, evaluated conditionally on the information available at time T, that is,

$$\begin{aligned} Z_T &:= \mathsf{E}\left[\int_0^\infty e^{-rt} f(A_t, S_t)\,\mathrm{d}t \middle| \mathcal{F}_T\right] \\ &= \mathsf{E}\left[\int_T^\infty e^{-rt} f(A_t, S_t)\,\mathrm{d}t \middle| \mathcal{F}_T\right] + \int_0^T e^{-rt} f(A_t, S_t)\,\mathrm{d}t \\ &= e^{-rT} v(A_T, S_T) + \int_0^T e^{-rt} f(A_t, S_t)\,\mathrm{d}t\,. \end{aligned}$$

[20]Here we mean $v : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ can be extended to a function that is continuously differentiable on some open set in $\mathbb{R}^2$ containing $\{a = 0\}$.

Note that $\{Z_t\}_{t\geq 0}$ is a martingale.[21] Indeed, for $0 \leq \tau \leq T$, we have

$$\begin{aligned}\mathsf{E}[Z_T|\mathcal{F}_\tau] &= \mathsf{E}\left[\mathsf{E}\left[\int_0^\infty e^{-rt} f(A_t, S_t)\,\mathrm{d}t \middle| \mathcal{F}_T\right]\middle| \mathcal{F}_\tau\right]\\ &= \mathsf{E}\left[\int_0^\infty e^{-rt} f(A_t, S_t)\,\mathrm{d}t \middle| \mathcal{F}_\tau\right]\\ &= Z_\tau.\end{aligned}$$

Because $K(\cdot, s)$ is piecewise continuous, we have $a \mapsto v(a, s)$ is piecewise C^2 on $\mathbb{R}_{++}$. Moreover, v is continuously differentiable on $\{a = 0\}$ by assumption. Therefore, we can apply Itô formula to $e^{-rT} v(A_T, S_T)$,[22] which gives

$$\begin{aligned}e^{-rT} v(A_T, S_T) - v(a, s) = &\int_0^T e^{-rt} (\mathsf{L}v - rv)(A_t, S_t)\,\mathrm{d}t\\ &- \int_0^T e^{-rt} \alpha(A_t, S_t) \frac{\partial v}{\partial a}(A_t, S_t)\,\mathrm{d}B_t\\ &+ \int_0^T e^{-rt} \left(\frac{\partial v}{\partial a} + \frac{\partial v}{\partial s}\right)(A_t, S_t)\,\mathrm{d}S_t\,.\end{aligned}$$

Because Z_T is a martingale starting from $v(a, s)$, we can conclude the process

$$\begin{aligned}&Z_T - v(a, s) + \int_0^T e^{-rt} \alpha(A_t, S_t) \frac{\partial v}{\partial a}(A_t, S_t)\,\mathrm{d}B_t\\ &= \int_0^T e^{-rt} \left((\mathsf{L}v - rv + f)(A_t, S_t)\right)\mathrm{d}t + \int_0^T e^{-rt} \left(\frac{\partial v}{\partial a} + \frac{\partial v}{\partial s}\right)(A_t, S_t)\,\mathrm{d}S_t\end{aligned}$$

is a continuous martingale starting from 0. Observe that the right-hand side is a finite variation process. As a consequence, Z_T must be 0 for all $T \geq 0$ almost surely. Moreover, note that the Lebesgue measure $\mathrm{d}t$ and the measure $\mathrm{d}S_T$ on $([0, T], \mathcal{B}([0, T]))$ are mutually singular almost surely. Indeed, the set $D := \{\, t \in [0, T] : A_t = 0 \,\}$ is a null set under Lebesgue measure $\mathrm{d}t$, and $[0, T]\backslash D$ is a null set under measure $\mathrm{d}S_T$ almost surely. Because almost surely the sum of the integrals on the right-hand side is zero and $\mathrm{d}t \perp \mathrm{d}S_T$, by Lebesgue decomposition theorem we know that each integrand is a.s. a.e. 0 on $[0, T]$ for all T.[23] This yields the Feynman-Kac formula (10) and the normal reflection condition (11).[24] □

[21] Assumption 1 guarantees all expectations in this proof are finite.

[22] See footnotes 27 and 28 for the validity of applying Itô's formula here.

[23] By a.e. 0 on $[0, T]$ we mean the integrands are 0 on all non-null sets w.r.t. the corresponding measure.

[24] We require $K(a, s) > 0$ on some neighborhood containing $\{a = 0\}$, as otherwise all measurable subsets of $[0, T]$ would be null sets w.r.t. $\mathrm{d}S_T\,(\omega)$, in which case we cannot conclude that v satisfies the normal reflection condition.

C Complete Information Setting

C.1 Proof of Lemma 1

Proof. Let $\hat{s}_W := \sup_{x\geq 0}\{W(x) - \delta x\}$. Note that $W_+(x) - y - \delta x$ has the same law as the Brownian motion starting from 0 with drift $\mu - \delta$ and volatility σ. Then it is well-known that $\hat{s}_{W_+} - y$ has the exponential distribution with mean $1/\lambda$ if $\lambda > 1$. Otherwise, if $\lambda \leq 1$, then $\hat{s}_{W_+} = +\infty$ with probability 1, which obviously leads to $\widehat{V}(a, s) = +\infty$.

Now suppose $\lambda > 1$. The (ex ante) value under complete information at state (a, s) can be calculated as

$$\begin{aligned}\widehat{V}(a, s) &= \mathsf{E}\left[\hat{v}(W)|W_+(0) = y, W(0) = s\right]\\ &= \mathsf{P}(\hat{s}_W \leq s)e^s + \mathsf{P}(\hat{s}_W > s)\mathsf{E}\left[e^{\hat{s}_W}\middle|\hat{s}_W > s\right]\\ &= e^s\left(\mathsf{P}(\hat{s}_W \leq s) + \mathsf{P}(\hat{s}_W > s)\mathsf{E}\left[e^{y-s}e^{\hat{s}_W - y}\middle|\hat{s}_W > s\right]\right),\end{aligned}$$

where $\hat{v}(W) = e^{\hat{s}_W}$ is proved in Lemma 8 below.

Note that $\hat{s}_W > s$ if and only if $\hat{s}_{W_+} - y > a$, and hence we have

$$\mathsf{P}(\hat{s}_W > s)\mathsf{E}\left[e^{y-s}e^{\hat{s}_W - y}\middle|\hat{s}_W > s\right] = e^{-a}\int_a^\infty e^z\lambda e^{-\lambda z}\,\mathrm{d}z = -\frac{\lambda}{1-\lambda}e^{-\lambda a}.$$

Therefore, we have $\widehat{V}(a, s) = e^s\left(1 - e^{-\lambda a} - \frac{\lambda}{1-\lambda}e^{-\lambda a}\right) =: e^s\widehat{U}(a)$ with $\widehat{U}(x) = 1 + e^{-\lambda a}/(\lambda - 1)$. $\square$

Lemma 8 (Average Ex-post Value Function under Complete Information)**.** *For a given Brownian path W, the average (ex-post) value under complete information is given by $\hat{v}(W) = e^{\hat{s}_W}$, where $\hat{s}_W = \sup_{x\geq 0}\{W(x) - \delta x\}$.*

Proof. For an arbitrary technology $x \geq 0$, denote by $K(x) = \{K_t\}_{t\geq 0}$ the cutoff strategy in which the players explore with full intensity until x is developed, and exploit x thereafter. In other words, $K_t = N\,\mathbf{1}_{[0,x/N)}(t)$. By the nature of the problem, it is obvious that we can focus on this class of cutoff strategies without loss.

