Strategic Exploration: Preemption and Prioritization*

Qingmin Liu† Yu Fu Wong‡

September 5, 2022

Abstract

This paper analyzes a model of strategic exploration in which competing players independently explore a set of alternatives. The model features a multiple-player multiple-armed bandit problem and captures a strategic trade-off between *preemption*—covert exploration of alternatives that the opponent will explore in the future—and *prioritization*—exploration of the most promising alternatives. Our results explain how the strategic trade-off shapes equilibrium behaviors and outcomes, e.g., in technology races between superpowers and R&D competitions between firms. We show that players compete on the same set of alternatives, leading to duplicated exploration from start to finish, and they explore alternatives that are a priori less promising before more promising ones are exhausted. The model also predicts that competition induces players to implement unreliable technologies too early, even though they should wait for the technologies to mature. Coordinated exploration is impossible even if the alternatives are equally promising, but it can emerge in equilibrium following a phase of preemptive competition if there is a short deadline. With asymmetric capacities of exploration, the weak player conducts *extensive* instead of *intensive* exploration—exploring as many alternatives as the strong player does but never fully exploring any.

*We thank Dilip Abreu, Nageeb Ali, Arjada Bardhi, Francis Bloch, Kalyan Chatterjee, Boyan Jovanovic, Navin Kartik, Nenad Kos, Rohit Lamba, George Mailath, Tatiana Mayskaya, Xin Meng, Marco Ottaviani, Jacopo Perego, Nikhil Vellodi, Siyang Xiong, Weijie Zhong, and workshop participants at Columbia, Penn, NYU, Bocconi, PSE, UCR, Michigan State, Essex, and Stony Brook Summer Conference for helpful comments. We thank the editor, Andrea Galeotti, and three referees for their excellent suggestions. This research is supported by NSF grant SES-1824328.

†Department of Economics, Columbia University. E-mail: ql2177@columbia.edu.

‡Department of Economics, Columbia University. E-mail: yw2942@columbia.edu.

Contents

1 Introduction

This paper studies strategic exploration in which competing players covertly explore a set of alternatives to find good candidates. The strategic tension is between preemption and prioritization. On one hand, each player would like to preempt his opponents by exploring the alternatives before they do. On the other hand, he would also like to prioritize the most promising alternatives given his capacity constraint. Our goal is to understand how the tension between preemption and prioritization shape competitors' strategic behavior and the overall discovery process and to develop analytical tools for this class of dynamic games.

This strategic tension appears in many applications of dynamic competition between firms and countries, where the alternatives can be research ideas, scientific experiments, business opportunities, etc. For instance, in a technology race between superpowers, the first country to discover the viable technology (such as nuclear weaponry, space technology, and more recently quantum computing) will gain a military, political, or economic advantage. The discovery and implementation necessitate exploring many different alternatives and conducting various experiments, not all of which are equally promising and very few of which will lead to a success. Therefore, each country must strategically allocate its resources over alternatives and over time.

A number of theoretically and empirically interesting questions can be formulated in these applications. Should players focus exclusively on the most promising route before moving on to less promising ones? Should players explore different alternatives to reduce duplicated explorations and alleviate competition? When is coordinated exploration possible? If a strong player is capable of exploring more alternatives over the same period of time, how would the asymmetry in capacity translate into asymmetry in their payoffs and probabilities of discovery? Should the weak player concentrate on fewer alternatives, or cover as much ground as the strong player although this would inevitably lead to insufficient exploration of some or all alternatives given the resource constraint? Would a player with negligible resources have a negligible impact on the competitor's exploration strategies and the overall process of discovery?

None of these questions is a priori obvious, but our model will provide clear predictions as consequences of the strategic tension between preemption and prioritization. We formulate the benchmark model in the simplest form. The unit interval represents the set of alternatives, of which at most one is good. Two players share a common prior of the good one over the set of alternatives. In continuous time, they each face a capacity constraint on the set of alternatives to explore per unit time. Whoever finds the good alternative first will receive a reward. The simplest model identifies the key elements of the strategic tension between preemption and prioritization. This allows us to see how other features, such as multiple good alternatives, short deadlines, asymmetric capacity constraints, multiple players, etc., add to the strategic interaction. For instance, the analysis and the main insights remain the same in an environment with gradual learning where the outcome of each alternative arrives stochastically at a rate controlled by resources allocated to the alternative. Unlike in the finite case, the continuum of alternatives makes the model tractable by eliminating the aggregate uncertainty of signal arrivals thanks to the law of large numbers. The model thus encapsulates a useful but underexplored multiple-player multiple-armed bandit problem.

The strategic exploration game features a unique and simple equilibrium. Instead of concentrating on the a priori most promising alternatives, the players explore an expanding set of alternatives of different prior probabilities in such a way that their posterior probabilities equalize. The strategy preempts the opponent's future explorations in order to maximize the option value of exploration and at the same time prioritizes alternatives with the highest posterior induced by the opponent's strategy in order to maximize the myopic value of discovery. Therefore, this "mutually greedy" strategy profile is the equilibration of preemption and prioritization. The strategy drives a wedge between equilibrium exploration and coordinated exploration that minimizes the time of discovery. Without preemption concerns, the fastest discovery would be achieved by prioritizing alternatives according to the *prior* probabilities. The uniqueness result rules out coordinated or specialized explorations even when the prior distribution is spread out or the reward from the good alternative increases over time.

Somewhat counterintuitively, the coordination failure can be mitigated if the deadline is short relative to the set of alternatives so that no single player can exhaustively explore all alternatives. The unique equilibrium in mutually greedy strategies of the benchmark model remains an equilibrium, but there are a continuum of equilibria with two phases: a competition phase in which the mutually greedy strategy profile is played to equalize the posteriors of a subset of alternatives, followed by a coordination phase in which the two players divide and prioritize this set. In the coordination phase, players no longer have incentives to preempt each other because the set of alternatives with the highest posterior induced by competition phase is large relative to the remaining time. Therefore, *competition creates the condition for coordination.* If the deadline is long, the most promising alternatives left from the competition stage will be exhausted before the deadline, and players will preempt their opponents in advance, unraveling the coordination phase.

When a "strong" player (she) has the capacity to explore more alternatives per unit time, the weak player (he) still explores alternatives with the highest posterior, but the strong player explores alternatives with unequal posterior and her strategy is no longer a greedy best response. The weak player always covers the same set of alternatives as the strong one, but he never explores any alternative with cumulative probability one, even if he can do so by focusing on a smaller set of alternatives. In other words, the weak player conducts an *extensive* exploration instead of an *intensive* exploration. We show that an edge in exploration capacity gives the strong player a disproportionately large payoff advantage due to an endogenous information advantage. When the exploration capacity is more asymmetric, the good alternative is discovered earlier in the first-order stochastic sense; nevertheless, the preemption incentive continues to play a non-vanishing role in slowing down discovery even when the weak player's capacity is vanishingly small.

We also study the strategic exploration game when more players enter the race. In the unique symmetric equilibrium, the players expand the set of alternatives to explore to equalize the posterior probabilities, just as in the benchmark setting. We show that the good alternative is discovered earlier with more players in the first-order stochastic dominance sense. However, there exist asymmetric equilibria with asymmetric payoffs even though the players are ex ante identical. Instead of competing over the same set of alternatives, the players specialize in different segments of alternatives in different groups.

Related literature

The paper connects several branches of active research. Optimal exploration of an unknown area is a well-known problem in operations research and computer science.[1] This literature has so far neglected the game-theoretic aspects of explorations. We do not consider the path dependence intrinsic to exploring physical locations and we assume away switching costs. In our model, the alternatives can be research ideas, scientific experiments, business opportunities, etc., all of which are of economic interest.

Although the underlying mechanisms are very different, our definition of "preemption" is not inconsistent with that in well-known preemption models, e.g., Fudenberg and Tirole (1985), Hendricks and Wilson (1992), Abreu and Brunnermeier (2003), and Hopenhayn and Squintani (2011). Players make a single irreversible preemption decision in these timing games, while they make such a decision for each alternative in our model. More importantly,

[1]It has applications in navigation algorithms and robotics, where inefficiency typically arises from the path dependence of exploration of physical locations. See, for example, the surveys by Kleinberg (1994) and Megow, Mehlhorn, and Schweitzer (2012).

we introduce prioritization among multiple alternatives to the preemption motive, such that the strategic tension gives rise to different equilibrium dynamics.

Fershtman and Rubinstein (1997) study a discrete-time finite-alternative search problem with preemption. As their analysis demonstrates, the discrete problem is intractable beyond a uniform prior. Under the uniform prior, however, the prioritization motive does not exist.[2] Matros, Ponomareva, Smirnov, and Wait (2019) consider a discrete-time continuum-alternative variant of Fershtman and Rubinstein (1997), but assume away the preemption motive, and find a qualitatively different equilibrium in pure strategies.[3] Bavly, Heller, and Schreiber (2022) consider a static search model where players have private information about the viability of different routes. Hence, these papers do not capture the dynamic trade-off between prioritization and preemption.

Chatterjee and Evans (2004) embed a two-alternative model of treasure hunting in a dynamic R&D game with Poisson bandits. Klein and Rady (2011) analyze a continuous-time model of a negative correlated bandit, in which one of the two arms contains a prize and two players share a common value instead of competing with each other. Our model features a negatively correlated bandit but, again, there is no preemption–prioritization trade-off in these two models.

The canonical models of strategic experimentation by Keller, Rady, and Cripps (2005) and Bolton and Harris (1999) capture a trade-off between exploration and exploitation in a multiple-player setting and are useful in many economic applications.[4] Most models in this literature feature a one-armed bandit, with one risky alternative and one safe default option, thus abstracting away the rich set of alternatives to explore and precluding interesting learning, search, and innovation processes in many applications. By eliminating aggregate uncertainties and facilitating the analysis of randomization, the multiple-player continuum-armed bandit formulated in this paper overcomes some of the analytical difficulties associated with finite-armed bandit problems that are largely intractable even when the arms are independent.

The strategic exploration game can also be viewed as a contest in which players choose

[2] Fershtman and Rubinstein (1997) allow players to choose their search capacities. They discuss the case where exactly one alternative has a different probability of success from the rest and make the observation that, under some parameter values, this alternative cannot always be searched first.

[3] They assume a uniform distribution, but this assumption is not the driving force. Instead, their model features a Bertrand-style competition so that the rent is dissipated completely. Their uniqueness fails, however, without the restriction to symmetric or Markovian strategies. Matros and Smirnov (2016) and de Roos, Matros, Smirnov, and Wait (2018) look into variants of this model with observable actions and with/without coordination.

[4] See also Manso (2011) and the large body of literature on contractual incentive provisions under experimentation. See Hörner and Skrzypacz (2016) for a survey.

what project to explore over time. Existing models in the literature of contests often study effort choices on a given project; see, e.g., Siegel (2009) and Fu, Lu, and Pan (2015) for recent developments.[5] Our model also differs from Hotelling's spatial competition models, e.g., Osborne and Pitchik (1986) and Ottaviani and Sørensen (2003), in that the good alternative (or location) is fixed, the topology of the alternatives is payoff-irrelevant, and the players choose which alternatives to explore dynamically.

2 Model

In this section, we formulate the strategic exploration game.

2.1 Setup

Two players explore a continuum of alternatives $x \in X := [0, 1]$ of which at most one is good. The prior distribution of the good alternative is given by a bounded density f that is strictly positive almost everywhere. Therefore, the prior probability that the good alternative exists is $\pi := \int_X f(x)dx \in (0, 1]$. The continuum of alternatives is an idealization of a large discrete set, and so two alternatives labeled 0.1 and 0.11 may not give similar outcomes.

The players explore over time horizon $[0, T]$ without observing each other's explorations, where $T \geq 1$. Each player faces a capacity constraint: he can explore up to a unit (Lebesgue) measure of alternatives per unit time. With some abuse of notation, we also let T denote as the set $[0, T]$ when no confusion arises.

The first player to find the good alternative exclusively claims a payoff of 1—and the two split it equally in the case of simultaneous discovery. In all other cases, their payoffs are normalized to 0. There is no temporal discounting.[6] Once the good alternative is found, the discovery is publicly announced and the game is over (alternatively, we may assume that the discovery is not made public but the reward is taken away from the good alternative once it has been discovered).

[5] The literature also studies optimal design of effort-maximizing contests; see, e.g., Moldovanu and Sela (2001, 2006) and Che and Gale (2003) for static problems and Bimpikis, Ehsani, and Mostagir (2019) and Halac, Kartik, and Liu (2017) for dynamic problems. Optimal design with strategic exploration will be an interesting direction for future research.

[6] The equilibrium construction of the benchmark model remains valid even if there is discounting. Appendix B.1 considers time preferences to capture growing payoff.

2.2 Strategy

We define the strategy space to describe how the players explore the alternatives over time. The strategy space demands formal treatments because the intuition of "exploring one alternative each period" inherited from a discrete problem does not extend to a continuum of alternatives in continuous time.[7] Our definition of strategy space exploits two ideas: the *outcome function approach* overcomes the indeterminacy of continuous-time strategies and the *distributional approach* handles the randomization.

2.2.1 Pure Strategy

A pure strategy $\sigma : T \times X \to \{0, 1\}$ specifies whether an alternative is explored by a certain time: an alternative $x \in X$ is explored at or before $t \in T$ if $\sigma(t, x) = 1$.

Definition 1. *A function $\sigma : T \times X \to \{0, 1\}$ is a* ***pure strategy*** *if it satisfies the following four conditions:*

1. *Initial condition: $\sigma(0, \cdot) = 0$;*
2. *Monotonicity and right-continuity: $\sigma(\cdot, x)$ is non-decreasing and right continuous for all $x \in X$;*
3. *Measurability: $\sigma(t, \cdot)$ is measurable for all $t \in T$;*
4. *Capacity constraint: $\int_X (\sigma(t, x) - \sigma(s, x))dx \leq t - s$ for all intervals $[s, t] \subset T$.*

The four conditions correspond to intuitive requirements for exploration activities.[8] The *initial condition* states that none of the alternatives has been explored at the beginning of the game. The *monotonicity* condition requires that, once an alternative has been explored, it is explored by any future time as well. The *right-continuity* property, similar to that of a cumulative distribution function, guarantees that the time at which an alternative $x \in X$ is explored $\tau(x) := \min\{t : \sigma(t, x) = 1\}$ is well defined. The *measurability* condition further

[7] First, this measure-preserving bijection between continuous time and the continuum of alternatives is not tractable. Second, a desirable definition should apply to an abstract set of alternatives, but the existence of measure-preserving bijections is not always guaranteed, let alone equilibria in such bijections. Third, the strategy should specify the set of alternatives explored at each moment in time. To avoid an uncountable union of measurable sets over time, we specify the outcome function that determines exploration activities.

[8] Simon and Stinchcombe (1989) point out the *indeterminacy* of the continuous-time strategy when it is written as a function of histories (including a player's own past actions). In our setting where the opponent's exploration is unobservable, the alternatives available for exploration for one player at each moment depend on his own exploration history. We overcome this issue by defining a strategy indirectly through an outcome function.

requires that $\tau : X \to T$ be a measurable function and $\tau^{-1}(t)$, the set of alternatives to be explored at time t, be a measurable set. It is the induced map τ^{-1} that instructs how the player should actually search.[9] Lastly, the *capacity constraint* describes how quickly a player can explore the space of alternatives. The maximum measure of alternatives explored per unit time is normalized to 1. We identify a strategy σ up to a stationary null set of X.

2.2.2 Distributional Strategy

We define distributional strategies to capture randomized explorations of a continuum of alternatives in continuous time. A distributional strategy $\rho : T \times X \to [0,1]$ specifies the probability that alternative $x \in X$ is explored by time $t \in T$. One should not confuse this notion with that of Milgrom and Weber (1985).[10]

Definition 2. *A function $\rho : T \times X \to [0,1]$ is a **distributional strategy** if it satisfies the following conditions:*

1. *Initial condition: $\rho(0, \cdot) = 0$;*
2. *Monotonicity and right-continuity: $\rho(\cdot, x)$ is non-decreasing and right continuous for all $x \in X$;*
3. *Measurability: $\rho(t, \cdot)$ is measurable for all $t \in T$;*
4. *Capacity constraint: $\int_X (\rho(t,x) - \rho(s,x))dx \leq t - s$ for all intervals $[s,t] \subset T$.*

The four conditions extend naturally from pure strategies. A distributional strategy ρ reduces to a pure strategy if $\rho(t,x) \in \{0,1\}$ as one can see by comparing Definition 1 and Definition 2.

Unlike the pure strategy, the distributional strategy specifies the cumulative probability of exploring each alternative but not the path of exploration. We prove a representation theorem (Theorem 7) that shows the outcome equivalence between distributional strategies and mixtures of pure strategies, and thus provides an indirect instruction to randomized explorations. We relegate the representation theorem, which relies on weak measurability and the Gelfand–Pettis integral, to Appendix A.8.

[9] The set of alternatives explored thus far $\{x : \sigma(t,x) = 1\} = \bigcup_{s \in [0,t]} \tau^{-1}(s)$ is measurable. Our outcome function approach circumvents the measurability of such an uncountable union of measurable sets when τ^{-1} is defined directly as in discrete problems.

[10] Abreu and Gul (2000) and Hendricks, Weiss, and Wilson (1988) use distributions to describe randomization in continuous-time games. In addition to the one-dimensional distribution of stopping times in prior work, the distributional strategy in this paper features the continuum of alternatives.

Remark 1. (**Interpretations of Randomization**) There are two more interpretations for the randomized strategies in addition to literal randomization. Randomized strategies can be viewed as the uncertainty entertained by the opponent.[11] Alternately, they can be interpreted as the cumulative resources spent on each alternative, which we formalize in Section 4.3.

2.3 Payoff

We compute the expected payoff of each player in a profile of distributional strategies. Let $-i$ denote player i's opponent. Given a profile of distributional strategies (ρ_i, ρ_{-i}), player i's expected payoff $u_i(\rho_i, \rho_{-i})$ is

$$\int_X \int_T f(x)(1-\rho_{-i}(t,x))d_t\rho_i(t,x)dx + \frac{1}{2}\int_X f(x)\sum_{t\in D_x}\Delta_t\rho_i(t,x)\Delta_t\rho_{-i}(t,x)dx. \tag{2.1}$$

The first term in (2.1) is player i's expected payoff from discovering the good alternative before his opponent. It reflects the probabilities of three events: alternative x is good with probability $f(x)$, the opponent $-i$ has not explored it yet with probability $1-\rho_{-i}(t,x)$, and player i explores that alternative instantaneously with probability $d_t\rho_i(t,x)$, where the time integral is the Lebesgue–Stieltjes integral with respect to the non-decreasing and right-continuous function $t \mapsto \rho_i(t,x)$.[12]

The second term in (2.1) is player i's expected payoff from simultaneously discovering the good alternative with his opponent. For each x, the set $D_x \subset T$ is the at most countable set of discontinuity points of both $\rho_i(\cdot,x)$ and $\rho_{-i}(\cdot,x)$. The function $\Delta_t\rho_i(t,x) := \rho_i(t,x) - \rho_i(t^-,x)$ is the jump measure of the distributional strategy $\rho_i(\cdot,x)$ on T, where $\rho_i(t^-,x) := \lim_{s\uparrow t}\rho_i(s,x)$. At alternative x, the probability of simultaneous discovery is $\sum_{t\in D_x}\Delta_t\rho_i(t,x)\Delta_t\rho_{-i}(t,x)$. The integral over X is well defined as the integrand can be written as the limit of measurable functions.

2.4 The One-Player Problem

Consider a useful benchmark of the single-player problem, which is equivalent to exogenously fixing $\rho \equiv 0$ for the other player in the two-player game. The incentive for preemption is absent in this case. There are many payoff-equivalent optimal search strategies. A **greedy strategy** is a distributional strategy that maximizes the myopic expected payoff at each

[11] See, e.g., Aumann (1987) for an exposition.

[12] The Lebesgue–Stieltjes measure is obtained from $\mu((s,t]) := \rho_i(t,x) - \rho_i(s,x)$ for all $0 \le s < t \le 1$.

moment in time, i.e., it prioritizes the most promising alternatives in the prior distribution. It will be a relevant benchmark for the competitive environment.[13]

For each $y \in [0, \infty)$, let $h(y) = \lambda(\{x : f(x) \geq y\})$ be the measure of alternatives whose prior densities are at least y, where λ is the Lebesgue measure. Therefore, h is non-increasing and left continuous with $\lim_{y \to +\infty} h(y) = 0$ and $h(0) = 1$. Intuitively, a greedy strategy explores alternatives sequentially according to their densities from the highest to the lowest. Formally, for any $x \in X$,

$$\rho(t, x) := \begin{cases} 1 & \text{if } t \geq h(f(x)), \\ 0 & \text{if } t < h(f(x)^+), \\ \frac{t - h(f(x)^+)}{h(f(x)) - h(f(x)^+)} & \text{if } t \in [h(f(x)^+), h(f(x))), \end{cases}$$

where $h(f(x)^+) := \lim_{y \downarrow f(x)} h(y)$. This greedy strategy ρ says that, when the prior density is constant over a positive measure set, the player randomizes uniformly over this set. This is clearly not the unique greedy strategy. When such a positive measure set does not exist, ρ specifies a deterministic exploration strategy and hence is uniquely determined.

