DISCUSSION PAPER SERIES

DP13483
(v. 2)

Experimentation, Learning, and Preemption

Emre Ozdenoren, Heidrun C. Hoppe-Wewetzer and Georgios Katsenos

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

Experimentation, Learning, and Preemption

Emre Ozdenoren, Heidrun C. Hoppe-Wewetzer and Georgios Katsenos

Discussion Paper DP13483
First Published 24 January 2019
This Revision 24 January 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Experimentation, Learning, and Preemption

Abstract

This paper offers a model of experimentation and learning with uncertain outcomes as suggested by Arrow (1969). Investigating a two-player stopping game, we show that competition leads to less experimentation, which extends existing results for preemption games to the context of experimentation with uncertain outcomes. Furthermore, we inquire about the extent of experimentation under two information settings: when the researchers share information about the outcomes of their experiments and when they do not share such information. We discover that the sharing of information can generate more experimentation and higher value for a relatively wide range of parameters. We trace this finding to the stronger ability to coordinate on the information obtained through experimentation when it is shared. Our model allows to shed light on recent criticism of the current scientific system.

JEL Classification: D83, O31

Keywords: N/A

Emre Ozdenoren - eozdenoren@london.edu
London Business School and CEPR

Heidrun C. Hoppe-Wewetzer - hoppe@mik.uni-hannover.de
Leibniz Universität Hannover and CEPR

Georgios Katsenos - katsenos@mik.uni-hannover.de
Leibniz Universität Hannover

Experimentation, Learning, and Preemption*

Heidrun Hoppe-Wewetzer† Georgios Katsenos‡

Emre Ozdenoren§

January 24, 2021

Abstract

We offer a model of experimentation and learning with uncertain outcomes, and show that competition leads to less experimentation, extending results for preemption games to experimentation with uncertain outcomes. We compare experimentation under two information settings: when the researchers share vs. keep private the information about the outcomes of their experiments. We discover that information sharing can generate more experimentation and higher welfare when uncertainty about the feasibility of a breakthrough is large; breakthroughs are rare even when they are feasible; and experiments frequently fail to produce results. Our results shed light on recent criticism of the scientific system. **Keywords:** Stopping game, experimentation, learning, preemption, multiarmed bandit problem **JEL Numbers**: D83, O31

*We thank Marina Halac, Ehud Kalai, Sven Rady, Yiman Sun, Stephan Thomsen, Julia Wirtz, and participants at the ESWC 2020 (Bocconi) for helpful comments. We are also grateful to Morton Kamien and Konstantin Wewetzer for discussions that inspired this work. Financial support of the German Science Foundation (DFG HO 3814/1-1) is gratefully acknowledged.

†Leibniz University Hannover, Department of Economics and CEPR, hoppe@mik.uni-hannover.de

‡Leibniz University Hannover, Department of Economics, katsenos@mik.uni-hannover.de

§London Business School and CEPR, eozdenoren@london.edu

1 Introduction

"Eureka!" moments may not happen frequently, but the prospects of discovery – establishing a new idea or observing something that no one has ever seen before – keep scientists going even when the rewards are uncertain. As emphasized by Arrow (1969), uncertainty about the likelihood of eventual success is an important feature of scientific inquiry. At each stage of the research process, *"something is learned with regard to the probability distribution of outcomes for future repetitions of the activity"* (Arrow, 1969). In fact, Arrow argues that the information gain from an experiment might be more important than its concrete output. Challenging earlier models of research and development, he calls for a more general formulation of research activity, including the case where the potential outcome is not known with certainty. Although uncertainty about outcomes is ubiquitous in science, very few formulations of this type have since been proposed in the literature (see, e.g., Halac et al. (2017) in the context of research contests).

In this paper, we offer a model combining uncertainty about research outcomes, as suggested by Arrow (1969), with another typical feature of research activity: the competition to be first. Scientists seek to establish priority by being first to publish an advance in knowledge and are concerned at being preempted in this by another scientist. As Fang and Casadevall (2012) put it, *"Since the earliest days of science, bragging rights to a discovery have gone to the person who first reports it"*.[1]

The main objective is to understand how the combination of learning about uncertain outcomes and preemption affects the duration of scientific inquiry and welfare. The model also allows us to analyze the fundamental issues underlying recent criticism of the current scientific system: Lawrence (2016), for instance, points to the attempts by university administrators to rank scientists against each other based on publications numbers and allocate funds and jobs accordingly. He argues that this

[1] For empirical evidence of this winner-takes-all rewards structure in science, see Hagstrom (1974), Newman (2009), and Sabatier and Chollet (2017).

practice is damaging modern science by increasingly enhancing the importance attached to being first: *"All of us (...) focus our research to produce enough papers to compete and survive. Thus, projects are published as soon as possible and many therefore resemble lab reports rather than fully rounded and completed stories. (...) I think this emphasis on article numbers has helped make papers poorer in quality."* Similarly, for biology and medicine, Broad (1981) observes that teams often settle for the *"least publishable unit"* - a practice that has come under fire for leading to research outcomes of lower quality overall. Adding to the criticism, the editors of *Nature* have recently urged scientists conducting laboratory studies to take greater care in their work, citing several types of *"avoidable errors"*, in terms of both methodology and presentation, that diminish the quality of the published output and make reproduction of the findings more difficult (Nature Publishing Group, 2012). In response to the critique, Fang and Casadevall (2012), among other scientists, advocate a system that offers greater collegiality, freer sharing of information, and cooperation.

To set the stage, we study the extent of experimentation in a two-player stopping game and compare it to its counterpart in a setting without competition, which corresponds to the cooperative problem. As we show in our benchmark result, cooperation indeed always leads to more experimentation and value. However, competition is almost always an inherent feature of scientific inquiry. This raises the important question of whether transparency and sharing of research progress leads to more or less experimentation and value in a competitive setting with uncertain research outcomes. The answer is not immediate since there are competing forces. On the one hand, keeping research progress private might soften the competitive preemption threat, but on the other hand, sharing of information may reduce the uncertainty about the likelihood of eventual success.

To better understand this trade-off, we compare experimentation under two information settings: when the researchers share information about the outcomes of their experiments and when they do not share such information. We find that the sharing of information generates more experimentation and higher welfare when uncertainty about the feasibility of a breakthrough is large; breakthroughs are rare even when they are feasible; and experiments frequently fail to produce results. In scientific research, we can approximate the probability of a breakthrough with the frequency of publishing a landmark paper, which appears to be quite low.[2] Hence, our result supports the view of Fang and Casadevall (2012), who are in favor of freer information sharing. This finding may be surprising, particularly in light of Hopenhayn and Squintani (2011), who show that secrecy may result in longer durations of experimentation by reducing the researcher's fear of being preempted. While there are several conflicting effects, we trace our results to the stronger ability to coordinate on the information obtained through experimentation when it is shared. This is one of the central insights of this paper.

Formally, we study a model in which two researchers running successive experiments decide at any point in time whether to stop and go forward with their best research finding thus far. Each experiment, with some probability, is successful, and the player receives a draw from some unknown distribution interpreted as the result of the experiment. With complementary probability, the experiment is unsuccessful and fails to produce any results. As we will see later, the probability of failed experiments is the key parameter that distinguishes public and private learning.[3] The unknown distribution of draws remains fixed throughout the game, either producing low-value draws with certainty or randomizing between low- and high-value draws. We interpret a low-value draw as a mundane result and a high-value one as a breakthrough result from the project. To capture the uncertainty about the potential of the project, we assume the researchers do not know which is the true distribution,

[2] See, e.g., Bornmann, Ye and Ye (2018).

[3] In natural science and many branches of social science, failure actually abounds (see, e.g., Mohs and Greig, 2017; Barwich, 2019 for empirical evidence). As Parkes (2019) notes: "*If we want to make new discoveries, that means taking a leap in the dark - a leap we might not take if we're too afraid to fail.*"

and they only share a prior belief about the feasibility of a high-value outcome.[4] The competition is winner-takes-all so researchers have an incentive to stop preemptively, i.e., *"publish their partial findings quickly, rather than dropping the bombshell of a completely solved problem on their surprised colleagues"* (Hagstrom, 1974).

We construct perfect Bayesian equilibria in symmetric threshold strategies. When the players can share information about their draws truthfully, we establish the existence of equilibria in which the two players share common beliefs about the potential of the project and remain in the game until either a draw of high value occurs or their beliefs about the possibility of such a draw become too pessimistic. The latter event occurs when the total number of low-value draws exceeds a certain threshold, with the consequence that the players decide to stop simultaneously in equilibrium.

Our analysis in the case of no information sharing is complicated because of the complexity of the belief structure. Each player has to form beliefs regarding the draws his opponent has received, and these beliefs and the player's own results determine in turn the player's belief about both the feasibility of a high-value outcome and the threat of preemption. In general, since the players' beliefs are private, it is difficult to track their evolution and, thus, to establish the existence of an equilibrium. The use of time as a public variable allows only for a partial simplification of the belief structure because each player's beliefs about the number of low-value draws the other player has obtained depends on the number of low-value draws the player has himself obtained, as well as the other player's equilibrium strategy. Despite this complication, we construct symmetric equilibria in strategies involving time-dependent thresholds. Each player experiments until he receives a high-value draw or accumulates too many low-value draws, although the threshold for the number of low value draws may vary non-monotonically over time.

[4]Note that this feature separates this paper from earlier research models in which players accumulate outcomes from a known distribution (e.g. Hopenhayn and Squintani, 2011; Bobtcheff et al. 2016). For a discussion of the related literature, see Section 7.

Comparing the duration of experimentation and the players' total welfare under common and private learning regimes, yields the following results. Without the possibility of failed experiments, common and private learning become identical. As the likelihood of failed experiments increases, common learning starts to generate more experimentation. In addition, if there is more uncertainty about the feasibility of a breakthrough or if breakthroughs are rare even when feasible, then common learning generates more experimentation. These results provide testable implications of our model.

2 Model

Two players, 1 and 2, engage in a stopping game of successive experiments, taking place in discrete time periods $t = 1, \ldots, T$. At the beginning of each period t, as long as the game continues, each player $i \in \{1, 2\}$ runs a new experiment. With probability $1 - r$, where $r \in (0, 1)$, player i's experiment is unsuccessful and fails to produce any valuable result. With probability r, the experiment is successful and provides new information about the common natural world. This is expressed by a draw x_t^i for player i in period t. We assume that $x_t^i \in \{L, H\}$, where $0 < L < H$. That is, a successful experiment either provides some partial finding (of value L) or yields an important discovery (of value H). Incremental improvements over time are neglected in our formulation in order to sharpen the focus on the players' incentives to keep going, even though experiments may fail, in the hope of making a significant discovery.

An inherent feature of experimentation is the uncertainty regarding the distribution of the draws. Specifically, we assume that the values x_t^i are distributed according to either

$$x_t^i = \begin{cases} H, & \text{with probability } q; \\ L, & \text{with probability } 1-q, \end{cases}$$

where $q \in (0,1)$, or

$$x_t^i \equiv L.$$

The distribution is chosen randomly (by *nature*) at the beginning of the game, with probabilities p and $1-p$ respectively, in a manner unobservable to the players, and remains the same throughout the game.[5] Conditional on the choice of distribution, the values x_t^i are independent across players and across periods. Thus, unless a draw of value H is obtained in an experiment, whether such an outcome is at all possible is unknown to the players.

We will consider two opposite cases regarding the observability of the players' experimentation outcomes: one in which each player can observe the draws of his opponent and the other in which each player can observe only his own draws.

At the end of each period t, each player i has to decide, after observing his own draw, x_t^i, and possibly his opponent's draw, x_t^j, whether to stop in that period or continue to period $t+1$. These actions are denoted by s or c, respectively. The two players make their decisions simultaneously, and the game continues until at least one player decides to stop.

We assume that the experiments of the two players are directly competitive: the player who stops first receives a payoff equal to the value of his best past draw, while his opponent receives nothing. This winner-takes-all assumption seems particularly suited for a model of rivalry among scientists (cf. Hagstrom, 1974; Lawrence, 2016).[6]

[5]Note that players are sampling from the same distribution. The assumption is made to focus on scientists seeking to identify facts about the common natural world rather than inventing new technologies.

[6]The assumption that preemption destroys all value to the second player simplifies the exposition but is not crucial to our results. Our analysis would apply as long as the claim of L by one player destroys some nontrivial part of the value that the other player can claim.

If both players decide to stop at the same time, then we assume that only one of them – each with probability 1/2 – actually succeeds and becomes the first mover.[7] The two players discount time by a common rate $\delta \in (0, 1)$ and they suffer no other cost for remaining active in the game.[8]

For each player i, a (private) history $h_t^i \in H_t^i$ at the time of his decision in period t consists of the following elements, depending on our observability assumption:

a. Player i's own past draws $x_\tau^i \in \{\emptyset, L, H\}$, for $\tau = 1, \ldots, t$, where $\emptyset$ denotes the occurrence of no draw;

b. Player j's past draws $x_\tau^j \in \{\emptyset, L, H\}$, for $\tau = 1, \ldots, t$, when their observation is possible;

c. Trivially, the two players' past decisions to continue, (c, c), for $\tau = 1, \ldots, t-1$.

A strategy of player i in period $t < T$ indicates whether the player stops or continues at the end of period t, for any possible time-t history. Hence, a period t strategy is a function

$$\sigma_t^i : H_t^i \longrightarrow \{ s, c \},$$

while player i's strategy for the entire game is a finite sequence of time-t strategies,

$$\sigma^i \;=\; \{\sigma_t^i\}_{t=1}^{T-1}.$$

We focus on pure strategies. Thus, each player i's strategy at time t partitions the set of the player's histories H_t^i into stopping and continuation regions, $\bar{H}_t^i$ and $H_t^i \setminus \bar{H}_t^i$.

Finally, our solution concept is that of the perfect Bayesian equilibrium.

---

[7]See Hoppe and Lehmann-Grube (2005) for a discussion of this tie-breaking rule in timing games.

[8]Our analysis extends with only slight modifications to the case in which there is a constant cost for each period a player is active. Since the presence of a discount factor suffices to make experimentation costly and to provide incentives to a player to stop experimenting even if he faces no preemption threat, we have chosen not to include such costs in our model.

3 The Single-Player Problem

We start our analysis by examining the benchmark case in which experimentation is carried out by only one player. This problem, adjusted for duplication of experiments by the two players, is equivalent to the cooperative problem.