The average payoff for strategy $K(x)$ can be calculated as

$$\hat{v}(W|K(x)) = \int_{x/N}^\infty re^{-rt}e^{W(x)}\,\mathrm{d}t = e^{W(x)-\delta x},$$

which gives

$$\hat{v}(W) = \sup_{x\geq 0}\hat{v}(W|K(x)) = e^{\hat{s}_W},$$

and the proof is complete.

We define the *first-best technology* to be any $\hat{x}_N \in \arg\max_{x\geq 0}\{W(x) - \delta x\}$, if this set is not empty given W. In such a case, the value $\hat{v}(W) < +\infty$ is achieved by strategy $K(\hat{x}_N)$.

□

D Long-run Outcomes

D.1 Proof of Lemma 2

Proof. This lemma can be easily derived from the results in Lehoczky (1977). □

D.2 Proof of Lemma 3

Proof. By the definition of $\hat{x}_N$ in the proof of Lemma 8, we have $W(\hat{x}_N) - \delta\hat{x}_N \geq W(x) - \delta x$ for all $x \geq 0$. We assume there exists a unique $\hat{x}_N$ without loss.[25] Let $\check{x} := \arg\max_{[0,\bar{x}]} W(x)$ denote the technology with quality $W(\check{x}) = \bar{s}$.[26] Note that if $a < \bar{a}$ we have $W(\bar{x}) = \bar{s} - \bar{a}$, otherwise we have $\bar{x} = 0$ and thus $W(\bar{x}) = s - a$.

Suppose that $\hat{x}_N > \bar{x}$. Then we have $W(\hat{x}_N) - \delta\hat{x}_N > \bar{s} - \delta\check{x} \geq \bar{s} - \delta\bar{x}$, which implies the event of $\sup_{x>\bar{x}}\{W(x) - \delta(x - \bar{x})\} > \bar{s}$. By the Markov property of the Brownian motion, this event occurs with probability $\mathsf{P}(M_\infty > \max\{a, \bar{a}\})$, where M_∞ denotes the global maximum of a Brownian motion starting from zero, with drift $\mu - \delta$ and volatility σ. It is well known that M_∞ has an exponential distribution with mean $-\frac{\sigma^2}{2(\mu-\delta)} = 1/\lambda$. Therefore, we have $q(\bar{a}) = 1 - \mathsf{P}_{as}(\hat{x}_N > \bar{x}) \geq 1 - \mathsf{P}(M_\infty > \max\{a, \bar{a}\}) \to 1$ as $\bar{a} \to +\infty$, since $\lambda > 1$ under Assumption 1. □

E Equilibrium Characterization

E.1 Proof of Lemma 4

Lemma 4 follows from Lemma 9 and 10 below.

Lemma 9 (Sufficiency)**.** *Given $\boldsymbol{k}_{\neg n} \in \mathcal{K}^{N-1}$, if $u_n : \mathbb{R}_+ \to \mathbb{R}$ satisfies Condition 1–4 in Lemma 4, then a piecewise right-continuous function $k_n^* : \mathbb{R}_+ \to [0,1]$ which maximizes the right-hand side of the HJB equation (5) at each continuity point of u_n'' is a best response against $\boldsymbol{k}_{\neg n}$, with u_n being the associated payoff function of player n.*

[25] This assumption holds almost surely under Assumption 1.

[26] By standard results, $\bar{x}(\bar{a})$ is almost surely finite under Assumption 1. See, e.g., equation (1.1) of Taylor (1975).

Proof. The proof is a standard verification argument. See, e.g., Fleming and Soner (2006), Theorem III.9.1. The role played by the linear growth condition in a standard proof is instead played by Assumption 1.

Given $\boldsymbol{k}_{\neg n} \in \mathcal{K}^{N-1}$, for any admissible control process $k_n = \{k_{n,t}\}_{t\geq 0} \in \mathcal{A}$, let

$$\mathsf{L} := \eta(a, s, k_{n,t})\frac{\partial}{\partial a} + \frac{1}{2}\alpha^2(a, s, k_{n,t})\frac{\partial^2}{\partial a^2},$$

with $\eta(a, s, k) = -\mu(k + K_{\neg n}(a))$ and $\alpha(a, s, k) = \sigma\sqrt{k + K_{\neg n}(a)}$. Let $f(a, s, k) = r(1-k)e^s$, and consider $v(a, s) = e^s u_n(a)$, where $u_n : \mathbb{R}_+ \to \mathbb{R}$ satisfies Conditions 1–4 in Lemma 4.

Usually, applying Itô's formula on v at $a > 0$ requires $u_n \in C^2(\mathbb{R}_{++})$. However, Itô's formula is still valid for C^1 functions with absolutely continuous derivatives, which is satisfied by u_n under Conditions 1 and 2.[27]

By applying Itô's formula on $e^{-rT}v(A_T, S_T)$,[28] for fixed $0 < T < \infty$ we have

$$\begin{aligned} e^{-rT}v(A_T, S_T) - v(a, s) = &\int_0^T e^{-rt}(\mathsf{L}v - rv)(A_t, S_t)\,\mathrm{d}t \\ &- \int_0^T e^{-rt}\alpha(A_t, S_t, k_{n,t})\frac{\partial v}{\partial a}(A_t, S_t)\,\mathrm{d}B_t \\ &+ \int_0^T e^{-rt}\left(\frac{\partial v}{\partial a} + \frac{\partial v}{\partial s}\right)(A_t, S_t)\,\mathrm{d}S_t\,. \end{aligned}$$

Since $\mathrm{d}S_t = 0$ whenever $A_t > 0$, and by Condition 3 we have $(\partial v/\partial a + \partial v/\partial s)(A_t, S_t) = 0$ when $A_t = 0$, the last term is identically zero. Moreover, because $\alpha(A_t, S_t, k_{n,t})$ and $\partial v/\partial a$ are bounded, the second term has mean zero. Taking expectation on both sides, for fixed $0 < T < \infty$ we have the Dynkin's formula

$$v(a, s) = -\mathsf{E}_{as}\left[\int_0^T e^{-rt}(\mathsf{L}v - rv)(A_t, S_t)\,\mathrm{d}t\right] + e^{-rT}\mathsf{E}_{as}[v(A_T, S_T)],$$

where the expectations on the right-hand side are finite for each $T < \infty$.

Note that HJB equation (5) implies

$$rv(a, s) \geq f(a, s, k_{n,t}) + \mathsf{L}v(a, s),$$

[27] See, e.g., Chung and R. J. Williams (2014), Remark 1, p. 187; Rogers and D. Williams (2000), Lemma IV.45.9, p. 105; or Strulovici and Szydlowski (2015), footnotes 73 and 78.