Example 1. *Consider a symmetric probability density function*

$$f(x) = \begin{cases} 4x & \text{if } x \leq \frac{1}{2}, \\ 4 - 4x & \text{if } x > \frac{1}{2}. \end{cases}$$

Then we have

$$h(y) = \begin{cases} 1 - \frac{y}{2} & \text{if } y \leq 2, \\ 0 & \text{if } y > 2, \end{cases}$$

and

$$\rho(t, x) = \begin{cases} 1 & \text{if } \frac{1-t}{2} \leq x \leq \frac{1+t}{2}, \\ 0 & \text{o.w.} \end{cases}$$

What might appear counterintuitive is that two alternatives $\{\frac{1-t}{2}, \frac{1+t}{2}\}$ *are explored at the same time* $t > 0$*, but it takes exactly one unit of time, instead of half units, to fully explore the unit interval* $X = [0, 1]$*. Intuitively, the capacity is divided so that the speed of exploration is halved.*

[13] A non-greedy strategy will not be robust to discounting or to a perturbation of the other player's strategy with a uniform randomization, $\rho(t, x) = \epsilon t$ for all $x \in X$ and $t \in T$, where $\epsilon > 0$ is small.

3 Equilibration of Preemption and Prioritization

We derive the unique Nash equilibrium of the strategic exploration game. A profile of distributional strategies (ρ_i, ρ_{-i}) is a **Nash equilibrium** if $u_i(\rho_i, \rho_{-i}) \geq u_i(\rho_i', \rho_{-i})$ for each $i \in \{1, 2\}$ and distributional strategy ρ_i'.

We shall argue that, due to preemption motives, no player can play a pure strategy in a Nash equilibrium. Facing any pure strategy ρ_{-i}, player i can stay "one-step-ahead" of his opponent by exploring at time t what his opponent will explore at time $t + \epsilon$. When ϵ is close to 0, player i's payoff from this response is close to π and his opponent's payoff is close to 0. Thus, player $-i$'s payoff is 0 in the putative equilibrium. However, the opponent $-i$ can always imitate player i's equilibrium strategy to guarantee a strictly positive payoff. Therefore, both players must randomize in any equilibrium.

But how should players randomize in general? They must consider not only the myopic value of each alternative (measured by its posterior density) but also the option value (which is determined by how intensively his opponent will explore certain alternatives in the future). The trade-off between prioritization and preemption thus emerge.

3.1 Belief Updating

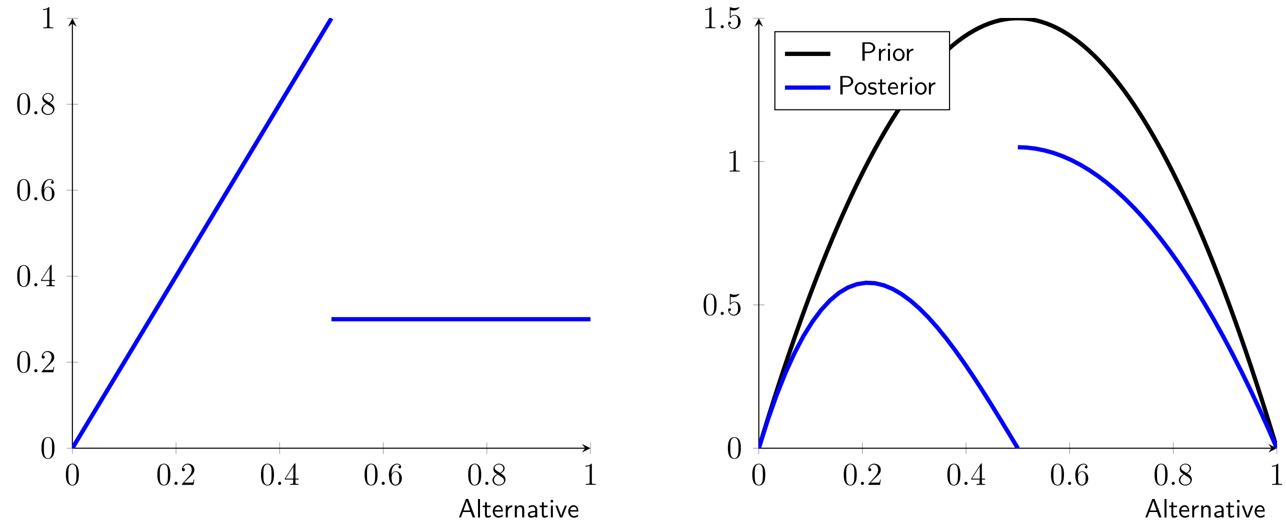
Belief updating is essential to characterizing equilibrium randomization. With prior density f and the opponent's distributional strategy ρ_{-i}, player i's posterior density that x is a good alternative right after t is

$$g_{-i}(t, x) := (1 - \rho_{-i}(t, x)) f(x). \tag{3.1}$$

We call $g_{-i}(t, x)$ player i's (unnormalized) **posterior distribution** over X at time t. We use the subscript "$-i$" because the posterior conditions *only* on the strategy of player $-i$. Note that the posterior does *not* take player i's own exploration into account; his posterior belief will be degenerate for alternatives that he has explored (see Section 4.3 for the case where exploration reveals the state only through gradual arrivals of conclusive signals.)

The initial condition of the distributional strategy in Definition 2 implies that $g_{-i}(0, x) = f(x)$. The monotonicity condition entails that the posterior $g_{-i}(t, x)$ is non-increasing in t for each $x \in X$. Intuitively, as the opponent $-i$ explores more alternatives over time, the posterior distribution is pushed lower and lower. Figure 3.1 illustrates the relationship between the distributional strategy, prior distribution, and posterior distribution.

We observe that the posterior equals the **flow payoff** when the probability of simulta-



(a) distributional strategy ρ_{-i} at a fixed time

(b) prior f and posterior g_{-i} at a fixed time

Figure 3.1: A distributional strategy ρ_{-i} and the posterior g_{-i}, at a fixed time.

neous discovery is zero. In that case, the expected payoff reduces to

$$u_i(\rho_i, \rho_{-i}) = \int_X \int_T (1 - \rho_{-i}(t,x)) f(x) d_t\rho_i(t,x) dx = \int_X \int_T g_{-i}(t,x) d_t\rho_i(t,x) dx. \tag{3.2}$$

Intuitively, the flow payoff of exploring alternative x at time t is the probability that x is good and the opponent has not explored it yet. A notable sufficient condition for zero probability of simultaneous discovery is that either ρ_i or ρ_{-i} is t-continuous.

3.2 Leveling Strategy

We shall construct a candidate equilibrium strategy such that the equilibrium posterior $g_i(t,x)$ levels the prior $f(x)$ over time as illustrated in Figure 3.2. As such, we shall call it the *leveling strategy*.

We first pin down the highest posterior as a function of time. By the definition of posterior distribution in Equation (3.1), player $-i$'s strategy ρ_{-i} and its induced posterior g_{-i} follow

$$\rho_{-i}(t,x) = 1 - \frac{g_{-i}(t,x)}{f(x)} \tag{3.3}$$

for all $t \in T$ and $x \in X$. We call a function $\bar{g} : T \to [0, \sup f]$ the **leveling function** if it

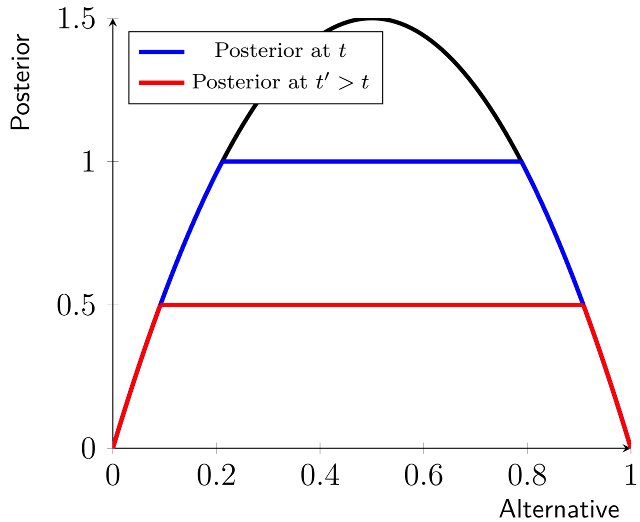


Figure 3.2: The posterior $g_i(t, x)$ levels the prior $f(x)$ over time.

satisfies

$$\int_X \left(1 - \frac{\bar{g}(t)}{f(x)}\right) \mathbf{1}_{\{f(x) \geq \bar{g}(t)\}}(x) dx = t \tag{3.4}$$

for all $t \in [0, 1]$ and $\bar{g}(t) = 0$ for $t > 1$. We then define the **leveling strategy** $\bar{\rho} : T \times X \to [0, 1]$ in terms of the leveling function as

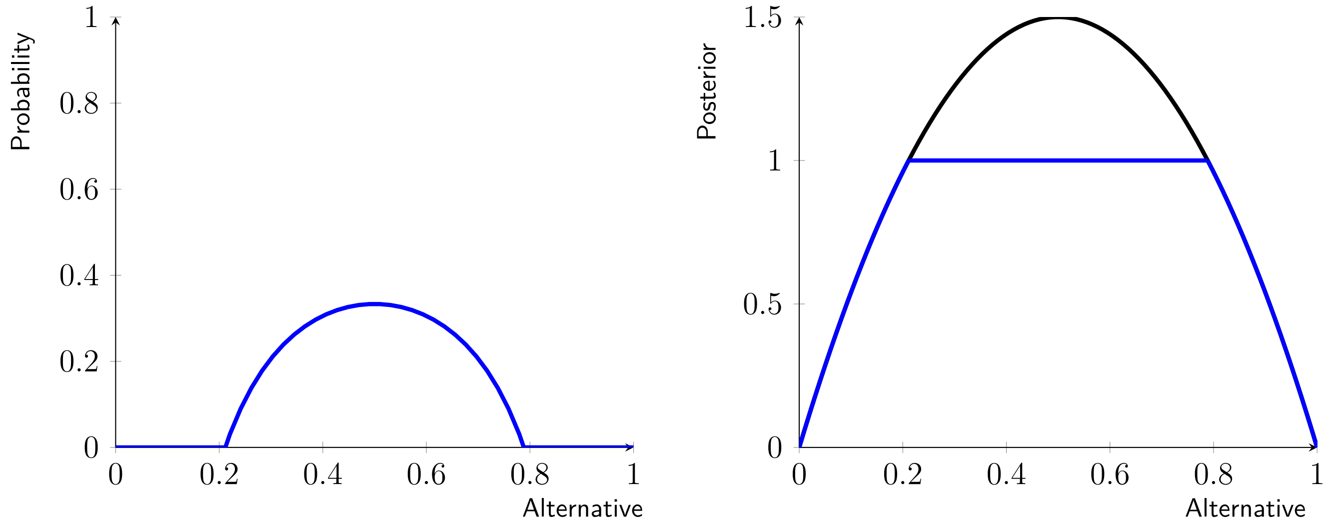
$$\bar{\rho}(t, x) := \left(1 - \frac{\bar{g}(t)}{f(x)}\right) \mathbf{1}_{\{f(x) \geq \bar{g}(t)\}}(x) \tag{3.5}$$

for all $t \in T$ and $x \in X$. As the integrand on the left-hand side of Equation (3.4) equals $\bar{\rho}$, Equation (3.4) corresponds to the binding capacity constraint in Definition 2. Therefore, by comparing Equation (3.3) and Equation (3.5), we know that the posterior induced by $\bar{\rho}$ at t achieves its maximum $\bar{g}(t)$ on $\{x \in X : f(x) \geq \bar{g}(t)\}$. Therefore the leveling strategy $\bar{\rho}$ levels the prior f over time.

Lemma 1. *The leveling function $\bar{g}$ exists and is unique, absolutely continuous, convex, and strictly decreasing on $[0, 1]$. The leveling strategy $\bar{\rho}$ is a well-defined distributional strategy.*

With some abuse of notation, we write the posterior density induced by the leveling strategy at time t as $\bar{g}(t, x) := (1 - \bar{\rho}(t, x))f(x)$ and call it the **leveling posterior** at t. We reiterate that it is player i's leveling strategy $\bar{\rho}$ that levels player $-i$'s posterior $\bar{g}$.

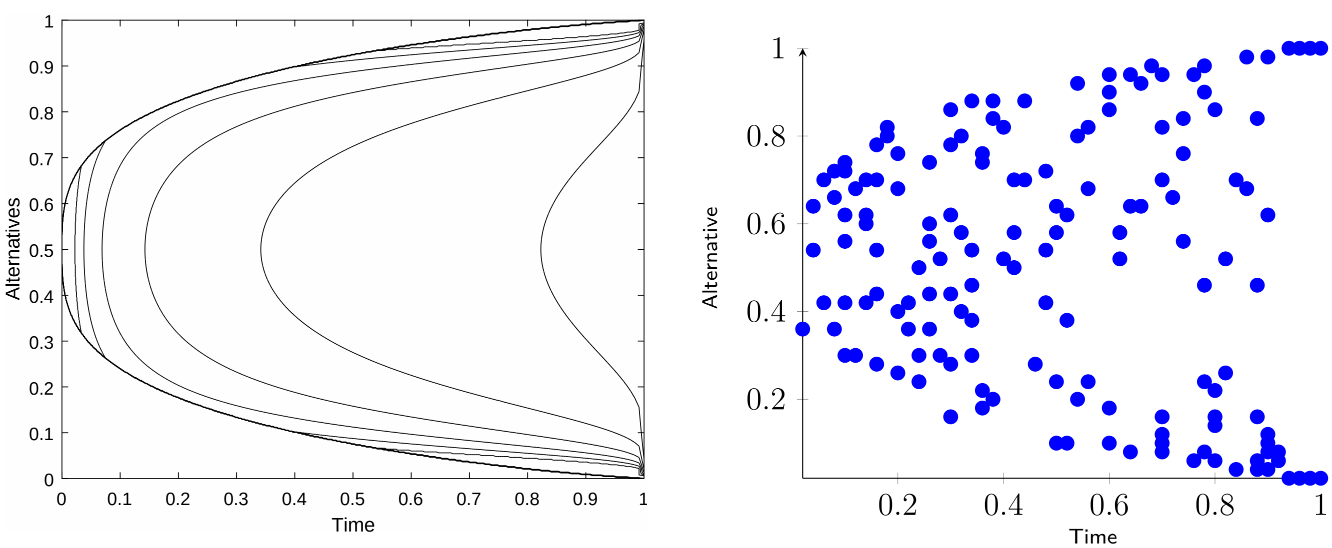
We demonstrate the relationship between the leveling strategy, the prior, and the leveling posterior in Figure 3.3, and illustrate the implementation of exploration over time Figure 3.4, where $\partial_t \bar{\rho}$ describes how intensive an alternative x is explored at time t.



(a) leveling strategy

(b) leveling posterior

Figure 3.3: The leveling strategy $\bar{\rho}$ and the posterior density $\bar{g}$, at a fixed time



(a) contour plot of the intensity of exploration $\partial_t \bar{\rho}$

(b) a discretized realization of exploration

Figure 3.4: Exploration over time according to the leveling strategy $\bar{\rho}$

3.3 Unique Equilibrium

We show that the symmetric leveling strategy profile is the unique Nash equilibrium of the strategic exploration game, and then discuss its economic implications.

Theorem 1. *The strategy profile $(\bar{\rho}, \bar{\rho})$ is the unique Nash equilibrium.*

The proof is contained in Appendix A.3. The fact that $(\bar{\rho}, \bar{\rho})$ is an equilibrium follows from its construction. Since the leveling strategy is t-continuous, the flow payoff of exploring an alternative equals the posterior. The equilibrium strategies are **mutually greedy** in that each player will explore only alternatives x with the highest posterior/myopic payoff at each time t, $\bar{g}(t)$, induced by the opponent's strategy to maximize his own myopic value of discovery. While myopic best responses are determined solely by the posterior beliefs (the prioritization motive), dynamic best responses in addition take into account the option value of exploration, or how quickly posteriors decline (the preemption motive). The leveling posterior of the most promising alternatives declines at the same rate, so the myopically optimal leveling strategy is indeed dynamically optimal against the opponent's leveling strategy.

A **coordinated exploration** prioritizes alternatives according to the prior density without preemption (i.e., the greedy exploration of the one-player problem described in Section 2.4, but with a combined capacity of 2). The strategic motives distort the equilibrium exploration in two ways from the coordinated exploration. First, the players prioritize the most promising alternatives a posteriori but not a priori. They explore many a priori less promising alternatives before exhausting the more promising ones. Second, the players preempt each other by duplicating the opponent's exploration in an extreme way. From start to finish, each player explores only the alternatives that his opponent could have already explored. Due to the two distortions, the equilibrium discovery time first-order stochastically dominates, i.e., is slower than, the coordinated counterpart. The instantaneous probability of discovery by one player at t equals the highest posterior $\bar{g}(t)$. Therefore, the probability that a discovery is made by time t is

$$P(t) = \int_0^t 2\bar{g}(s)ds. \tag{3.6}$$

For uniqueness, we note that the leveling strategy $\bar{\rho}$ guarantees a payoff of $\pi/2$ regardless of the opponent's strategy, i.e., both players can guarantee half of the total payoff. Therefore, it suffices to find, against each non-leveling strategy, a deviation with a payoff above $\pi/2$. As shown in Figure 3.2, the leveling strategy explores at full capacity such that the posterior or flow payoff decreases uniformly over time. For any other strategy, there must exist an interval of time over which the posterior declines faster for one set of alternatives and slower for another. One can then modify the leveling strategy to preempt that strategy by prioritizing the former set at the expense of the latter, in the spirit of the "one-step-ahead" strategy to achieve a higher payoff.

We conclude this section by highlighting the robustness of the model.

Remark 2. (**Robustness to Modeling Details**) The model is parsimonious to capture the essence of the strategic tension, independent of some modeling details. Theorem 1 continues to hold verbatim in the following environments: (a) the first discoverer enjoys a larger but not exclusive share of the prize, or payoff sharing is arbitrary in the case of simultaneous discovery, since in both cases the unique equilibrium strategy is t-continuous; (b) the space of alternatives is multidimensional, because of the one-to-one correspondence between the leveling function and the leveling strategy (of course, actual exploration activities depend on the space of alternatives); (c) the players value the weighted sum of prizes from multiple (or from a continuum of) independent and identically distributed good alternatives.

Remark 3. (**Robustness to Time Preferences**) The equilibrium in leveling strategies in Theorem 1 is robust to the introduction of exponential discounting, but we do not know its uniqueness. In many applications, the reward is increasing in time. For example, a technology may become safer and more reliable over time, and its premature exploitation may sometimes lead to adverse outcomes. This corresponds to a strictly increasing reward function $\beta : [0, T] \to \mathbb{R}_+$. When a player discovers the good alternative at time t, he enjoys a payoff of $\beta(t)$. Without competition, players would prefer waiting to exploring a premature technology. Nevertheless, Theorem 8 in Online Appendix B.1 shows that the equilibrium in Theorem 1 is the unique Nash equilibrium even in this case. The result demonstrates the payoff effect and information advantage behind the strategic tension between preemption and prioritization. If an early exploration leads to a discovery, the player enjoys the current payoff but eliminates the possibility of a later discovery that can be much more valuable. However, if early explorations fail, the player can preempt his opponent in future explorations by concentrating his capacity on the remaining alternatives. We show that the information advantage dominates the payoff effect so that the players cannot coordinate on delayed explorations. Moreover, they reap the full information advantage by prioritizing the most promising alternatives a posteriori. Therefore, the unique equilibrium features the mutually greedy leveling strategy.

Remark 4. (**Multiple Players**) The equilibrium in leveling strategies can be extended to an environment with more than two players, where the players expand the set of alternatives to explore in such a way that equalizes the posterior probabilities. Theorem 9 in Online Appendix B.3 shows that this is the unique symmetric equilibrium. Theorem 10 show that the good alternative is discovered earlier with more players in the first-order stochastic dominance sense. However, there exist asymmetric equilibria with asymmetric payoffs even though the players are ex ante identical: instead of competing over the same set of alternatives, the players form different groups that specialize in different segments of alternatives; competition occurs within groups and coordination occurs across groups.

3.4 Impact of Prior Beliefs

We study how the prior distribution affects the equilibrium exploration, and show that the good alternative is discovered more quickly if the prior is less evenly distributed.

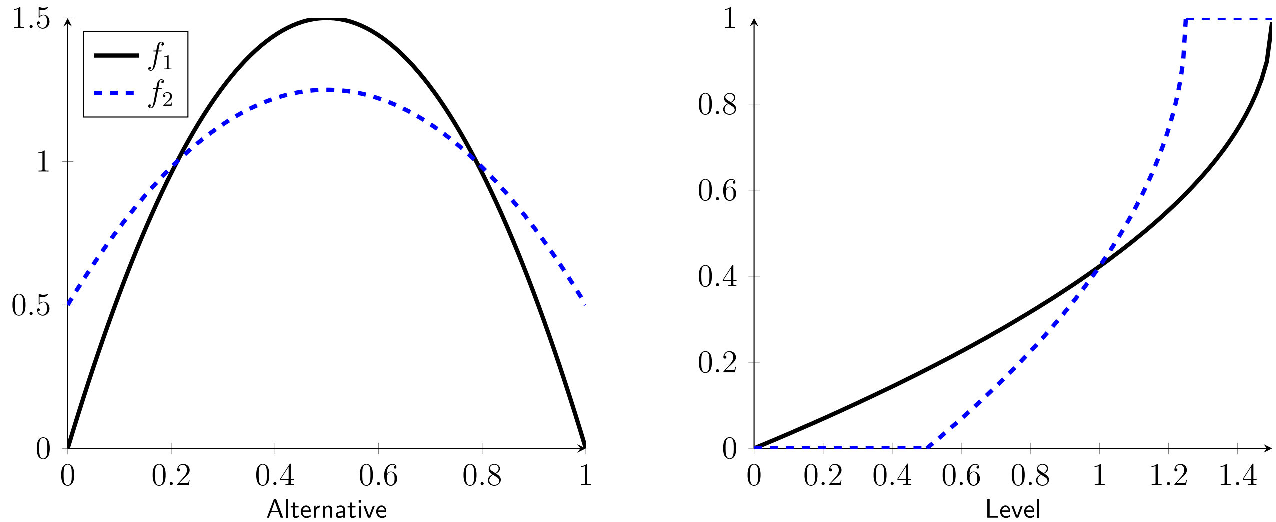
With the probability of existence of the good alternative π fixed, we vary the evenness of the prior distribution. Let λ be the Lebesgue measure. For any prior distribution f, let $\lambda \circ f^{-1}$ be the pushforward measure over $\mathbb{R}_+$. Note that $\lambda \circ f^{-1}(\mathbb{R}_+) = \lambda([0,1]) = 1$, so the pushforward measure is a probability measure of the prior density. Its expectation is $\int_{\mathbb{R}_+} y d\lambda \circ f^{-1} = \int_{[0,1]} f(x) d\lambda = \pi$. For example, if the good alternative is uniformly distributed over X, then $f(x)$ is a constant and $\lambda \circ f^{-1}$ assigns probability 1 to a single point. This is the case where the good alternative is most evenly distributed over $X = [0,1]$. We can then capture the evenness of a prior distribution by its pushfoward measure. Figure 3.5 illustrates the partial order of evenness.