Clearly, the player will not stop before obtaining at least one draw and will not continue after obtaining a draw of H. Hence, the problem reduces to choosing whether to stop experimenting, claiming a value of L, or to continue at a cost of $(1-\delta)L$ for each additional period to potentially increase this value by $\delta(H-L)$.

Given the player's uncertainty about the feasibility of H, the problem of this section takes the form of a so-called *multiarmed bandit problem.* That is, experimentation can be thought of as a sequence of plays on a slot machine that has multiple arms, where each arm corresponds to a different but unknown probability distribution of payoffs. In our setting, the player must choose between a *sure arm*, i.e., exiting the game and obtaining the retirement payoff, and a *risky arm*, i.e., continuing to receive draws, the profitability of which he can investigate by selecting it.

The expected payoff from continuing to the next period depends on the player's belief about the distribution from which he draws. The player becomes more pessimistic that a draw of value H is feasible each time he receives a new draw of L. In particular, if the player has received $n \geq 1$ draws of L, then the player believes that he draws from the first distribution with probability

$$p(n) = \frac{(1-q)\, p(n-1)}{1-q\, p(n-1)}, \tag{1}$$

defined recursively, with $p(0)=p$. The sequence $\{p(n)\}_{n=0}^{\infty}$ is decreasing, since we have $p(n)/p(n-1)<1$, for all $n \in \mathbb{N}$. Therefore, the expected value of staying in the game one more period, i.e., the value of using the stochastic arm one more time, weakly decreases as the game progresses. Hence, this is the *deteriorating case* of the multiarmed bandit problem, in which the optimal policy takes a simple cutoff form (see Bertsekas, 2001, p.69): the player should use the stochastic arm (for at least one more period) if and only if the expected payoff from its next use exceeds the immediate retirement payoff, that is, as long as the number n of L draws that the player has obtained does not exceed a certain threshold.

To calculate that threshold, we can write the continuation value, $V(n)$, when the player has $n \geq 1$ draws of L as

$$V(n) = \max\{\, L,\, \delta\, (rp(n)qH + r(1-p(n)q)V(n+1) + (1-r)V(n))\, \}$$

The player can always obtain L by stopping immediately today. If he continues, there are three possibilities. With probability $rp(n)q$, he draws H and stops. With probability $r(1-p(n)q)$, he draws L so that he has $n+1$ draws of L, and the value of his continuation problem is $V(n+1)$. Finally, with probability $1-r$, he does not receive a draw and the value of his continuation problem remains $V(n)$.

At the threshold, player finds it optimal to continue with n draws of L but to stop with $n+1$ such draws. Then, the value function becomes

$$V(n) = \delta\, (rp(n)qH + r(1-p(n)q)L + (1-r)V(n))$$

Hence,

$$V(n) = \frac{\delta rp(n)qH + \delta r(1-p(n)q)L}{1-\delta+\delta r}$$

Let $\hat{N}$ be the largest n such that $V(n) \geq L$ or, equivalently, the smallest n such that $V(n) < L$

$$\hat{N} = \min\{\, n \in \mathbb{N} \,:\, \delta\, p\,(n)\; rq\,(H-L) \;<\; (1-\delta)\, L\,\} \tag{2}$$

Then, the optimal rule is to continue experimentation as long as $n_t^i < \hat{N}$ and $t \leq T-1$; and to stop otherwise. In particular, since $p\,(n) \to 0$ as $n \to \infty$, player i will stop experimenting after receiving a finite number of L draws.

Finally, to obtain a proper benchmark for our subsequent analysis of the impact of rivalry, we slightly modify the single-player case and allow the player to receive up to 2 draws in each period. This modification is necessary to account for the mere duplication of experiments in the setting with rivalry between two players. In this case, given the player's beliefs $p(n_t^i)$ at the end of period t, the probability that the player obtains *at least* one draw of H in the period $t+1$ is

$$p^H\left(n_t^i\right) \;=\; p(n_t^i)\,\left[\,1-(1-rq)^2\,\right] \tag{3}$$

Our previous analysis implies that the single player will continue experimentation in periods $t = 1, \ldots, T-1$, until he receives either a draw of H or $n_t^i \geq N^*$ draws of L, where

$$N^* = \min\{\, n \in \mathbb{N} \,:\, \delta\, p^H(n)\,(H-L) \;<\; (1-\delta)\,L\,\} \tag{4}$$

In the sequel, we consider the impact of competition on experimentation when outcomes are observed publicly or privately and compare these cases with each other as well as with the above single-player benchmark.

4 Common Learning

We now examine the two players' interaction. In this section, we assume that each player is fully informed of the experimental results of his rival. Players may have this information for various reasons. For example, they may be able to observe each others experiments or there may be truthful communication between the players.

In this environment, in every period $t \geq 1$, the two players share *common beliefs* about the feasibility of an H outcome. If no draw of H has been obtained, these beliefs are expressed by the probability $p(n_t)$, where n_t is the total number of L draws obtained by the two players up to period t, determined recursively, according to equation (1) in the single-player problem. Hence, the probability that at least one draw of H is obtained by either player in the next period if both players continue to it, is $p^H\,(n_t)$, defined by equation (3).

We construct a symmetric perfect Bayesian equilibrium in which experimentation terminates prior to the final period T if one or both players receive an H draw or if the total number of L draws reaches a certain threshold. Like in the single player case, with common learning, we show that each player's continuation payoff decreases as the number of L draws obtained (and jointly observed) by the two players increases. Therefore, each player's optimal stopping strategy must take the form of a threshold rule on the total number of L draws; this threshold is obtained by solving the Bellman equation describing each player's continuation problem.

At any time $t < T$, a player will not stop without having obtained at least one draw (of L or H) and will not continue if he has already obtained a draw of H. Thus, in the sequel, while analyzing the players' continuation and stopping incentives, we can restrict attention to a player who has already obtained a draw of L but no draw of H. We consider two cases, depending on whether both or only one of the players has received draws in the past.

First, suppose that by the time of the continuation or stopping decision in period t, each player has received at least one draw of L, that is, $n_t^i, n_t^j \geq 1$. Let $n_t = n_t^i + n_t^j$ be the total number of L draws the two players have obtained. Suppose also that player j stops experimentation if and only if $n_t \geq n$, for some threshold value $n \geq 1$. Then, for $n_t < n$, player i's value in period t is

$$\begin{aligned} V_t^i(n_t) \;=\; & \max\{ L, \delta\, [\, p^H(n_t)(H/2) + r^2(p(n_t)(1-q)^2 + (1-p(n_t)))V_{t+1}^i(n_t+2) \\ & + 2r(1-r)(1-p(n_t)q)V_{t+1}^i(n_t+1) + (1-r)^2 V_{t+1}^i(n_t)\,]\, \} \end{aligned} \tag{5}$$

To understand this value function, note that player i can receive L by preempting the other player in the current period. If player i continues experimenting, then, with probability $p^H(n_t)$, one or both players receive an H draw; in this case, since the game is symmetric, player i receives an expected payoff of $H/2$. With probability $r^2[p(n_t)(1-q)^2 + (1-p(n_t))]$, both players receive L draws, for a continuation value $V_{t+1}^i(n_t+2)$. In addition, with probability $2r(1-r)(1-p(n_t)q)$, one player receives an L draw and the other player does not receive any draw, for a continuation value $V_{t+1}^i(n_t+1)$. Finally, with probability $(1-r)^2$, neither player receives a draw, for a continuation payoff $V_{t+1}^i(n_t)$.

At the threshold, for $n_t = n - 1$, we have $V^i_{t+1}(n_t + 1) = V^i_{t+1}(n_t + 2) = L/2$ and $V^i_t(n_t) = V^i_{t+1}(n_t)$ so that player i's value function becomes

$$V^i_t(n_t) = \delta p^H(n_t)\frac{H-L}{2} + \delta r(1-\frac{r}{2})L + \delta(1-r)^2 V^i_t(n_t)$$

Hence,

$$V^i_t(n_t) = \frac{\delta p^H(n_t)\frac{H-L}{2} + \delta r(1-\frac{r}{2})L}{1-\delta(1-r)^2},$$

which represents player i's continuation payoff when the two players have obtained a total of n_t draws, and player j will stop as soon as another draw occurs.

Notice that the last expression is independent of the player identity i and the time t; it depends only the total number of draws n_t. Since the belief $p\,(n)$ is decreasing in n and goes to 0 as $n \to \infty$, player i's expected gain is also decreasing in the number n_t, with limit equal to $[\delta r(1-\frac{r}{2}))/(1-\delta(1-r)^2]L < L$. Thus, by requiring that $V^i_t(n_t) > L$, player i's preemption value, we obtain the threshold number of draws

$$N_1 = \min\left\{ n \geq 2 : \frac{\delta p^H(n)}{2}\,(H-L) < \left(1 - \frac{\delta}{2}\left[(1-r)^2 + 1\right]\right) L \right\} \tag{6}$$

If the total number of L draws the two players have obtained is $n_t < N_1$, then a player will prefer to continue experimenting, given that his opponent plans to continue experimenting for at least one more period. Clearly, the threshold N_1 can only be reached in periods $t \geq T_1 = (1/2)N_1$. Prior to time T_1, independently of the number of L draws obtained, the two players will not have any incentive to preempt one another. In particular, if $T_1 \geq T$, the players will not stop prior to the final period T unless they receive a draw of H.

Second, suppose only a single player, say i, has received all draws obtained up to time t. In this case, player i's value, denoted by $V^i_t(n_t, 0)$, is

$$\begin{aligned} V^i_t(n_t,0) = \max\{L,\ \delta\,[p^H(n_t)\frac{H}{2} &+ r^2(p(n_t)(1-q)^2 + (1-p(n_t)))V^i_{t+1}(n_t+2) \\ + r(1-r)(1-p(n_t)q)V^i_{t+1}(n_t+1,0)\ &+\ r(1-r)(1-p(n_t)q)V^i_{t+1}(n_t+1) \\ &+ (1-r)^2 V^i_{t+1}(n_t,0)]\} \end{aligned}$$

Again, at the threshold at which either player stops when another draw occurs, we have $V^i_{t+1}(n_t+2) = V^i_{t+1}(n_t+1) = L/2$, $V^i_{t+1}(n_t+1, 0) = L$ and $V^i_t(n_t, 0) = V^i_{t+1}(n_t, 0)$ so that the above value function becomes

$$\begin{aligned} V^i_t(n_t, 0) &= \delta p^H(n_t)\frac{H-L}{2} + \delta\left[r(1-\frac{r}{2}) + \frac{r}{2}(1-r)(1-p(n_t)q)\right] L \\ &+ \delta(1-r)^2 V^i_t(n_t, 0) \end{aligned}$$

Hence,

$$V^i_t(n_t, 0) = \frac{\delta p^H(n_t)\frac{H-L}{2} + \delta\left[r(1-\frac{r}{2}) + \frac{r}{2}(1-r)(1-p(n_t)q)\right] L}{1-\delta(1-r)^2}$$

which represents player i's continuation payoff when he has obtained a total of n_t draws, player j has obtained no draw, and the two players will stop as soon as another draw occurs. The extra term in front of L expresses the additional payoff that player i will receive if he stops with a value of L and player j receives no draw in period $t+1$.

For parameters $H/L < (3-2rq)/(2-rq)$, we have $V^i_t(n_t, 0) < \delta\, L$, for all $n_t \geq 1$, so that experimentation ends after the first draw. Otherwise, for $H/L \geq (3-2rq)/(2-rq)$, it is easy to check that the last expression for $V^i_t(n_t, 0)$ is decreasing in n_t; as $p(n)$ goes to zero, this expression approaches a limit that is less than L. Thus, in a manner analogous to N_1, by requiring that $V^i_t(n_t, 0) > L$ we can define the threshold number of draws

$$N_2 = \min\Big\{n \geq 1 : \tfrac{\delta p^H(n)}{2}\,(H-L) - \tfrac{\delta p(n) q r}{2}\,(1-r)\,L < \left(1-\delta\left(1-\tfrac{r}{2}\right)L\right)\Big\} \quad (7)$$

That is, when $n_t < N_2$, a player in such a situation will have no incentives to abandon experimentation, given that his opponent plans to continue experimenting for at least one more period. The earliest time that this threshold can be reached is $T_2 = N_2$. It is easy to see that $N_1 \leq N_2$, as the argument requires, since a player's incentive to continue experimenting is stronger, given the same amount of information, when his opponent is less likely to stop.

Consider the *threshold strategy* $\sigma^* = \{\sigma_t^*\}_{t=1}^{T-1}$, prescribing to player i the following behavior in each period t:

- Player i stops in period t, if

  a. Player i has drawn H in some period $t' \leq t$; or

  b. Player j has received a draw in some period $t' \leq t$, and $n_t^i + n_t^j \geq N_1$; or

  c. Player j has received no draw in periods $t' \leq t$, and $n_t^i \geq N_2$.

- Otherwise, player i continues.

Clearly, the strategy σ^* is fully characterized by the thresholds N_1 and N_2, which remain constant over time.

Proposition 1 *The strategy profile* (σ^*, σ^*) *constitutes a perfect Bayesian equilibrium.*[9]

The equilibrium has a simple structure. The players remain in the game prior to the final period T until either a draw of high value occurs or their beliefs about the possibility of such a draw become too pessimistic. Since the players share common beliefs about the potential of the project, the latter event occurs when the *total* number of low-value draws exceeds a certain threshold. Consequently, in equilibrium, the players decide to stop simultaneously.

The game admits other equilibria in which the players stop experimenting after obtaining a total of $N' < N_1$ draws of L or after reaching a certain time T', where N' and T' are exogenously set. To see this, note that in such equilibria, because of the possibility of preemption, each player's decision to stop experimentation earlier forces his rival also to stop. However, it is interesting to note that experimentation resulting in more than N_1 or N_2 draws of L turns out to be impossible.

[9] All proofs are in Appendix B.

Corollary 1 *There exists no perfect Bayesian equilibrium involving experimentation that can generate more draws than the strategy σ^*.*

Comparing the single-player problem to the two-player one, we obtain we following result:

Corollary 2 *The maximal experimentation duration is longer in the case of one player than in any perfect Bayesian equilibrium of the two-player case.*

The corollary states that a single agent experiments longer than an agent facing competition, even if the latter has received all draws that have been obtained so far. Thus, the threat of preemption leads to a decrease in the total amount of experimentation. Since the one-player problem, adjusted for the duplication of experiments of two players, is equivalent to the social planner's problem, we conclude that, in the two-player case, experimentation terminates too early from a welfare point of view.[10]

5 Private Learning

We now turn our attention to the case in which the two players cannot observe one another's experimental outcomes. Instead, in each period, each player has to form beliefs about the draws of his opponent, depending on the duration of experimentation, the stopping strategy his opponent has been using, and significantly, the draws he has received himself. Naturally, these beliefs affect the two players' continuation or stopping incentives, via their calculations about the likelihood of an H outcome as well as about the possibility that the other player stops in the current or next period.