[28] Note that S_t is a finite variation process and hence both the quadratic variation $\langle S, S\rangle_t$ and the covariation $\langle A, S\rangle_t$ are identically zero. See Shreve (2004) Section 7.4.2 for a non-technical treatment.

and therefore we have

$$v(a,s) \geq \mathsf{E}_{as}\left[\int_0^T e^{-rt} f(A_t, S_t, k_{n,t})\,\mathrm{d}t\right] + e^{-rT}\mathsf{E}_{as}[v(A_T, S_T)]. \tag{12}$$

Let $T \to \infty$, we have

$$v(a,s) \geq \liminf_{T\to\infty} \mathsf{E}_{as}\left[\int_0^T e^{-rt} f(A_t, S_t, k_{n,t})\,\mathrm{d}t\right] + \liminf_{T\to\infty} e^{-rT}\mathsf{E}_{as}[v(A_T, S_T)].$$

Because $\int_0^T e^{-rt} f(A_t, S_t, k_{n,t})\,\mathrm{d}t$ is increasing in T, as the integrand is non-negative, we can apply either the monotone convergence theorem or Fatou's lemma to have

$$\begin{aligned} v(a,s) &\geq \mathsf{E}_{as}\left[\int_0^\infty e^{-rt} f(A_t, S_t, k_{n,t})\,\mathrm{d}t\right] + \liminf_{T\to\infty} e^{-rT}\mathsf{E}_{as}[v(A_T, S_T)] \\ &= v_n(a, s|k_n, \boldsymbol{k}_{\neg n}) + \liminf_{T\to\infty} e^{-rT}\mathsf{E}_{as}[v(A_T, S_T)] \\ &\geq v_n(a, s|k_n, \boldsymbol{k}_{\neg n}). \end{aligned}$$

Now we repeat the argument by replacing $k_{n,t}$ with $k^*_{n,t} = k^*_n(A_t)$. Inequality (12) becomes equality, and for $0 < T < \infty$ we have

$$v(a,s) = \mathsf{E}_{as}\left[\int_0^T e^{-rt} f(A_t, S_t, k^*_{n,t})\,\mathrm{d}t\right] + e^{-rT}\mathsf{E}_{as}[v(A_T, S_T)].$$

As the first term on the right-hand side is increasing in T, the second term must be decreasing, and hence

$$\begin{aligned} v(a,s) &= \mathsf{E}_{as}\left[\int_0^\infty e^{-rt} f(A_t, S_t, k^*_{n,t})\,\mathrm{d}t\right] + \lim_{T\to\infty} e^{-rT}\mathsf{E}_{as}[v(A_T, S_T)] \\ &= v(a, s|k^*_n, \boldsymbol{k}_{\neg n}) + \lim_{T\to\infty} e^{-rT}\mathsf{E}_{as}[v(A_T, S_T)]. \end{aligned}$$

We finish the proof by showing $\lim_{T\to\infty} e^{-rT}\mathsf{E}_{as}[v(A_T, S_T)] = 0$ as follows.

Because $A_t \geq 0$, we have

$$\begin{aligned} \mathsf{E}_{as}[v(A_T, S_T)] &\leq \mathsf{E}_{as}[v(0, S_T)] \\ &= u(0)\mathsf{E}_{as}[e^{S_T}] \\ &\leq u(0)\mathsf{E}_{0s}[e^{S_T}] \\ &= u(0)e^s\mathsf{E}_{0s}[e^{S_T - s}] \\ &\leq u(0)e^s\mathsf{E}_{0s}[e^{M_{NT}}], \end{aligned}$$

where $M_T = \max_{0\leq t\leq T}\{\mu t + \sigma B_t\}$. The last inequality comes from the fact that $S_T - s \leq M_{NT}$ almost surely. Indeed, given that $A_0 = 0$, the process $\{S_T - s\}_{T\geq 0}$ can be viewed as $\{M_T\}_{T\geq 0}$ with time change in the sense that $S_T - s = M_{T'}$ with $T' = \int_0^T K_t \,\mathrm{d}t$. Because $K_t \in [0, N]$, we have $T' \leq NT$ and hence $M_{T'} \leq M_{NT}$.

Then from the Math Appendix in the Online Appendix, we have

$$\mathsf{E}_{0s}[e^{M_{NT}}] \leq C_1 e^{(\mu+\sigma^2/2)NT} + C_2$$

for some $C_1, C_2 \geq 0$, and therefore

$$e^{-rT}\mathsf{E}_{as}[v(A_T, S_T)] \leq C_1 e^{(-r+(\mu+\sigma^2/2)N)T} + C_2 e^{-rT}.$$

The right-hand side goes to 0 as $T \to \infty$ if $(\mu + \sigma^2/2)N < r$, which is precisely Assumption 1.

□

Lemma 10 (Necessity)**.** *In any MPE $\boldsymbol{k} \in \mathcal{K}^N$, for each player $n \in \{1, \ldots, N\}$, her payoff function $u_n(\cdot|\boldsymbol{k})$ satisfies Conditions 1–4 in Lemma 4, and her equilibrium strategy $k_n(a)$ maximizes the right-hand side of the HJB equation (5) at each continuity point of u_n''.*

Proof. Condition 1: On the stopping region where $K(a) = 0$, it is trivial that the payoff functions are smooth. On the region where $K(a)$ is bounded away from zero, players' payoff functions are once continuously differentiable by standard results. Because the intensity of exploration in any MPE is bounded away from zero on any compact subset of $[0, \bar{a})$, as we argue in Section 4.2, the only part left to prove is the smooth pasting condition, which states that the payoff functions in any MPE must be once continuously differentiable at the stopping threshold $\bar{a}$. This is proved in Lemma 11 below.

Condition 2: By standard results, players' payoff functions are twice continuously differentiable at each point at which all players' strategies are continuous. Condition 2 thus follows from our piecewise continuity assumption on players' strategies.

Condition 3 follows from Lemma 6.

Condition 4 follows directly from the dynamic programming principle.

□

Lemma 11 (Smooth Pasting Condition)**.** *In any MPE with stopping threshold $\bar{a} > 0$, each player's normalized payoff function is continuously differentiable at $\bar{a}$.*

Proof. Given any MPE $\boldsymbol{k} = (k_1, \ldots, k_N)$, consider the region $[0, \bar{a})$ on which the intensity of exploration is positive. Clearly $u_n'(\bar{a}-)$, the left derivative of the equilibrium payoff function of player n at $\bar{a}$, cannot be positive, because otherwise we would have $u_n(a) < 1$ for a immediately to the left of $\bar{a}$, contradicting the fact that k_n is a best response.

Suppose by contradiction that u_n violates the smooth pasting condition at $\bar{a}$ with $u_n'(\bar{a}-) < 0$. Consider a deviation strategy $\tilde{k}_n$ in which $\tilde{k}_n(a) = 1$ on $[\bar{a} - \epsilon, \bar{a} + \epsilon)$ for some small $\epsilon > 0$, and $\tilde{k}_n = k_n$ otherwise, with $\tilde{u}_n$ denoting the associated payoff function of player n. Moreover, write $w_n := (\ln(u_n))' = u_n'/u_n$ and $\tilde{w}_n := (\ln(\tilde{u}_n))' = \tilde{u}_n'/\tilde{u}_n$. Notice that under this deviation the intensity of exploration $\tilde{k}_n + \sum_{l \neq n} k_l$ is bounded away from zero on $[0, \bar{a} + \epsilon)$ so that $\tilde{u}_n$ is once continuously differentiable on this region, which implies $\tilde{w}_n$ is continuous on $[0, \bar{a} + \epsilon)$. Also note that because the strategy profile remains unchanged on $[0, \bar{a} - \epsilon)$, the normal reflection condition $w_n(0) = \tilde{w}_n(0) = -1$ and the Feynman-Kac equation (6) imply that w_n and $\tilde{w}_n$ coincide on $[0, \bar{a} - \epsilon)$.[29] From $u_n'(\bar{a}-) < 0$ we know that $w_n(\bar{a}-) < 0$ as well, and therefore by the continuity of $\tilde{w}_n$ we can choose ϵ small enough so that $\tilde{w}_n < 0$ on $[\bar{a} - \epsilon, \bar{a} + \epsilon)$. This implies $\tilde{u}_n > 1 = u_n$ on $[\bar{a}, \bar{a} + \epsilon)$, because $\tilde{u}_n(a) = \exp\left(\int_{\bar{a}+\epsilon}^{a} \tilde{w}_n(z)\, \mathrm{d}z\right)$ for $a < \bar{a} + \epsilon$. Therefore, the deviation strategy $\tilde{k}_n$ leads to a higher payoff on $[\bar{a}, \bar{a} + \epsilon)$ than the equilibrium strategy k_n, which is a contradiction.