Definition 3. *Let f_1 and f_2 be two prior distributions. We say that f_2 is **more even than** f_1 if $\lambda \circ f_1^{-1}$ is a mean-preserving spread of $\lambda \circ f_2^{-1}$.*

The good alternative is discovered more quickly if the prior distribution is less even. The players concentrate their exploration, which increases preemptive duplications but prioritizes more promising alternatives. We show that the prioritization effect dominates.

Theorem 2. *If f_2 is more even than f_1, then the distribution of the equilibrium discovery time associated with f_2 first-order stochastically dominates that associated with f_1; i.e., the good alternative is discovered earlier with f_1 than with f_2.*

The comparative statics is intuitive given our equilibrium characterization but it is not obvious a priori. One may have anticipated Theorem 2 since it also applies to coordinated exploration. We would like to highlight our result that the players cannot coordinate on less duplicative explorations *regardless of the prior distribution*. A priori, one might expect the players to specialize in distinct sets of alternatives to speed up equilibrium exploration when the prior distribution is more even. It would be more difficult to coordinate when



(a) prior density functions

(b) pushforward measures

Figure 3.5: Two distributions f_1 and f_2, where f_2 is more even than f_1.

the prior is less even because of the strengthened incentives to prioritize the most promising alternatives. We have shown, however, that such coordination is impossible *for any prior distribution*. By ruling out all possible coordination, our equilibrium characterization makes it straightfoward to prove the monotonicity in discovery time.

4 Extensions

Having analyzed the strategic exploration in the simplest setting, we shall extend the analysis to three settings of interest. These extensions are by no means exhaustive, but they will demonstrate how the strategic tension between preemption and prioritization plays out in richer settings of search and learning.

4.1 Short Deadlines

The model assumes that the deadline is sufficiently long so that a player can exhaustively explore all alternatives. Theorem 1 shows that coordination is impossible to attain. Can a short deadline $T < 1$ make coordination possible? The answer is in the affirmative but there is a limit to the scope of coordination.

We start developing the intuition from the mutually greedy strategy profile over $[0, T]$. Suppose that players level each other's posteriors up till $T - \epsilon$ for some small $\epsilon > 0$. At this moment each player faces a set of unexplored alternatives with the highest posteriors.

Since $T < 1$, this set is large relative to the remaining time ϵ, which is very short. The two players can divide the set into two disjoint segments, one for each player to search exclusively. No player has an incentive to preempt his opponent anymore, as long as his own segment contains enough alternatives to explore before the deadline. Therefore, *competition creates the condition for coordination*. This line of argument suggests that an equilibrium play consists of two phases: a phase of *competition* for a sufficiently long period of time followed by a phase of *coordination*.

We must also complement the above reasoning by the following two observations. First, the mutually greedy strategy profile is still an equilibrium in which the players level their posteriors till the deadline, completely crowding out the coordination phase. This equilibrium does not unravel backwards from the deadline simply because, if his opponent uses the leveling strategy, a player has no incentive to use a different strategy. For the same reason, in an equilibrium with two phases, a player does not have an incentive to enter the coordination phase unilaterally even if the competition phase has created a sufficiently large set of alternatives with the highest posterior.

Second, for long deadlines, there does not exist an equilibrium with a coordination phase as shown in Theorem 1. The reason is that, at any point in time $t < 1$, the remaining alternatives (let alone the alternatives with the highest posterior) are few relative to the remaining time and, hence, there would not be enough alternatives for the two players to split. Suppose to the contrary that there were a coordination phase, both players would exhaust their exploration of the most promising alternatives before the deadline, and they would have no choice but to continue to explore less promising alternatives; however, a player would covertly explore his opponent's segment early on. Therefore, the preemption incentive unravels the coordination phase under long deadlines.

Theorem 3. *For the strategic exploration game with deadline $T < 1$, there exists $T^*(T) \in [0, T)$ such that the following properties are satisfied:*

1. *If (ρ_1, ρ_2) is a Nash equilibrium, then there exists $t^* \in [T^*(T), T]$ that divides the equilibrium play into two phases:*
   (a) *Competition: for all $t \in [0, t^*]$, $\rho_1(t, \cdot) = \rho_2(t, \cdot) = \bar{\rho}(t, \cdot)$; i.e., players play the leveling strategy.*
   (b) *Coordination: for all $t \in (t^*, T]$, if $\rho_i(T, x) - \rho_i(t^*, x) > 0$ then $g_{-i}(t, x) = \overline{g}(t^*)$ and $\rho_{-i}(T, x) - \rho_{-i}(t^*, x) = 0$; i.e., players explore only alternatives with the posterior $\overline{g}(t^*)$ and they do not duplicate each other's search after t^*.*

2. *For any $t^* \in [T^*(T), T]$, there exists a Nash equilibrium (ρ_1, ρ_2) that satisfies (a) and (b) above.*

Furthermore, $\lim_{T\to 1} T^*(T) = 1$ *and* $\lim_{T\to 0} T^*(T) = 0$.

The proof in Appendix A.5 characterizes the critical threshold as

$$T^*(T) = \min\left\{t \in [0,T] : \int \frac{\overline{g}(t)}{f(x)} \mathbf{1}_{f(x)\geq \overline{g}(t)} dx \geq 2(T-t)\right\}. \tag{4.1}$$

Since $\lim_{T\to 1} T^*(T) = 1$, the coordination phases in all equilibria vanish as the deadline approaches 1, when each player can exhaust all alternatives. Since $T^*(T) < T$, there must be a continuum of equilibria with a coordination phase with transition times $t^* \in [T^*(T), T)$. The equilibrium with transition time $t^* = T^*(T)$ has the longest coordination phase. The leveling equilibrium (with transition time $t^* = T$) has the most duplicated search. Property (b) in Theorem 3 does not uniquely determine a distributional strategy, because there are multiple ways to partition alternatives with the posterior $\overline{g}(t^*)$ into two segments and players' searches over their own segments are flexible.

We shall use the uniform distribution to demonstrate the role of deadlines.

Example 2. *Suppose $f \equiv 1$. Then the leveling strategy $\bar{\rho}$ is given by $\bar{\rho}(t,x) = t$.*

- *For long deadlines $T \geq 1$, the unique equilibrium is the leveling equilibrium $(\bar{\rho}, \bar{\rho})$ by Theorem 1.*

- *For very short deadlines $T \in (0, \frac{1}{2}]$, the transition to coordination can start immediately $T^*(T) = 0$. The two extremes are the equilibrium with $t^* = 0$ (coordination starts immediately) and the equilibrium with $t^* = T$ (the leveling equilibrium). The fact that $T^*(T) = 0$ and hence equilibrium coordination can start immediately is a knife-edge case, which, by Theorem 3, only occurs when the prior density plateaus for a set of alternatives with a measure of at least $2T$, so that the two players do not need to compete over the course of the game.*

- *For moderately short deadlines $T \in (\frac{1}{2}, 1)$, the earliest transition time is $T^*(T) = 2T - 1$. To see this, note that in the equilibrium where players use the leveling strategy till $T^*(T)$, the set of alternatives X will be divided equally between the two players for them to explore over the remaining time $T - T^*(T)$. In the competition phase prior to $T^*(T)$, each player has explored a measure of $\frac{1}{2}T^*(T)$ in each segment, leaving a measure of $\frac{1}{2} - \frac{1}{2}T^*(T)$ for the player. Therefore, $T - T^*(T) = \frac{1}{2} - \frac{1}{2}T^*(T)$ and, hence, $T^*(T) = 2T - 1$.*

4.2 Asymmetric Capacity

In this section, we analyze the strategic tension between prioritization and preemption under asymmetric capacity, and uncover a number of new interesting implications veiled in the symmetric problem. We shall assume a long deadline $T \geq 1$ as in the benchmark model.

The two players have different capacities: player 1 can explore measure 1 of alternatives per unit time but player 2 can explore only measure $\alpha \in (0, 1]$. Player 1 (the "strong" player, she) is more capable or more resourceful than player 2 (the "weak" player, he) at exploration. We refer to α as player 2's *capacity*. Thus, player 2's distributional strategy $\rho_2^\alpha : T \times X \to [0, 1]$ should satisfy the new capacity constraint

$$\int_X (\rho_2^\alpha(t,x) - \rho_2^\alpha(s,x))dx \leq \alpha(t-s) \text{ for all } [s,t] \subset T \tag{4.2}$$

as well as the first three conditions in Definition 2.

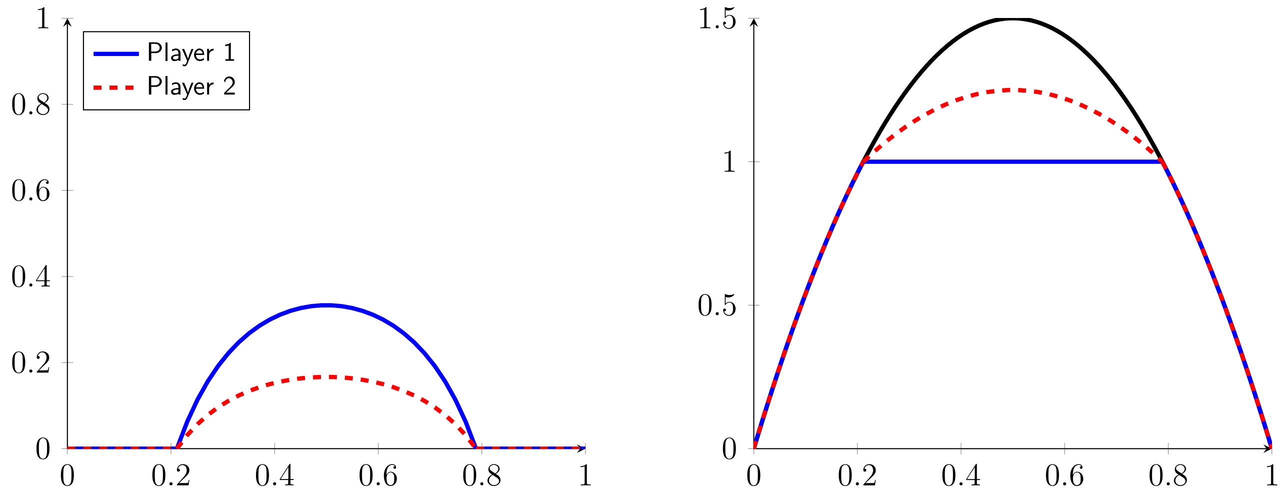
4.2.1 Unique Equilibrium

We first introduce the equilibrium strategy of the weak player. Consider a distributional strategy ρ_2^α of the weak player $\rho_2^\alpha(t,x) = \alpha\bar{\rho}(t,x)$ for all $t \in [0,1]$ and $x \in X$. The corresponding posterior $g(t,x) = (1 - \alpha\bar{\rho}(t,x))f(x)$ decreases uniformly just like the leveling strategy but at a slower speed. The fractional strategy $\alpha\bar{\rho}$, however, no longer levels the posterior, and explores any given alternative only with probability α by $t = 1$. Figure 4.1 demonstrates the distinction of $\bar{\rho}$ and $\alpha\bar{\rho}$.

In the unique equilibrium with asymmetric capacity, the strong player plays the leveling strategy $\bar{\rho}$ and the weak player plays the fractional strategy $\alpha\bar{\rho}$. To develop intuition for this candidate equilibrium, we explain how it is related to the unique equilibrium in the symmetric case.

First, although the two players differ in their capacity of exploration, they randomize over the same expanding set of alternatives. This results from the preemption motive: if the weak player concentrates on a smaller set of alternatives, the strong player can preempt him by exploring this set more intensively. Consequently, the weak player cannot explore the same set of alternatives as intensively as the strong player. In the candidate equilibrium, the weak player never explores any alternative with probability more than α, although a priori player 2 can choose to explore a subset of them with probability greater than α.[14]

[14] If we model each alternative as a Poisson process with the arrival rate controlled by resource allocated to the alternative (see Section 4.3), the weak player would cover as many alternatives as the strong player



(a) strategy profile $(\bar{\rho}, \alpha\bar{\rho})$

(b) posterior densities

Figure 4.1: The strategy profile $(\bar{\rho}, \alpha\bar{\rho})$ and the corresponding posterior densities, at a fixed time. In this example, player 2 is half as capable as player 1, i.e., $\alpha = 1/2$.

Therefore, the preemption motive drives the weak player to conduct *extensive*, but not *intensive*, explorations.

Second, unlike the symmetric case, the asymmetric capacity drives a wedge between the myopic payoff and the dynamic option value of exploration for the strong player. Although the weak player cannot equalize the posterior faced by the strong player over the common support as shown in Figure 4.1, the induced posterior declines uniformly at a constant rate by the construction of $\alpha\bar{\rho}$. The equal option values makes the strong player's leveling strategy dynamically optimal, but not myopically optimal. Nevertheless, the strong player's leveling strategy $\bar{\rho}$ and its induced leveling posterior make the weaker player's extensive exploration both myopically and dynamically optimal, as in the symmetric case. To summarize, the strong player's strategy is leveling but not greedy, while the weaker player's strategy is greedy but not leveling.

Theorem 4. *The profile of distributional strategies $(\bar{\rho}, \alpha\bar{\rho})$ gives the unique Nash equilibrium outcome of the game with asymmetric players. Player 1's equilibrium payoff is $\left(1 - \frac{1}{2}\alpha\right)\pi$ and player 2's equilibrium payoff is $\frac{1}{2}\alpha\pi$.*

We note that the strong player is able to explore all alternatives at $t = 1$, so the weak player's exploration activities afterwards is irrelevant. Therefore, the equilibrium outcome is uniquely pinned down by $(\bar{\rho}, \alpha\bar{\rho})$ even if we want to assume that the weak player continues to search after $t = 1$.

does, but would spend less of the resources on each alternative.

The strong player enjoys a disproportionately larger share of the payoff, $(2-\alpha):\alpha$ than of the capacity, $1:\alpha$. It is as if player 1 monopolizes a fraction $1-\alpha$ of the total surplus and then splits the remaining fraction evenly with player 2. For example, if the strong player is twice as fast as the weak player $\alpha=\frac{1}{2}$, the payoff share is $(\frac{3}{4},\frac{1}{4})$. The strong player's payoff is three times as much as the weak player's. By comparison, in a three-player game in which the more resourceful player is split into two equal selves, the payoff share will be $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ in the symmetric three-player equilibrium (see Online Appendix B.3 for the extension). The excess payoff of player 1 beyond the sum of her two selves is due to the pooled information of the two: knowing which alternatives have been explored by herself, player 1 does better as one big player than as an ensemble of smaller selves who may duplicate each other's explorations.

The connection between the asymmetric case and the symmetric case is as follows. For every strategy ρ_2^α of player 2, define $\rho_2 := \rho_2^\alpha/\alpha$. It is easy to verify that $\rho_2 : T\times X \to [0,1/\alpha]$ satisfies the four conditions of Definition 2. It differs from a distributional strategy in its codomain $[0,1/\alpha]$ instead of $[0,1]$. We shall call $\rho_2 : T\times X \to [0,1/\alpha]$ a **normalized strategy**. Players' payoffs from the strategy profile (ρ_1,ρ_2^α) can be rewritten as payoffs from (ρ_1,ρ_2) as follows:

$$u_1(\rho_1,\rho_2^\alpha) = (1-\alpha)\pi + \alpha u_1(\rho_1,\rho_2); \tag{4.3}$$

$$u_2(\rho_2^\alpha,\rho_1) = \alpha u_2(\rho_2,\rho_1). \tag{4.4}$$

Therefore, the payoff functions under asymmetric capacity are increasing affine transformations of those with a normalized strategy of player 2. Thus, the game with asymmetric capacity is strategically equivalent to the game with a normalized strategy, and the existence and uniqueness of the Nash equilibrium in the game with asymmetric players will follow from the existence and uniqueness in the game with normalized strategies. But the latter game is not quite the same as the symmetric game because of the codomain of the normalized strategy ρ_2; i.e., it is not a priori clear that $\rho_2(1,\cdot)=1$ in equilibrium. This gap is closed using the following proof strategy. We decompose the maximization over normalized strategies into two components: the (normalized) probability of exploration by the end of the game $\rho_2(1,\cdot)$, and the dynamic implementation of the exploration given this probability. We show that, for any probability of exploration, a generalized leveling strategy is optimal for player 2, and his payoff is uniquely maximized at $\rho_2(1,\cdot)=1$ given the leveling strategy.

4.2.2 Discovery Time

We investigate how equilibrium exploration depends on the asymmetry in capacity and show that the asymmetry speeds up the process of discovery. We fix the total capacity to 2 (as in the symmetric case), and vary the asymmetry between the two players. Formally, let $\gamma \in [1, 2)$ and consider the strategic exploration game in which the strong player has capacity γ and the weak player has capacity $2 - \gamma \in (0, 1]$.

We compute the distribution of discovery time in the unique equilibrium under asymmetric capacity. Applying a change of variables in Theorem 4, we know that the strong player plays $\rho_1(t, x) = \bar{\rho}(\gamma t, x)$, which levels the posterior $\bar{g}(\gamma t)$, and the weak player plays $\rho_2(t, x) = \frac{1-\gamma}{\gamma}\bar{\rho}(\gamma t, x)$, which does not level the posterior. For $t \geq 1/\gamma$, the strong player has exhausted all alternatives so the probability of discovery by t is $P_\gamma(t) = \pi$. For $t \in [0, 1/\gamma]$, the probability of discovery by t is given by

$$P_\gamma(t) = \int_X (f(x) - \bar{g}(\gamma t))\, \mathbf{1}_{\{f(x) \geq \bar{g}(\gamma t)\}} dx + (2 - \gamma)t\bar{g}(\gamma t).$$

The two terms correspond to a hypothetical sequential exploration. The first term is the probability of discovery by the strong player if she were to level the posterior before the weak player explored anything. The second term is the probability of discovery by the weak player if he were to expend his cumulative capacity $(2 - \gamma)t$ on the leveled posterior.

Because the strong player always levels the posterior with all her capacity, the remaining capacity of the weak player replicates some of her exploration. As the strong player enjoys a larger share of capacity, there is less duplication and thus a faster discovery time. This property is illustrated in Figure 4.2 and formally established in Theorem 5.

Theorem 5. *The distribution of discovery time is decreasing in γ in the first-order stochastic dominance sense.*

As the strong player controls almost the total capacity, the duplication effect vanishes but the equilibrium discovery remains discontinuously slower than the coordinated exploration (Figure 4.2) due to the preemption motive. To avoid preemption by the weak player, the strong player must randomize exploration according to the leveling strategy even as $\gamma \to 2$. The randomization prevents her from prioritizing the most promising alternatives, either a priori or a posteriori, and thus slows down discovery compared to the fastest, coordinated exploration.[15]

[15] Nonetheless, there is no discontinuity in payoffs. Indeed, when the strong player controls the total

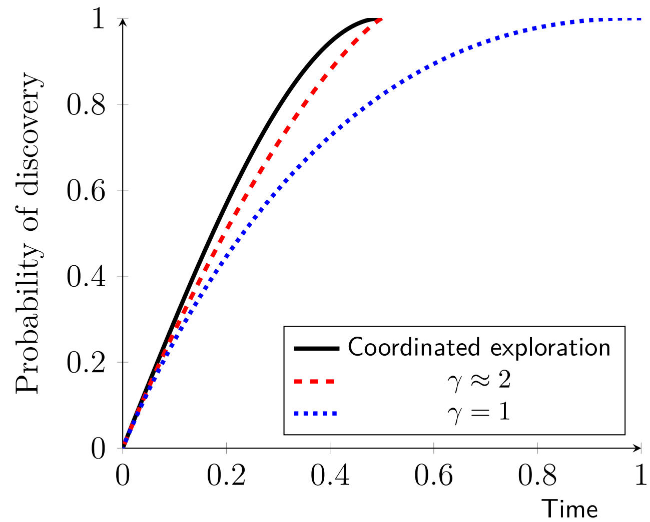


Figure 4.2: Distribution of discovery time for the coordinated exploration and equilibria with different divisions of capacity.