In general, the beliefs of player i at time t take the form of a probability distribution over the feasible histories of the game, in particular, over the history components that are privately observed by player j. In analyzing the stopping decision of player i in period t, when he has received no draw of H, we can assume that player j has

[10]See the definition of N^* in Section 3.

received no draw of H either. Consequently, the beliefs of player i reduce to a probability distribution over the number of L draws, n_t^j, that player j has received up to period t.[11]

Since the probability of drawing L depends on the distribution from which the two players draw, player i's beliefs about n_t^j need to take into account his own private information, that is, the number n_t^i of L draws he has received.[12] In addition, player i needs to condition his beliefs upon any information he can infer from player j's decisions not to stop in any earlier period, in connection to the strategy s^j.[13] The following result shows that the players' beliefs are positively correlated, that is, each player's beliefs about the draws of his opponent stochastically increase in the number of his own draws.

Lemma 1 *Suppose that player j follows the strategy s^j and that player i has obtained $n_t^i = n^i$ draws of L by period t. Then, at the end of period t, conditionally on player j having received no draw of H, player i believes that $n_t^j = n^j$ with probability*

$$p_t(n^j, n^i, s^j) \;=\; \frac{h_t(n^j, s^j)\, r^{n^j}\,(1-r)^{t-n^j}\,[p(1-q)^{n^i+n^j} + (1-p)]}{\sum_{n=0}^{t} h_t(n, s^j)\, r^n\,(1-r)^{t-n}\,[p(1-q)^{n^i+n} + (1-p)]},$$

where $h_t(n^j, s^j) \leq \binom{t}{n^j}$ is the number of histories of player j consistent with $n_t^j = n^j$, the stopping constraints of strategy s^j, and the hypothesis that no draw of H has occurred.

In addition, for any $\tilde{n}^i > n^i$, the distribution $p_t(\,\cdot, \tilde{n}^i, s^j)$ first-order stochastically dominates the distribution $p_t(\,\cdot, n^i, s^j)$.

---

[11] As Lemma 3 below will show, the timing of the players' draw arrivals is irrelevant in equilibrium.

[12] For example, with a parameter $q \approx 1$, at the end of period $t = 0$, player i believes that H is feasible with probability approximately equal to p or 0, if, respectively, $n_t^i = 0$ or $n_t^i = 1$. Consequently, he believes that $n_t^j = 1$ with probability approximately equal to $(1-p)\,r$ or r, depending on whether $n_t^i = 0$ or $n_t^i = 1$.

[13] In particular, if player j follows a strategy s_j characterized by stopping thresholds $\{N_t^j\}_{t=0}^{T-1}$, then player i will condition his beliefs at period t upon $n_{t'}^j < N_{t'}^j$, for all $t' < t$.

We construct equilibria in symmetric threshold strategies, that is, each player stops in period t if either he obtains a draw of H or the number of L draws he has received exceeds a certain threshold N_t, depending on that period.

For such strategies, each player's beliefs are stochastically increasing in each threshold of his opponent.

Lemma 2 *Let s^j and $\hat{s}^j$ be two threshold strategies for player j such that $N^j_\tau \leq \hat{N}^j_\tau$ for all $\tau < t$. Then, for all n^i_t, the distribution $p_t(\cdot, n^i_t, \hat{s}^j)$ describing player i's beliefs about n^j_t at time t, conditional on player j having received no draw of H, first-order stochastically dominates the distribution $p_t(\cdot, n^i_t, s^j)$.*

Therefore, under private learning, the problem of calculating a player's best response to a stopping strategy with decreasing thresholds is no longer a monotone decreasing one. For example, if player j's thresholds in periods $t-1$ and t are $N^j_{t-1} > N^j_t$, then at the end of period $t+1$, player i updates his beliefs about n^j_{t+1} in a manner that can make H more likely to be feasible and stopping by player j less likely to occur in that period.[14] Consequently, player i's expected payoff from continuing to the next period may increase, despite the decrease in player j's threshold; thus, the methods used in the case of common learning are no longer applicable. Instead, we calculate the players' best-response strategies by proceeding backwards from period T.

To guarantee the existence of an equilibrium in nontrivial symmetric strategies[15], we introduce Condition 1 upon the parameters of the model, presented in Appendix A.

[14] As a simple example, suppose at $t = 1$, player j always continues. At $t = 2$, player i may stop with one L believing that player j might have received an L draw, which makes player i pessimistic that H is feasible. Now, suppose instead that at $t = 1$ player j stops with one L. At $t = 2$, player i may now continue with one L because player i knows player j has not received a draw (or otherwise she would have stopped) so player i is more optimistic that H is feasible.

[15] Trivially, there is always an equilibrium in which each player stops in each period, independently of the draws he has received.

This implies that player i's best-response cutoff at time t is monotonically increasing in player j's cutoff N_t^j, for any $t < T$. At $t = T-1$, Condition 1 simplifies to:

$$\delta\left[p(2T)\left[1-(1-rq)^2\right](H-L)+L\right] \;\geq\; L$$

To understand the condition, suppose that player j has n_{T-1}^j draws of L and switches from a strategy $s_{T-1}(n_{T-1}^j)$ of stopping in period $T-1$ to a strategy $\hat{s}_{T-1}(n_{T-1}^j)$ of continuing in period T, with all other elements of his strategy remaining the same. Consequently, player i's payoff calculations involve a lower probability of player j stopping in period $T-1$ but also a lower expected payoff from experimentation, conditional on the game reaching period T, because of more pessimistic beliefs. Condition 1 implies that player i's gain from the switch in player j's strategy is greater when he follows a strategy $\hat{s}_{T-1}(n_{T-1}^i)$ of continuing than when he follows a strategy $s_{T-1}(n_{T-1}^i)$ of stopping at the end of period $T-1$, for all n_{T-1}^j and n_{T-1}^i; eventually it allows player i's best-response strategy in period $T-1$ to be monotonically increasing in the threshold N_{T-1}^j of player j in period $T-1$.

More generally, in any period $t < T$, suppose that player j has n_t^j draws of L and changes his strategy at time t from stopping to continuing and his continuation strategy from $\{s_\tau^j\}_{\tau=t+1}^{T-1}$ to $\{\hat{s}_\tau^j\}_{\tau=t+1}^{T-1}$.[16] Then player i's calculations about the benefits of further experimentation should involve not only more pessimistic beliefs, if the game reaches period $t+1$, but also a potential loss from the change in player j's continuation strategy. Condition 1 requires that even under the worst-case scenario about the switch $\{s_\tau^j\}_{\tau>t}$ to $\{\hat{s}_\tau^j\}_{\tau>t}$, player i will benefit more from the change in player j's strategy if he continues at time t rather than if he stops.

Although Condition 1 is stronger than necessary, when it fails, a symmetric equilibrium may not exist even for short time horizons. For example, when $\delta = 0.9$, $p = 0.8$, $q = 0.9$, $H = 8$, and $L = 1$, and two-periods ($T = 2$), each player's strategy reduces to deciding whether to stop or to continue with one draw of L at the end of

[16] In period $T-1$, a change in player j's continuation strategy is not possible.

period $t = 1$. If $r \in (0.237, 0.242)$, then each player is better off stopping against an opponent who continues and continuing against an opponent who stops; therefore, there is no symmetric equilibrium.[17]

Lemma 3 *For any $T \in \mathbb{Z}^+$, if Condition 1 holds, then each player i's best response to any threshold strategy $\{N_t^j\}_{t=1}^{T-1}$ of player j is also a threshold strategy $\{N_t^i\}_{t=1}^{T-1}$.*

The mutual optimality of the threshold strategies is rather intuitive. With a higher number of L draws, player i becomes less willing to continue experimentation, for three reasons. First, independently of his opponent's presence, the extra draws of L have a negative effect upon player i's beliefs regarding the feasibility of H. Second, with another player experimenting in parallel, player i's pessimism about H is reinforced by the knowledge that the other player has not succeeded either; independently of any preemption threat, in particular, when player j will not stop unless he obtains H, player i's pessimism increases at a higher rate when he has received a higher number of L draws.[18] Third, considering also the opponent's stopping strategy, player i 's fear of being preempted by the other player increases with each additional draw of L that he receives. In total, since the draws of L have only negative effects upon a player's expectations and payoffs, if player i is better off stopping with a certain number of L draws, then he will be better off stopping also with any higher number of such draws.

Suppose that player j follows a strategy σ_j characterized by thresholds $\{N_t^j\}_{t=1}^{T-1}$. Then, at the end of each period t, player i's expected gain from continuing to period $t+1$ (and subsequently using his optimal continuation strategy) rather than stopping

[17] In this example, q takes a relatively high value so that player j's decision to continue with one draw of L has a relatively large negative effect upon player i's beliefs about the feasibility of H, conditional on the game reaching period T.

[18] It is straightforward to calculate the probability that H is feasible, conditionally on n_t^i draws of L for player i and no draw of H for player j, and to show that the rate at which this probability decreases in the the experimentation duration t is increasing in n_t^i.

at period t, when he has obtained n_t^i draws of L, is

$$\Delta V_t \;=\; \Delta V_t(n_t^i \,|\, \sigma^j)$$

defined recursively by equations (B.1)–(B.6) in the proof of Lemma 3, with player i's beliefs about player j's draws being the ones induced from strategy σ^j via Lemma 1.

For any $T \in \mathbb{Z}^+$, a strategy σ with thresholds $\{N_t\}_{t=1}^{T-1}$ will be part of a symmetric equilibrium if and only if in each period $t < T$, we have

$$\Delta V_t(n_t^i \,|\, \sigma) \begin{cases} > 0, & \text{if } n_t^i < N_t; \\ \leq 0, & \text{if } n_t^i \geq N_t. \end{cases}$$

The following results asserts that such a symmetric equilibrium exists.

Proposition 2 *For any $T \in \mathbb{Z}^+$, if Condition 1 holds, then there exists a symmetric perfect Bayesian equilibrium in threshold strategies $\{N_t\}_{t=1}^{T-1}$.*

To describe the way the thresholds N_t are determined, consider a player who has received $n_t^i = N$ draws of L by period t and who knows that his opponent will stop in that period if and only if he has also obtained $n_t^j \geq N_t^j = N$ draws of L. An increase in the number N has two effects upon the continuation incentives of that player: a positive one, stemming from the increase in N_t^j and the higher probability that his opponent will continue to the next period; and a negative one, stemming from the increase in n_t^i and the lower probability that H is feasible. As N increases, the second effect becomes more important. Eventually, either it comes to dominate the first effect, for a threshold $N_t \leq t+1$, or the two players choose always to continue experimenting for at least one more period.

6 Comparison of Common and Private Learning

In this section, we compare the duration of experimentation and the players' total welfare under common and private learning regimes.[19] Our results indicate that com-

[19] As we show welfare is typically but not always higher with longer experimentation.

mon learning generates more experimentation when q is either low or high, r is low and p is low. Private learning, on the other hand, generates more experimentation when q is intermediate, and r and p are high. In scientific research, often there is a great deal of uncertainty about the feasibility of a breakthrough (low p); breakthroughs are rare even when they are feasible (low q); and experiments frequently fail to produce results (low r). Hence, our findings suggest that common learning would generate more experimentation than would private learning in scientific research.

Throughout this section, when there are multiple equilibria, we focus on the highest welfare equilibrium. As we noted before, the optimal experimentation duration and welfare are equal under both regimes when failed experiments are not possible. This is because when $r = 1$, under private learning, in each period, each player knows with certainty the number of L draws his opponent has received.

When the arrival of draws is uncertain, i.e., for $r < 1$, the comparison becomes interesting since common and private learning are no longer equivalent. We first compare the two regimes for the two-period case where we can solve equilibria in closed form. To compare the two observability regimes analytically, in the case of two periods, We provide a complete analytical characterization in the next proposition.

Proposition 3 *Suppose $T = 2$, and*

$$p(2)\,[1-(1-rq)^2]\;\frac{H-L}{L}\;\;\geq\;\;\frac{1-\delta}{\delta} \tag{8}$$

holds.[20] *Then the comparison of the most efficient equilibria under common and private learning depends on the following three conditions:*

$$p(2)\,[1-(1-rq)^2]\;\frac{H-L}{L} \geq \frac{2-\delta}{\delta} \tag{9}$$

$$p(1)\,[1-(1-rq)^2]\;\frac{H-L}{L}\;+\;(1-r)[1-p(1)\,rq] < \frac{2-\delta}{\delta} \tag{10}$$

[20]This is Condition 1 in Appendix A adapted to the case of two periods; it is sufficient for equilibrium existence under private learning.

$$[p_0(0,1)\,p(1) + (1 - p_0(0,1))\,p(2)]\;[1 - (1 - rq)^2]\;\frac{H - L}{L} + \; p_0(0,1)\,(1 - r)\,[1 - p(1)\,rq] \;<\; \frac{2 - \delta}{\delta} \qquad (11)$$

where $p(\cdot)$ *and* $p_1(0,1)$ *are defined respectively by equation* (1) *and Lemma 1.*[21]

a. *If condition* (9) *holds, then common and private learning result in the same outcomes and payoffs, with each player continuing to period* $T = 2$ *unless he receives* H*.*

b. *If condition* (10) *holds, then common and private learning result in the same outcomes and payoffs, with each player stopping as soon as he receives a draw.*

c. *Otherwise, if conditions* (9) *and* (10) *do not hold, under common learning, the two players stop in period* $t = 1$*, if they both receive a draw of* L*; else, they continue to period* $T = 2$*. In this case, common learning generates more experimentation than private learning if and only if condition* (11) *holds. In addition, under condition* (11)*, common learning results in higher expected payoffs.*

Figure 1 illustrates the analytical result in the two-period case, for parameters $\delta = 0.9$, $H = 8$, $L = 1$, and $p = 0.6$. To understand the figure, let us first review possible optimal strategies under the two regimes. Since players always stop with a draw of H, when describing the various cases, we ignore this possibility. In each case, we indicate in parentheses the combinations of parameters q and r in Figure 1 for which it arises in equilibrium.