□

F Properties of MPE

F.1 Proof of Proposition 2

Proof. Suppose to the contrary that there is an MPE where player 1 chooses exploitation at all states. Let $\bar{a} > 0$ denote the state at which all exploration stops. Then on $(0, \bar{a})$ player 1's payoff function u_1 is continously differentiable and solves the free-rider ODE $u(a) = 1 + K(a)\beta(a, u)$ with value matching $u(\bar{a}) = 1$ and smooth pasting condition

[29] This claim is not affected by the possible discontinuities of the strategies on $[0, \bar{a})$, as the payoff functions are at least once continuously differentiable on this region.

$u'(\bar{a}) = 0$. It can then be easily verified that $u_1(a) = 1$ is the unique solution, which is a contradiction because the normal reflection condition (4) is violated. □

F.2 Proof of Proposition 3

Proof. As the individual payoff functions are bounded from below by the single-agent payoff function U_1^*, it is clear that the stopping threshold $\bar{a}$ in any MPE is weakly larger than the single-agent cutoff a_1^*. Suppose by contradiction that $\bar{a} = a_1^*$. Assume without loss of generality that all players' strategies are continuous on $(a_1^* - \epsilon, a_1^*)$ for some $\epsilon > 0$, so that each player's payoff function is twice continuously differentiable on this region. Note that for each player n, both u_n and U_1^* satisfy value matching and smooth pasting at a_1^*. Then from equation (6), we have for each player n,

$$\rho u_n''(a_1^*-) = \frac{k_n(a_1^*-)}{K(a_1^*-)} \leq 1$$

and $\rho U_1^{*\prime\prime}(a_1^*-) = 1$. As $\rho u_n''(a_1^*-) < 1$ for at least one player, her payoff lower bound $u_n \geq U_1^*$ is then violated at the states immediately to the left of $\bar{a}$. □

F.3 Proof of Proposition 4

Proof. Suppose by contradiction that there is an MPE where all players use cutoff strategies. Let player 1 be the one who uses the strategy with the largest cutoff $\bar{a}$. Then no other player uses the same cutoff $\bar{a}$ as player 1 for the following reason. Suppose to the contrary that player 2 uses a strategy with the same cutoff $\bar{a}$, then both player 1 and 2 must have a payoff strictly greater than 1 and lower than 2 at the states immediately to the left of $\bar{a}$, as their payoff functions solve the explorer ODE $u(a) = K\beta(a, u)$ for some $K > 1$ with initial condition $u(\bar{a}) = 1$ and $u'(\bar{a}) = 0$. As a result, exploration with full intensity cannot be optimal for both players at the states immediately to the left of cutoff $\bar{a}$ according to our characterization of best responses. Therefore, player 1 would be the lone explorer with $u_1 > 1$ on $(\bar{a} - \epsilon, \bar{a})$ for some $\epsilon > 0$, whereas all other players free-ride on this region.

Moreover, it is easy to see that player 1's payoff is weakly lower than the others, because the region on which she collects flow payoffs is the smallest among all players. However, on $(\bar{a} - \epsilon, \bar{a})$ the payoff function u_n of player $n \neq 1$ satisfies the free-rider ODE $u = 1 + \beta(a, u)$ with the same initial conditions as player 1. This yields the unique solution $u_n = 1 < u_1$ on $(\bar{a} - \epsilon, \bar{a})$, a contradiction.

□

G Symmetric MPE

G.1 Proof of Proposition 5

From Lemma 4, it is not difficult to verify that the strategy profile in Corollary 6 constitutes an equilibrium, and that our proposition properly summarizes the equilibrium payoff function in that corollary.

Uniqueness follows directly from symmetry and Lemma 4. One can check by explicit calculations that, under Assumption 1, the equilibrium in Corollary 6 is the only symmetric s-invariant strategy profile such that the associated common payoff function satisfies Conditions 1–4 in Lemma 4.

The comparison between the stopping threshold and the cooperative cutoffs follows from Lemma 12.

H Comparative Statics

Lemma 12 (Welfare Comparison)**.** *Consider normalized payoff functions* $u_i : \mathbb{R}_+ \to [1, +\infty)$ *for* $i = 1, 2$, *with stopping thresholds* $\bar{a}_i := \sup\{ a \geq 0 \mid u_i(a) > 1 \} < +\infty$. *Suppose for each* $i = 1, 2$, u_i *satisfies the following conditions:*

1. C^1 *on* $(0, \bar{a}_i)$ *with* $u_1'(\bar{a}_1) = u_2'(\bar{a}_2) \leq 0$;
2. *piecewise* C^2 *on* $\mathbb{R}_{++}$;
3. *there exists some function* $g_i : (1, +\infty) \times \mathbb{R}_- \to \mathbb{R}_{++}$ *with* $g_i(z_1, z_2)/z_2$ *nondecreasing in* z_1 *and nonincreasing in* z_2 *such that*
   (a) $u_i''(a) = g_i(u_i(a), u_i'(a))$ *for all* $a \in (0, \bar{a}_i)$ *at which* $u_i''(a)$ *is continuous;*
   (b) $u_i'(a) + g_i(u_i(a), u_i'(a)) > 0$ *for all* $a \in (0, \bar{a}_i)$.

If $g_1 \geq g_2$, *then* $\bar{a}_1 \leq \bar{a}_2$ *and* $u_1 \leq u_2$. *Additionally, if for some* $\breve{u} \in (1, u_2(0)]$ *we have* $g_1 = g_2$ *on* $(\breve{u}, u_2(0)) \times \mathbb{R}_-$ *and for all* $z_2 \leq 0$ *at least one of the following two conditions holds:*

1. $g_1(\breve{u}-, z_2) > g_2(\breve{u}-, z_2)$;
2. $\frac{\partial g_1(\breve{u}-, z_2)}{\partial z_1} < \frac{\partial g_2(\breve{u}-, z_2)}{\partial z_1}$ *and* $\frac{\partial g_1(\breve{u}-, z_2)}{\partial z_2} \geq \frac{\partial g_2(\breve{u}-, z_2)}{\partial z_2}$;

then $\bar{a}_1 < \bar{a}_2$, *and* $u_1 < u_2$ *on* $[0, \bar{a}_2)$.

Proof. See the Online Appendix. □

Remark. Almost all the payoff functions in this paper fulfill the first three conditions in the Lemma above under Assumption 1 with value matching and smooth pasting at $\bar{a}$. These include

- the cooperative solution: $g(z_1, z_2) = \frac{z_1}{N\rho} + \theta z_2$;
- the symmetric MPE: $g(z_1, z_2) = \max\{z_1/N, 1\}/\rho + \theta z_2$;
- the average payoff in the asymmetric MPE:

$$g(z_1, z_2) = \max\{z_1/N, \min\{1, z_1 - (1 - 1/N)\}\}/\rho + \theta z_2;$$

- the average payoff in the simple asymmetric MPE of Proposition 10:

$$g(z_1, z_2) = \begin{cases} \frac{z_1-(1-K/N)}{K\rho} + \theta z_2, & \text{if } K < z_1 \le K+1 \text{ for } K = 1, \dots, N-1, \\ \frac{z_1}{N\rho} + \theta z_2, & \text{if } z_1 \ge N. \end{cases}$$

H.1 Proof of Corollary 1

Proof. Here we prove the limit results only. Other results can be easily derived from explicit calculations. Write $\hat{w} = (\ln \widehat{U})'$ and $w^* = (\ln U^*)'$. It is not difficult to verify that $\hat{w}' = -\left(\frac{1}{N\rho} - \theta\right)\hat{w} - \hat{w}^2$ on $\mathbb{R}_{++}$; $w^{*\prime} = \frac{1}{N\rho} + \theta w^* - w^{*2}$ on $(0, a_N^*)$ and $w^* = 0$ on $[a_N^*, +\infty)$.