4.2.3 Endogenous Capacities

The exact characterization of equilibrium payoffs in Theorem 4 facilitates the study of capacity investment in an augmented game. Suppose that player i can choose a capacity α_i at a constant marginal cost c_i before entering the strategic exploration game, where player 2 faces a higher marginal cost $c_2 > c_1 > 0$. It follows from Theorem 4 that, for $\alpha_1 \geq \alpha_2$, the two players' equilibrium payoffs from the strategic exploration game (before paying the cost) are $(1 - \frac{1}{2}\alpha)\pi$ and $\frac{1}{2}\alpha\pi$, respectively, where $\alpha := \alpha_2/\alpha_1$. Therefore, the equilibrium capacities (α_1^*, α_2^*) satisfy the following conditions:

$$\alpha_1^* \in \operatorname{argmax}_{\alpha_1 \geq 0} \left(1 - \frac{\alpha_2}{2\alpha_1}\right)\pi - c_1\alpha_1,$$

$$\alpha_2^* \in \operatorname{argmax}_{\alpha_2 \geq 0} \left(\frac{\alpha_2}{2\alpha_1}\right)\pi - c_2\alpha_2.$$

It can be shown that $\alpha^* = c_1/c_2$. Player 1 earns a strictly positive net profit $(1 - c_1/c_2)\pi$ while player 2 dissipates his return from exploration through capacity investment. In the limit of $c_1 = c_2 > 0$, both players' net payoffs are zero.

If the cost of investment is $c_i(\alpha_i)^2$, where $c_2 > c_1 > 0$, then a similar line of argument shows that $\alpha^* = (c_1/c_2)^{\frac{1}{2}}$. The two players' net payoffs are $\left(1 - \frac{3}{4}(c_1/c_2)^{\frac{1}{2}}\right)\pi > 0$ and

capacity, all exploration strategies are optimal. As in many known results, discontinuity is often due to different orders of convergence. In our analysis, we take the limit of asymmetry in the model that is the patient, no-discounting limit.

$\frac{1}{4}\left(c_1/c_2\right)^{\frac{1}{2}}\pi > 0$, respectively. In the limit of $c_1 = c_2 > 0$, both players' net payoffs are $\frac{1}{4}\pi > 0$.

Research personnel is a major component of exploration capacity in applications such as the technology race between superpowers, and migration of research personnel is one way capacity can change. Although the number of researchers can be assumed to be relatively stable, researchers migrate from one country to the other. To this end, suppose that the total exploration capacity is 1 and the weak player has a fraction $\theta < \frac{1}{2}$ of the total capacity. By Theorem 4, the weak player's equilibrium payoff share is $\frac{1}{2}\frac{\theta}{1-\theta}$. The elasticity of his payoff share with respect to his capacity share is $\frac{1}{1-\theta} > 2$. Thus, migration always has an outsized impact on the weak player. For the strong player, the elasticity of her payoff share with respect to her own capacity share $1-\theta$ is $\frac{1}{3(1-\theta)-1}$, and unit elasticity is attained when her capacity share is $\frac{2}{3}$. Hence, although the strong player always benefits from the migration of researchers, the scale depends on her existing capacity.

4.3 Poisson Learning

In this section, we consider the strategic exploration game where each player finds out whether an alternative is good gradually instead of instantaneously as in the benchmark model.

Each player allocates his resources among the set of alternatives in order to find the good alternative. We again assume a long deadline $T \geq 1$. A conclusive signal on the alternative arrives at a Poisson rate proportional to the flow rate of resources on the alternative, and reveals whether that alternative is good or not.[16] Let $r(t, x)$ be the cumulative amount of resources a player has spent on alternative $x \in X$ by time $t \in T$ *conditional on no signal arrival before* t. The probability of a signal arrival by time t is thus $1 - e^{-r(t,x)}$. If $r(\cdot, x)$ is differentiable in t, the time derivative $\partial_t r(t, x)$ is the arrival rate of the potentially non-stationary Poisson process associated with alternative x.

Analogous to a distributional strategy, a function $r : T \times X \to \mathbb{R}_+ \cup \{\infty\}$ is a *resource allocation strategy* if it satisfies the following four conditions:

1. Initial condition: $r(0, x) = 0$ for all $x \in X$;
2. Monotonicity and right-continuity: $r(\cdot, x)$ is increasing and right-continuous for all $x \in X$;

[16] In other words, the arrival process is independent of the state of the alternative and the absence of signals per se leads to no belief updating, as in Akcigit and Liu (2015). See also Mayskaya and Nikandrova (2022) for a model where the arrival rate is state dependent.

3. Measurability: $r(t, \cdot)$ is measurable for all $t \in T$;

4. Capacity constraint: $\int_X \left(e^{-r(s,x)} - e^{-r(t,x)}\right) dx \leq t - s$ for all $[s, t] \subset T$.

The first three conditions extend directly from distributional strategies but we shall elaborate on the capacity constraint. The actual expenditure on alternative x by time t is not the same as $r(t, x)$ because the latter conditions on no signal arrival; in fact, no more resources will be spent once a signal arrives. The expected amount of resources expended on x by t is

$$\int_0^{r(t,x)} e^{-q} dq = 1 - e^{-r(t,x)},$$

where e^{-q} is the probability of no signal arrival given the cumulative resources q. As the arrival of Poisson signals is independent across alternatives, the law of large numbers implies that the actual resources expended across all alternatives equals its expectation.[17] The capacity constraint therefore means that the aggregate amount of resources available to each player per unit time are normalized to 1.

The continuum of alternatives is essential to the straightforward capacity constraint. By contrast, one needs to specify resource allocation *for each history of signal arrivals* if there are finitely many alternatives only. In that case, the stochastic signal arrivals impart aggregate uncertainty about the remaining alternatives. The capacity constraint must restrict the actual resource expended, which no longer equals its expectation, for each realization of signal arrivals.

We derive the unique equilibrium of the strategic exploration game with Poisson learning by noting a one-to-one correspondence between resource allocation strategies and distributional strategies. Given player i's resource allocation strategy r_i, the probability of signal arrival from alternative x by time t is $\rho_i(t, x) := 1 - e^{-r_i(t,x)} \in [0, 1]$. With this one-to-one relationship between r_i and ρ_i, it is immediate that r_i is a resource allocation strategy if and only if ρ_i is a distributional strategy that satisfies the four conditions in Definition 2. Player i's expected payoff from a profile (r_i, r_{-i}) is the same as $u_i(\rho_i, \rho_{-i})$ as defined in Equation (2.1). Therefore, the strategic exploration game with Poisson learning is isomorphic to the main model with instantaneous arrival. Define a resource allocation strategy

$$\bar{r} := -\log\left(1 - \bar{\rho}\right),$$

[17] As more signals arrive, the player can concentrate his resource on the remaining alternatives, expediting the arrival of Poisson signals over those alternatives. Therefore, he is able to learn about all alternatives at $t = 1$ with $r(1, x) = \infty$ almost everywhere.

where $\bar{\rho}$ is the leveling strategy. The following result is immediate from Theorem 1.

Corollary 1. *In the strategic exploration game with Poisson learning, the profile $(\bar{r}, \bar{r})$ is the unique Nash equilibrium in resource allocation strategies.*

We note that Remark 2 remains valid here: the equilibrium characterization in resource allocation strategies is invariant to payoff-sharing rules, the space of alternatives, and the multiplicity of good alternatives. It is clear that all previous results in distributional strategies apply to the resource allocation strategies, mutandis mutatis.

A Proofs

A.1 Proof of Lemma 1

Proof. Since $\bar{g}$ is a constant function on $t > 1$, it suffices to prove the lemma on $t \in [0, 1]$.

For $y \in [0, \sup f]$, let

$$h(y) := \int_X \left(1 - \frac{y}{f(x)}\right) \mathbf{1}_{\{f(x) \geq y\}}(x) dx. \tag{A.1}$$

For $x \in X$, the integrand $\left(1 - \frac{y}{f(x)}\right) \mathbf{1}_{\{f(x)\geq y\}}(x)$ is decreasing in y and strictly so for $f(x) > y$, which has positive measure for $y < \sup f$. Thus, h is strictly decreasing. In addition, the integrand is continuous, and therefore h is continuous by the dominated convergence theorem. The convexity of h also follows from that of the integrand.

The function h is continuous and strictly decreasing with $h(0) = 1$ and $h(\sup f) = 0$. Therefore, there exists a unique, continuous, and strictly decreasing function $\bar{g} = h^{-1}$ that solves Equation (3.4). Since h is strictly decreasing, its inverse $\bar{g}$ is also convex. The absolute continuity of $\bar{g}$ follows from its continuity and convexity.

We verify that $\bar{\rho}$ is a well-defined distributional strategy. It is straightforward to check that $\bar{\rho}$ satisfies the initial condition. The function $(x, y) \mapsto \left(1 - \frac{y}{f(x)}\right) \mathbf{1}_{\{f(x)\geq y\}}$ is continuous and decreasing in y. Together with the continuity and monotonicity of $\bar{g}$, this property implies that $\bar{\rho}$ is continuous in t and satisfies the monotonicity and right-continuity condition. The function is also measurable in x and hence $\bar{\rho}$ satisfies the measurability condition. Finally, $\bar{\rho}$ respects the capacity constraint by Equation (3.4). $\square$

A.2 Auxiliary Game

To prove the theorems that characterize equilibria in the benchmark model and its extensions, we first develop some general tools. We define an auxiliary game called the timing game in which players' strategy is the timing of exploration, but not the probability of exploration by the end of the game.

A.2.1 The timing game and leveling strategies

Consider two functions $\Delta\rho_1 : X \to [0,1]$ and $\Delta\rho_2 : X \to [0, 1/\alpha]$ for some $\alpha \in (0,1]$ such that $\int_X \Delta\rho_i dx = T$ where $T \leq 1$.[18] We interpret $\Delta\rho_i(x)$ as the probability that x is explored before the deadline. To accommodate the asymmetric case, we allow the probability of exploration of the weak player to have an enlarged domain.

Definition 4 (Timing game)**.** *For $T \leq 1$ and $(\Delta\rho_1, \Delta\rho_2)$, the timing game is a game in which player 1 plays a distributional strategy, player 2 plays a normalized strategy subject to terminal condition $\rho_i(T, \cdot) = \Delta\rho_i(\cdot)$, and the payoff function is given by Equation (2.1).*

The timing game is a constant sum game since the sum of payoffs depends only on the probabilities of exploration.

For any timing game, we define $\Delta\rho_{\min} := \min\{\Delta\rho_1, \Delta\rho_2\}$ and $t^* := \int_X \Delta\rho_{\min} dx \in [0, T]$. We define the leveling strategy in a timing game, which generalizes the leveling strategy for the benchmark case $T = 1$ and $\Delta\rho_1, \Delta\rho_2 \equiv 1$.

Definition 5 (Leveling strategy in timing game)**.** *Function $\bar{g} : [0, t^*] \to [0, \sup_x f\Delta\rho_{\min}]$ is the leveling function if*

$$\int_X \left(\Delta\rho_{\min}(x) - \frac{\bar{g}(t)}{f(x)} \right) \mathbf{1}_{\{f(x)\Delta\rho_{\min}(x) \geq \bar{g}(t)\}}(x) dx = t \text{ for all } t \in [0, t^*].$$

Function $\bar{\rho} : [0, t^*] \times X \to [0,1]$ is the leveling strategy if

$$\bar{\rho}(t, x) := \left(\Delta\rho_{\min}(x) - \frac{\bar{g}(t)}{f(x)} \right) \mathbf{1}_{\{f(x)\Delta\rho_{\min}(x) \geq \bar{g}(t)\}}(x) \text{ for all } t \in [0, t^*] \text{ and } x \in X.$$

Lemma 2. *The leveling function $\bar{g}$ exists and is unique, absolutely continuous, convex, and strictly decreasing. The leveling strategy $\bar{\rho}$ is a well-defined strategy on $[0, t^*]$.*

[18] This T is not the same as the deadline of the original strategic exploration game, which can be strictly larger than 1.

Proof. If $\Delta\rho_{\min} \equiv 0$, $t^* = 0$ so the unique $\bar{g}(0) = 0$ satisfies the lemma trivially and so does the leveling strategy $\bar{\rho}(0, \cdot) = 0$.

Otherwise, $\Delta\rho_{\min} \not\equiv 0$ so $t^* > 0$. For $y \in [0, \sup f\Delta\rho_{\min}]$, let

$$h(y) := \int_X \left(\Delta\rho_{\min}(x) - \frac{y}{f(x)}\right) \mathbf{1}_{\{f(x)\Delta\rho_{\min}(x)\geq y\}}(x)dx. \tag{A.2}$$

For $x \in X$, the integrand $\left(\Delta\rho_{\min}(x) - \frac{y}{f(x)}\right) \mathbf{1}_{\{f(x)\Delta\rho_{\min}(x)\geq y\}}(x)$ is decreasing in y and strictly so for $f(x)\Delta\rho_{\min}(x) > y$, which has positive measure for $y < \sup f\Delta\rho_{\min}$. Thus, h is strictly decreasing. In addition, the integrand is continuous, and therefore h is continuous by the dominated convergence theorem. The convexity of h also follows from that of the integrand.

The function h is continuous and strictly decreasing with $h(0) = t^*$ and $h(\sup f\Delta\rho_{\min}) = 0$. Therefore, there exists a unique, continuous, and strictly decreasing function $\bar{g} := h^{-1}$ that solves Equation (3.4). Since h is strictly decreasing, its inverse $\bar{g}$ is also convex. The absolute continuity of $\bar{g}$ follows from its continuity and convexity.

We verify that $\bar{\rho}$ is a well-defined distributional strategy. It is straightforward to check that $\bar{\rho}$ satisfies the initial condition. The function $(x, y) \mapsto \left(\Delta\rho_{\min}(x) - \frac{y}{f(x)}\right) \mathbf{1}_{\{f(x)\Delta\rho_{\min}(x)\geq y\}}$ is continuous and decreasing in y. Together with the continuity and monotonicity of $\bar{g}$, this property implies that $\bar{\rho}$ is continuous in t and satisfies the monotonicity and right-continuity condition. The function is also measurable in x and hence $\bar{\rho}$ satisfies the measurability condition. Finally, $\bar{\rho}$ respects the capacity constraint by the definition of $\bar{g}$ in Definition 5. $\square$

We extend $\bar{g}$ to $(t^*, T]$ at value 0. With a slight abuse of notation, we call a strategy $\rho : [0, T] \times X \to [0, 1]$ leveling if $\rho|_{[0,t^*]\times X} = \bar{\rho}$. It is obvious that a t-continuous leveling strategy exists.

A.2.2 Equilibria of timing game

We show that the equilibria of the timing game are the leveling strategy profiles.

Theorem 6. *Strategy profile (ρ_1, ρ_2) is an equilibrium of the timing game if and only if ρ_1 and ρ_2 are both leveling.*

Overview. A leveling strategy profile is a Nash equilibrium of the timing game because it maximizes the myopic payoff, just as in the benchmark model. For the timing game, we generalize the posterior distribution as $g_i(t, x) := f(x)\left(\Delta\rho_{\min}(x) - \rho_i(t, x)\right)$.

All equilibria are leveling strategy profiles due to three lemmas. Lemma 3 states that, in equilibrium over $[0, t^*]$, each player can search only over the set of alternatives, which we call the upper contour set of $f\Delta\rho_{\min}$, that the leveling strategy randomizes over. Otherwise, his payoff against the leveling strategy will be lower than his equilibrium payoff.

Lemma 4 is key to Theorem 6. It states that, in equilibrium, the posterior declines fastest on the upper contour set. If instead the posterior declined slower in some subset of the upper contour set than in another subset for a period of time, the opponent could devise a *modified leveling strategy* that searches the former in place of the latter just before the period, and vice versa just after the period. The modification generalizes the "one-step-ahead" strategy in Section 3. The opponent's strategy would then preempt the player's strategy and yield a higher payoff than the leveling strategy, which cannot be true in equilibrium in a constant-sum game.

Lemma 4 has two useful implications. Corollary 2 establishes the t-continuity of the equilibrium strategy on $[0, t^*]$. If the posterior were discontinuous over some alternatives at some point, the posterior of all alternatives in the upper contour set would have to decline discontinuously. This would violate the capacity constraint. By applying Lemma 4 twice, we then obtain Corollary 3: the decrease in the posterior must be equal across the upper contour set.

Lemma 5 computes the equilibrium posterior within the upper contour set. As the decrease in the posterior is constant across the set of alternatives according to Corollary 3, the posterior is pinned down by the capacity constraint and the initial condition, which is exactly the leveling posterior defined by the leveling strategy.

Proof of Theorem 6. The payoff function can be written as

$$
\begin{aligned}
&u_i(\rho_i, \rho_{-i}) \\
&= \int_X \int_T f(1-\rho_{-i}) d_t\rho_i dx + \frac{1}{2}\int_X f \sum_{t\in D_x} \Delta_t\rho_i \Delta_t\rho_{-i} dx \\
&= \int_X f(1-\Delta\rho_{\min})\Delta\rho_i dx + \int_X\int_T f\left(\Delta\rho_{\min} - \rho_{-i}\right) d_t\rho_i dx + \frac{1}{2}\int_X f \sum_{t\in D_x} \Delta_t\rho_i\Delta_t\rho_{-i} dx \\
&\leq \int_X f(1-\Delta\rho_{\min})\Delta\rho_i dx + \int_X\int_T f\left(\Delta\rho_{\min} - \rho_{-i}^-\right) d_t\rho_i dx,
\end{aligned}
$$

where the inequality follows from an identity of Stieltjes integral

$$\int_T \rho_{-i}^- d_t\rho_i + \sum_{t\in D_x} \Delta_t\rho_i\Delta_t\rho_{-i} = \int_T \rho_{-i} d_t\rho_i.$$

Note that the integrand in the second utility term motivates the more general definition of the posterior distribution.

We verify that any leveling strategy profile (ρ_1, ρ_2) is a Nash equilibrium. Suppose that ρ_{-i} is leveling. Over $[0, t^*]$, it is t-continuous and hence $g^-_{-i} = g_{-i}$. The posterior g_{-i} is maximized at value $\bar{g}$ on H. Therefore, the leveling strategy ρ_i, which searches only on H, attains the maximum myopic payoff. Over $(t^*, T]$, the posterior is maximized at value 0 on $\{\Delta\rho_{\min} = \Delta\rho_{-i}\}$. Therefore, the leveling strategy ρ_i, which searches only on $\{\Delta\rho_{\min} < \Delta\rho_i\}$, attains the maximum myopic payoff.

Formally, for $x \in X$, let $\kappa_x \in \Delta(T)$ be the Lebesgue–Stieltjes measure induced by $\rho_i(\cdot, x)$. Then $\kappa_x([0, t]) = \rho_i(t, x)$. For any $t \in T$, we have

$$\int_X \kappa_x([0,t])dx = \int_X \rho_i(t,x)dx = t, \tag{A.3}$$

where the last equality follows from the capacity constraint of the distributional strategy ρ_i. Thus $\int_X \kappa_x dx \in \Delta(T)$ is the Lebesgue measure by the Caratheodory extension theorem. For any ρ_i, the payoff from $[0, t^*]$ is bounded by that of $\bar{\rho}_i$:

$$\begin{aligned}\int_X \int_{[0,t^*]} \bar{g}(t,x) d_t\rho_i dx &\leq \int_X \int_{[0,t^*]} \bar{g}(t) d_t\rho_i dx \\ &= \int_X \int_{[0,t^*]} \bar{g}(t) d\kappa_x dx \\ &= \int_{[0,t^*]} \bar{g}(t) d\left(\int_X \kappa_x dx\right) \\ &= \int_{[0,t^*]} \bar{g}(t) dt \\ &= \int_X \int_{[0,t^*]} \bar{g}(t) d_t\bar{\rho}_i dx, \end{aligned} \tag{A.4}$$

where the first equality is by the definition of Lebesgue–Stieltjes integration and the second equality follows from Fubini's theorem. The payoff from $(t^*, T]$ follows a similar inequality from the same argument.

Since ρ_i attains the maximum myopic payoff for all $t \in [0, T]$, it is a best response to ρ_{-i}. The equilibrium payoff is

$$\begin{aligned} u_i(\rho_i, \rho_{-i}) &= \int_X f(\Delta\rho_i - \Delta\rho_{\min})\Delta\rho_i dx + \int_X \int_{[0,t^*]} \bar{g} d_t\bar{\rho} dx \\ &= \int_X f(1 - \Delta\rho_{\min})\Delta\rho_i dx + \frac{1}{2}\int_X f(\Delta\rho_{\min})^2 dx \end{aligned}$$

$$= \int_X f\left((1-\Delta\rho_{\min})\Delta\rho_i + \frac{1}{2}(\Delta\rho_{\min})^2\right) dx. \tag{A.5}$$

All equilibria must give the same payoffs because the timing game is a constant-sum game.

For $t \in [0, t^*)$, denote $H(t) := \{x \in X : f(x)\Delta\rho_{\min} \geq \bar{g}(t)\}$ as the upper contour set of f. For $t = t^*$, define $H(t^*) := \lim_{s\uparrow t^*} H(s)$. We denote its complement by $H^C(t) := X \backslash H(t)$.

Lemma 3. *Let (ρ_1, ρ_2) be a Nash equilibrium. Then, for all $t_0 \in [0, t^*]$ and $i \in \{1, 2\}$, $\rho_i(t_0, x) = 0$ for $x \in H^C(t_0)$ almost everywhere.*

Proof. The statement for $t_0 = 0$ follows from the initial condition. Suppose that there exist a time $t_0 \in (0, t^*]$ and a positive-measure set $A \subset H^C(t_0)$ such that $\rho_i(t_0, x) > 0$ for all $x \in A$. Then the payoff of player i against a t-continuous leveling strategy of player $-i$ is strictly below the equilibrium payoff:

$$\begin{aligned}
& u_i(\rho_i, \bar{\rho}_{-i}) - u_i(\bar{\rho}_i, \bar{\rho}_{-i}) \\
&= \int_X \int_T g_{-i}(t,x) d_t\rho_i(t,x)dx - \int_X \int_T \bar{g}(t,x) d_t\bar{\rho}_i(t,x)dx \\
&= \int_X \int_T g_{-i}(t,x) d_t\rho_i(t,x)dx - \int_X \int_T \bar{g}(t) d_t\rho_i(t,x)dx \\
&\leq \int_A \int_{[0,t_0]} (g_{-i}(t,x) - \bar{g}(t))\, d_t\rho_i(t,x)dx \\
&< 0.
\end{aligned}$$

The second equality is due to the Fubini argument in Equation (A.4). The weak inequality follows from $g_{-i}(t,x) \leq \bar{g}(t)$ for all $t \in T$ and $x \in X$, and the strict one from the supposition that $g_{-i}(t,x) < \bar{g}(t)$ for all $t \leq t_0$ and $\rho_i(t_0, x) > 0$ for all $x \in A$. □

For $i \in \{1, 2\}$, $t \in T$, and $x \in X$, denote $g_i(t^-, x) := \lim_{s\uparrow t} g_i(s, x)$.