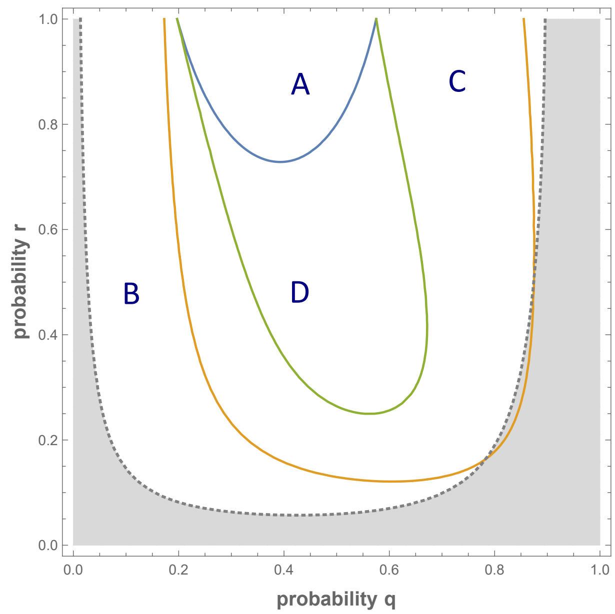
Under private learning, there are two cases:

Case PL1: Players always continue (in areas A and D).

Case PL2: Each player stops with a draw of L (in areas B and C).

[21] Since $p(1) > p(2)$, the LHS in condition (9) is smaller than the LHS in condition (10), so that the two inequalities cannot hold simultaneously. In addition, the LHS in condition (11) is a convex combination of the LHS in conditions (9) and (10), weighted according to the players' beliefs $p_1(0,1)$ and $p_1(1,1)$; therefore, condition (11) must fail/hold when condition (9)/(10) holds.

Figure 1: Parameter values are set at $\delta = 0.9$, $H = 8$, $L = 1$, and $p = 0.6$, $T = 2$. Condition (8), ensuring equilibrium existence under private learning, holds in the white area.



Under common learning there are three cases:

Case CL1: Players always continue (in area A).

Case CL2: Players stop if they both receive draws of L; otherwise, they continue (in areas C and D).

Case CL3: Either player stops with a draw of L regardless of the other player's draw (in area B).

Hence, in areas A or B, the length of experimentation is the same under both regimes; in area D, private learning generates more experimentation than common learning; and in area C, common learning generates more experimentation than private learning.[22]

[22] As shown in the proof of Proposition 3, when condition (11) fails, in the areas A+D in Figure 1, under private learning, an equilibrium exists even if condition (8) fails, with each player continuing to $T = 2$ unless he receives H in period $t = 1$. It is only for parameters for which (11) holds, in the

Since private learning softens the threat of preemption, it can lead to more experimentation for a range of parameters. Using Figure 1, we see that private learning generates more experimentation than common learning when q has intermediate values and r is high enough. However, despite the possibility of preemption, common learning generates more experimentation than private learning when q is either low or high enough. To see why this is the case, note that, under private learning, when q is low, players stop with a single L because they believe that obtaining H with the next draw is very unlikely. On the other hand, when q is high, obtaining an L leads players to update their beliefs drastically and believe that a breakthrough is not feasible (because if it were, they would have received an H with high probability given that q is high). This leads them to stop immediately. Under common learning, however, there is a range for the parameter q in which players would continue with a single L and stop only if they observe two Ls. Hence, when q is in this range, common learning generates more experimentation. Put differently, independent learning leads to coordination failures when players stop with a single L under private learning but continue with a single L and stop if they both receive Ls under common learning. In addition, such coordination failures become more likely when r is low. Indeed, for low values of r, common learning dominates private learning for all values of q.

For $T = 2$, if it is more likely that a breakthrough is feasible, softening preemption becomes more important, and private learning generates more experimentation for a wider set of parameters. Graphically, in Figure 1, as p increases, areas B+C contract, while areas A+D expand. We state this formally in the next proposition.

Proposition 4 *The set of probability parameters q and r such that common learning generates more experimentation and higher payoffs than private learning is decreasing with respect to set inclusion $\subseteq$ in the probability p.*

When common learning generates more experimentation, it necessarily results in higher welfare, as it is closer to the single-player optimum. It is interesting to notice,

areas B+C, that an equilibrium may not exist without condition (8) being satisfied.

though, that common learning can result in higher welfare even in cases in which it generates less experimentation, if conditions (9)-(11) do not hold, where the solution to the cooperative problem is to experiment until obtaining $N^* = 2$ draws of L. For such parameters, the failure under private learning to aggregate the two players' information may result in excessive experimentation.[23]

For more than two periods, an analytical comparison of common and private learning is complicated because of the large number of cases that need to be considered. Instead, we provide numerical examples showing that the conclusions from the two-period case are robust to increasing the number of periods. Figure 2 shows the length of experimentation and welfare as q varies from 0.05 to 1 in increments of 0.05 for $T = 5$. We see that the conclusions from $T = 2$ extend to more periods as private learning produces longer experimentation only for intermediate values of q (specifically, when $q \approx 0.6$).

Figure 2 demonstrates that the length of experimentation under both private and common learning is non-monotone in q. For values near $q \approx 0$, the probability of obtaining H is too small, even for high values of the parameters p and r; thus, the two players stop experimenting as soon as they can claim a value of L, without incurring any experimentation cost. As q increases, on the one hand, the likelihood of a successful draw increases, if H is indeed feasible; on the other hand, the players' beliefs about the feasibility of H decrease at a faster rate with each unsuccessful draw. For intermediate values of q, the first effect dominates so that the corresponding equilib-

[23]For an illustration, let $\delta = 0.9$, $H = 8$, $L = 1$, $p = 0.9$, $q = 0.9$, $r = 0.1$, and $T = 2$. Then, a single player will keep experimenting until he obtains 2 draws of L. Under common learning, the two players will stop at $t = 1$ if and only if they both obtain L draws; thus, the equilibrium achieves the single-player optimal experimentation outcome. Under private learning, each player will continue to period $T = 2$ with either 0 or 1 draw of L; thus, in equilibrium, it is possible that the two players continue experimenting beyond the single-player optimal stopping threshold. The expected duration/payoff of experimentation for a single player and for common learning is lower/higher than the expected duration/payoff under private learning.

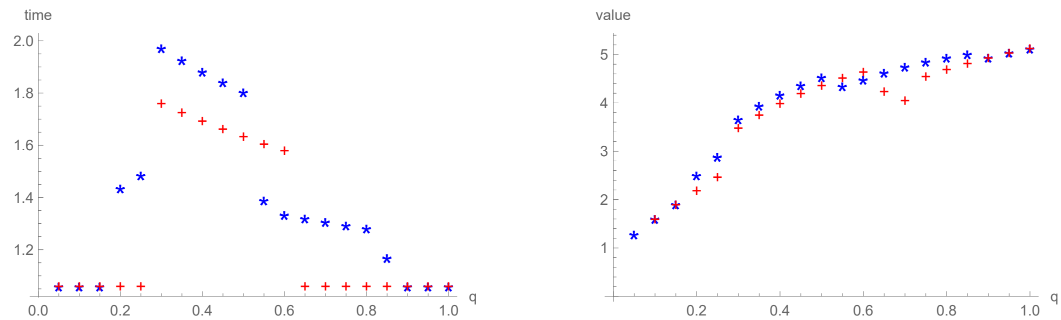


Figure 2: Experimentation length and value for common ($*$) and private ($+$) learning when $\delta = 0.9$, $H = 8$, $L = 1$, $r = 0.75$, and $p = 0.6$, $T = 5$.

ria achieve the greatest amount of experimentation. Eventually, however, the second effect becomes more important so that the players start adopting tighter thresholds and experimenting less. At the extreme, for values near $q \approx 1$, the players stop experimenting after the first draw since a single draw of L suffices for their beliefs to become too pessimistic.[24]

Surprisingly, although higher values of q correspond to experiments that are more likely to result in H, when it is feasible, the players' expected payoffs are not monotonically increasing in q. As Figure 2 illustrates, for both common and private learning, since increasing q eventually results in equilibrium strategies that involve less experimentation, there are regions where the players' expected payoffs are decreasing in q.

Figure 3 shows the length of experimentation and the players' welfare, as r varies from 0.05 to 1, in increments of 0.05, for $T = 5$. Once again, the conclusions from $T = 2$ extend to more periods as private learning produces longer experimentation for relatively large values of r, and the outcomes converge when r is close to 1.

[24]It should be noted that the effect of q upon the players' experimentation value is less clear, as higher values of q imply a higher probability of obtaining H for the same experimentation strategies.

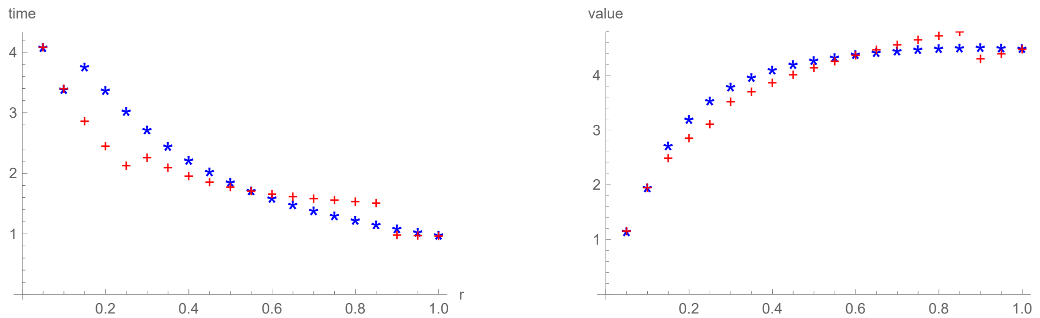


Figure 3: Experimentation length and value for common ($*$) and private ($+$) learning, when $\delta = 0.9$, $H = 8$, $L = 1$, $p = 0.6$ $q = 0.6$, $T = 5$.

The setting of the last example allows us to consider the way that information sharing affects the two players' preemption motives. As the value of r increases, the probability of successful experimentation in the next period increases for the same beliefs about the feasibility of H. Under common learning, this is the only effect upon the players' payoff calculations, and for values of r that are not too low (so that a player would prefer to stop as soon as he can claim L), the two players stop when their beliefs about H drop too much relative to their stopping value, that is, when they obtain 2 draws of L in total. In particular, each player knows how close his opponent is to terminating experimentation. Under private learning, however, the increase in the value of r has two adverse effects, stemming from the increase in each player's belief about the number of L draws of his opponent. First, each player's belief in the feasibility of H decreases; second, for any threshold strategy, each player thinks that his opponent is closer to stopping and preempting him. For very low values of r, because of the difficulty of obtaining another draw, each player stops after the first draw. In the opposite case, for very high values of r, in each period, each player is sufficiently sure that his opponent has received a draw, so again, he stops with one draw, with his payoff calculations approximating those under common learning. In between, for intermediate values of r, in the first period, each player is always willing to continue to the next period since the probability that the other player has received a draw is not too high for weaker preemption motives in comparison to the situation with common learning. However, this calculation leads to the reverse conclusion in the second period, with each player stopping if he can claim L, since the probability of the other player having obtained a draw increases for stronger preemption motives. Eventually, as the various effects operate in opposite directions, a general comparison over the entire time horizon is not feasible, but as the previous example indicates, any efficiency gains from softer preemption motives under private learning are rather limited.

7 Literature Review

Our paper is related to two bodies of work where the distinction between the two is the possibility of preemption. In preemption games, players decide when to terminate the game, given a first-mover advantage in the payoffs. They can seek to obtain a larger prize by moving late but also have the opportunity to accept a smaller prize, and by doing so, they prevent all others from obtaining any prize at all.

The first body of work features preemption in the sense that we just described, but does not deal with uncertainty and learning about research outcomes. Hopenhayn and Squintani (2011) consider a preemption game in which two players randomly receive new information over time, interpreted as innovation increments.[25] They find that private information about each player's state tends to soften the fear of being preempted, resulting in longer expected durations in equilibrium, which is in contrast to our findings. The main element differentiating our setting from that of Hopenhayn and Squintani is the presence of uncertainty about the potential of experimentation. In our model, the players draw from an unknown distribution, essentially experimenting with a multiarmed bandit, unlike in Hopenhayn and Squintani, where the players accumulate outcomes from a known distribution. Thus, in our problem there are

[25] A similar model has been introduced by Lippman and Mamer (1993).

gains from sharing information regarding the draws that the players obtain, in terms of learning about the unknown distribution, that are not present in the Hopenhayn and Squintani model. As a consequence, common learning can lead to more efficient outcomes than private learning in our model in contrast to Hopenhayn and Squintani. Bobtcheff et al. (2017) consider preemption in a model where two researchers privately have a breakthrough idea and decide how long to let their ideas mature before disclosing them. In their model, the arrival time of a breakthrough is a random variable, but its value and the returns to maturation are known with certainty. In our model, the feasibility of a high-value breakthrough is uncertain. This distinction matters for learning dynamics and welfare. Here, researchers learn about both the threat of being preempted and the project's potential value, and there are gains from sharing this information that have no counterpart in their setup. Other preemption games in the context of research activity are investigated, for instance, by Hoppe and Lehmann-Grube (2005) and Bobtcheff and Mariotti (2012). Like Bobtcheff et al. (2017), these studies consider preemption under deterministic payoffs.[26]

The other body of work deals with experimentation and learning in stopping games without the threat of preemption as in the multi-armed bandit models (for a recent survey of this literature, see Hörner and Skrzypacz (2016).[27] Keller et al. (2005), Keller and Rady (2010), and Klein and Rady (2011), for instance, examine two-armed bandit models in which players must allocate resources to a risky project and a safe option.[28] The risky project is characterized by uncertainty about the

[26]Boyarchenko and Levendorskii (2014) examine a stochastic version of Fudenberg and Tirole's (1985) preemption game, but learning about an uncertain distribution is not an issue. For an early study of the timing of innovations under rivalry, see, e.g., Kamien and Schwartz (1972).

[27]In our setting, the stopping and continuation decisions correspond, respectively, to settling for a sure arm and trying a stochastic arm. Note that in our model, a player's stopping decision affects the value of both arms for the other player.