Suppose that $\theta \le -1$. Then on $(0, a_N^*)$, both ODEs above converge to $w' - \theta w + w^2 = 0$ as $N \to +\infty$. Because the normal reflection condition (4) implies $\hat{w}(0) = w^*(0) = -1$, we must have $\lim w^* = \lim \hat{w}$ on $(0, a_N^*)$. However, notice that $\lim \hat{w}(a) < 0$ for all $a \ge 0$. Therefore, we must have $a_N^* \to +\infty$ by the continuity of w^* on $\mathbb{R}_+$. This, together with the convergence of w^* and $\hat{w}$, implies $\lim U^*(a) = \lim \widehat{U}(a)$ for each $a \ge 0$. In particular for $\theta = -1$, from Lemma 1 we know $\widehat{U}(a) \to +\infty$ as $N \to +\infty$, and hence $U^*(a) \to +\infty$ for each $a \ge 0$. Then the results for the case of $\theta > -1$ follow from the monotonicity of $U^*(a)$ in θ, which can be easily verified by Lemma 12.

□

H.2 Proof of Corollary 3

Proof. Since $r = \sigma^2/(2\rho)$, the comparative statics with respect to r directly follow from the following proof of the comparative statics with respect to ρ. Recall that the common payoff function $U_\rho^\dagger$ in the symmetric MPE satisfies ODE $u'' = g(u, u')$ on $(0, \tilde{a})$ where

$g(z_1, z_2) = \max\{z_1/N, 1\}/\rho + \theta z_2$. Then the comparative statics of $U^\dagger_\rho$ and the stopping threshold $\tilde{a}_\rho$ follow directly from Lemma 12.

Because $U^\dagger_\rho(a)$ is decreasing on $[0, \tilde{a})$ and increasing in ρ for each $a \in [0, \tilde{a})$, the full-intensity threshold $a^\dagger_\rho = U^{\dagger-1}_\rho(N)$ is weakly increasing in ρ.

Lastly, because $k^\dagger_\rho(a) = (U^\dagger_\rho(a) - 1)/(N - 1)$ on $(a^\dagger_\rho, \tilde{a}_\rho)$, the comparative statics results above imply that $k^\dagger_\rho(a)$ is weakly increasing in ρ for all $a \geq 0$.

□

H.3 Proof of Corollary 4

Proof. For the binding case, similar to the proof of Corollary 3, the comparative statics of $U^\dagger_N$ and $\tilde{a}_N$ with respect to N are immediate from Lemma 12.

For the non-binding case, Lemma 12 implies that both $U^\dagger_N$ and $\tilde{a}_N$ are constant over N, as $g(z_1, z_2) = 1/\rho + \theta z_2$ does not depend on N. Then it is obvious that $k_N(x) = (U^\dagger_N(a) - 1)/(N - 1)$ is weakly decreasing in N for each $a \geq 0$. □

H.4 Proof of Corollary 5

For the sake of notational convenience, we are going to prove the results in terms of $\rho = \sigma^2/(2r) > \hat{\rho}$ for some $\hat{\rho}$, which is equivalent to $r < \hat{r} = \sigma^2/(2\hat{\rho})$. We let $\hat{\rho} = (\theta - \ln(1 + \theta))/\theta^2$ if $\theta > -1$.[30]

To show the statement for the case that $\rho > \hat{\rho}$ and $\theta > -1$, we first show in Lemma 13 below that $|a^*_N - a^\dagger_N| \to \Delta$ for some $\Delta < +\infty$ as $N \to \overline{N}$. Because we know from Corollary 1 that $a^*_N \to +\infty$, we conclude that $a^\dagger_N \to +\infty$, and hence $\tilde{a}_N \to +\infty$ as well. The convergence of $q_N(\tilde{a}_N)$ then follows from Lemma 3.

After showing that $\lim_{N \to \overline{N}} \left\| U^*_N / U^\dagger_N \right\|_\infty < +\infty$ in Lemma 14, we can then conclude $U^\dagger_N \to +\infty$ from the fact that $U^*_N \to +\infty$ from Corollary 1.

The opposite case that $\rho \leq \hat{\rho}$ or $\theta \leq -1$ follows from Lemma 15 below.

Lemma 13. *If $\rho > \hat{\rho}$, we have $|a^*_N - a^\dagger_N| \to \Delta \in (0, +\infty)$ as $N \to \overline{N}$.*

Proof. Note that when $\rho > \hat{\rho}$, the expression of $\tilde{a}$ for the non-binding case in Corollary 6 cannot be applied. Therefore, the symmetric MPE must belong to the binding case for all $N \in (1, \overline{N})$. Moreover, for $\rho > \hat{\rho}$, we can verify that ι_N in Corollary 6 converges to some $\iota_{\overline{N}} \in (0, 1)$. Then from explicit calculations we know that $\gamma_2 \to 1 + \theta > 0$ and

[30] If $\theta = 0$ we set $\hat{\rho} = 1/2$, the limit of the right-hand side as $\theta \to 0$.

$\gamma_1 \to -1$ as $N \to \overline{N}$, and therefore we have

$$\begin{aligned}\Delta &:= \lim_{N\to\overline{N}} |a_N^* - a_N^\dagger| \\ &= \lim_{N\to\overline{N}} \frac{1}{\gamma_2 - \gamma_1} \ln\left(\frac{1+\iota/\gamma_2}{1+\iota/\gamma_1}\right) \\ &= \frac{1}{2+\theta} \ln\left(\frac{1+\iota_{\overline{N}}/(1+\theta)}{1-\iota_{\overline{N}}}\right) \\ &\in (0, +\infty).\end{aligned}$$

□

Lemma 14. *If $\rho > \hat{\rho}$, we have* $\lim \left\|U_N^*/U_N^\dagger\right\|_\infty < +\infty$ *as $N \to \overline{N}$.*

Proof. We first show that $\left\|U^*/U^\dagger\right\|_\infty = U^*(0)/U^\dagger(0)$ by showing $U^*(a)/U^\dagger(a)$ is nonincreasing in a.

Write $w^* = (\ln U^*)'$ and $w^\dagger = (\ln U^\dagger)'$. It is not difficult to verify that

$$w^{*\prime} = \frac{1}{N\rho} + \theta w^* - w^{*2}, \quad \text{on } (0, a^*),$$

$$\text{and} \quad w^{\dagger\prime} = \begin{cases} \frac{1}{N\rho} + \theta w^\dagger - w^{\dagger 2}, & \text{on } (0, a^\dagger), \\ \frac{1}{U^\dagger \rho} + \theta w^\dagger - w^{\dagger 2}, & \text{on } (a^\dagger, \tilde{a}). \end{cases}$$

Because $U^\dagger < N$ on $(a^\dagger, \tilde{a})$, we have $w^{\dagger\prime} \geq w^{*\prime}$. Then from the normal reflection condition $w^*(0) = w^\dagger(0) = -1$ we have $w^* \leq w^\dagger$. Therefore, $w^*(a) - w^\dagger(a) = \big(\ln\big(U^*(a)/U^\dagger(a)\big)\big)' \leq 0$, which implies $\ln\big(U^*(a)/U^\dagger(a)\big)$ is nonincreasing in a and hence so is $U^*(a)/U^\dagger(a)$. Therefore, we have $\left\|U^*/U^\dagger\right\|_\infty = U^*(0)/U^\dagger(0)$.