Lemma 4. *Let (ρ_1, ρ_2) be a Nash equilibrium. For $0 < t_0 < t_1 \leq t_2 < t^*$ and $i \in \{1, 2\}$,*

$$g_i(t_2, x_A) - g_i(t_1^-, x_A) \geq g_i(t_2, x_B) - g_i(t_1^-, x_B)$$

for $x_A \in X$ and $x_B \in H(t_0)$ almost everywhere.

Proof. For $x_A \in H^C(t_2)$ almost everywhere, Lemma 3 implies $g_i(t_2, x_A) - g_i(t_1^-, x_A) = 0$ and hence the inequality follows from the monotonicity condition.

Suppose that there exist positive-measure sets $A \subset H(t_2)$ and $B \subset H(t_0)$ such that $g_i(t_2, x_A) - g_i(t_1^-, x_A) < g_i(t_2, x_B) - g_i(t_1^-, x_B)$ for all $x_A \in A, x_B \in B$. Note that $f\Delta\rho_{\min} > 0$

on $A \cup B$. Without loss of generality, assume that $g_i(t_2, x_A) - g_i(t_1^-, x_A) < a < g_i(t_2, x_B) - g_i(t_1^-, x_B)$ for some $a < 0$, $\operatorname{ess\,inf}_{A\cup B} f\Delta\rho_{\min} > 0$, and $\int_A \frac{1}{f} = \int_B \frac{1}{f} > 0$; if this is not the case, replace the sets by some positive-measure subsets. Fix $\epsilon \in (0,1)$. We proceed in two steps.

Step 1: A Modified Leveling Strategy.

Let $\epsilon_1 := (t_1 - t_0)\epsilon > 0$. Let $\epsilon_2 > 0$ be a solution to $\bar{g}(t_1 - \epsilon_1) - \bar{g}(t_1) = \bar{g}(t_2) - \bar{g}(t_2 + \epsilon_2)$. For sufficiently small ϵ, it exists and is unique by the continuity and monotonicity of $\bar{g}$. In addition, $0 < t_1 - \epsilon_1$ and $t_2 + \epsilon_2 < t^*$. The left- and right-differentiability of $\bar{g}$ from Lemma 2 imply that $\epsilon_1 \partial_t^- \bar{g}(t_1) = \epsilon_2 \partial_t^+ \bar{g}(t_2) + o(\epsilon)$. Let $\Delta_1 t := [t_1 - \epsilon_1, t_1]$ and $\Delta_2 t := [t_2, t_2 + \epsilon_2]$.

Consider the following *modified leveling strategy* $\tilde{\rho}_{-i}$, which, compared to the leveling strategy, explores A at the expense of B over $\Delta_1 t$, and vice versa over $\Delta_2 t$.

If $t \in \Delta_1 t$, let

$$\tilde{\rho}_{-i}(t,x) := \begin{cases} \bar{\rho}(t,x) + \frac{\bar{g}(t_1-\epsilon_1)-\bar{g}(t)}{f(x)}, & \text{if } x \in A; \\ \bar{\rho}(t_1 - \epsilon_1, x), & \text{if } x \in B. \end{cases}$$

If $t \in (t_1, t_2)$, let

$$\tilde{\rho}_{-i}(t,x) := \begin{cases} \bar{\rho}(t,x) + \frac{\bar{g}(t_1-\epsilon_1)-\bar{g}(t_1)}{f(x)}, & \text{if } x \in A; \\ \bar{\rho}(t,x) - \frac{\bar{g}(t_1-\epsilon_1)-\bar{g}(t_1)}{f(x)}, & \text{if } x \in B. \end{cases}$$

If $t \in \Delta_2 t$, let

$$\tilde{\rho}_{-i}(t,x) := \begin{cases} \bar{\rho}(t_2 + \epsilon_2, x), & \text{if } x \in A; \\ \bar{\rho}(t,x) - \frac{\bar{g}(t_1-\epsilon_1)-\bar{g}(t_1)}{f(x)} + \frac{\bar{g}(t_2)-\bar{g}(t)}{f(x)}, & \text{if } x \in B. \end{cases}$$

If $x \notin A \cup B$ or $t \notin [t_1 - \epsilon_1, t_2 + \epsilon_2]$, let $\tilde{\rho}_{-i}(t,x) := \bar{\rho}(t,x)$.

Note that the modified strategy $\tilde{\rho}_{-i}$ is a strategy for player $-i$, and that, in particular, it satisfies the capacity constraint because $\int_A \frac{1}{f} = \int_B \frac{1}{f}$. It can be verified to be t-continuous.

Step 2: Payoffs from the Modified Leveling Strategy.

Observe that the difference in strategies is $d_t\tilde{\rho}_{-i} - d_t\bar{\rho} = -\frac{1}{f} d_t\bar{g}$ on A and $d_t\tilde{\rho}_{-i} - d_t\bar{\rho} = \frac{1}{f} d_t\bar{g}$ on B over $\Delta_1 t$, and vice versa over $\Delta_2 t$. It is zero otherwise. The utility difference of the modified leveling strategy compared to the leveling strategy, $u_{-i}(\tilde{\rho}_{-i}, \rho_i) - u_{-i}(\bar{\rho}_{-i}, \rho_i)$, is

$$-\int_A \frac{1}{f} \int_{\Delta_1 t} g_i d_t\bar{g} dx + \int_B \frac{1}{f} \int_{\Delta_1 t} g_i d_t\bar{g} dx \quad + \int_A \frac{1}{f} \int_{\Delta_2 t} g_i d_t\bar{g} dx - \int_B \frac{1}{f} \int_{\Delta_2 t} g_i d_t\bar{g} dx. \quad \text{(A.6)}$$

For the first term in (A.6), we perform a change of variable to get

$$\begin{aligned}
&-\int_A \frac{1}{f(x)} \int_{\Delta_1 t} g_i(t,x) d\bar{g}(t,x) dx \\
=&-\epsilon_1 \int_A \frac{1}{f(x)} \int_{[0,1]} g_i(t_1 - s\epsilon_1) \partial_t^- \bar{g}(t_1 - s\epsilon_1) ds dx \\
=&-\epsilon_1 \partial_t^- \bar{g}(t_1) \int_A \frac{1}{f(x)} g_i(t_1^-, x) dx + o(\epsilon),
\end{aligned}$$

where the second equality is due to the dominated convergence theorem. The equation states that, over the short time interval $\Delta_1 t$, both g_i and $\partial_t^- \bar{g}_{-i}$ can be taken as constants with respect to time. The same can be applied to the other three terms.

The payoff difference $u_{-i}(\tilde{\rho}_{-i}, \rho_i) - u_{-i}(\bar{\rho}_{-i}, \rho_i)$ can thus be written as

$$\begin{aligned}
&-\left(\int_A \frac{1}{f(x)} g_i(t_1^-, x) dx - \int_B \frac{1}{f(x)} g_i(t_1^-, x) dx\right) \epsilon_1 \partial_t^- \bar{g}(t_1) \\
&+\left(\int_A \frac{1}{f(x)} g_i(t_2, x) dx - \int_B \frac{1}{f(x)} g_i(t_2, x) dx\right) \epsilon_2 \partial_t^+ \bar{g}(t_2) + o(\epsilon) \\
=&\left(-\int_A \frac{1}{f(x)} \left(g_i(t_2, x) - g_i(t_1^-, x)\right) dx + \int_B \frac{1}{f(x)} \left(g_i(t_2, x) - g_i(t_1^-, x)\right) dx\right) \partial_t^+ \bar{g}(t_2) \epsilon_2 + o(\epsilon).
\end{aligned}$$

By supposition,

$$\begin{aligned}
&-\int_A \frac{1}{f(x)} \left(g_i(t_2, x) - g_i(t_1^-, x)\right) dx + \int_B \frac{1}{f(x)} \left(g_i(t_2, x) - g_i(t_1^-, x)\right) \\
>&-a \int_A \frac{1}{f(x)} dx + a \int_B \frac{1}{f(x)} dx \\
=&\, 0.
\end{aligned}$$

Therefore, there exists $\epsilon > 0$ sufficiently small such that, against ρ_i, the modified leveling strategy $\tilde{\rho}_{-i}$ yields a strictly higher payoff than leveling strategy does, which guarantees the maxmin payoff. $\square$

The first corollary below establishes the t-continuity of the equilibrium strategy on $[0, t^*]$. According to Lemma 4, the posterior decreases fastest on the upper contour set H. If the strategy were discontinuous, the posterior for those alternatives would have to decline discontinuously, which would violate the capacity constraint.

Corollary 2. *In any Nash equilibrium, player i's strategy ρ_i is t-continuous on $[0, t^*]$.*

Proof. The statement for $t = 0$ follows from the monotonicity and right-continuity condition, and the statement for $t = t^*$ is without loss of generality because the set $\{x \in X : \rho_i(t^*, x) - \rho_i((t^*)^-, x) > 0\}$ is null. It is trivial for $x \in H^C(t)$ because $\rho_i(t, x) = 0$ by Lemma 3.

Suppose there exists positive-measure set $B \subset X$ such that ρ_i, or equivalently g_i, is not t-continuous on $(0, t^*) \times B$. Without loss of generality, there exist $b < 0$ and $\epsilon \in (0, t^*)$ such that $B \subset H(\epsilon)$ and, for all $x \in B$, there exists $t_x \in (\epsilon, t^*)$ satisfying

$$g_i(t_x, x) - g_i(t_x^-, x) \leq b.$$

The compactness of $[\epsilon, t^*]$ implies that, for any $\delta > 0$, there exist $\underline{t}_\delta, \bar{t}_\delta \in (\epsilon, t^*)$, where $\underline{t}_\delta < \bar{t}_\delta$ and $\bar{t}_\delta - \underline{t}_\delta < \delta$, and a positive-measure subset $B_\delta \subset B$ such that

$$g_i(\bar{t}_\delta, x) - g_i(\underline{t}_\delta^-, x) \leq b.$$

Lemma 4 implies that $\rho_i(\bar{t}_\delta, x) - \rho_i(\underline{t}_\delta^-, x) \geq -b/f(x)$ for all $x \in H(\epsilon)$, a positive-measure set. The capacity constraint implies that

$$\delta = \int_X \rho_i(\bar{t}_\delta, x) - \rho_i(\underline{t}_\delta^-, x) dx \geq \int_{H(\epsilon)} \rho_i(\bar{t}_\delta, x) - \rho_i(\underline{t}_\delta^-, x) dx \geq \int_{H(\epsilon)} \frac{-b}{f(x)} dx,$$

which yields a contradiction as $\delta \downarrow 0$. □

Corollary 3. *In any Nash equilibrium, for* $0 < t_1 < t_2 \leq t^*$,

$$g_i(t_2, x_A) - g_i(t_1, x_A) = g_i(t_2, x_B) - g_i(t_1, x_B)$$

for $x_A, x_B \in H(t_1)$ *almost everywhere.*

Proof. Assume that $t_2 < t^*$. For any $t \in (t_1, t_2)$, $H(t_1) \subset H(t)$. Lemma 4 thus gives the equality

$$g_i(t_2, x_A) - g_i(t, x_A) = g_i(t_2, x_B) - g_i(t, x_B)$$

for $x_A, x_B \in H(t_1)$ almost everywhere. The statement is obtained by taking a countable sequence $t \uparrow t_1$, noting that g_i is t-continuous by Corollary 2.

The boundary case $t_2 = t^*$ follows similarly by taking a countable sequence $t_2 \uparrow t^*$. □

Lemma 5. *In any Nash equilibrium, for* $i \in \{1, 2\}$ *and* $t \in [0, t^*]$, $g_i(t, x) = \bar{g}(t)$ *for* $x \in H(t)$ *almost everywhere.*

Proof. Once the statement is proven for $t \in (0, t^*]$, it extends to the endpoint $t = 0$ because of the monotonicity and right-continuity condition.

For $t \in (0, t^*]$, define $\tilde{g}_i(t) := \sup_{x \in H(t)} g_i(t, x) - g_i(t^*, x)$. The t-continuity of g_i on $[0, t^*] \times X$ implies that $\tilde{g}_i(t^*) = 0$. Corollary 3 with $t_2 = t^*$ implies that

$$g_i(t, x) - g_i(t^*, x) = \tilde{g}_i(t) \tag{A.7}$$

for all $x \in H(t)$ almost everywhere.

We now derive the right-derivative $\partial_t^+ \tilde{g}_i$. For $0 < t_1 < t_2 < 1$, the capacity constraint yields

$$\begin{aligned} t_2 - t_1 &= \int_{H(t_1)} \rho_i(t_2, x) - \rho_i(t_1, x) dx + \int_{H(t_2) \backslash H(t_1)} \rho_i(t_2, x) - \rho_i(t_1, x) dx \\ &= - \left(\tilde{g}_i(t_2) - \tilde{g}_i(t_1) \right) \int_{H(t_1)} \frac{dx}{f(x)} - \int_{H(t_2) \backslash H(t_1)} \frac{g_i(t_2, x) - g_i(t_1, x)}{f(x)} dx \\ &\leq - \left(\tilde{g}_i(t_2) - \tilde{g}_i(t_1) \right) \left(\int_{H(t_1)} \frac{dx}{f(x)} + \int_{H(t_2) \backslash H(t_1)} \frac{dx}{f(x)} \right). \end{aligned} \tag{A.8}$$

The inequality is due to Lemma 4. Rearranging terms, we have

$$0 \geq \frac{\tilde{g}_i(t_2) - \tilde{g}_i(t_1)}{t_2 - t_1} \geq - \left(\int_{H(t_1)} \frac{dx}{f(x)} + \int_{H(t_2) \backslash H(t_1)} \frac{dx}{f(x)} \right)^{-1} \geq - \frac{\bar{g}(t_2)}{|H(t_2)|} > -\infty,$$

where the third inequality is due to the definition of H. The function $\tilde{g}_i$ is Lipschitz and thus absolutely continuous on $(0, 1)$.

Take $t_2 \downarrow t_1$. Since $H(t_2) \downarrow H(t_1)$ in the set-inclusion sense, the dominated convergence theorem states that $|H(t_2) \backslash H(t_1)| \downarrow 0$. The second term in Equation (A.8) is dominated by

$$\int_{H(t_2) \backslash H(t_1)} \frac{|g_i(t_2, x) - g_i(t_1, x)|}{f(x)} dx \leq \frac{|\tilde{g}_i(t_2) - \tilde{g}_i(t_1)| \, |H(t_2) \backslash H(t_1)|}{\bar{g}\left(\frac{1}{2}(t_1 + 1) \right)} = o(t_2 - t_1).$$

The right-derivative of $\tilde{g}_i$ is thus given by

$$\partial_t^+ \tilde{g}_i(t_1) = \lim_{t_2 \downarrow t_1} \frac{\tilde{g}_i(t_2) - \tilde{g}_i(t_1)}{t_2 - t_1} = - \int_{H(t_1)} \frac{dx}{f(x)}.$$

Since $\bar{g}$ also satisfies the first two lemmas and the two corollaries, an analogous calculation shows that

$$\partial_t^+ \bar{g}(t) = -\int_{H(t)} \frac{dx}{f(x)} = \partial_t^+ \tilde{g}_i(t).$$

Therefore, $\bar{g} = \tilde{g}_i + C$ for some constant $C \in \mathbb{R}$. The boundary condition at $t = t^*$ is $\lim_{t\uparrow 1} \tilde{g}_i(t) = \bar{g}(t^*) = 0$, which implies that $C = 0$.

On $H(t^*) = \{x \in X : \Delta\rho_{\min}(x) > 0\}$ almost surely, the other boundary condition at $t = \bar{g}^{-1}(f(x)\Delta\rho_{\min}(x)) < t^*$ shows that

$$g_i(t^*, x) = \lim_{s\downarrow t} g_i(s, x) - \bar{g}(t, x) = f(x)\Delta\rho_{\min}(x) - f(x)\Delta\rho_{\min}(x) = 0.$$

This establishes the desired result. □

Lemma 3 and Lemma 5 imply that, for all $t \in [0, t^*]$, $g_i(t, \cdot) = \bar{g}(t, \cdot)$ almost everywhere. There exists a full measure set over which the equality holds for all $t \in [0, 1] \cap \mathbb{Q}$. The theorem then follows from the monotonicity and right-continuity condition.

□

A.3 Proof of Theorem 1

It is sufficient to consider strategies and deviations ρ that satisfy $\rho(1, x) = 1$ for all $x \in X$. These strategies explore all alternatives for sure, and do so at full capacity by $t = 1$. For any strategy ρ_i that does not satisfy the restriction, there exist multiple strategies ρ_i''s that satisfy it with $\rho_i' \geq \rho_i$ by exploring the alternatives with higher probability and/or earlier. Since the payoff is monotonic in ρ_i, we have $u_i(\rho_i', \rho_{-i}) \geq u_i(\rho_i, \rho_{-i})$ so it is sufficient to consider these deviations only.

If $u_i(\rho_i', \rho_{-i}) > u_i(\rho_i, \rho_{-i})$, then ρ_i' is a profitable deviation for i so (ρ_i, ρ_{-i}) cannot be an equilibrium. Otherwise, ρ_i' does not change player i's payoff $u_i(\rho_i', \rho_{-i}) = u_i(\rho_i, \rho_{-i})$. The added/expedited explorations by ρ_i' do not lead to any additional probability of discovery so ρ_i' does not change opponent $-i$'s payoff either. Formally, the sum of the changes in the payoff equals the change in the probability of discovery

$$(u_i(\rho_i', \rho_{-i}) - u_i(\rho_i, \rho_{-i})) + (u_{-i}(\rho_i', \rho_{-i}) - u_{-i}(\rho_i, \rho_{-i})) = \int_X (1 - \rho_{-i}(T, x)) (\rho_i'(T, x) - \rho_i(T, x)) \, dx$$

from integration by parts for the Lebesgue–Stieltjes integral. Since $\rho' \geq \rho$, the change in $-i$'s payoff is non-positive and the change in total probability is nonnegative. Zero change in i's payoff thus implies zero change in $-i$'s payoff. If (ρ_i, ρ_{-i}) is an equilibrium, then so is (ρ_i', ρ_{-i}). The multiplicity of ρ_i' will then imply the multiplicity of equilibria within the restricted class of strategies. By contraposition, uniqueness within this class of strategy profiles will imply uniqueness for all profiles.

For this class, the strategic exploration game is a timing game with $T = 1$, $\Delta\rho_1 = \Delta\rho_2 = 1$, and $\alpha = 1$. The equilibria are leveling strategy profiles by Theorem 6. Since $t^* = 1$, there is no degeneracy on $(t^*, 1]$, so the leveling equilibrium is unique.

A.4 Proof of Theorem 2

We first show a technical lemma.

Lemma 6. *Let v_1 and v_2 be continuous, strictly decreasing, and convex functions from $[0,1]$ to $\mathbb{R}_+$ with $v_1(1) = v_2(1) = 0$ and $\int_0^1 v_1(s)ds = \int_0^1 v_2(s)ds$. Let v_1^{-1} and v_2^{-1} be their respective inverses, with the extension of value 0 outside of their domains. Then $\int_0^t v_1(s)ds \geq \int_0^t v_2(s)ds$ for all $t \in [0,1]$ if and only if $\int_0^z v_1^{-1}(y)dy \leq \int_0^z v_2^{-1}(y)dy$ for all $z \in \mathbb{R}_+$.*

Proof. We prove the inequality $\int_0^t v_1(s)ds \leq \int_0^t v_2(s)ds$ for all $t \in T$ under the hypothesis that $\int_0^z v_1^{-1}(y)dy \geq \int_0^z v_2^{-1}(y)dy$ for all $z \in \mathbb{R}_+$. The converse is analogous.

Since v_m, $m \in \{1, 2\}$, is strictly decreasing and convex, it has a strictly negative derivative almost everywhere on $(0, 1)$. Performing the integration by parts and then a change of variables, we have

$$\int_0^t v_m(s)ds = tv_m(t) - \int_0^t s dv_m(s) = tv_m(t) + \int_{v_m(t)}^{\infty} v_m^{-1}(y)dy. \tag{A.9}$$

Taking $t = 1$ (and hence $v_m(t) = 0$), we obtain

$$\int_0^1 v_m(s)ds = \int_0^{\infty} v_m^{-1}(y)dy. \tag{A.10}$$

Equation (A.9) and Equation (A.10) combine to yield

$$\int_0^t v_m(s)ds = tv_m(t) + \int_0^1 v_m(s)ds - \int_0^{v_m(t)} v_m^{-1}(y)dy. \tag{A.11}$$

Thus the desired inequality holds on the set $S := \{t \in T : v_1(t) = v_2(t)\}$ because it follows from Equation (A.11) that

$$\int_0^t (v_1(s) - v_2(s))ds = \int_0^{v_1(t)} (v_2^{-1}(y) - v_1^{-1}(y))dy \leq 0.$$

The set S is closed by the continuity of v_1 and v_2, and it contains $t = 1$ by assumption. The inequality holds trivially at $t = 0$. Denote $S^* := S \cup \{0\}$.