[28]Private information about the realized rewards and the possibility of communication via cheap talk is considered by Heidhues et al. (2015) in a multi-armed bandit model.

arrival rate of rewards. Players learn about this arrival rate over time by observing each other's actions and rewards. However, there is no advantage from disclosing an experimentation result ahead of the opponent, which is exactly the opposite of what is assumed in our paper. Akcigit and Liu (2015) also study a two-armed bandit problem. In their paper, two players begin experimenting with a risky arm that results in either a good outcome or a dead end. At any point, a player can privately and irreversibly switch to a safe arm. A good outcome from the risky arm is public, but a dead end is observed in private. Only a single player can obtain a reward from a given arm. The paper identifies the channels for inefficient experimentation. Aside from the different focus, the key difference between our paper and that of Akcigit and Liu (2015) stems from the lack of preemption in their framework. Without the threat of preemption, information sharing is always superior to private experimentation. A second more technical difference is that the evolution of beliefs is simpler in their setting since, for any strategies, players can only become pessimistic over time, a property that is not present in our problem.

Moscarini and Squintani (2010) consider a two-player model of experimentation with private information and learning about the arrival rate of an invention.[29] At each point in time, each player decides whether to stop experimentation or not. When a player stops before an innovation arrives, he earns nothing. As a consequence, preemption is not possible.[30] By contrast, the possibility of preemption gives rise to different learning dynamics in our model: each player's beliefs regarding the position of his opponent are used to estimate not only the likelihood of achieving a high-value outcome but also the probability of being preempted with a low-value result.

Halac et al. (2017) study innovation contests when there is uncertainty about the

[29]Private signals in patent races have been introduced by Reinganum (1982). Choi (1991) considers a winner-take-all race in which participants have imperfect but symmetric information about the arrival rate of the R&D process.

[30]Also related is the two-armed bandit model of strategic experimentation with private information by Das (2017), who does not consider the possibility of preemption with a low-value outcome.

feasibility of a successful innovation. There is a principal who designs a contest to maximize the probability of obtaining a successful innovation and several researchers who engage in costly experimentation for a fixed number of periods. The principal allocates a fixed prize among the researchers and chooses a prize-sharing scheme and a disclosure policy. There are two possible prize-sharing schemes. In winner-take-all contests, the player who achieves a success first receives the entire prize. In equal-sharing contests, players who achieve success, regardless of the order of achieving it, split the prize equally. There are also two possible disclosure policies. In public contests, whenever a player achieves success all players are informed. In private (or hidden) contests, this information is only revealed after the contest. Note that the fact that a contest is winner-take-all does not mean that preemption is possible. This is because players cannot stop with anything less than a success, and even then, experimentation can continue after one of the players obtains success either because the contest is private or because there is equal sharing. In contrast, in our model, with or without information sharing, experimentation stops as soon as one of the players reveals either a low- or a high-value success. The key result in Halac et al. (2017) is that an equal-sharing private contest can be strictly better for the principal than any other contest. This result favoring private learning crucially depends on two factors, the sharing rule in the contest and the certainty about the amount of negative information obtained, conditional on lack of success in the past. By contrast, in our paper, optimality of no information sharing holds in winner-take-all contests and is driven by preemption.

One paper that falls within the intersection of the two bodies of literature, dealing with preemption and learning about uncertain research outcomes, is Spatt and Sterbenz (1985). The authors show that preemption shortens experimentation. There are two crucial differences from our paper. First, in every period, there is a single public draw, and second, there are no failed experiments. Thus, while there is no possibility of private learning in their setting, our paper compares private and common learning.

8 Conclusion

We have examined the effects of rivalry upon experimentation and learning in a stopping game in which the players acquire information over time about the distribution of their potential payoffs. A key innovation in our setting is that experiments are not always successful and sometimes do not return any useful results.

Under the assumption of public observation of the players' experimentation results, we have constructed a perfect Bayesian equilibrium in threshold strategies; the two players continue experimenting, trying to obtain a high-value outcome, until their beliefs about its feasibility become too pessimistic. Because of the possibility of preemption, the length of experimentation is shorter than socially optimal.

If the players cannot observe one another's results, i.e., under private learning, they need to form beliefs about the experimentation outcomes of their opponent and eventually about the feasibility of a high-value outcome. In our setting, these beliefs can be quite complex because they depend not only on the length of time the players have been experimenting but also on the number of successful experiments. Despite this complexity, we provide conditions for the existence of equilibria in strategies involving nonmonotone time-variant thresholds.

Information sharing is important since it can be influenced by policy, and our paper sheds light on which information-sharing regime, public or private, generates longer experimentation horizons and greater value for scientists. The received wisdom on this is that private learning generates longer experimentation horizons because it softens the preemption threat. Our simulations show that this intuition is incomplete, and common learning generates longer experimentation under a wide range of parameters. We trace this outcome to the players' inability to coordinate on their information under private learning. A player who does not observe his opponent's results and, due to unsuccessful experimentation, who does not himself have many results might still believe that his opponent has run many successful experiments and obtained more results. This situation would push the player to prematurely stop experimenting.

References

[1] Akcigit, U. and Liu, Q. (2015), The role of information in innovation and competition, *Journal of European Economic Association* 14, 828-870

[2] Arrow, K.J. (1969), Classification notes on the production and transmission of technological knowledge, *American Economic Review* 59, 29-35

[3] Barwich, A.-S. (2019), The value of failure in science: The story of grandmother cells in neuroscience, *Frontiers in Neuroscience* 13:1121. doi: 10.3389/fnins.2019.01121

[4] Bertsekas, D. (2001), *Dynamic Programming and Optimal Control*, 2nd Edition, Athena Scientific, Belmont, MA

[5] Bobtcheff, C., Bolte, J., and Mariotti, T. (2017), Researcher's Dilemma, *Review of Economic Studies* 84, 969-1014

[6] Bobtcheff, C. and Mariotti, T. (2012), Potential competition in preemption games, *Games and Economic Behavior* 75, 53-66

[7] Bonatti, A. and Hörner, J. (2011), Collaborating, *American Economic Review* 101, 632–63

[8] Bornmann, L., Ye, A. and Ye, F. (2018), Identifying landmark publications in the long run using field-normalized citation data, *Journal of Documentation* 74, 278–88

[9] Boyarchenko, S. and Levendorskii (2014), Preemption games and Lévy uncertainty, *Games and Economic Behavior* 88, 354-380

[10] Broad, W.J. (1981), The publishing game: Getting more for less, *Science* 211, 1137-1139

[11] Choi, J.P. (1991), Dynamic R&D competition under hazard rate uncertainty, *RAND Journal of Economics* 22, 596-610

[12] Das, K. (2017), Strategic experimentation with competition and private arrival of information, Working paper, University of Exeter

[13] Fang, F.C. and Casadevall, A. (2012), Intense competition among scientists has gotten out of hand, *Scientific American* 307(2), 13-17

[14] Fudenberg, D. and Tirole, J. (1985), Preemption and rent equalization in the adoption of new technology, *Review of Economic Studies* 52, 383-401

[15] Hagstrom, W.O. (1974), Competition in science, *American Sociological Review* 39, 1-18

[16] Halac, M., Kartik, N. and Liu, Q. (2017), Contests for experimentation, *Journal of Political Economy* 125, 1523-1569

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

[18] Hopenhayn, H.A. and Squintani, F. (2011), Preemption games with private information, *Review of Economic Studies* 78, 667-692

[19] Hoppe, H.C. and Lehmann-Grube, U. (2005), Innovation timing games: A general framework with applications, *Journal of Economic Theory* 121, 30-50.

[20] Hörner, J. and Skrzypacz. A. (2016), Learning, experimentation and information design, Working paper, Yale University

[21] Kamien, M. and Schwartz, N. (1972), Timing of innovations under rivalry, *Econometrica* 40, 43-60

[22] Keller, G. and Rady, S. (2010), Strategic experimentation with poisson bandits, *Theoretical Economics* 5, 275-311

[23] Keller, G., Rady, S. and Cripps, M. (2005), Strategic experimentation with exponential bandits, *Econometrica* 73, 39-68

[24] Klein, N. and Rady, S. (2011), Negatively correlated bandits, *Review of Economic Studies* 78, 693-732

[25] Lawrence, P.A. (2016) The last 50 years: mismeasurement and mismanagement are impeding scientific research, *Current Topics in Development Biology*, in press

[26] Lippman, S.A. and Mamer, J.W. (1993), Preemptive innovation, *Journal of Economic Theory* 61, 104-119

[27] Newman, M.E.J. (2009), The first-mover advantage in scientific publication, *EPL (Europhysics Letters)* 86, 68001

[28] Mohs, R.C. and Greig, N.H. (2017), Drug discovery and development: Role of basic biological research, *Alzheimers Dement (N Y)* 3, 651-657

[29] Moscarini, G. and Squintani, F. (2010), Competitive experimentation with private information: The survivor's curse, *Journal of Economic Theory* 145, 639-660

[30] Nature Publishing Group (2012), Must Try Harder, Editorial, *Nature* 483, 509

[31] Parkes, E. (2019), Scientific progress is built on failure, *Nature,* Comments and Opinion (10 January 2019) doi: 10.1038/d41586-019-00107-y

[32] Reinganum, J.F. (1982), A dynamic game of R and D: Patent protection and competitive behavior, *Econometrica* 50, 671-688

[33] Sabatier, M. and Chollet, B. (2017), Is there a first mover advantage in science? Pioneering behavior and scientific production in nanotechnology, *Research Policy* 46, 522-533

[34] Spatt, C.S. and Sterbenz, F.P. (1985), Learning, preemption, and the degree of rivalry, *RAND Journal of Economics* 16, 84-92

Appendix A: Statement of Condition 1

We use the following condition to show the existence of an equilibrium in non-trivial symmetric strategies.

Condition 1 *The parameters δ, r, p, q, H, L and T are such that*

$$p_t(N,1,\underline{s})\ \left[p(2N)\left[1-(1-rq)^2\right]\ -\ \frac{1-\delta}{\delta}\,\frac{L}{H-L}\right]\ \geq$$

$$\sum_{n<N} p_t(n,1,\underline{s})\ p(n+1)\ \left[(1-rq)^2\ -\ (1-rq)^{2(T-t)}\right],$$

for all $N \leq t+1$, for all $t < T$, where $\underline{s}$ is the strategy with thresholds $N_\tau = 1$, for $\tau < t-N$, and $N_\tau = \tau-(t-N)$, for $\tau \geq t-N$.

The strategy $\underline{s}$ in Condition 1 is "minimal" among the threshold strategies for which $n_t^j \geq N$ with positive probability; that is, if s^j is a threshold strategy such that $p_t(N\,|\,1,s^j) > 0$, then $N_\tau \geq \underline{N}_\tau$, for all $\tau < t$. Therefore, by Lemma 2, the inequality in Condition 1 extends to all such thresholds strategies s^j.

Appendix B: Proof of Results

Proof of Proposition 1:

Clearly a player will not stop experimenting without having obtained at least one draw; and that he will not continue experimenting after obtaining H, the maximal value which he can claim. So, it suffices to examine the incentives of a player to stop with a draw of L.

Suppose that player j follows the strategy σ^* described above. We show that player i is better off continuing if and only if the number of L draws that the two players have obtained by time t is $n_t < N_1$, for the case in which player j has received at least one draw in the past; and similarly, if and only if $n_t^i < N_2$, for the case in which only player i has received draws in the past. Because of the recursive definition of player i's continuation payoff, we proceed by means of induction on the number of draws.

First, in period t, for any $t < T$, suppose that player j has obtained at least one draw of L. When $n_t \geq N_1$, because of player j's decision to stop, player i is better off also stopping. Let $n_t = N_1 - 1$ and consider player i's payoff from continuing to period $t + 1$. As argued in the text, we have $V^i_{t+1}(n_t + 1) = V^i_{t+1}(n_t + 2) = L/2$ and $V^i_t(n_t) = V^i_{t+1}(n_t)$, so that player i's continuation payoff becomes

$$V^i_t(N_1 - 1) \; = \; \frac{\delta p^H(N_1 - 1)\frac{H-L}{2} + \delta r(1 - \frac{r}{2})L}{1 - \delta(1 - r)^2} \, ;$$

and by the definition of N_1, it follows that $V^i_t(N_1 - 1) \geq L$, so that player i is better off continuing to the next period. Similarly, for $n_t = N_1 - 2$, we have $V^i_{t+1}(n_t + 1) \geq L > L/2$, $V^i_{t+1}(n_t + 2) = L/2$ and $V^i_t(n_t) = V^i_{t+1}(n_t)$, so that player i's continuation payoff becomes

$$V^i_t(N_1 - 2) \; > \; \frac{\delta p^H(N_1 - 2)\frac{H-L}{2} + \delta r(1 - \frac{r}{2})L}{1 - \delta(1 - r)^2} \, ;$$

therefore, since the probability $p^H(\cdot)$ is decreasing, we have $V^i_t(N_1 - 2) > V^i_t(N_1 - 1) \geq L$, so that induction starts.

Now, let $n_t < N_1 - 2$, if feasible, and suppose that $V^i_{t+1}(n) \geq L$, for $n = n_t + 1, ..., N_1 - 1$, for the induction hypothesis. A straightforward replication of the argument for $n_t = N_1 - 2$ shows that

$$V^i_t(n_t) \; > \; \frac{\delta p^H(n_t)\frac{H-L}{2} + \delta r(1 - \frac{r}{2})L}{1 - \delta(1 - r)^2},$$

so that $V^i_t(n_t) > V^i_t(N_1 - 1) \geq L$, completing the induction.

Second, suppose that player j has obtained no draw up to period t, for $t < T$. As argued in the text, if $H/L < (3 - 2rq)/(2 - rq)$, we have $V^i_t(n_t, 0) < \delta\, L$, for all $n_t \geq 1$, so that $N_2 = 1$, i.e., experimentation ends after the first draw. Otherwise, for $H/L \geq (3 - 2rq)/(2 - rq)$, an inductive argument similar to that of the previous case shows that $V^i_t(n_t, 0) \geq L$, for all $n_t < N_2$, so that player i is better off continuing to the next period.

It remains to show that player i will stop with $n_t \geq N_2$ draws. In this case, notice that player i's optimal strategy is the solution to a multi-armed bandit problem, with state variable n_t^i, initial state N_2, random transitions determined by the arrival of new draws, with the game ending when either player j obtains a draw or H is obtained. Since the probability $p^H(n_t)$ is decreasing, player i's gain from continuing for exactly one more period

$$U_t^i(n_t) \;=\; \delta p^H(n_t)\frac{H-L}{2} + \delta\left[r(1-\frac{r}{2}) + \frac{r}{2}(1-r)(1-p(n_t)q)\right]L + \delta(1-r)^2 L$$

is also decreasing in n_t, so that this is the *deteriorating case* of that problem.[31] Therefore, as in the single-player case, player i's optimal strategy takes the form of a one-step policy, according to which player i shall stop experimenting if and only if $U_t^i(n_t) < L$, that is, when $n_t \geq N_2$.