Moreover, because $w^*(0) = w^\dagger(0) = -1$ and $w^{*\prime} = w^{\dagger\prime}$ on $(0, a_N^\dagger)$, we have $w^* = w^\dagger$ on $(0, a_N^\dagger)$ and hence $U^*/U^\dagger$ is constant on $(0, a_N^\dagger)$. Therefore, we can write

$$\frac{U_N^*(0)}{U_N^\dagger(0)} = \frac{U_N^*(a_N^\dagger)}{U_N^\dagger(a_N^\dagger)} = \frac{1}{N}\frac{1}{\gamma_2 - \gamma_1}\left(\gamma_2 e^{-\gamma_1(a^* - a^\dagger)} - \gamma_1 e^{-\gamma_2(a^* - a^\dagger)}\right).$$

Since we know that $\gamma_2 \to 1+\theta > 0$ and $\gamma_1 \to -1$, we have

$$\left\|\frac{U_N^*}{U_N^\dagger}\right\|_\infty = \frac{U_N^*(0)}{U_N^\dagger(0)} \to \frac{1}{\overline{N}(2+\theta)}\left((1+\theta)e^{\Delta} + e^{-(1+\theta)\Delta}\right) < +\infty.$$

□

Lemma 15. *If either $\rho \leq \hat{\rho}$ or $\theta \leq -1$, then there exists $\underline{N} > 1$ such that $a_N^\dagger = 0$ and $U_N^\dagger = U_{\underline{N}}^\dagger < +\infty$ for all $N \geq \underline{N}$.*

Here we do not impose Assumption 1 in Lemma 15. Therefore, this lemma states that when N reaches $\underline{N}$ as increased from 1, the symmetric equilibrium falls into the non-binding category and continues to be an equilibrium for all $N > \underline{N}$, even when Assumption 1 is violated for large N in the case of $\theta > -1$. As a consequence, we have $\tilde{a}_N = \tilde{a}_{\underline{N}}$ for all $N \geq \underline{N}$, and thus the stopping threshold $\tilde{a}_N$ is bounded from above as $N \to +\infty$.[31] This implies $|U_N^*(a) - U_N^\dagger(a)|$ is bounded away from zero in the limit, because for large enough N we have that $U_N^\dagger(a)$ is constant over N, and that $U_N^*(a)$ is increasing in N for each $a \geq 0$.

Moreover, since $\tilde{a}_N$ is constant over N for large enough N, the amount of exploration $\bar{x}(\tilde{a}_N)$ is also constant for large N. The limit result for $q_N(\tilde{a}_N)$ then follows from the fact that the first-best technology $\hat{x}_N$ is nondecreasing in N for each given Brownian path W and each initial state (a, s).

Next we prove the above lemma.

Proof of Lemma 15. Suppose either $\rho \leq \hat{\rho}$ or $\theta \leq -1$. We construct $U_{\underline{N}}^\dagger$ according to the closed-form expression of the equilibrium payoff function for the non-binding case in Corollary 6. This is possible only when either $\rho \leq \hat{\rho}$ or $\theta \leq -1$, as otherwise the expression for $\tilde{a}$ in that corollary is not well-defined. Let $\underline{N} = U_{\underline{N}}^\dagger(0)$, and for all $N \geq \underline{N}$, let $k_N^\dagger = (U_{\underline{N}}^\dagger - 1)/(N-1)$.

We now verify that the strategy profile $\boldsymbol{k}_N^\dagger$ with each player playing $k_N^\dagger$ constitutes a symmetric MPE with non-binding resource constraints in the N-player exploration game for all $N \geq \underline{N}$, with $U_{\underline{N}}^\dagger$ being the associated payoff function. Note that we cannot apply Lemma 4 directly because it relies on Assumption 1. Nevertheless, from the proof of Lemma 9, we know that $U_{\underline{N}}^\dagger$ is an upper bound on player n's achievable payoffs against $\boldsymbol{k}_{\neg n}^\dagger$ (this fact does not rely on Assumption 1). Moreover, $U_{\underline{N}}^\dagger$ is indeed the payoff function associated with $\boldsymbol{k}_N^\dagger$, since $U_{\underline{N}}^\dagger$ is the only function that satisfies both of the properties in Lemma 6. As a result, the upper bound on the payoff functions when the player plays against $\boldsymbol{k}_{\neg n}^\dagger$ is achieved by $k_N^\dagger$, and therefore $\boldsymbol{k}_N^\dagger$ is a symmetric MPE.

□

[31] More precisely, when Assumption 1 is violated and the uniqueness of the symmetric MPE is not guaranteed, here we mean there exists a sequence of symmetric MPE indexed by N, with $\tilde{a}_N$ bounded from above as $N \to +\infty$.

I Asymmetric MPE

I.1 Simple MPE

Here we present a class of asymmetric MPE similar to the simple MPE in Section 6.1 of Keller, Rady, and Cripps (2005). The construction is almost identical to the asymmetric MPE of Proposition 6, and therefore the proof is omitted.

Proposition 10 (Asymmetric Simple MPE)**.** *The N-player exploration game admits simple Markov perfect equilibria with M thresholds $0 =: \bar{a}_{M+1} < \bar{a}_M < \cdots < \bar{a}_1 < \bar{a}_0 := +\infty$ for some $M \in \{1, \ldots, N\}$, such that on $[\bar{a}_{K+1}, \bar{a}_K)$ exact K players take turns exploring on consecutive subintervals.*

The average payoff function $\bar{u}$ is strictly decreasing on $[0, \bar{a}_1]$, twice continuously differentiable on $\mathbb{R}_{++}$ except for the states $\{\bar{a}_K\}_{K=1}^{M}$. Moreover, we have $\lfloor \bar{u}(0) \rfloor = M$, and $\bar{u}$ solves the ODE $u = (1 - K/N) + K\rho\,(u'' - \theta u')$ on $[\bar{a}_{K+1}, \bar{a}_K)$ for each $K \in \{0, 1, \ldots, M\}$. The payoff function of each player u_n is strictly decreasing on $[0, \bar{a}_1]$, once continuously differentiable on $\mathbb{R}_{++}$, and satisfies $u_n(\bar{a}_K) = \bar{u}(\bar{a}_K) = K$ for $K \in \{1, \ldots, M\}$.

I.2 Proof of Proposition 6

Step 1: Construction of the average payoff function.

Let $\tilde{u} : \mathbb{R}_- \to \mathbb{R}$ be once continuously differentiable and solves

$$\max\{u(a)/N, \min\{1, u(a) - (1 - 1/N)\}\} = \beta(a, u), \tag{13}$$

with initial conditions $u(0) = 1$ and $u'(0) = 0$.[32] Let $a^\flat > 0$ be such that $\tilde{u}(-a^\flat) + \tilde{u}'(-a^\flat) = 0$.[33]

Now we let the average payoff function $\bar{u}(a) = \tilde{u}(a - a^\flat)$ for $a \le a^\flat$ and $\bar{u}(a) = 1$ for $a \ge a^\flat$. We can check that $\bar{u}$ satisfies value matching $\bar{u}(a^\flat) = 1$, smooth pasting $\bar{u}'(a^\flat) = 0$, and the normal reflection condition $\bar{u}(0) + \bar{u}'(0) = 0$. Let $a^\sharp = \bar{u}^{-1}(2 - 1/N)$ and $a^\ddagger = \bar{u}^{-1}(N)$ where $\bar{u}^{-1}(u) = \inf\{a \ge 0 \mid \bar{u}(a) \le u\}$. Under such construction we have $0 \le a^\ddagger \le a^\sharp < a^\flat$ and $\bar{u} : \mathbb{R}_+ \to \mathbb{R}$ is continuously differentiable and satisfies, on $(a^\flat, +\infty)$, $\bar{u} = 1$; on $(a^\sharp, a^\flat)$, solves $u = 1 - 1/N + \beta(a, u)$; on $(a^\ddagger, a^\sharp)$, solves $1 = \beta(a, u)$; on $(0, a^\ddagger)$, solves $u = N\beta(a, u)$.