For any $t \notin S^*$, define two endpoints $\underline{t} := \max\{s \in S^* : s < t\}$ and $\bar{t} := \min\{s \in S^* : s > t\}$. They are well defined because S^* is closed. The difference $v_1(s) - v_2(s)$ has the same sign over $(\underline{t}, \bar{t})$ by continuity, so its integral $\int_0^{t'} (v_1(s) - v_2(s))ds$ is monotonic over the same interval. As the desired inequality holds at the endpoints, it actually holds over the entire interval, and at t in particular. □

Proof of Theorem 2. Let $\bar{g}_m$ be the equilibrium leveling function for prior f_m and h_m be its inverse, $m \in \{1, 2\}$, where

$$h_m(y) = \int_X \left(1 - \frac{y}{f_m(x)}\right) \mathbf{1}_{\{f_m(x) \geq y\}}(x)dx. \tag{A.12}$$

As the probability of simultaneous discovery is zero, the flow probabilities of discovery are $2\bar{g}_1$ and $2\bar{g}_2$, respectively. The desired conclusion is $\int_0^t 2\bar{g}_1(s)ds \geq \int_0^t 2\bar{g}_2(s)ds$ for all $t \in Y$. Since $\bar{g}_1$ and $\bar{g}_2$ satisfy the assumptions of Lemma 6, it suffices to show that their inverses satisfy $\int_0^z h_1(y)dy \leq \int_0^z h_2(y)dy \quad \forall z \in \mathbb{R}_+$. By Equation (A.12), the Fubini theorem, and then a change of variables, the integral can be written as

$$\begin{aligned}
\int_0^z h_m(y)dy &= \int_0^z \int_X \left(1 - \frac{y}{f_m(x)}\right) \mathbf{1}_{\{f_m(x) \geq y\}}(x, y)dxdy \\
&= \int_X \int_0^z \left(1 - \frac{y}{f_m(x)}\right) \mathbf{1}_{\{f_m(x) \geq y\}}(x, y)dydx \\
&= \int_0^\infty \int_0^z \left(1 - \frac{y}{w}\right) \mathbf{1}_{\{w \geq y\}}(w, y)dyd\lambda \circ f_m^{-1} \\
&= \int_0^\infty I(w, z)d\lambda \circ f_m^{-1},
\end{aligned}$$

where the integrand is given by

$$I(w, z) = \begin{cases} z - \frac{1}{2}\left(\frac{z}{w}\right)z, & \text{if } w \geq z, \\ \frac{w}{2}, & \text{if } w < z. \end{cases}$$

The result follows because $I(\cdot, z)$ is an increasing concave function and $\lambda \circ f_1^{-1}$ is a mean-preserving spread of $\lambda \circ f_2^{-1}$. $\square$

A.5 Proof of Theorem 3

We first define the set of transition times as

$$\theta(T) := \left\{ t \in [0, T] : \int \frac{\overline{g}(t)}{f(x)} \mathbf{1}_{f(x) \geq \overline{g}(t)} dx \geq 2(T - t) \right\}.$$

Lemma 7. *$\theta(T)$ is a non-degenerate closed interval containing T. Moreover,* $\min \theta(T) \to 1$ *as $T \to 1$.*

Proof. By the definition of $\overline{g}$, time t belongs to $\theta(T)$ if and only if

$$\int \mathbf{1}_{f(x) \geq \overline{g}(t)} dx \geq 2T - t \, .$$

The left-hand side is right-continuous and increasing in t because $\overline{g}$ is continuous and decreasing. It attains $\lambda\{f \geq \overline{g}(T)\} = T + \int \frac{\overline{g}(T)}{f(x)} \mathbf{1}_{f(x) \geq \overline{g}(T)} dx > T$ at $t = T$. The right-hand side is continuous and decreasing in t and attains T at $t = T$. Therefore, θ is a non-degenerate closed interval containing T.

Moreover, for any fixed $t \in (0, 1)$, the left-hand side is strictly less than one but the right-hand side is strictly larger than one for sufficiently large T. The second statement then follows. $\square$

Proof of Theorem 3. We define the lowest transition required by Theorem 3 as

$$T^*(T) := \min \theta(T) = \min \left\{ t \in [0, T] : \int \frac{\overline{g}(t)}{f(x)} \mathbf{1}_{f(x) \geq \overline{g}(t)} dx \geq 2(T - t) \right\}.$$

Therefore, $\theta(T) = [T^*(T), T]$ by Lemma 7.

We first show the second part of the theorem. For $t^* \in [T^*, T]$, define

$$x_1 := \inf \left\{ x' : \int_0^{x'} \frac{\overline{g}(t^*)}{f(x)} \mathbf{1}_{f(x) \geq \overline{g}(t^*)} dx = T - t^* \right\};$$
$$x_2 := \sup \left\{ x' : \int_{x'}^1 \frac{\overline{g}(t^*)}{f(x)} \mathbf{1}_{f(x) \geq \overline{g}(t^*)} dx = T - t^* \right\}.$$

We have $x_1 \leq x_2$ by the definition of T^*. Recall that $H(t) = \{x : f(x) \geq \overline{g}(t)\}$ is the upper contour set of $\overline{g}$. We write $X_1 := (0, x_1) \bigcap H(t)$ and $X_2 := (x_2, 1) \bigcap H(t)$. Therefore,

$X_1 \bigcap X_2 = \emptyset$. For $i = 1, 2$ and $t \in (t^*, T]$, define

$$\rho_i(t, x) = \begin{cases} \overline{\rho}(t^*, x) + \frac{t-t^*}{T-t^*}\left(1 - \overline{\rho}(t^*, x)\right), & \text{if } x \in X_i, \\ \rho(t^*, x), & \text{if } x \notin X_i. \end{cases}$$

It is straightforward to verify that ρ_i is a well-defined distributional strategy. We write

$$\delta_i(x) := \rho_i(T, x) - \rho_i(t^*, x).$$

By the definition of X_i, if $\delta_i(x) > 0$, then $\overline{g}(t^*) = g_{-i}(t, x)$ for all $t > t^*$ and $\delta_{-i}(x) = 0$. Moreover, it is an equilibrium because both players maximize the myopic payoff.

We now show the first part of the theorem. We argue that such an equilibrium exists only if $t \in [T^*(T), T]$. If the players compete during $[0, t^*]$ and coordinate during $(t^*, T]$ in an equilibrium, the capacity constraint must be binding because, for $i = 1, 2$ and $t \in [0, T]$, there exists some alternative with a positive myopic return. We then have

$$\begin{aligned} 2(T - t^*) &= \int \delta_1(x)\mathbf{1}_{\delta_1(x)>0}dx + \int \delta_2(x)\mathbf{1}_{\delta_2(x)>0}dx \\ &\leq \int \delta_1(x)\mathbf{1}_{f(x)\geq\overline{g}(t^*)}dx + \int \delta_2(x)\mathbf{1}_{f(x)\geq\overline{g}(t^*)}dx \\ &\leq \int \frac{\overline{g}(t^*)}{f(x)}\mathbf{1}_{\delta_1(x)>0}\mathbf{1}_{f(x)\geq\overline{g}(t^*)}dx + \int \frac{\overline{g}(t^*)}{f(x)}\mathbf{1}_{\delta_2(x)>0}\mathbf{1}_{f(x)\geq\overline{g}(t^*)}dx \\ &\leq \int \frac{\overline{g}(t^*)}{f(x)}\mathbf{1}_{f(x)\geq\overline{g}(t^*)}dx, \end{aligned}$$

where the equality follows from the binding capacity constraint, the first weak inequality holds because in the coordination phase $\delta_i(x) > 0$ implies that $f(x) \geq \overline{g}(t^*)$, the second weak inequality follows from the leveling strategy over $[0, t^*]$, and the last inequality holds because in the coordination phase $\delta_i(x) > 0$ implies that $\delta_{-i} = 0$. Therefore, $t^* \geq T^*(T)$ by the definition of $T^*(T)$.

We prove that all equilibria take the prescribed form using the timing game. For equilibrium (ρ_1, ρ_2), define $\Delta\rho_1 := \rho_1(T, \cdot)$ and $\Delta\rho_2 := \rho_2(T, \cdot)$. We first restrict attention to strategies that exhaust the capacity $\int \Delta\rho_1 dx = \int \Delta\rho_2 dx = T$. Consider the timing game with deadline T, $(\Delta\rho_1, \Delta\rho_2)$, and $\alpha = 1$. Since the strategies are feasible in the timing game, they are not profitable deviations in (ρ_1, ρ_2), which is thus an equilibrium of the timing game.

Theorem 6 gives the equilibrium payoff of the timing game as

$$u_i = \int_X f\left(\Delta\rho_i - \Delta\rho_{\min}\Delta\rho_i + \frac{1}{2}\left(\Delta\rho_{\min}\right)^2\right)dx \tag{A.13}$$

where $\Delta\rho_{\min} := \min\{\Delta\rho_1, \Delta\rho_2\}$. The integrand is the utility from exploring x. It is $\mathcal{C}^1$ and concave in $\Delta\rho_i(x)$. The marginal utility $f(x)\left(1-\Delta\rho_{\min}(x)\right)$ is common among both players, and decreasing in both $\Delta\rho_i(x)$ and $\Delta\rho_{-i}(x)$.

We show by contraposition that $(\Delta\rho_1, \Delta\rho_2)$ is an equilibrium in a game in which the players choose $\Delta\rho_i$ that exhausts the capacity and receive the timing-game payoff in Equation (A.13). Without loss of generality, suppose that player 1 has a profitable deviation $\Delta\rho_1'$ in this game. Let ρ_1' be a strategy of player 1 corresponding to $(\Delta\rho_1', \Delta\rho_2)$. Since the timing game is a constant sum game, ρ_1' guarantees the maxmin payoff in Equation (A.13). Thus, ρ_1' is a profitable deviation in the game with deadline T, contradicting the equilibrium (ρ_1, ρ_2).

We obtain the equilibria of the game of $\Delta\rho$'s by computing the best response correspondence. For fixed $\Delta\rho_{-i}$, player i faces a concave maximization problem subject to the capacity constraint $\int_X \Delta\rho_i dx = T$. Therefore, for any equilibrium (ρ_1, ρ_2), there exists the shadow value of capacity $\lambda_i \geq 0$ such that the first-order condition holds: $\lambda_i = f(x)\left(1-\Delta\rho_{\min}(x)\right)$ for $\Delta\rho_i(x) \in (0,1)$, and the complementary slackness conditions hold: $\lambda_i \geq f(x)\left(1-\Delta\rho_{\min}(x)\right)$ if $\Delta\rho_i = 0$, and $\lambda_i \leq f(x)\left(1-\Delta\rho_{\min}(x)\right)$ if $\Delta\rho_i(x) = 1$.

We show that the shadow value equalizes for the two players.

Lemma 8. $\lambda_1 = \lambda_2$.

Proof. Without loss of generality, suppose $\lambda_1 > \lambda_2 \geq 0$. For $x \in X$ such that $f(x) < \lambda_1$, the marginal payoff is strictly below λ_1 and hence $\Delta\rho_1(x) = 0$. For $x \in X$ such that $f(x) \geq \lambda_1$, we claim that $\Delta\rho_2(x) > \Delta\rho_1(x)$. Suppose that $\Delta\rho_1(x) \geq \Delta\rho_2(x)$. We have $\Delta\rho_2(x) > 0$; otherwise, the marginal utility is above the shadow cost for player 2: $f(x) \geq \lambda_1 > \lambda_2$. Since $\Delta\rho_1(x) \geq \Delta\rho_2(x) > 0$, the complementary slackness condition gives $\lambda_1 \leq f(1-\Delta\rho_{\min}) = \lambda_2$, a contradiction.

Alternatives in $\{x : f(x) \geq \lambda_1\}$ have a positive measure because player 1 explores only this set. With $\Delta\rho_1 = 0 \leq \Delta\rho_2$ on $\{x : f(x) < \lambda_1\}$ and $\Delta\rho_1 < \Delta\rho_2$ on $\{x : f(x) \geq \lambda_1\}$, the capacity constraint is $T = \int_X \Delta\rho_1 dx = \int_{f \geq \lambda_1} \Delta\rho_1 dx < \int_{f \geq \lambda_1} \Delta\rho_2 dx \leq T$, a contradiction. □

Let $\lambda \geq 0$ be the common shadow value.

Lemma 9. *If $f(x) < \lambda$, then $\Delta\rho_1(x) = \Delta\rho_2(x) = 0$. If $f(x) \geq \lambda$, $\lambda = f(x)(1-\Delta\rho_{\min}(x))$.*

Proof. If $f(x) < \lambda$, the marginal payoff is strictly less than the shadow value $f(x)(1 - \Delta\rho_{\min}(x)) \leq f(x) < \lambda$ and hence $\Delta\rho_1(x) = \Delta\rho_2(x) = 0$. If $f(x) \geq \lambda$, we show the equality by contraposition. Suppose $\lambda < f(1 - \Delta\rho_{\min}(x))$. Then $\Delta\rho_1(x) = \Delta\rho_2(x) = 1$, but then $f(x)(1 - \Delta\rho_{\min}(x)) = 0 < \lambda$, a contradiction. Suppose $\lambda > f(x)(1 - \Delta\rho_{\min}(x))$. Then $\Delta\rho_1(x) = \Delta\rho_2(x) = 0$, but then $f(x)(1 - \Delta\rho_{\min}(x)) = f(x) \geq \lambda$, a contradiction. $\square$

The equilibrium strategy profile of the game of $\Delta\rho$'s is then given by Theorem 6. Define $t^* := \int \Delta\rho_{\min} dx \in [0, T]$. On $[0, t^*]$, the players level $f\Delta\rho_{\min} = f - \lambda$. Since λ is a constant, the strategy is equivalent to leveling f itself. At t^*, the posterior satisfies $\bar{g}(t^*) = f(x)(1 - \rho_i(t^*, x)) = f(x)(1 - \Delta\rho_{\min}(x)) = \lambda$ on $\{x : f(x) \geq \lambda\}$. For such x, $\rho_1(t^*, x) = \rho_2(t^*, x) = \Delta\rho_{\min}(x)$ implies $\rho_i(T, x) - \rho_i(t^*, x) = 0$ for at least one i. It follows that $\rho_i(T, x) - \rho_i(t^*, x) > 0$ implies $\rho_{-i}(T, x) - \rho_{-i}(t^*, x) = 0$. Moreover, $\rho_i(T, x) - \rho_i(t^*, x) > 0$ implies $g_{-i}(t, x) = f(x)(1 - \rho_{-i}(t, x)) = f(x)(1 - \Delta\rho_{\min}(x)) = \lambda = \bar{g}(t^*)$.

Since we have verified that the equilibria of the game of $\Delta\rho$'s remain equilibria in the game with deadline T, it remains to rule out (ρ_1, ρ_2) in which some strategy does not exhaust the capacity. Following the discussion in Appendix A.3, (ρ_1, ρ_2) is an equilibrium only if its full-capacity version (ρ_1', ρ_2') is also an equilibrium that gives the same payoffs. However, in such an equilibrium (ρ_1', ρ_2'), the myopic payoff is strictly positive and, hence, no (ρ_1, ρ_2) gives the same payoffs.

$\square$

A.6 Proof of Theorem 4

The idea of the proof of Theorem 4 is to normalize the strategy of the weak player to obtain a timing game, and then maximize his payoff over the probability of exploration.

For every strategy ρ_2^α of player 2, define $\rho_2 := \rho_2^\alpha / \alpha$. It is easy to verify that $\rho_2 : T \times X \to [0, 1/\alpha]$ satisfies the four conditions of Definition 2. It differs from a distributional strategy in its codomain $[0, 1/\alpha]$ instead of $[0, 1]$. We shall call $\rho_2 : T \times X \to [0, 1/\alpha]$ a *normalized strategy*. Players' payoffs from the strategy profile (ρ_1, ρ_2^α) can be rewritten as payoffs from (ρ_1, ρ_2) as follows:

$$
\begin{aligned}
u_1(\rho_1, \rho_2^\alpha) &= \int_X \int_T (1 - \alpha\rho_2(t, x)) f(x) d_t\rho_1(t, x) dx + \frac{1}{2} \int_X f(x) \sum_{t \in D_x} \Delta_t\rho_1(t, x) \Delta_t \alpha\rho_2(t, x) dx \\
&= (1 - \alpha) \int_X \int_T f(x) d_t\rho_1(t, x) dx \\
&\quad + \alpha \int_X \int_T (1 - \rho_2(t, x)) f(x) d_t\rho_1(t, x) dx + \frac{\alpha}{2} \int_X f(x) \sum_{t \in D_x} \Delta_t\rho_1(t, x) \Delta_t\rho_2(t, x) dx
\end{aligned}
$$

$$= (1-\alpha)\pi + \alpha u_1(\rho_1, \rho_2). \tag{A.14}$$

$$u_2(\rho_2^\alpha, \rho_1) = \int_X \int_T \alpha \left(1 - \rho_1(t,x)\right) f(x) d_t\rho_2(t,x)dx + \frac{\alpha}{2} \int_X f(x) \sum_{t \in D_x} \Delta_t\rho_1(t,x)\Delta_t\rho_2(t,x)dx$$

$$= \alpha u_2(\rho_2, \rho_1). \tag{A.15}$$

Therefore, the payoff functions under asymmetric capacity are increasing affine transformations of the payoff functions with a normalized strategy of player 2. Thus, the game with asymmetric capacity is strategically equivalent to the game with a normalized strategy.

We first argue that the candidate is a Nash equilibrium in the normalized game. The strong player 1 has the same set of strategies in the normalized game as in the benchmark game. Since the profile is a Nash equilibrium in the benchmark, she has no profitable deviation in the normalized game. The weak player 2 has no profitable deviations by the myopic argument because of the leveling posterior g_1. Both players enjoy maxmin payoff $\pi/2$.

We continue to show the uniqueness of the equilibrium. Let us consider strategies that exhaust the capacity $\rho_1(1,\cdot) = 1$. Among those strategies, any equilibrium of the normalized game is an equilibrium of the timing game with the corresponding normalized probability of exploration $\rho_2(1,\cdot)$, i.e., a timing game with $T = 1$, $\Delta\rho_1 = 1$, $\Delta\rho_2 = \rho_2(1,\cdot)$, and $\alpha \in (0,1]$. Theorem 6 implies that the equilibria of this timing game are the leveling strategy profiles, and the weak player's equilibrium payoff is

$$u_2(\rho_1, \rho_2) = \int_X f\left((1 - \min\{1, \Delta\rho_2\})\Delta\rho_2 + \frac{1}{2}\min\{1, \Delta\rho_2\}^2\right) dx.$$

Consider the function of $\Delta\rho_2(x)$:

$$(1 - \min\{1, \Delta\rho_2\})\Delta\rho_2 + \frac{1}{2}\min\{1, \Delta\rho_2\}^2 .$$

It is maximized on $\Delta\rho_2(x) \in [1, 1/\alpha]$. If $\Delta\rho_2 > 1$ for a positive measure set, the capacity constraint implies that $\Delta\rho_2 < 1$ for another positive measure set, which yields a strictly lower payoff. Therefore, the equilibrium payoff of the normalized game $\pi/2$ can only be achieved with $\Delta\rho_2 = 1$ almost everywhere with the corresponding leveling profile $(\bar{\rho}, \bar{\rho})$, which is unique because $t^* = 1$.

Since the equilibrium is unique among strategies that exhaust the capacity, it is also unique among all normalized strategies following the discussion in Appendix A.3.

A.7 Proof of Theorem 5

Proof. It suffices to show that $P_\gamma(t)$ is increasing in γ for all $t \in T$. By differentiating Equation (3.4) with respect to time, we obtain

$$\bar{g}'(t) = -\left(\int_X f(x)^{-1}\mathbf{1}_{\{f(x)\geq \bar{g}(t)\}}dx\right)^{-1}.$$

The Lipschitz term due to the changing domain of integration vanishes because $1 - \frac{\bar{g}(t)}{f(x)} = 0$ on $\{x \in X : f(x) = \bar{g}(t)\}$.

As the leveling function $\bar{g}$ is absolutely continuous, the probability of discovery, $P_\gamma(t)$, is absolutely continuous with respect to γ, and for γ-almost everywhere,

$$\begin{aligned}
\partial_\gamma P_\gamma(t) =& -t\bar{g}'(\gamma t)\int_X \mathbf{1}_{\{f(x)\geq \bar{g}(\gamma t)\}}dx - t\bar{g}(\gamma t) + (2-\gamma)t^2\bar{g}'(\gamma t) \\
=& \frac{t}{\int_X f(x)^{-1}\mathbf{1}_{\{f(x)\geq \bar{g}(\gamma t)\}}dx}\left(\int_X \mathbf{1}_{f(x)\geq \bar{g}(\gamma t)}dx - \int_X \frac{\bar{g}(\gamma t)}{f(x)}\mathbf{1}_{\{f(x)\geq \bar{g}(\gamma t)\}}dx - (2-\gamma)t\right) \\
=& \frac{t}{\int_X f(x)^{-1}\mathbf{1}_{\{f(x)\geq \bar{g}(\gamma t)\}}dx}\left(\gamma t - (2-\gamma)t\right) \\
=& \frac{2(\gamma-1)t^2}{\int_X f(x)^{-1}\mathbf{1}_{\{f(x)\geq \bar{g}(\gamma t)\}}dx} \\
\geq& \, 0,
\end{aligned}$$

where the third equality follows from Equation (3.4). □

A.8 Implementation of Distributional Strategy

We introduce a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to describe randomization. In addition, weak measurability is used to accommodate continuous time and continuous space, and the Gelfand–Pettis integral is used, which extends the Lebesgue integral to functional spaces.[19]

Definition 6. *A* ***mixed strategy*** *on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a function* $\sigma : \Omega \times T \times X \to \{0, 1\}$ *that satisfies the following conditions:*

1. *Initial condition:* $\sigma(\omega, 0, \cdot) = 0$ *for all* $\omega \in \Omega$*;*
2. *Monotonicity and right-continuity:* $\sigma(\omega, \cdot, x)$ *is non-decreasing and right continuous for all* $\omega \in \Omega$ *and* $x \in X$*;*

[19] See Talagrand (1984) for an exposition.