Proof of Corollary 1:

In any equilibrium, if the game ends following a history in which both players have received draws and no draw of H has been obtained, then the two players must stop simultaneously; otherwise, the preempted player would be able to profit by deviating to stopping earlier. Therefore, for such histories, each player's incentives to continue or to stop experimentation are described by the inequality in the definition of the threshold N_1, showing that the two players will stop experimenting if the total number of draws reaches that threshold.

In addition, following histories in which player i has received all draws, experimentation will last the longest if the opponent does not stop prior to receiving at least one draw. In this case, if the total number of draws exceeds the threshold N_1, by our previous argument, player j will stop as soon as he receives his first draw. Therefore, player i's problem reduces to the one analyzed in the proof of Proposition 1, so that he will not continue experimenting after he obtains N_2 draws of L.

Proof of Corollary 2:

Comparing the inequalities in (4) and (7), defining the thresholds N^* and N_2, we

[31] See Bertsekas (2001), Vol. II, Section 1.5.

find that a player's gain from continuing experimenting for exactly one more period is larger when he is alone.

Proof of Lemma 1:

At the end of period t, consider the joint event in which the two players have observed respectively histories h_t^i and h_t^j involving n_t^i and n_t^j draws of L and no draw of H. The probability of this event is

$$P(h_t^i, h_t^j) \;=\; r^{n_t^i+n_t^j}\,(1-r)^{2t-n_t^i-n_t^j}\,[\,p(1-q)^{n_t^i+n_t^j}+(1-p)\,]$$

Aggregating over all time-t histories h_t^j involving n_t^j draws of L, no draw of H, and satisfying the continuation constraints of the strategy s^j for all periods up to time $t-1$, we get

$$P(h_t^i, n_t^j, s^j) \;=\; h_t(n_t^j, s_t^j)\; r^{n_t^i+n_t^j}\,(1-r)^{2t-n_t^i-n_t^j}\,[\,p(1-q)^{n_t^i+n_t^j}+(1-p)\,],$$

where $h_t(n_t^j, s_t^j) \leq \binom{t}{n_t^j}$ is the total number of such histories.

Therefore, player i's belief that $n_t^j = n^j$ is given by the conditional probability

$$\begin{aligned} p_t(n_t^j, n_t^i, s^j) \;&=\; P(n_t^j \mid h_t^i, s^j) = \frac{P(h_t^i, n_t^j, s^j)}{\sum_{n=0}^t P(h_t^i, n, s^j)} \\ &=\; \frac{h_t(n_t^j, s^j)\, r^{n_t^j}\,(1-r)^{t-n_t^j}\,[p(1-q)^{n_t^i+n_t^j}+(1-p)]}{\sum_{n=0}^t h_t(n, s^j)\, r^n\,(1-r)^{t-n}\,[p(1-q)^{n_t^i+n}+(1-p)]}, \end{aligned}$$

with the second equality being obtained by canceling equal terms.

To explore the monotonicity of the beliefs $p_t(n_t^j, n_t^i, s^j)$ with respect to the variable n_t^i, notice that

$$\begin{aligned} \frac{dp_t}{dn_t^i}(n_t^j, n_t^i, s^j) \;&=\; \\ &\frac{\ln(1-q)\;\, h_t(n_t^j, s^j)\, r^{n_t^j}\,(1-r)^{t-n_t^j}}{\sum_{n=0}^t h_t(n, s^j)\, r^n\,(1-r)^{t-n}\,[p(1-q)^{n_t^i+n}+(1-p)]} \\ &\times\; \sum_{n=0}^t h_t(n, s^j)\; r^n\,(1-r)^{t-n}\,p\,(1-p)\,(1-q)^{n_t^i}\,[(1-q)^{n_t^j}-(1-q)^n] \end{aligned}$$

Therefore, since $\ln(1-q) \leq 0$,

$$\frac{dp_t}{dn_t^i}(n_t^j, n_t^i, s^j) \gtreqless 0 \quad\Longleftrightarrow\quad \sum_{n=0}^t h_t(n, s^j)\; r^n\,(1-r)^{t-n}\,[(1-q)^{n_t^j}-(1-q)^n] \lesseqgtr 0,$$

The sum is independent of n_t^i, decreasing in n_t^j, positive for $n_t^j = 0$, negative for $n_t^j = t$. Hence, for every t and s^j, there is a value $\bar{n}_t^j$ such that

$$\frac{dp_t}{dn_t^i,}(n_t^j, n_t^i, s^j) \gtreqless 0 \iff n_t^j \gtreqless \bar{n}_t^j$$

Let $\tilde{n}_t^i > n_t^i$. To show that

$$\sum_{n_t^j=0}^{n} [\, p_t(n_t^j, n_t^i, s^j) - p_t(n_t^j, \tilde{n}_t^i, s^j)\,] \;\geq\; 0, \quad \text{for all } n = 0, 1, ...t,$$

as required for first-order stochastic dominance, notice that

$$p_t(n_t^j, \tilde{n}_t^i, s^j) \gtreqless p_t(n_t^j, n_t^i, s^j) \iff n_t^j \gtreqless \bar{n}_t^j.$$

Therefore, the sum is positive for values $n \leq \bar{n}_t^j$. For values $n \geq \bar{n}_t^j$, we have

$$\sum_{n_t^j=0}^{n} [\, p_t(n_t^j, n_t^i, s^j) - p_t(n_t^j, \tilde{n}_t^i, s^j)\,] \;=\; -\sum_{n_t^j=n+1}^{t} [\, p_t(n_t^j, n_t^i, s^j) - p_t(n_t^j, \tilde{n}_t^i, s^j)\,]$$

so that again the sum is positive, as required.

Proof of Lemma 2:

Since first-order stochastic dominance is a transitive relation, so that our argument can proceed from s^j to $\hat{s}^j$ in a threshold-by-threshold, it suffices to show the result for strategies s^j and $\hat{s}^j$ such that $N_\tau^j = \hat{N}_\tau^j$, for $\tau \neq t_0$, and $N_\tau^j < \hat{N}_\tau^j$, for $\tau = t_0$, for some time $t_0 < t$.

Given two threshold strategies s^j and $\hat{s}^j$ that differ only at time $t_0 < t$, with $N_{t_0}^j < \hat{N}_{t_0}^j$, by Lemma 1, for all $M \leq t$, we have

$$P[n_t^j \leq M \,|\, n_t^i, \hat{s}^j] \;-\; P[n_t^j \leq M \,|\, n_t^i, s^j] \;= \\ \sum_{m=0}^{M} \left[\frac{h_t(m, \hat{s}^j)\, \bar{p}(m, n_t^i)}{\sum_{n=0}^{t} h_t(n, \hat{s}^j)\, \bar{p}(n, n_t^i)} \;-\; \frac{h_t(m, s^j)\, \bar{p}(m, n_t^i)}{\sum_{n=0}^{t} h_t(n, s^j)\, \bar{p}(n, n_t^i)} \right],$$

with the expression $\bar{p}(m, n_t^i) \;=\; r^m\,(1-r)^{t-m}\,[p(1-q)^{n_t^i+m} + (1-p)]$ being used to simplify the notation. Therefore, for all $M \leq t$,

$$P[n_t^j \leq M \,|\, n_t^i, \hat{s}^j] \;-\; P[n_t^j \leq M \,|\, n_t^i, s^j] \;\leq\; 0$$

as required for for the result, if and only if

$$\sum_{m=0}^{M}\sum_{n=0}^{t}\bar{p}(m,n_t^i)\,\bar{p}(n,n_t^i)\;\left[h_t(m,\hat{s}^j)\,h_t(n,s^j)\;-\;h_t(m,s^j)\,h_t(n,\hat{s}^j)\right]\;\leq\;0$$

or, after canceling equal terms, if and only if

$$\sum_{m=0}^{M}\;\sum_{n=M+1}^{t}\;\bar{p}(m,n_t^i)\,\bar{p}(n,n_t^i)\;\left[h_t(m,\hat{s}^j)\,h_t(n,s^j)\;-\;h_t(m,s^j)\,h_t(n,\hat{s}^j)\right]\;\leq\;0$$

Therefore, it suffices to show that

$$h_t(m,\hat{s}^j)\,h_t(n,s^j)\;-\;h_t(m,s^j)\,h_t(n,\hat{s}^j)\;\leq\;0,$$

for all $m, n \leq t$ such that $m \leq M < n$.

Notice that for all strategies s with thresholds $\{N_\tau\}_{\tau=1}^{t-1}$ and any time $t_0 < t$, we have

$$h_t(k,s)\;=\;\sum_{l=0}^{k}\;h'_{t_0}[l,(N_\tau)_{\tau=1}^{t_0}]\;h_{t-1-t_0}[k-l,(N_\tau-l)_{\tau=t_0+1}^{t-1}]$$

where $h'_{t_0}[l,(N_\tau)_{\tau=1}^{t_0}]$ is the number of player j's histories at the end of period t_0 such that player j has received l draws of L and no draw of H and such that $n_\tau^j < N_\tau$ for all $\tau \leq t_0$.

Therefore, it suffices to show that

$$\sum_{k=0}^{n_t^j}\;h'_{t_0}[k,(\hat{N}_\tau^j)_{\tau=1}^{t_0}]\;h_{t-1-t_0}[n_t^j-k,(\hat{N}_\tau^j-k)_{\tau=t_0+1}^{t-1}]$$

$$\times\;\sum_{l=0}^{n}\;h'_{t_0}[l,(N_\tau^j)_{\tau=1}^{t_0}]\;h_{t-1-t_0}[n-l,(N_\tau^j-l)_{\tau=t_0+1}^{t-1}]\;\;-$$

$$\sum_{k=0}^{n_t^j}\;h'_{t_0}[k,(N_\tau^j)_{\tau=1}^{t_0}]\;h_{t-1-t_0}[n_t^j-k,(N_\tau^j-k)_{\tau=t_0+1}^{t-1}]$$

$$\times\;\sum_{l=0}^{n}\;h'_{t_0}[l,(\hat{N}_\tau^j)_{\tau=1}^{t_0}]\;h_{t-1-t_0}[n-l,(\hat{N}_\tau^j-l)_{\tau=t_0+1}^{t-1}]\;\;\leq\;\;0$$

Since $\hat{N}^j_\tau = N^j_\tau$, for all $\tau > t_0$, this reduces to showing (after again canceling equal terms)

$$\sum_{k=0}^{m} \sum_{l=m+1}^{n} h_{t-1-t_0}[m-k, (N^j_\tau - k)_{\tau=t_0+1}^{t-1}]\; h_{t-1-t_0}[m-l, (N^j_\tau - l)_{\tau=t_0+1}^{t-1}]$$
$$\times \begin{bmatrix} h'_{t_0}[k, (\hat{N}^j_\tau)_{\tau=1}^{t_0}]\; h'_{t_0}[l, (N^j_\tau)_{\tau=1}^{t_0}] - \\ h'_{t_0}[k, (N^j_\tau)_{\tau=1}^{t_0}]\; h'_{t_0}[l, (\hat{N}^j_\tau)_{\tau=1}^{t_0}] \end{bmatrix} \;\leq\; 0$$

for all $m, n \leq t$ such that $m \leq M < n$.

For $m < N^j_{t_0}$, we have $h'_{t_0}[k, (\hat{N}^j_\tau)_{\tau=1}^{t_0}] = h'_{t_0}[k, (N^j_\tau)_{\tau=1}^{t_0}]$, for all $k \leq m$, so that the inequality follows from the fact that $h'_{t_0}[l, (N^j_\tau)_{\tau=1}^{t_0}] \leq h'_{t_0}[l, (\hat{N}^j_\tau)_{\tau=1}^{t_0}]$, for all $l \geq 0$.

Finally, for $m \geq N^j_{t_0}$, we have $h'_{t_0}[l, (N^j_\tau)_{\tau=1}^{t_0}] = 0$, for all $l \geq m+1$, so that the expression on the left-hand-side of the inequality involves only non-positive terms.

Proof of Lemma 3:

We argue by means of backwards induction, in periods $T-1, T-2, \ldots, 1$, showing in each period, first, that player i's optimal strategy at the end of the period takes the form of a threshold rule; and second, that player i's expected payoff from following his optimal strategy is decreasing in the number of L draws he has obtained that far.

Throughout our argument we condition on player j having obtained no draw of H by the time of player i's decision; otherwise, player i's decision is irrelevant for his payoff. For the sake of brevity, we drop this condition from our notation.

Given any $T \in \mathbb{Z}^+$, suppose that player j's strategy s^j is such that he stops in periods $t < T$ if and only if $n^j_t \geq N^j_t$, for some sequence of thresholds $\{N^j_t\}_{t=1}^{T-1}$.