[32] It is straightforward to verify that $\tilde{u}$ is strictly decreasing.

[33] It can be shown that under Assumption 1 there exists a unique $0 < a^\flat < +\infty$.

Step 2: Construction of the players' payoff functions and strategies.

For each player n, on $[a^\flat, +\infty)$, let $k_n(a) = 0$ and $u_n = \bar{u} = 1$; on $[0, a^\sharp)$, if not empty, let $k_n(a) = \min\{1, (\bar{u}(a) - 1)/(N - 1)\}$ and $u_n = \bar{u}$.

Next, consider any partition $a^\sharp = a_1 < a_2 < \cdots < a_m < a_{m+1} = a^\flat$ of interval $[a^\sharp, a^\flat]$. For each subinterval $[a_j, a_{j+1}]$ in the partition $\{a_j\}_{j=1}^{m+1}$, we use Algorithm 1 to construct payoff function u_n and strategy k_n for each player n.

For each subinterval $[a_j, a_{j+1}]$, Algorithm 1 calls procedure ActionAssignment, which first calls function Split to split $[a_j, a_{j+1}]$ into three subintervals according to Lemma 16 below, and lets the player with the lowest index available freeride on the subinterval $[a_{M-}, a_{M+}]$ in the middle, and explore on the rest two subintervals at both ends. In such a way, the strategy k_n and payoff function u_n of this player are defined on $[a_j, a_{j+1}]$, and she is then labeled as unavailable. Then ActionAssignment is called recursively on these three subintervals, preserving the total intensity $K = 1$ by allocating intensity 0 on the subintervals at both ends, and intensity 1 on $[a_{M-}, a_{M+}]$, with $\bar{u}$ on these subintervals being replaced by $\bar{u}_{\neg n}$, which is the average payoff function among the rest of the available players.

Lemma 16 ensures function Split partitions any interval $[a_L, a_R]$ into three subintervals in a unique way such that u_n, and therefore also $\bar{u}_{\neg n}$, have the same values and derivatives as $\bar{u}$ at the end points a_L and a_R.

The termination of Algorithm 1 is trivial because each time ActionAssignment is called, the strategy of one of the players is assigned and she is thereafter removed from the set of the available players. In the case where the allocation of strategies is unambiguous—that is, either the task of exploration has already been assigned or is to be assigned to the last available player—splitting the interval is unnecessary, and the corresponding strategies are set at once for all these players.

When Algorithm 1 terminates, the strategies and the constructed payoff functions of each player are defined on $[a^\sharp, a^\flat]$. Also note that $\bar{u}$ from Step 1 is indeed the average of the constructed payoffs of all the players, and the input requirement for $\bar{u}$ in procedure Split is satisfied in each call to the procedure. These conditions are maintained by line 22 during each call of ActionAssignment. Lastly, u_n is once continuously differentiable on $\mathbb{R}_{++}$ for each n, since no "kink" is created in Algorithm 1.

Step 3: Ensuring mutually best responses.

Because each u_n is once differentiable on $\mathbb{R}_{++}$, is twice differentiable except for the switch points, and satisfies normal reflection condition (4), our characterization of MPE in Lemma 4 states that $(k_1, \ldots, k_N)$ constitutes an MPE with u_n being the associated

Algorithm 1 Payoff construction

Require: $a^{\sharp} = a_1 < \ldots < a_{m+1} = a^{\flat}$, $\bar{u}$ from Step 1, and a finite set of players $N = \{1, \ldots, |N|\}$

Ensure: equilibrium strategy profile $\{k_n\}_{n=1}^{|N|}$ is defined on $[a_1, a_{m+1}]$, and $\{u_n\}_{n=1}^{|N|}$ is the set payoff functions corresponding to strategy profile $\{k_n\}_{n=1}^{|N|}$

for all $j \in \{1, \ldots, m\}$ **do**
  ACTIONASSIGNMENT(N, 1, a_j, a_{j+1}, $\bar{u}$)
end for

function SPLIT($|I|$, a_L, a_R, $\bar{u}$) ▹ input and output according to Lemma 16

Require: $|I|$: number of available players

Require: $\bar{u}$: solves $u = 1 - 1/|I| + \rho(u'' - \theta u')$ on (a_L, a_R)

Ensure: $\check{u} \in C^1((a_L, a_R))$ solves $u = \rho(u'' - \theta u')$ on $(a_L, a_{M-}) \cup (a_{M+}, a_R)$, and $u = 1 + \rho(u'' - \theta u')$ on (a_{M-}, a_{M+}), with the same values and derivatives as $\bar{u}$ at a_L and a_R

  return $(a_{M-}, a_{M+}, \check{u})$
end function

procedure ACTIONASSIGNMENT(I, κ, a_L, a_R, $\bar{u}$)
  ▹ allocate aggregate intensity $\kappa \in \{0, 1\}$ to a set of available players I
  if $\kappa = 0$ **then** ▹ let all available players free-ride
    for all $n \in I$ **do**
      $k_n|_{[a_L, a_R)} \leftarrow 0$
      $u_n|_{[a_L, a_R)} \leftarrow \bar{u}$
    end for
  else if $\kappa = 1 = |I| =: |\{n\}|$ **then** ▹ let the only available player explore
    $k_n|_{[a_L, a_R)} \leftarrow 1$
    $u_n|_{[a_L, a_R)} \leftarrow \bar{u}$
  else ▹ let one of the $|I| \geq 2$ available player explore
    $(a_{M-}, a_{M+}, \check{u}) \leftarrow$ SPLIT($|I|, a_L, a_R, \bar{u}$)
    $n \leftarrow \min I$
    $k_n|_{[a_L, a_{M-}) \cup [a_{M+}, a_R)} \leftarrow 1$
    $k_n|_{[a_{M-}, a_{M+})} \leftarrow 0$
    $u_n|_{[a_L, a_R)} \leftarrow \check{u}$
    $\bar{u}_{\neg n}|_{[a_L, a_R)} \leftarrow \frac{|I|\bar{u} - u_n}{|I| - 1}$ ▹ the average payoff among the rest of the players

    ACTIONASSIGNMENT($I \backslash \{n\}$, κ, a_{M-}, a_{M+}, $\bar{u}_{\neg n}$)
    ACTIONASSIGNMENT($I \backslash \{n\}$, $\kappa - 1$, a_L, a_{M-}, $\bar{u}_{\neg n}$)
    ACTIONASSIGNMENT($I \backslash \{n\}$, $\kappa - 1$, a_{M+}, a_R, $\bar{u}_{\neg n}$)
  end if
end procedure

payoff function of player n if u_n solves the HJB equation

$$u_n(a) = 1 + K_{\neg n}(a)\beta(a, u_n) + \max_{k_n \in [0,1]} k_n\{\beta(a, u_n) - 1\}$$

for each continuity point of u_n'', and the constructed strategy maximizes the right-hand side.

In other words, we need to verify for all $a \geq 0$, $k_n(a) = 1$ if $\beta(a, u_n) > 1$, and $k_n(a) = 0$ if $\beta(a, u_n) < 1$. Then the HJB is satisfied by construction.