3. *Measurability: mapping $x \mapsto \sigma(\cdot, t, x) \in L^2(\Omega)$ is weakly measurable for all $t \in T$;*

4. *Capacity constraint: $\int_X (\sigma(\cdot, t, x) - \sigma(\cdot, s, x))dx \leq t - s$ for all intervals $[s, t] \subset T$, where the integral is the Gelfand–Pettis integral.*

The initial condition and the monotonicity and right-continuity condition are the realization-by-realization generalizations of their counterparts in Definition 1 of pure strategies. The measurability condition and the capacity constraint in Definition 6 for mixed strategies, however, are weaker than their counterparts. They must hold when averaged over any measurable event in Ω that has a positive probability under $\mathbb{P}$, but not necessarily at each $\omega \in \Omega$. If Ω is a singleton, a mixed strategy reduces to a pure strategy as defined in Definition 1.

With realization $\omega \in \Omega$, an alternative $x \in X$ is explored at or before time $t \in T$ if and only if $\sigma(\omega, t, x) = 1$, analogously to the pure strategy case. The stochastic time at which alternative x is searched is a random variable on Ω given by $\tau(\omega, x) = \min\{t : \sigma(\omega, t, x) = 1\}$.

Theorem 7. *1. For every mixed strategy $\sigma : \Omega \times T \times X \to \{0, 1\}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the function ρ, defined by $\rho(t, x) := \mathbb{E}[\sigma(\cdot, t, x)]$ for $t \in T$ and $x \in X$, is a distributional strategy that represents σ; i.e., the probability of an alternative $x \in X$ being explored by $t \in T$ under σ is $\rho(t, x)$.*

2. For every distributional strategy ρ, there exists a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a mixed strategy $\sigma : \Omega \times T \times X \to \{0, 1\}$ that implements ρ, i.e., $\mathbb{E}[\sigma(\cdot, t, x)] = \rho(t, x)$ for all $t \in T$ and $x \in X$.

Proof. The first part of Theorem 7 follows directly from the definitions of the weak measurability and the weak integral, so its proof is omitted.

We show the second part by construction. By the Kolmogorov extension theorem, there exists a probability triple $(\Omega, \mathcal{F}, \mathbb{P})$ in which random variables $r_x \sim U(0, 1)$ are i.i.d. across $x \in X$. Define candidate mixed strategy $\sigma(\omega, t, x) := \mathbf{1}_{\{r_x(\omega) \leq \rho(t,x)\}}(\omega, t, x)$. By construction, it satisfies the initial condition and the monotonicity and right-continuity condition, and implements the search density.

Fix $t \in [0, 1]$. We shall show that $x \mapsto \sigma(\cdot, t, x) \in L^2(\Omega)$ is weak-integrable over the Lebesgue measure with integral $\int_X \rho(t, x)dx$. The dual space of $L^2(\Omega)$ is isomorphic to itself by the Riesz representation theorem. Every element $Z \in L^2(\Omega)$ operates on $Y \in L^2(\Omega)$ via $ZY = \mathbb{E}[ZY]$.

Since $\sigma(\cdot, x, t) \in \{0, 1\}$, its variance is bounded by $1/4$. The pairwise independence of $\{\sigma(\cdot, t, x) : x \in X\}$ implies that $\{\sigma(\cdot, t, x) - \rho(t, x) : x \in X\}$ is an orthogonal set in $L_2(\Omega)$.

By the Bessel theorem, we have that for, any countable collection $\{x_n\}$,

$$\frac{1}{4}\mathbb{E}[Z^2] \geq \sum_{n=1}^{\infty} \frac{\left(\mathbb{E}[Z(\sigma(\cdot,t,x_n)-\rho(t,x_n))]\right)^2}{4Var[\sigma(\cdot,t,x)]} \geq \sum_{n=1}^{\infty} \left(\mathbb{E}[Z(\sigma(\cdot,t,x_n)-\rho(t,x_n))]\right)^2,$$

which implies that $\mathbb{E}[Z(\sigma(\cdot,t,x)-\rho(t,x))]=0$, or $\mathbb{E}[Z\sigma(\cdot,t,x)]=\mathbb{E}[Z]\rho(t,x)$, everywhere except on a countable set. Therefore, the function $\sigma(\omega,t,\cdot)$ is weakly measurable and has weak-integral $\int_X \rho(t,x)dx$, satisfying the capacity constraint. $\square$

The following example shows that the mixed-strategy implementation is not unique.

Example 3. *Under the Lebesgue probability space on* $\Omega=[0,1]$*, both mixed strategies* $\sigma_1(\omega,x,t) := \mathbf{1}_{\{\mathrm{frac}(x-\omega)\leq t\}}$ *and* $\sigma_2(\omega,x,t) := \mathbf{1}_{\{\mathrm{frac}(x+\omega)\leq t\}}$*, where* $\mathbf{1}$ *denotes the indicator function and* $\mathrm{frac}(y)=y-\lfloor y\rfloor$ *denotes the fractional part of* y*, implement the same distributional strategy* $\rho(t,x)=t$*. Intuitively, according to the mixed strategy* σ_1*, a player searches to the right starting from* $x=\omega$*, where* ω *is drawn uniformly from the interval* $[0,1]$*, and continues at* $x=0$ *after reaching* $x=1$*, while according to* σ_2*, a player searches in the other direction starting from the same starting point. The two mixed strategies correspond to the same uncertainty faced by the opponent.*

References

Abreu, Dilip and Markus K Brunnermeier (2003). "Bubbles and crashes". In: *Econometrica* 71.1, pp. 173–204.

Abreu, Dilip and Faruk Gul (2000). "Bargaining and reputation". In: *Econometrica* 68.1, pp. 85–117.

Akcigit, Ufuk and Qingmin Liu (2015). "The role of information in innovation and competition". In: *Journal of the European Economic Association* 14.4, pp. 828–870.

Aumann, Robert J (1987). "Correlated equilibrium as an expression of Bayesian rationality". In: *Econometrica*, pp. 1–18.

Bavly, Gilad, Yuval Heller, and Amnon Schreiber (2022). "Social welfare in search games with asymmetric information". In: *Journal of Economic Theory* 202, pp. 105–462.

Bimpikis, Kostas, Shayan Ehsani, and Mohamed Mostagir (2019). "Designing dynamic contests". In: *Operations Research* 67.2, pp. 339–356.

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

Chatterjee, Kalyan and Robert Evans (2004). “Rivals’ search for buried treasure: competition and duplication in R&D”. In: *RAND Journal of Economics*, pp. 160–183.

Che, Yeon-Koo and Ian Gale (2003). “Optimal design of research contests”. In: *American Economic Review* 93.3, pp. 646–671.

de Roos, Nicolas, Alexander Matros, Vladimir Smirnov, and Andrew Wait (2018). “Shipwrecks and treasure hunters”. In: *Journal of Economic Dynamics and Control* 90, pp. 259–283.

Fershtman, Chaim and Ariel Rubinstein (1997). “A simple model of equilibrium in search procedures”. In: *Journal of Economic Theory* 72.2, pp. 432–441.

Fu, Qiang, Jingfeng Lu, and Yue Pan (2015). “Team contests with multiple pairwise battles”. In: *American Economic Review* 105.7, pp. 2120–40.

Fudenberg, Drew and Jean Tirole (1985). “Preemption and rent equalization in the adoption of new technology”. In: *The Review of Economic Studies* 52.3, pp. 383–401.

Halac, Marina, Navin Kartik, and Qingmin Liu (2017). “Contests for experimentation”. In: *Journal of Political Economy* 125.5, pp. 1523–1569.

Hendricks, Kenneth, Andrew Weiss, and Charles Wilson (1988). “The war of attrition in continuous time with complete information”. In: *International Economic Review*, pp. 663–680.

Hendricks, Kenneth and Charles Wilson (1992). “Equilibrium in preemption games with complete information”. In: *Equilibrium and Dynamics*. Springer, pp. 123–147.

Hopenhayn, Hugo A and Francesco Squintani (2011). “Preemption games with private information”. In: *The Review of Economic Studies* 78.2, pp. 667–692.

Hörner, Johannes and Andrzej Skrzypacz (2016). “Learning, experimentation and information design”. In: *Working Paper, Yale University and Stanford University.*

Keller, Godfrey, Sven Rady, and Martin Cripps (2005). “Strategic experimentation with exponential bandits”. In: *Econometrica* 73.1, pp. 39–68.

Klein, Nicolas and Sven Rady (2011). “Negatively correlated bandits”. In: *The Review of Economic Studies* 78.2, pp. 693–732.

Kleinberg, Jon M (1994). “On-line Search in a Simple Polygon”. In: *SODA*. Vol. 94, pp. 8–15.

Manso, Gustavo (2011). “Motivating innovation”. In: *The Journal of Finance* 66.5, pp. 1823–1860.

Matros, Alexander, Natalia Ponomareva, Vladimir Smirnov, and Andrew Wait (2019). “Search without observability”. In: *Working Paper, University of Sydney.*

Matros, Alexander and Vladimir Smirnov (2016). “Duplicative search”. In: *Games and Economic Behavior* 99, pp. 1–22.

Mayskaya, Tatiana and Arina Nikandrova (2022). “The Dark Side of Transparency: When Hiding in Plain Sight Works”. In: *HSE Working Paper*.

Megow, Nicole, Kurt Mehlhorn, and Pascal Schweitzer (2012). “Online graph exploration: New results on old and new algorithms”. In: *Theoretical Computer Science* 463, pp. 62–72.

Milgrom, Paul R and Robert J Weber (1985). “Distributional strategies for games with incomplete information”. In: *Mathematics of operations research* 10.4, pp. 619–632.

Moldovanu, Benny and Aner Sela (2001). “The optimal allocation of prizes in contests”. In: *American Economic Review* 91.3, pp. 542–558.

— (2006). “Contest architecture”. In: *Journal of Economic Theory* 126.1, pp. 70–96.

Osborne, Martin J and Carolyn Pitchik (1986). “The nature of equilibrium in a location model”. In: *International Economic Review*, pp. 223–237.

Ottaviani, Marco and Peter Sørensen (2003). “Forecasting and rank order contests”. In: *Working Paper, University of Copenhagen*.

Siegel, Ron (2009). “All-pay contests”. In: *Econometrica* 77.1, pp. 71–92.

Simon, Leo K. and Maxwell B. Stinchcombe (1989). “Extensive form games in continuous time: Pure strategies”. In: *Econometrica* 57, pp. 1171–1214.

Talagrand, Michel (1984). “Pettis integral and measure theory”. In: *Memoirs of the American Mathematical Society*.

B Online Appendix

B.1 Increasing Payoffs and Time Preferences

In this section, we show that the strategy profile $(\bar{\rho}, \bar{\rho})$ is the unique Nash equilibrium even if the payoff from a later discovery is larger. In doing so, we demonstrate the payoff effect and information advantage behind the strategic tension between preemption and prioritization.

Increasing payoff over time is a salient feature in many applications. For example, two competing companies work on a drug with a rising price, or two countries compete on a technology that will become safer in the future. Our result shows that even with arbitrarily strong incentives to wait on exploration, the players fail to coordinate and explore greedily as in the case of stationary payoff.

We model the increasing payoff by a strictly increasing function $\beta : [0, T] \to \mathbb{R}_+$ common among the two players. When a player discovers the good alternative at time t, he enjoys payoff $\beta(t)$ which increases over time. The expected payoff of player i is thus

$$\int_X \int_T \beta(t) f(x)(1 - \rho_{-i}(t, x)) d_t \rho_i(t, x) dx + \frac{1}{2} \int_X f(x) \sum_{t \in D_x} \beta(t) \Delta_t \rho_i(t, x) \Delta_t \rho_{-i}(t, x) dx.$$

The benchmark model corresponds to the limiting case $\beta \equiv 1$. The time preference accommodates not only a continuously increasing payoff studied in preemption games, but also a discontinuous payoff, e.g., the return to discovering a new drug jumps discontinuously when a complementary patent expires.

We show that the players fail to coordinate on slower or delayed explorations to take advantage of the increasing payoff, even when the increment can be arbitrarily large and arrive arbitrarily soon.

Theorem 8. *In the strategic exploration game with time preference β, the profile $(\bar{\rho}, \bar{\rho})$ is the unique Nash equilibrium.*

The prioritization and preemption motives shape the unique equilibrium. Under an increasing reward, early exploration can result in a payoff effect and an information advantage. If it leads to a discovery, the player enjoys the current payoff but eliminates the possibility of a later discovery, which can be much more valuable. However, if early exploration fails, the player can preempt his opponent in future explorations by concentrating his capacity on the remaining alternatives. We show that the information advantage dominates the payoff effect so the players cannot coordinate on delayed explorations. Moreover, they reap the full information advantage by prioritizing the most promising alternatives a posteriori. Therefore, the unique equilibrium features the greedy leveling strategy.

B.2 Proof of Theorem 8

The proof of Theorem 8 follows from five lemmas. Lemma 10 establishes symmetry in any equilibrium. The idea is that, when one player imitates his opponent's strategy, the duplicated search will delay discovery and hence increases the total payoff, half of which goes to the deviating player. Lemma 11 then proves the t-continuity of the equilibrium strategy by considering a generalized one-step-ahead deviation. Lemma 12 shows that the capacity constraint binds in equilibrium. It highlights the tradeoff between earlier and later explorations. Although the payoff increases over time, the marginal payoff vanishes as the posterior goes to zero. If both players delay their exploration, each one would have incentive to expedite the final explorations, unraveling the delay. Finally, Lemma 13 and Lemma 14 establishes a version of Equation (3.3) by a similar argument and completes the proof using Lemma 2.

Lemma 10. *The strategies are symmetric $\rho_1 = \rho_2$ and satisfy the terminal condition $\rho_i(T,x) = 1$ a.e. in any equilibrium.*

Proof. Suppose $\rho_1 \neq \rho_2$. We shall show that at least one player can profitably deviate by imitating his opponent's strategy so (ρ_1, ρ_2) not an equilibrium. It suffices to show that the sum of changes in payoff, for player 1 to play ρ_1 and player 2 to play ρ_2, is positive.

We may restrict attention to strategies ρ_i that satisfies the terminal condition. Suppose not. There exists strategy $\rho_i' \geq \rho_i$ such that $\rho_i'(T,x) = 1$ a.e.—it searches more to satisfy the terminal condition. If the payoff increases $u_i(\rho_i', \rho_{-i}) > u_i(\rho_i, \rho_{-i})$, then ρ_i' is a profitable deviation. Else, the payoff remains the same $u_i(\rho_i', \rho_{-i}) = u_i(\rho_i, \rho_{-i})$. The additional search always duplicates the opponent's search so the payoff of the opponent also remains the same $u_{-i}(\rho_i', \rho_{-i}) = u_{-i}(\rho_i, \rho_{-i})$. Since ρ_i and ρ_i' are payoff equivalent, we may replace ρ_i by ρ_i' when constructing a deviation.

We first consider a continuous time preference. The sum of changes in payoff is

$$
\begin{aligned}
&\int_X \int_T f\beta(1-\rho_2) d_t\left(\rho_2-\rho_1\right) dx + \frac{1}{2}\int_X f \sum_{t\in D_x} \beta \Delta_t \left(\rho_2-\rho_1\right) \Delta_t \rho_2 dx \\
&\quad + \int_X \int_T f\beta(1-\rho_1) d_t\left(\rho_1-\rho_2\right) dx + \frac{1}{2}\int_X f \sum_{t\in D_x} \beta \Delta_t \left(\rho_1-\rho_2\right) \Delta_t \rho_2 dx \\
&= \int_X \int_T f\beta(\rho_1-\rho_2) d_t(\rho_2-\rho_1) dx + \frac{1}{2}\int_X f \sum_{t\in D_x} \beta \left(\Delta_t(\rho_2-\rho_1)\right)^2 dx
\end{aligned}
$$

$$
\begin{aligned}
&= -\frac{1}{2}\int_X \int_T f\beta d_t(\rho_1-\rho_2)^2 dx \\
&= \frac{1}{2}\int_X \left(\int_T f(\rho_1-\rho_2)^2 d_t\beta - f\beta(\rho_1-\rho_2)^2\Big|_0^T\right) dx \\
&= \frac{1}{2}\int_X \int_T f(\rho_1-\rho_2)^2 d_t\beta dx > 0\,.
\end{aligned}
$$

The second equality follows from a change of variable and the third from the integration by parts for Lebesgue-Stieltjes integration. The boundary term in the fourth equality vanishes because of the initial condition and terminal condition. The integral term is strictly positive since $\rho_1 \neq \rho_2$ and β is strictly increasing.

We then consider a possibly discontinuous time preference. Let $\{\beta_n\}_{n\in\mathbb{N}}$ be a sequence of uniformly bounded, continuous time preference that converges to β a.e.. The dominated convergence theorem implies that the sum of changes in payoff is

$$
\lim_{n\to\infty} \frac{1}{2}\int_X \int_T f(\rho_1-\rho_2)^2 d_t\beta_n dx \frac{1}{2}\int_X \int_T f(\rho_1-\rho_2)^2 d_t\beta dx > 0\,.
$$

□

For the rest of the proof, we denote ρ as the (symmetric) strategy in an equilibrium.

Lemma 11. *ρ is t-continuous.*

Proof. Suppose ρ is not t-continuous. We shall show that both players can profitably deviate to a generalized one-step-ahead strategy that preempt the opponent's discontinuous search.

We first define $\nu(t) := \int_X \sum_{D_x\cap[0,t]} \Delta_t\rho\, dx$ for $t \in T$ as the cumulative capacity expended on discontinuous search. It is nondecreasing due to the monotonicity condition and Lipschitz continuous due to the capacity constraint. By hypothesis, $\nu(T) > 0$. We denote $\nu^{-1}(r) := \min\{t \in T : \nu(t) = r\}$ for $r \in [0, \nu(T)]$ as the left-inverse of ν.

We then construct the generalized one-step ahead strategy. Let $\epsilon \in (0, \nu(T))$. Define ρ':

$$
\rho' := \begin{cases} \rho - \sum_{D_x\cap[0,t]} \Delta_t\rho + \sum_{D_x\cap[\nu^{-1}(\epsilon),\nu^{-1}(\nu(t)+\epsilon)]} \Delta_t\rho & \forall t \in [0, \nu^{-1}(\nu(T)-\epsilon)) \\ \rho - \sum_{D_x\cap[0,\nu^{-1}(\epsilon)]} \Delta_t\rho + \sum_{D_x\cap[0,\nu^{-1}(\nu(t)+\epsilon-\nu(T))]} & \forall t \in [\nu^{-1}(\nu(T)-\epsilon), T] \end{cases}.
$$

It is straightforward to verify that ρ' is a strategy. It expediates discontinuous search over $[\nu^{-1}(\epsilon), T]$ from t to $t'(t) := \nu^{-1}(\nu(t)-\epsilon) < t$ at the expense of delay for search over $[0, \nu^{-1}(\epsilon)]$ in the spirit of the one-step ahead strategy. Note that $t'(t) \uparrow \nu^{-1}(\nu(t)) \leq t$ as $\epsilon \downarrow 0$.

We finally show that ρ' is a profitable deviation for sufficiently small ϵ. Since the continuous part of ρ' is the same as ρ, we only need to quantify the discontinuous part. The payoff of ρ from discontinuous part is

$$\int_X \sum_{t\in D_x} f(x)\beta(t)\left(1-\rho(t,x)+\frac{1}{2}\Delta_t\rho(t,x)\right)\Delta_t\rho(t,x)dx\,.$$

The payoff of ρ' from the discontinuous part satisfies

$$\begin{aligned}
&\int_X \sum_{t'\in D'_x} f(x)\beta(t')\left((1-\rho(t',x))\Delta_t\rho'(t',x)+\frac{1}{2}\Delta_t\rho(t',x)\Delta_t\rho'(t',x)\right)dx\\
\geq &\int_X \sum_{t\in D_x\cap[0,\nu^{-1}(\nu(T)-\epsilon)]} f(x)\beta(t'(t))(1-\rho(t'(t),x))\Delta_t\rho(t,x)dx\\
\to &\int_X \sum_{t\in D_x} f(x)\beta(t)\left(1-\rho(t^-,x)\right)\Delta_t\rho(t,x)dx \quad \text{as } \epsilon\downarrow 0\,.
\end{aligned}$$

The inequality follows because the payoff from simultaneous discovery is nonnegative and $\Delta_t\rho(t,x)=\Delta_t\rho'(t'(t),x)$. The convergence follows from the dominated convergence theorem because $t'(t)\uparrow t$, ν-a.e. and $\beta(t^-)=\beta(t)$ a.e.. Since $1-\rho(t^-,x)=1-\rho(t,x)+\Delta_t\rho(t,x)$, the change in payoff converges to

$$\frac{1}{2}\int_X \sum_{t\in D_x} f(x)\beta(t)\left(\Delta_t\rho(t,x)\right)^2 dx>0\,.$$

Therefore, there exists a sufficiently small ϵ such that ρ' is a profitable deviation. □

Lemma 12. *The capacity constraint binds for ρ on $t\in[0,1]$.*

Proof. Let the flow rate of capacity used be $\gamma(t):=\frac{d}{dt}\int_X\rho(t,x)dx$. It exists and $\gamma(t)\leq 1$ due to the capacity constraint. Suppose the capacity constraint does not bind, i.e., $\gamma(t)<1$ for a positive-measure set of time $A\subset[0,1]\subset T$. Then there exists a positive-measure set of alternatives $B\subset X$ such that $\rho(1,x)<1$ for all $x\in B$. Recall λ is the Lebesgue measure on X. By selecting subsets of A and B, we have $\int_A(1-\gamma(t))dt=\lambda(B)$, $\sup A<1-\delta$, and $\rho(1,x)<1-\delta$ for all $x\in B$ for some $\delta\in(0,1)$.