Moving backwards in the periods of the game, suppose that player i has obtained $n^i_{T-1} > 0$ draws of L by the end of period $T-1$.[32] Then player i's expected payoff from continuing to the last period T, conditionally on player j having obtained n^j_{T-1}

[32]When $n^i_{T-1} = 0$, player i has an incentive to continue into period T, independently of h^j_{T-1}.

draws of L and on the game actually reaching period T, is

$$U_T(n^i_{T-1}|n^j_{T-1},s^j) = \begin{cases} \frac{1}{2}\delta\,[p^H(n^i_{T-1}+n^j_{T-1})(H-L)+L], & n^j_{T-1}>0; \\ \frac{1}{2}\delta\,[p^H(n^i_{T-1})(H-L)+L] + & \\ \frac{1}{2}\delta\,[1-r\,p(n^i_{T-1})\,q]\,(1-r)\,L, & n^j_{T-1}=0. \end{cases} \tag{B.1}$$

Therefore, conditionally on n^j_{T-1}, player i's expected gain from continuing to period T instead of stopping in period $T-1$ is

$$\Delta V_{T-1}(n^i_{T-1}\,|\,n^j_{T-1},s^j) \;=\; \begin{cases} -L/2, & n^j_{T-1}\geq N^j_{T-1}; \\ U_T(n^i_{T-1}\,|\,n^j_{T-1},s^j)\;-\;L, & n^j_{T-1}<N^j_{T-1}. \end{cases}$$

Finally, player i's (unconditional) expected gain from continuing instead of stopping is

$$\Delta V_{T-1}(n^i_{T-1}\,|\,s^j) \;=\; \sum_{n^j_{T-1}=0}^{T-1} p_{T-1}(n^j,n^i_{T-1},s^j)\,\Delta V_{T-1}(n^i_{T-1}\,|\,n^j_{T-1},s^j), \tag{B.2}$$

Under Condition 1, the function $\Delta V_{T-1}(\,\cdot\,|\,\cdot,s^j)$ is decreasing in n^j_{T-1}.[33] In addition,

$$\begin{aligned}\Delta V_{T-1}(n^i_{T-1}\,|\,0,s^j) \;&=\; \\ &(1/2)\delta\,p(n^i_{T-1})\,rq\,[(2-rq)H-(3-r-rq)L]-L+(1/2)\delta\,(2-r)L\end{aligned}$$

Therefore, for parameters $H/L<(3-r-rq)/(2-rq)$, we have $\Delta V_{T-1}(n^i_{T-1}\,|\,0,s^j)<0$, so that $\Delta V_{T-1}(n^i_{T-1}\,|\,n^j_{T-1},s^j)<0$, for all $n^i_{T-1}\geq 1$, $n^j_{T-1}\geq 0$. In this case, player i's expected gain from continuing is $\Delta V_{T-1}(n^i_{T-1}\,|\,s^j)<0$, for all $n^i_{T-1}\geq 1$, implying that player i is best-off stopping if he has at least one draw of L. Otherwise, for parameters $H/L\geq(3-r-rq)/(2-rq)$, the function $\Delta V_{T-1}(\,\cdot\,|\,\cdot,s^j)$ is decreasing

[33]For all n^i_{T-1}, since the probability $p^H(n^i_{T-1}+n^j_{T-1})$ is decreasing in n^j_{T-1}, the payoff $U_T(n^i_{T-1}\,|\,n^j_{T-1},s^j)$ is also decreasing in n^j_{T-1}. Condition 1 ensures that $U_T(n^i_{T-1}\,|\,n^j_{T-1},s^j)-L>-L/2$, for all $n^j_{T-1}<N^j_{T-1}$, for all N^j_{T-1}.

also in n^i_{T-1}. In this case, for $\tilde{n}^i_{T-1} > n^i_{T-1}$, we have

$$\begin{aligned}
\Delta V_{T-1}(\tilde{n}^i_{T-1} \,|\, s^j) &= \sum_{n^j_{T-1}=0}^{T-1} p_{T-1}(n^j_{T-1}, \tilde{n}^i_{T-1}, s^j)\, \Delta V_{T-1}(\tilde{n}^i_{T-1} \,|\, n^j_{T-1}, s^j) \\
&\leq \sum_{n^j_{T-1}=0}^{T-1} p_{T-1}(n^j_{T-1}, n^i_{T-1}, s^j)\, \Delta V_{T-1}(n^i_{T-1} \,|\, n^j_{T-1}, s^j) \\
&= \Delta V_{T-1}(n^i_{T-1} \,|\, s^j),
\end{aligned}$$

with the inequality being obtained from the fact that the probability distribution $p_{T-1}(\,\cdot, \tilde{n}^i_{T-1}, s^j)$ first-order stochastically dominates the distribution $p_{T-1}(\,\cdot, n^i_{T-1}, s^j)$. Hence, player i's incentive to continue to period T is decreasing in the number n^i_{T-1} of L draws he has received, implying that his best response in period $T-1$ takes the form of a threshold rule, N^i_{T-1}.

To complete the first step of the induction, notice that player i's expected payoff from choosing to continue to period T,

$$V^c_{T-1}(n^i_{T-1} \,|\, s^j) \;=\; \sum_{n^j_{T-1}=0}^{N^j_{T-1}-1} p_{T-1}(n^j_{T-1}, n^i_{T-1}, s^j)\, U_T(n^i_{T-1} \,|\, n^j_{T-1}, s^j)$$

is decreasing in n^i_{T-1}, since the distribution $p_{T-1}(\,\cdot, n^i_{T-1}, s^j)$ is first-order stochastically increasing in n^i_{T-1} and the payoff $U_T(\, n^i_{T-1} \,|\, n^j_{T-1}, s^j)$ is decreasing in n^i_{T-1} and n^j_{T-1}. In addition, player i's payoff from stopping in period $T-1$,

$$V^s_{T-1}(n^i_{T-1} \,|\, s^j) \;=\; (L/2) \;+\; \sum_{n^j_{T-1}=0}^{N^j_{T-1}-1} p_{T-1}(n^j_{T-1}, n^i_{T-1}, s^j)\ (L/2),$$

is also decreasing in n^i_{T-1}, because of stochastic dominance. Therefore, player i's optimal payoff at the end of period $T-1$,

$$V^*_{T-1}(n^i_{T-1} \,|\, s^j) \;=\; \max\{\, V^c_{T-1}(n^i_{T-1} \,|\, s^j),\, V^s_{T-1}(n^i_{T-1} \,|\, s^j)\,\} \tag{B.3}$$

is decreasing in n^i_{T-1}.

Proceeding to periods $t = T-2, T-3, \ldots, 1$, suppose that player i's optimal continuation strategy in period $t+1$ takes the form of a threshold rule $\{N^i_\tau\}_{\tau=t+1}^{T-1}$, depending only on the strategy s^j; and that his optimal payoff at the end of period $t+1$,

$$V^*_{t+1}(n^i_{t+1} \,|\, s^j) \;=\; V_{t+1}[n^i_{t+1} \,|\, s^j, (N^i_\tau)_{\tau=t+1}^{T-1}]$$

is decreasing in n^i_{t+1} (induction hypothesis).

At the beginning of period $t+1$, player i's expected payoff from drawing in that period and then following the optimal continuation strategy $\{N^i_\tau\}_{\tau=t+1}^{T-1}$ is

$$\begin{aligned} & U^*_{t+1}(n^i_t \,|\, s^j) = U_{t+1}[n^i_t \,|\, s^j, (N^i_\tau)_{\tau=t+1}^{T-1}] \\ = \;\; & \hat{p}_H(n^i_t \,|\, s^j)\; (1/2)\, H + \; [1 - \hat{p}^H_t(n^i_t \,|\, s^j)]\; \hat{p}^L_t(n^i_t \,|\, s^j)\; V^*_{t+1}(n^i_t + 1 \,|\, s^j) \qquad \text{(B.4)} \\ & + [1 - \hat{p}^H_t(n^i_t \,|\, s^j)]\; [1 - \hat{p}^L_t(n^i_t \,|\, s^j)]\; V^*_{t+1}(n^i_t \,|\, s^j) \end{aligned}$$

where

$$\hat{p}^H_t(n^i_t \,|\, s^j) \;=\; \sum_{n^j_t=0}^{t} \hat{p}_t(n^j_t, n^i_t, s^j)\, p_H(n^j_t + n^i_t)$$

is player i's belief at the beginning of period $t+1$ that at least one draw of H will be obtained in that period,

$$p^L_t(n^i_t \,|\, s^j) \;=\; \sum_{n^j_t=0}^{t} p'_t(n^j_t, n^i_t, s^j)\, \frac{[\,1 - p(n^j_t + n^i_t) + p(n^j_t + n^i_t)\,(1-q)\,(1-rq)\,]\; r}{1 - p(n^j_t + n^i_t) + p(n^j_t + n^i_t)\,(1-rq)^2}$$

is player i's belief at the beginning of period $t+1$ that he will draw L in that period, conditional on neither player drawing H, with

$$p'_t(n^j_t, n^i_t, s^j) \;=\; \frac{h'_t(n^j_t, s^j)\, r^{n^j_t}\, (1-r)^{t-n^j_t}\, [p(1-q)^{n^i_t + n^j_t} + (1-p)]}{\sum_{n=0}^{t} \hat{h}_t(n, s^j)\, r^n\, (1-r)^{t-n}\, [p(1-q)^{n^i_t + n} + (1-p)]},$$

defined in a manner analogue to $p_t(n^j_t, n^i_t, s^j)$, being the probability that player j has obtained n^j_t draws of L by the end of period t, conditional on n^i_t and on the constraints of the stopping strategy s^j, including the one at the end of period t.[34]

[34]In particular, $h'_t(n^j_t, s^j) \leq \binom{t+1}{n^j_t}$ is the number of histories of player j consistent with with player j having obtained n^j_t draws of L and the constraints of the stopping strategy s^j in periods $1, 2, \ldots, t$. Notice that these constraints include the hypothesis that no draw of H has occurred.

Arguing as in Lemma 1, it can be shown that the distribution $p'_t(\,\cdot, n^i_t, s^j)$ first-order stochastically increases in n^i_t. Therefore, the probabilities $\hat{p}^H_t(n^i_t \,|\, s^j)$ and $\hat{p}^L_t(n^i_t \,|\, s^j)$ are respectively decreasing and increasing in n^i_t. In addition, $V^*_{t+1}(\,\cdot\,|\, s^j)$ is decreasing (from the induction hypothesis) and $V^*_{t+1}(n^i_{t+1} \,|\, s^j) \leq (1/2)H$, for all $n^i_{t+1} \geq 0$. Hence, the payoff $U^*_{t+1}(n^i_t \,|\, s^j) = U_{t+1}[n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}]$ is decreasing in n^i_t.

At the end of period t, player i's expected gain from choosing to continue rather than to stop is

$$\begin{aligned}
&\Delta V_t(n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}) \qquad\qquad \text{(B.5)}\\
&\quad = P[n^j_t < N^j_t \,|\, n^i_t, s^j]\; [\,U_{t+1}[n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}] - L\,]\\
&\qquad + P[n^j_t \geq N^j_t \,|\, n^i_t, s^j]\; (-L/2)\\
&\quad = P[n^j_t < N^j_t \,|\, n^i_t, s^j]\; [\,U_{t+1}[n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}] - L/2\,] \;-\; L/2
\end{aligned}$$

Using again the fact that an increase in n^i_t results in a stochastic dominant distribution for the unknown variable n^j_t, along with the fact that $U^*_{t+1}(\,\cdot\,|\, s^j)$ is decreasing, it follows that player i's gain $\Delta V_t[n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}]$ is decreasing in n^i_t, so that player i's best-response strategy in period t takes the form of a threshold rule, N^i_t.

Finally, since the probability $P[n^j_t < N^j_t \,|\, n^i_t, s^j]$ and the expected payoff $U_{t+1}(n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1})$ are decreasing in n^i_t, it follows that the payoffs

$$V^c_t[n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}] \;=\; P[n^j_t < N^j_t \,|\, n^i_t, s^j]\; U_{t+1}[n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}],$$

$$V^s_t[n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}] \;=\; (L/2) \;+\; P[n^j_t < N^j_t \,|\, n^i_t, s^j]\; (L/2)$$

and

$$\begin{aligned}
V^*_t(n^i_t \,|\, s^j) \;&=\; V_t[n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}]\\
&=\; \max\{\, V^c_t[n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}],\; V^s_t[n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}]\,\} \qquad \text{(B.6)}
\end{aligned}$$

are decreasing in n^i_t, completing the induction.

Proof of Proposition 2:

Similar to the proof of Lemma 3, we condition our continuation payoff calculations on player j having obtained no draw of H by the time of player i's decision.

In the continuation game starting at the end of period T, it is clear that the strategy profile in which each player stops immediately constitutes an equilibrium, independently of the players' strategies up to that period and associated beliefs.

In period $T-1$, suppose that the two players have followed symmetric strategies s' with stopping thresholds $\{N_t\}_{t=0}^{T-2}$ prior to that period; and that player j follows a threshold N^j_{T-1} in that period. If player i has obtained $n^i_{T-1} > 0$ draws of L, then his expected gain from continuing to period T instead of stopping in period $T-1$ is given by equations (B.1) and (B.2) in the proof of Lemma 3.[35]

For parameters $H/L < (3-r-rq)/(2-rq)$, as argued in the proof of Lemma 3, we have $\Delta V_{T-1}(n^i_{T-1}, \,|\, n^j_{T-1}, s', N^j_{T-1}) < 0$, for all $n^i_{T-1} \geq 1$, $n^j_{T-1} \geq 0$, so that player i's continuation gain is $\Delta V_{T-1}(n^i_{T-1} \,|\, s', N^j_{T-1}) < 0$, for all $n^i_{T-1} \geq 1$. In this case, there are two equilibria for the continuation game, with thresholds either $N_{T-1} = 0$ or $N_{T-1} = 1$.[36]

For parameters $H/L \geq (3-r-rq)/(2-rq)$, again as argued in the proof of Lemma 3, the payoff $\Delta V_{T-1}(n^i_{T-1} \,|\, s', N^j_{T-1})$ is decreasing in the number of draws n^i_{T-1}. In addition, under Condition 1, the payoffs $\Delta V_{T-1}(n^i_{T-1}, \,|\, n^j_{T-1}, s', N^j_{T-1})$ and, therefore, $\Delta V_{T-1}(n^i_{T-1} \,|\, s', N^j_{T-1})$ are increasing in player j's threshold N^j_{T-1}. Hence, the threshold characterizing player i's best-response strategy in period $T-1$, given by

$$BR^i_{T-1}(N^j_{T-1} \,|\, s') \;=\; \max\{\, n = 1, 2, \ldots, T-1 :\; \Delta V_{T-1}(n \,|\, s', N^j_{T-1}) \;>\; 0 \,\} \;+\; 1,$$

with $BR^i_{T-1}(s^j) = 1$ when the set is empty, is an increasing function of the threshold

[35]Notice that player i's beliefs regarding the number of draws of his opponent, n^j_{T-1}, are independent of his opponent's continuation strategy, in particular, of the threshold N^j_{T-1}.

[36]The case of stopping even without draws, $N_{T-1} = 0$, is trivially present for all parameters.

N^j_{T-1} in the strategy s^j.[37]

The set $\{1, 2, ..., T\}$ is a lattice with respect to the order $\geq$, complete because of finiteness. Therefore, since the function $BR^i_{T-1}(\,\cdot \mid s')$ is increasing in the variable N^j_{T-1}, it has at least one fixed point. Hence, for each symmetric strategy $s' = \{N_t\}_{t=1}^{T-2}$ prior to period $T-1$, we can define the players' common threshold at time $T-1$ as the maximal fixed point of $BR^i_{T-1}(\,\cdot \mid s')$.

Moving backwards to periods $t = T-2, T-3, \ldots,\ 1$, suppose that for each symmetric strategy profile with stopping thresholds $\{N_\tau\}_{\tau=1}^t$ up to the end of period t, there is a symmetric equilibrium $s''[(N_\tau)_{\tau=1}^t]$ for the continuation game starting in period $t+1$, with thresholds that depend on $\{N_\tau\}_{\tau=1}^t$. (induction hypothesis).