On $(a^\flat, +\infty)$, $u_n(a) = 1$ gives $\beta(a, u_n) = 0 < 1$, therefore $k_n(a) = 0$ is optimal. On $[0, a^\sharp)$, the argument is exactly the same as in the symmetric MPE. Lastly, on $(a^\sharp, a^\flat)$, we have $K = 1$ by construction. The monotonicity of $\bar{u}$ and the construction of u_n implies $1 \leq u_n(a) < \bar{u}(a^\sharp) = 2 - 1/N$ for each player n. Our construction of u_n ensures $k_n(a) = 1$ if and only if $u_n(a) = \beta(a, u)$, and $k_n(a) = 0$ if and only if $u_n(a) = 1 + \beta(a, u)$ for each $x \in (a^\sharp, a^\flat)$. Suppose $\beta(a, u_n) > 1$, which implies $1 + \beta(a, u_n) > 2 > u_n(a)$. Then by construction it must be $u_n(a) = \beta(a, u)$, and hence the constructed strategy $k_n(a) = 1$ is optimal. On the contrary, if $\beta(a, u_n) < 1$, which implies $\beta(a, u_n) < u_n(a)$, then by construction it must be $u_n(a) = 1 + \beta(a, u_n)$, and hence the constructed strategy $k_n(a) = 0$ is optimal.

Comparison with symmetric MPE

Note that $\bar{u}$ solves ODE

$$u'' = \max\{u/N, \min\{1, u - (1 - 1/N)\}\}/\rho + \theta u'$$

on $(0, a^\flat)$, while the average payoff function $U^\dagger$ of symmetric MPE solves

$$u'' = \max\{u/N, 1\}/\rho + \theta u'$$

on $(0, \tilde{a})$, with value matching and smooth pasting at $a^\flat$ and $\tilde{a}$, respectively. Obviously the right-hand side in the first equation is weakly smaller. We can then check the second condition for the strict inequalities in Lemma 12 with $\breve{u} = 2 - 1/N$ and conclude that $\tilde{a} < a^\flat$ and $\bar{u} > U^\dagger$ on $[0, a^\flat)$. Similarly, we can also compare the first equation with the ODE in the cooperative problem $u'' = u/(N\rho) + \theta u'$ and conclude $a^\flat < a_N^*$.

□

Lemma 16 (Split). *Given a strictly decreasing function $\bar{u} : [a_L, a_R] \to [u_L, u_R]$, with $\bar{u}(a_L) = u_L$ and $\bar{u}(a_R) = u_R \geq 1$, which satisfies the average payoff ODE $u = f + \rho(u'' - \theta u')$ on (a_L, a_R) with $0 < f < 1$ and $\rho > 0$, there exist $a_L < a_{M-} < a_{M+} < a_R$,* *and a function* $\check{u} : [a_L, a_R] \to [u_L, u_R]$ *continuously differentiable on* (a_L, a_R) *such that* $\check{u}(a_L) = \bar{u}(a_L)$*;* $\check{u}'(a_L) = \bar{u}'(a_L)$*;* $\check{u}(a_R) = \bar{u}(a_R)$*;* $\check{u}'(a_R) = \bar{u}'(a_R)$*;* $\check{u}$ *solves the explorer ODE* $u = \rho(u'' - \theta u')$ *on* $(a_L, a_{M-}) \cup (a_{M+}, a_R)$ *and solves the free-rider ODE* $u = 1 + \rho(u'' - \theta u')$ *on* (a_{M-}, a_{M+})*.*

Proof. See the Online Appendix. □

I.3 Proof of Proposition 7

Proof. The average payoff $\hat{u}$ in any MPE must be once continuously differentiable and satisfy the ODE $u'' = g(u, u')$ for some g respecting Feynman-Kac equation (6) whenever $\hat{u} > 1$ and $\hat{u}''$ is continuous. For $N = 2$, the characterization of best responses gives the following minimal requirements on g. For each $1 < u < 2$, $g(u, u')$ can only be either $\frac{u-1/2}{\rho} + \theta u'$ (one player explores and the other freerides) or $\frac{1}{\rho} + \theta u'$ (both players choose a common interior allocation); for each $u > 2$, g can only be either $\frac{u-1/2}{\rho} + \theta u'$ (one player explores and the other freerides) or $\frac{u}{2\rho} + \theta u'$ (both players explore). The asymmetric MPE we construct in Proposition 6 adopts $g(u, u') = \max\{u/2, \min\{1, u - 1/2\}\}/\rho + \theta u'$, which is the minimal g that satisfies these constraints, and therefore achieves the highest payoff among all MPE by Lemma 12. □

I.4 Proof of Proposition 8

Proof. Choose a partition $\{a_j\}_{j=1}^{m+1}$ of $[a^\sharp, a^\flat]$ in the construction of the equilibria in Proposition 6, so that

$$\max_{1 \le j \le m} \left|\bar{u}(a_{j+1}) - \bar{u}(a_j)\right| \le \delta := \frac{1}{2} \max_{a^\sharp \le a \le \tilde{a}} \{\bar{u}(a) - U^\dagger(a)\},$$

and $|a_m - a_{m+1}| < \epsilon$. Because the average payoff $\bar{u}$ and the players' payoff functions are monotone and coincide at the endpoints of the subintervals $[a_j, a_{j+1}]$, we have $|u_n(a) - \bar{u}(a)| \le \delta$ for $a \in [a^\sharp, \tilde{a}]$ for all player n. Therefore, we have $u_n \ge \bar{u} - \delta > U^\dagger$ on $[a^\sharp, \tilde{a})$. Also, as $u_n(a_m) = \bar{u}(a_m) > 1$ for all player n, we have $u_n > 1 = U^\dagger$ on $[\tilde{a}, a_m] = [\tilde{a}, a^\flat - \epsilon]$. Lastly, $u_n = \bar{u} > U^\dagger$ on $[0, a^\sharp)$, if it is not empty.

□

I.5 Proof of Proposition 9

Proof. Suppose $\theta < -1$. Then Assumption 1 is always satisfied and the asymmetric MPE in Proposition 6 exist for all N.

Denote by $\bar{u}_N$ the average payoff function in the asymmetric MPE of Proposition 6 for an N-player exploration game. Because $\bar{u}_N$ satisfies the ODE $u = 1 - 1/N + \rho(u'' - \theta u')$ whenever $1 < u < 2 - 1/N$, we verify that $\bar{w}_N := \ln(\bar{u}_N)'$ satisfies the second-order ODE

$$w'' - \theta w' + 3ww' + w(w^2 - \theta w - 1/\rho) = 0 \tag{14}$$

on $(a^\sharp, a^\flat)$, with initial conditions $w(a^\flat) = 0$ and $w'(a^\flat) = \frac{1}{N\rho}$, and $\bar{w}_N = 0$ on $[a^\flat, +\infty)$. By the continuity of the solution to ODE (14) in the initial data (see, e.g., Theorem 2.14 of Barbu (2016)), as $N \to +\infty$, w_N converges to 0 on $(a^\sharp, a^\flat)$, which is the unique solution to ODE (14) with initial conditions $w(a^\flat) = 0$ and $w'(a^\flat) = \lim \frac{1}{N\rho} = 0$. Suppose by contradiction that $a^\flat$ is bounded from above as $N \to +\infty$. This implies for sufficiently large N we have $\bar{u}_N(a) = \exp\left(\int_{a^\flat}^{a} w_N\right) < 2 - 1/N$ for all $a \in [a^\sharp, a^\flat]$, and thus by construction we have $a^\sharp = 0$. Then the fact that $w_N \to 0$ on $(0, a^\flat)$ contradicts $w_N(0) = -1$ imposed by the normal reflection condition (4).

Now suppose $\theta \geq -1$. If the constructed asymmetric MPE exist for all $N > 1$, then $a^\flat \to +\infty$ as $N \to +\infty$. This is because for fixed $N > 1$ and $a \geq 0$, $\bar{u}_N(a)$ is nondecreasing in θ, and thus the claim follows from the case of $\theta < -1$ as we have shown above. Otherwise, it can be shown that as $N \to \frac{1}{\rho(1+\theta)}$ we have $\bar{u}_N \to +\infty$ and $a^\flat \to +\infty$ by an argument similar to the proof of Corollary 5 for the symmetric equilibrium.

□