We construct a profitable deviation to use the excess capacity early at the expense of search at a later time. Let $\epsilon\in(0,\delta)$. Define ρ' by

$$\rho'(t,x):=\min\left\{\rho(t,x)+\epsilon\frac{\int_{A\cap[0,t]}(1-\gamma(t))dt}{\int_A(1-\gamma(t))dt},1\right\}\,.$$

It is straightforward to verify that ρ' is a strategy. It uses ϵ fraction of the excess capacity to search B during A. The additional search on $x \in B$ precludes the last bit of search over $[t_x, T]$ where $t_x := \min\{t \in T : \rho(t,x) = 1-\epsilon\}$. Note that $t_x > 1$ because $\rho(1,x) < 1-\delta$.

The change in payoff is

$$\begin{aligned}
&\int_B \int_A f(x)\beta(t)(1-\rho(t,x))\epsilon \frac{(1-\gamma(t))dt}{\int_A (1-\gamma(t))dt} dx - \int_B \int_{[t_x,T]} f(x)\beta(t)(1-\rho(t,x))d_t\rho(t,x)dx \\
\geq\, &\epsilon\delta \frac{\int_B \int_A f(x)\beta(t)(1-\gamma(t))dtdx}{\int_A (1-\gamma(t))dt} - \epsilon^2 \left(\sup_T \beta\right) \int_B f(x)dx \,.
\end{aligned}$$

In the first term, we have $1-\rho(t,x) \geq 1-\rho(t,x) \geq \delta$. In the second term, we have $1-\rho(t_x,x) \leq \epsilon$ and $\int_{[t_x,T]} d_t\rho(t,x) = \rho(T,x) - \rho(t_x,x) = \epsilon$. As $\epsilon \downarrow 0$, the positive, linear term dominates the negative, quadratic term so the change in payoff is positive for sufficiently small ϵ. $\square$

We define the highest posterior as $\tilde{g}(t) := \sup_{x \in X} g(t,x) = \sup_{x\in X} f(x)(1-\rho(t,x))$. Inherited from ρ, it is nonincreasing and right-continuous. Let $H(t) := \{x \in X : f(x) \geq \tilde{g}(t)\}$ be the upper contour set of $\tilde{g}$.

Lemma 13. *$\rho(t,x) = 0$ for $x \in H^C(t)$ a.e. for all $t \in T$.*

Proof. Suppose there exists $t_0 \in (0,1)$ and positive-measure set $A \subset H^C(t_0)$ such that $\rho(t_0,x) > 0$ for all $x \in A$. We shall construct a profitable deviation that searches a different set of alternatives B with posterior close to $\tilde{g}(t_0)$ instead of A during $[0,t_0]$. This would contradict the hypothesis that (ρ,ρ) is an equilibrium.

The definition of $\tilde{g}$ implies that there exists positive-measure set $B \subset X$ such that $g(t_0,B) > \tilde{g}(t_0) - \delta/2$ for some $\delta > 0$. By selecting a sufficiently small δ and subsets of A and B, we also have $f(A) < \tilde{g}(t_0) - \delta$, $\rho(t_0,B) < 1-\delta$, and $\int_A \rho(t_0,A)dx = \mu(B)$.

Let $\epsilon \in (0,\delta)$. Define ρ' by

$$\rho'(t,x) := \begin{cases} (1-\epsilon)\rho(t,x) & \forall x \in A, t \in [0,t_0] \\ \rho(t,x) - \epsilon\rho(t_0,x) & \forall x \in A, t \in (t_0,T] \\ \min\left\{\rho(t,x) + \epsilon\frac{\int_A \rho(t,x)dx}{\int_A \rho(t_0,x)dx}, 1\right\} & \forall x \in B \\ \rho(t,x) & \forall x \notin A \cup B \end{cases}.$$

It is straightforward to verify that ρ' is a strategy. It uses fraction ϵ of capacity expended on A during $[0,t_0]$ to search B, and idles its capacity on $x \in B$ over $[t_x,1]$ where $t_x :=$

$\min\{t \in T : \rho(t,x) = \epsilon\}$.

The change in payoff is

$$\begin{aligned}
&\epsilon \int_B \int_{[0,t_0]} \beta(t) g(t,x) d_t \frac{\int_A \rho(t,x)dx}{\int_A \rho(t_0,x)dx} dx - \epsilon \int_A \int_{[0,t_0]} \beta(t) g(t,x) d_t \rho(t,x) dx \\
&- \int_B \int_{[t_x,T]} \beta(t) g(t,x) d_t \rho(t,x) dx \\
&> \epsilon \int_B \int_{[0,t_0]} \beta(t) \left(\tilde{g}(t_0) - \frac{\delta}{2}\right) d_t \frac{\int_A \rho(t,x)dx}{\int_A \rho(t_0,x)dx} dx - \epsilon \int_A \int_{[0,t_0]} \beta(t)(\tilde{g}(t_0) - \delta) d_t \rho(t,x) dx \\
&- \epsilon^2 \left(\sup_T \beta\right) \int_B f(x) dx \\
&= \frac{\delta}{2} \epsilon \int_{[0,t_0]} \beta(t) d_t \left(\int_A \rho(t,x)dx\right) - \epsilon^2 \left(\sup_T \beta\right) \int_B f(x) dx\,.
\end{aligned}$$

In the inequality, we use $g(t,B) \geq g(t_0,B) \geq \tilde{g}(t_0) - \delta$ for the first term, $g(t,A) < f(A) < \tilde{g}(t_0) - \delta$ for the second term, and $g(t,x) \leq f(x)$ for the third term. In the equality, we use $\mu(B) = \int_A \rho(t_0,x)dx$. The linear coefficient is positive because β is positive a.e. and $\int_A \rho(t,x)dx$ is a positive and absolutely continuous measure on $[0,t_0]$. Therefore, the linear term dominates the quadratic term so the difference in payoff is positive as $\epsilon \downarrow 0$. $\square$

Lemma 14. *$\rho(t,x) = 1 - \frac{\tilde{g}(t)}{f(x)}$ for $x \in H(t)$ a.e. for all $t \in [0,1]$.*

Proof. By definition, $\{x \in X : g(t,x) > \tilde{g}(t)\}$ is a null set for all $t \in \mathbb{Q} \cap [0,1]$. Therefore, we may focus on $\{x \in X : g(t,x) \leq \tilde{g}(t) \quad \forall t \in T\}$ which is of full measure by the right-continuity of ρ and $\tilde{g}$. Because $\tilde{g}$ is nondecreasing and right-continuous while $g(\cdot,x)$ is continuous, the difference $\tilde{g}(\cdot) - g(\cdot,x)$ is lower semi-continuous.

Suppose there exists $t_0 \in (0,1)$ and positive measure set $A \subset H(t_0)$ such that $\rho(t_0,x) > 1 - \frac{\tilde{g}(t_0)}{f(x)}$ or equivalently $\tilde{g}(t_0) - g(t,x) > 0$ for all $x \in A_0$. Without loss of generality, we further have $\tilde{g}(t_0) - g(t_0,x) > \eta$ for some $\eta > 0$. We shall construct a deviation that searches alternatives with posterior close to $\tilde{g}$ instead of A_0.

We first identify time interval $[t_1,t_2]$ over which ρ searches some set A with posterior that is far from the maximum. Let $t^x := \max\{t \in (0,t_0) : \tilde{g}(t) - g(t,x) \leq \eta/2\}$ for $x \in A_0$. It is well-defined because $\tilde{g}(\cdot) - g(\cdot,x)$ is lower semi-continuous. We have $\tilde{g}(t) - g(t,x) \geq \eta/2$ for all $t \in [t^x,t_0]$. Observe that $g(t^x,x) - g(t_0,x) \geq (\tilde{g}(t_0) - g(t_0,x)) - (\tilde{g}(t^x) - g(t^x,x)) \geq \eta/2$. By selecting a sufficiently small η and a subset of A_0, we further have $\rho(t_0,x) - \rho(t^x,x) \geq \eta/2$. As we have shown the properties for all $x \in A_0$, there exists $[t_1,t_2] \subset [0,t_0]$, positive-measure set $A \subset A_0$, and $\delta > 0$ such that $\tilde{g}(t) - g(t,A) > \delta$ and $\rho(t_2,A) - \rho(t_1,A) > \delta$.

We then identify a set B posterior of which is close to the maximum. By definition of $\tilde{g}$, there exists positive-measure set $B \subset X$ such that $\tilde{g}(t) - g(t, B) < \delta/2$ for all $t \in [t_1, t_2]$. By selecting a sufficiently small δ and subsets of A and B, we further have $\rho(t_2, B) < 1 - \delta$ and $\int_A \rho(t_2, x) - \rho(t_1, x) dx = \mu(B)$.

We now construct the deviation strategy. Let $\epsilon > 0$. Define ρ' by

$$\rho'(t,x) := \begin{cases} \rho(t,x) & \forall x \notin A \cup B \text{ or } t \in [0, t_1) \\ \rho(t_1, x) + (1-\epsilon)\left(\rho(t,x) - \rho(t_1,x)\right) & \forall x \in A, t \in [t_1, t_2) \\ \rho(t,x) - \epsilon\left(\rho(t_2,x) - \rho(t_1,x)\right) & \forall x \in A, t \in [t_2, T] \\ \min\left\{\rho(t,x) + \epsilon \frac{\int_A (\rho(t,x) - \rho(t_1,x))dx}{\int_A (\rho(t_2,x) - \rho(t_1,x))dx}, 1\right\} & \forall x \in B, t \in [t_1, T] \end{cases}.$$

It is straightforward to verify that ρ' is a strategy. It expends fraction ϵ of the capacity on A during $[t_1, t_2]$ to search B, and idles its capacity on $x \in B$ over $[t_x, 1]$ where $t_x := \min\{t \in T : \rho(t,x) = \epsilon\}$.

The change in payoff is

$$\begin{aligned}
&\epsilon \int_B \int_{[t_1,t_2]} \beta(t) g(t,x) d_t \frac{\int_A \left(\rho(t,x) - \rho(t_1,x)\right) dx}{\int_A \left(\rho(t_2,x) - \rho(t_1,x)\right) dx} dx - \epsilon \int_A \int_{[t_1,t_2]} \beta(t) g(t,x) d_t \rho(t,x) dx \\
&- \int_B \int_{t_x, T} \beta(t) g(t,x) d_t \rho(t,x) dx \\
>& \epsilon \int_B \int_{[t_1,t_2]} \beta(t) \left(\tilde{g}(t) - \frac{\delta}{2}\right) d_t \frac{\int_A \left(\rho(t,x) - \rho(t_1,x)\right) dx}{\int_A \left(\rho(t_2,x) - \rho(t_1,x)\right) dx} dx - \epsilon \int_A \int_{[t_1,t_2]} \beta(t) \left(\tilde{g}(t) - \delta\right) d_t \rho(t,x) dx \\
&- \epsilon^2 \left(\sup_T \beta\right) \int_B f(x) dx \\
=& \frac{1}{2} \epsilon \delta \int_{[t_1,t_2]} \beta(t) d_t \left(\int_A \rho(t,x) dx\right) - \epsilon^2 \left(\sup_T \beta\right) \int_B f(x) dx\,.
\end{aligned}$$

In the inequality, we use $g(t, B) > \tilde{g}(t) - \delta/2$ for the first term, $g(t, A) < \tilde{g}(t) - \delta$ for the second term, and $\beta(t) \leq \beta(T)$, $g(t,x) \leq f(x)\epsilon$, and $\int_{[t_x,T]} d_t \rho(t,x) = \epsilon$ for the third term. Since β is positive a.e. and $\int_A \rho(t,x)dx$ is a positive and absolutely continuous measure on $[t_1, t_2]$, the linear coefficient is positive. The linear term dominates the quadratic term so the difference in payoff is positive as $\epsilon \downarrow 0$. $\square$

Lemma 12 states that the capacity constraint binds. Therefore, Lemma 13 and Lemma 14 defines ρ as the leveling strategy $\bar{\rho}$ by Lemma 2.

B.3 Multiple Players

At last, we analyze the strategic exploration game with more than two players, and derive the unique symmetric equilibrium similar to the leveling equilibrium. We also show the existence of asymmetric equilibria by an example.

The setup generalizes naturally for multiple players. There are $n > 2$ players each of whom faces the same capacity constraint. Therefore, the set of strategies available to each of them is still given by Definition 2. Denote the distributional strategy of player i by ρ_i. The probability that x is searched up to time t by at least one of player i's opponents is then $\rho_{-i}(t,x) := 1 - \prod_{j \neq i}(1 - \rho_j(t,x))$. As in the case with two players, the posterior induced by ρ_{-i} is $g_{-i}(t,x) := (1 - \rho_{-i}(t,x))f(x)$. With this notation, the payoff of player i is again given by Equation (2.1).

We consider the symmetric strategy profile such that the posterior g_{-i} is leveling for every player i. More precisely, let $\bar{g} : T \to [0, \sup f]$ be the leveling function defined implicitly by

$$\int_X \left(1 - \left(\frac{\bar{g}(t)}{f(x)}\right)^{\frac{1}{n-1}}\right) \mathbf{1}_{\{f(x) \geq \bar{g}(t)\}}(x)dx = t \tag{B.1}$$

for $t \in [0,1]$ and $\bar{g}(t) = 0$ for $t > 1$. The proof of existence and uniqueness of $\bar{g}$ is analogous to the proof of Lemma 2. By Equation (B.1), the leveling strategy

$$\bar{\rho}(t,x) := \left(1 - \left(\frac{\bar{g}(t)}{f(x)}\right)^{\frac{1}{n-1}}\right) \mathbf{1}_{\{f(x) \geq \bar{g}(t)\}}(x) \tag{B.2}$$

is a well-defined distributional strategy.

Theorem 9. *The profile of distributional strategies $(\bar{\rho}, ..., \bar{\rho})$ is the unique symmetric Nash equilibrium of the game with n players.*

Theorem 9 characterizes the unique equilibrium among the class of symmetric equilibria, but does not establish the uniqueness of *Nash equilibrium* which does not hold for $n > 2$. We present an example of an asymmetric equilibrium in which symmetric players enjoy unequal equilibrium payoffs.

Example 4. *Take the uniform prior $f \equiv 1$, $T = 1$, and $n = 5$ players. Partition the alternatives $X = [0,1]$ into two halves, $X_1 := [0, \frac{1}{2})$ and $X_2 := [\frac{1}{2}, 1]$, and the players into two groups: {1,2} and {3,4,5}. For each player in the first group, the strategy is given as* *follows:*

$$\rho_1(t,x) = \begin{cases} \min\{2t, 1\}, & \textit{if } x \in X_1, \\ \max\{2t-1, 0\}, & \textit{if } x \in X_2. \end{cases}$$

For each player in the second group, the strategy is given as follows:

$$\rho_2(t,x) = \begin{cases} \min\{2t, 1\}, & \textit{if } x \in X_2, \\ \max\{2t-1, 0\}, & \textit{if } x \in X_1. \end{cases}$$

That is, each player in the first group explores uniformly over the left half X_1 until the alternatives are exhausted at $t = \frac{1}{2}$, and then the other half X_2. Each player in the second group explores in the reverse order. It can be verified that $(\rho_1, \rho_1, \rho_2, \rho_2, \rho_2)$ is a Nash equilibrium of the 5-player game. Since the discovery must occur before $t = \frac{1}{2}$, the equilibrium exploration is different from the one described in Theorem 9 which has a full support $T = [0, 1]$. Moreover, despite symmetric capacities, the equilibrium payoffs are asymmetric: player 1 and 2 enjoy an expected payoff of $\frac{1}{4}$ while player 3, 4, and 5 have an expected payoff of $\frac{1}{6}$.

Multiple equilibria arise because more than one opponent can preempt any given player. Whenever a player explores an alternative in equilibrium, at least one of his opponents will explore the same alternative to preempt that player. With only two players, the equilibrium exploration is unique because each player faces only one opponent. With more than two players, however, multiple opponents can preempt any given player. For $t \in \left[0, \frac{1}{2}\right]$ in Example 4, player 1 is preempted by player 2 on the left half and by all other players on the right half.[20] Other players face similar situations which sustain the non-leveling equilibrium.

In the unique symmetric equilibrium, the good alternative is discovered more quickly as the number of players increases because of the increased total capacity despite the additional duplication.

Theorem 10. *In the class of symmetric equilibria, the distribution of discovery time is decreasing in n in the first-order stochastic dominance sense.*

The result is rather intuitive; the non-trivial part is the ranking of discovery times in the first-order stochastic dominance, as the distributions of discovery time for all n players share the same support.

[20] The strategies over $t \in \left[\frac{1}{2}, 1\right]$ are irrelevant because the posterior of every player of every alternative is zero.

B.4 Proof of Theorem 9

Proof. Verifying that the strategy profile is a Nash equilibrium is analogous to the arguments in the proof of Theorem 6 and we shall not replicate the proof here. In any symmetric strategy profile, the myopic payoff is strictly positive so the players must exhaust their capacities in equilibrium. Therefore, in any symmetric Nash equilibrium, the payoff of each player is π/n by symmetry.

We first show that, for any strategy ρ, the leveling strategy $\bar{\rho}$ guarantees player i payoff π/n when all other players employ ρ. Since the leveling strategy is t-absolutely continuous, the payoff of player i equals to the time integral of flow payoffs by Fubini's theorem

$$\begin{aligned} u\left(\bar{\rho}, \rho_{-i}\right) &= \int_X \int_T f(x)(1-\rho_{-i}(t,x))\partial_t \bar{\rho}(t,x) dt dx \\ &= \int_T \int_X f(x)(1-\rho_{-i}(t,x))\partial_t \bar{\rho}(t,x) dx dt\,. \end{aligned}$$

As the leveling profile gives payoff π/n for all players, it suffices to show that, for all $t \in (0,1)$, the leveling strategy $\bar{\rho}(t,\cdot)$ minimizes the flow payoff of $\bar{\rho}$

$$\min_{\rho(t,\cdot)} \int_X f(x)(1-\rho(t,x))^{n-1}\partial_t \bar{\rho}(t,x) dx \tag{B.3}$$

subject to the capacity constraint at t. Without the constraint $\rho(t,\cdot) \leq 1$, the relaxed problem is equivalent to the first-order condition

$$(n-1)f(x)\partial_t \bar{\rho}(t,x)(1-\rho(t,x))^{n-2} = C$$

for $\{x \in X : \rho(t,x) > 0\}$ almost everywhere, together with the complementary slackness condition

$$(n-1)f(x)\partial_t \bar{\rho}(t,x) \leq C$$

for $\{x \in X : \rho(t,x) = 0\}$ almost everywhere, for some Lagrange multiplier $C \geq 0$. It is then straightforward to show that the leveling strategy $\rho(t,\cdot)$ solves the two conditions and hence the minimization problem.

From here, uniqueness can be shown along the idea of Theorem 1 by constructing a modified leveling strategy that yields strictly higher payoff, as in the case of $n = 2$. We provide here a shorter proof that takes advantage of the strict convexity of minimization problem (B.3) when $n > 2$. With duplication among other players, they can no longer achieve the minimum payoff of player i by any other strategies.

We proceed to show that the leveling strategy $\bar{\rho}$ yields payoff strictly above π/n when all other players employ $\rho \neq \bar{\rho}$. Since the strategy is not leveling, there exist time interval (t_0, t_1) and positive-measure set $A \subset X$ such that $\rho(t,x) \neq \bar{\rho}(t,x)$ for all $t \in (t_0, t_1)$ and $x \in A$. The strict convexity of minimization problem (B.3) implies that the minimizer $\bar{\rho}(t,\cdot)$ is unique. Therefore, the flow payoff of $\bar{\rho}$ is strictly above the minimum over (t_0, t_1).

For any symmetric strategy profile $(\rho, ..., \rho)$ for $\rho \neq \bar{\rho}$, all players have a profitable deviation to $\bar{\rho}$. Therefore, the strategy profile is not an equilibrium. □

B.5 Proof of Theorem 10

Proof. Suppose $n' > n$. Let $\bar{g}'$ and $\bar{g}$ be the leveling function associated with n' players and n players respectively. Parallel to the proof of the benchmark case, the probabilities of simultaneous discovery are both zero, and the flow probabilities of discovery are $n'\bar{g}'$ and $n\bar{g}$ respectively. By Lemma 6, it suffices to prove the stochastic order between their inverses $h'(\cdot/n')$ and $h(\cdot/n)$.

By the Fubini theorem, the integral can be written as

$$\begin{aligned}
\int_0^y h\left(\frac{z}{n}\right) dz &= \int_0^y \int_X \left(1 - \left(\frac{z}{nf(x)}\right)^{\frac{1}{n-1}}\right) \mathbf{1}_{\{nf(x)\geq z\}}(x,z) dx dz \\
&= \int_X \int_0^y \left(1 - \left(\frac{z}{nf(x)}\right)^{\frac{1}{n-1}}\right) \mathbf{1}_{\{nf(x)\geq z\}}(x,z) dz dx \\
&= \int_X I(n, y, f(x)) dx,
\end{aligned}$$

where the integrand is given by

$$I(n, y, f(x)) = \begin{cases} y - \frac{n-1}{n}\left(\frac{y}{nf(x)}\right)^{\frac{1}{n-1}} y, & \text{if } n \geq \frac{y}{f(x)}; \\ f(x), & \text{if } n < \frac{y}{f(x)}. \end{cases}$$

We obtain the desired inequality by noting that $I(n, y, f(x))$ is decreasing in n. □