Suppose that the two players have followed a symmetric threshold strategy s' up to the end of period $t-1$ and let player j change, first, his threshold in period t from N to $N+1$, and second, his continuation strategy from $s''(s', N)$ to $s''(s', N+1)$.

If player i has n^i_t draws of L, then his expected gain from continuing rather than stopping at the end of period t, against a strategy $s(s', M) = [s', M, s''(s', M)]$ of player j, is

$$\Delta V_t[(n^i_t \mid s(s'^j_t \geq M \mid n^i_t, s')(-L/2) + P(n^j_t < M \mid n^i_t, s')\ [\,U_{t+1}[n^i_t \mid s(s', M)] - L\,],$$

where $U_{t+1}(n^i_t \mid s(s', M))$, defined recursively by equations (B.1)–(B.6) in the proof of Lemma 3, is player i's optimal expected payoff in the continuation game starting in period $t+1$, conditional on period $t+1$ being reached, with player j following a strategy $s(s', M)$. Since player j's continuation strategy $s''(s', M)$ is part of a symmetric equilibrium for that game, given (s', M), notice that the payoff $U_{t+1}(n^i_t \mid s(s', M))$ is achieved with player i also following the continuation strategy $s''(s', M)$.

[37]If $\tilde{N}^j_{T-1} > N^j_{T-1}$, then we have $\Delta V_{T-1}(n \mid s',\, \tilde{N}^j_{T-1}) > \Delta V_{T-1}(n \mid s',\, N^j_{T-1})$, for all $n = 1, 2, \ldots, T$, implying that $\{n \in \mathbb{N} : \Delta V_{T-1}(n \mid s',\, \tilde{N}^j_{T-1}) > 0\} \supseteq \{n \in \mathbb{N} : \Delta V_{T-1}(n \mid s',\, N^j_{T-1}) > 0\}$ and, therefore, that the best response is $BR^i_{T-1}(\tilde{N}^j_{T-1} \mid s') \geq BR^i_{T-1}(N^j_{T-1} \mid s')$, as required.

When player j switches from $s(s', N)$ to $s(s', N+1)$, we have

$$\Delta V_t[n_t^i \,|\, s(s', N+1)] - \Delta V_t[n_t^i \,|\, s(s', N)] = p_t(N, n_t^i, s')\ (-L/2)$$
$$+\ P(n_t^j \leq N \,|\, n_t^i, s')\ U_{t+1}[n_t^i \,|\, s(s', N+1)] -\ P(n_t^j \leq N-1 \,|\, n_t^i, s')\ U_{t+1}[n_t^i \,|\, s(s', N)]$$

Since player i cannot gain from deviating from $s''(s', N+1)$ to the strategy of surely stopping in period $t+1$, against $s''(s', N+1)$, in the continuation game following $(s', N+1)$, we have

$$U_{t+1}[n_t^i \,|\, s(s', N+1)] \quad \geq$$
$$\sum_{n_t^j=0}^{N} \frac{p_t(n_t^j, n_t^i, s')}{P(n_t^j \leq N \,|\, n_t^i, s')}\ (1/2)\, \delta\, [\, p(n_t^i + n_t^j)\, (1 - (1-rq)^2)\, (H - L)\ +\ L\,]$$

In addition, in the continuation game following (s', N), we have

$$U_{t+1}[n_t^i \,|\, s(s', N)] \quad \leq$$
$$\sum_{n_t^j=0}^{N-1} \frac{p_t(n_t^j, n_t^i, s')}{P(n_t^j \leq N-1 \,|\, n_t^i, s')}\ (1/2)\, \delta\, [\, p(n_t^i + n_t^j)\, (1 - (1-rq)^{2(T-t)})\, (H - L)\ +\ L\,]$$

that is, player i's optimal expected payoff cannot exceed what could be achieved if the two players shared L or H after performing maximal costless experimentation in the time remaining until final period T.

Therefore, after some rearrangement of the terms, we have

$$\Delta V_t[n_t^i \,|\, s(s', N+1)] - \Delta V_t[n_t^i \,|\, s(s', N)] \quad \geq$$
$$p_t(N, n_t^i, s')\ (1/2)\, [\, \delta\, p(n_t^i + N)\, (1 - (1-rq)^2)\, (H - L)\ -\ (1-\delta)\, L\,]$$
$$-\ \sum_{n_t^j=0}^{N-1} p_t(n_t^j, n_t^i, s')\ (1/2)\, \delta\, p(n_t^i + n_t^j)\, [\, (1-rq)^2 - (1-rq)^{2(T-t)}\,]\, (H - L)$$

In addition, since the function $p(\cdot)$ is decreasing, we have

$$\Delta V_t[n_t^i \,|\, s(s', N+1)] - \Delta V_t[n_t^i \,|\, s(s', N)] \;\; \geq$$

$$p_t(N, n_t^i, s'^2)\,(H-L) \,-\, (1-\delta)\,L\,]$$

$$- \sum_{n_t^j=0}^{N-1} p_t(n_t^j, n_t^i, s')\;(1/2)\,\delta\,p(n_t^j)\,[\,(1-rq)^2 - (1-rq)^{2(T-t)}\,]\,(H-L)$$

Thus, for player i's expected gain from continuing at the end of period t to be

$$\Delta V_t[n_t^i \,|\, s(s', N+1)] \;\; \geq \;\; \Delta V_t[n_t^i \,|\, s(s', N)]$$

it is sufficient that

$$p_t(N, n_t^i, s')\;\;\left[p(2N)\,[1-(1-rq)^2] \;-\; \tfrac{1-\delta}{\delta}\;\tfrac{L}{H-L}\right]$$

$$+ \;\; \sum_{n_t^j=0}^{N-1} p_t(n_t^j, n_t^i, s_t'^j)\,[(1-rq)^{2(T-t)} \;-\; (1-rq)^2] \;\; \geq \;\; 0$$

The expression on the left-hand-side is the expectation of a function increasing in n_t^j with respect to a distribution of n_t^j that is stochastically increasing in n_t^i, so it achieves its minimal value for $n_t^i = 1$. Hence, the above inequality follows directly from Condition 1.

Hence, under Condition 1, for each strategy s' prior to period t, for each n_t^i, player i's expected gain $\Delta V_t[n_t^i \,|\, s(s', N_t^j)]$ from continuing instead of stopping at the end of period t is increasing in the threshold N_t^j parameterizing player j's continuation strategy $s''(s', N_t^j)$. Thus, for each strategy s' prior to period t, the threshold N_t^i parameterizing player i's best-response continuation strategy $s''(s', N_t^i))$ in period t,

$$BR_t^i(N_t^j \,|\, s') \;=\; \max\{\, n = 1, 2, \ldots, t:\; \Delta V_t[n \,|\, s(s', N_t^j)] \;>\; 0\,\} \;+\; 1,$$

with $BR_t^i(N_t^j \,|\, s') = 1$ when the set is empty, is an increasing function of the threshold N_t^j in player j' strategy $[s', N_t^j, s''(s', N_t^j)]$.

The set $\{1, 2, ..., t+1\}$ of possible thresholds in period t is a lattice with respect to the order $\geq$, complete because of finiteness. Therefore, since the function $BR^i_t(\cdot \mid s')$ is increasing in N^j_t, it has at least one fixed point.

For each symmetric threshold strategy s' prior to period t, we define the players' common threshold N_t at period t as the maximal fixed point of $BR^i_{T-1}(\cdot \mid s')$; and by construction, the continuation strategy $(N_t, s''(s', N_t))$ forms a symmetric equilibrium for the game starting at period t, when the two players have the beliefs induced by the strategy s' that they have followed prior to that period.

The argument concludes when it defines a threshold N_1 for the first period of the game, with the impled strategy $[N_1, s''(N_1)]$ forming a symmetric perfect Bayesian equilibrium for the entire game.

Proof of Proposition 3:

In the case of common learning, suppose that player i has obtained $n^i_1 = 1$ draw of L in period $t = 1$ and faces an opponent who will continue to period $T = 2$, the last period of the game. If player j has obtained $n^j_1 = 1$ draw of L in period $t = 1$, then player i's expected payoff from continuing to period $T = 2$ is

$$v_1(1,1) \;=\; \delta\,[\,L/2 \;+\; p^H(2)\,(H-L)/2\,]$$

If player j has obtained $n^j_1 = 0$ draw of L in period $t = 1$, then player i's expected payoff from continuing to period $T = 1$ is

$$v_1(1,0) \;=\; \delta\,[\,L/2 \;+\; p^H(1)(H-L)/2 \;+\; [1-p(1)\,rq]\,(L/2)\,]$$

Since all terms are positive and $p^H(1) > p^H(2)$, it follows that $v_1(1,0) > v_1(1,1)$.

Using some simple algebraic manipulations, it is easy to check that the inequalities $v_1(1,1) \geq L$ and $v_1(0,1) < L$ are equivalent respectively to conditions (9) and (10).

Now, consider the strategy in which a player continues at the end of period $t = 1$, independently of the number of draws he and his opponent have. For this strategy to be part of a symmetric equilibrium, it is necessary and sufficient that $v_1(1,0) \geq L$

and $v_1(1,1) \geq L$, a condition that reduces to $v_1(1,1) \geq L$, which is equivalent to condition (9).

Similarly, consider the strategy in which a player continues at the end of period $t = 1$ if and only if he has received no draw. For this strategy to be part of a symmetric equilibrium, it is necessary and sufficient that $v_1(1,1) < L$ and $v_1(1,0) < L$, a condition that reduces to $v_1(0,1) < L$, which is equivalent to condition (10).

Finally, consider the strategy in which a player continues at the end of period $t = 1$ if and only if either he or his opponent has failed to obtain a draw. For this strategy to be part of a symmetric equilibrium, it is necessary and sufficient that $v_1(1,1) < L$ and that $v_1(1,0) \geq L$, i.e., that conditions (9) and (10) both fail.

Looking at the corresponding setting under private learning, when condition (8) holds, by Proposition 2, there must exist at least one Bayesian equilibrium.

Suppose that player i has obtained one draw of L in period $t = 1$ and faces an opponent who will continue to period $T = 2$ unless he obtains H. Then player i's expected payoff from continuing (and stopping) at $T = 2$, conditional on his opponent having not obtained H, is

$$v_1(1) \;=\; p_1(0,1)\, v_1(1,0) \;+\; [1 - p_1(0,1)]\, v_1(1,1)$$

Using the expressions for $v_1(1,0)$ and $v_1(1,1)$ and applying some simple algebraic manipulations, it is easy to show that the inequality $v_1(1) < L$ is equivalent to condition (11).

When $v_1(1) \geq L$, the strategy profile in which each player continues to period $T = 2$ unless he obtains H forms a symmetric equilibrium under private learning.

When $v_1(1) < L$, this strategy profile is no longer an equilibrium. In this case, player i's expected payoff from continuing with one draw of L to period $T = 2$ against an opponent who will stop as soon as he receives one draw is

$$u_1(1) \;=\; p_1(0,1)\, v_1(1,0)$$

Therefore, the strategy profile in which each player stops if he obtains a draw at $t = 1$

forms a symmetric equilibrium under private learning if and only if

$$u_1(1) \;<\; p_1(0,1)\,L \;+\; [1-p_1(0,1)]\,(L/2)$$

which is true when conditions (8) and (11) hold.

We conclude the proof by comparing the equilibria under common and private learning.

Under condition (9), we have that $v_1(1,0) \geq L$ and $v_1(1,1) \geq L$, so that $v_1(1) \geq L$. Therefore, in both settings, the two players continue to period $T = 2$ unless they obtain H and then stop, for the same equilibrium outcomes.

Similarly, under condition (10), we have $v_1(1,0) < L$ and $v_1(1,1) < L$, so that $v_1(1) < L$. In both settings, each player stops either as soon as he obtains a draw, again for the same equilibrium outcomes.

Finally, if conditions (9) and (10) both fail, we have $v_1(1,1) < L$ and $v_1(1,0) \geq L$, so, under common learning the two players stop at $t = 1$ if and only if they both obtain draws. Under private learning, when condition (11) holds, the game will stop in period $t = 1$ even with a single draw, for a shorter expected experimentation horizon. On the other hand, when condition (11) fails, the game will continue to period $t = T = 2$ unless H is obtained, for a longer expected experimentation horizon.

When condition (11) holds, less experimentation under private learning implies also lower expected payoffs, since the generated welfare is respectively increasing / decreasing in N, the total number of L draws that the players obtain by the time they stop experimenting.

Proof of Proposition 4

For any probability parameters $r, q \in [0,1]$, we need to show that if condition (11) is satisfied for some probability $p \in [0,1]$, then it is also satisfied for all probvabilities $p' \leq p$. For this, we need that the LHS in inequality (11) is increasing in p. Equivalently, we show that the continuation payoff $v_1(1) = v_1(1;p)$, defined in the proof of Proposition 3, is increasing in p.

Suppose first that $H/L > 3/2$. Then the continuation payoffs $v_1(1,1;p)$ and $v_1(1,0;p)$ are both increasing in p, with $v_1(1,0;p) \geq v_1(1,1;p)$, for all $p \in [0,1]$. In addition, the beliefs $p_1(0,1) = p_1(0,1;p)$ are increasing in p. Therefore,

$$\begin{aligned}
\tfrac{\partial}{\partial p} v_1(1;p) \;&=\; \tfrac{\partial}{\partial p} p_1(0,1;p)\;[v_1(1,0;p) - v_1(1,1;p)] \\
&\quad +\; p_1(0,1;p)\;\tfrac{\partial}{\partial p} v_1(1,0;p) \;+\; [1 - p_1(0,1;p)]\;\tfrac{\partial}{\partial p} v_1(1,1;p) \quad > \; 0,
\end{aligned}$$

since all terms are positive, so that $v_1(1;p)$ is increasing in p.

Finally, when $H/L < 3/2$, then

$$\begin{aligned}
& p(1)\,[1 - (1 - rq)^2]\,(H - L)/L \;+\; (1 - r)[1 - p(1)\,rq] \\
&\qquad < \; p(1)\,[1 - (1 - rq)^2]\;,(1/2) \;+\; (1 - r)[1 - p(1)\,rq] \\
&\qquad = \; 1 \;-\; r\;[1 - p(1)\,rq\,(1 - q/2)] < 1 < (2 - \delta)/\delta,
\end{aligned}$$

so that condition (10) and therefore condition (11) are satisfied for all probabilities $p \in [0,1]$, for the result to hold trivially.