DISCUSSION PAPER SERIES

DP13483

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
Published 24 January 2019
Submitted 23 January 2019

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 22, 2019

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.

Keywords: Stopping game, experimentation, learning, preemption, multi-armed bandit problem

JEL Numbers: D83, O31

*We thank 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 Management, hoppe@mik.uni-hannover.de

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

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

"Ever tried. Ever failed. No matter. Try again. Fail again. Fail better."

Samuel Beckett (Worstward Ho, 1983)

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, p.31). 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.

This paper offers a model of experimentation with uncertain outcomes, as suggested by Arrow (1969), and competition.[1] The basic features of this environment resemble those of multi-armed bandit models, as in Keller et al. (2005), Keller and Rady (2010), and Klein and Rady (2011), except that the experimenter in our model faces the threat of being preempted by a competitor in disclosing an experimental outcome. Assuming that the first mover receives recognition while the other gains nothing, the model captures another essential feature of scientific inquiry, the importance of being first (for empirical evidence of this winner-takes-all rewards structure in science, see Hagstrom, 1974, Newman, 2009, and Sabatier and Chollet, 2017). Thus, the competition in our model takes the form of a preemption game, as in Lippman and Mamer (1993), Hopenhayn and Squintani (2011), and Bobtcheff et al. (2017).

The main objective is to understand how the combination of learning about uncertain outcomes and preemption affects the duration of experimentation activity and welfare. Specifi-

[1] "Experimentation" in our paper specifically refers to the scientific procedures undertaken to make a discovery, whereas we understand by "research" the systematic activity undertaken to increase the stock of knowledge.

cally, 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 social planner's problem. [2] We find that competition leads to less experimentation. This finding extends existing results for preemption games to the context of experimentation with uncertain outcomes.[3] 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 sharing of information can generate more experimentation and higher value for a relatively wide range of parameters. In fact, our analysis indicates that the absence of information sharing produces better expected outcomes only in limited cases. This finding is 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 affecting the comparison in our model, 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.

Our model sheds light on recent criticism of the current scientific system and derive potentially important implications for science policy. Lawrence (2016), a biologist at the University of Cambridge, for instance, criticizes university administrators for adopting numerical productivity measures, such as the number of publications, in order to rank researchers, one against another, and allocating funding and jobs accordingly. As he argues, the associated competitive pressure on scientists may not produce a socially desirable outcome: *"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) observed that teams often settle for the so-called *least publishable unit*. This practice has come under fire for, among other reasons, leading to research outcomes of lower quality overall.

Adding to the criticism, the editors of *Nature* have recently urged scientists conducting

[2]The social planner's problem takes the form of a single-player experimentation problem adjusted so as to account for the duplication of experiments by two players.

[3]Preemption games have been studied, among others, by Fudenberg and Tirole (1985), Hoppe and Lehmann-Grube (2005), Hopenhayn and Squintani (2011).

laboratory studies to take greater care in their work, citing several types of *"avoidable errors"*, in terms of both methodology and presentation, which diminish the quality of the published output and make the reproduction of the findings more difficult (Nature Publishing Group, 2012). Thus, Fang and Casadevall (2012) and other scientists propose a comprehensive re-structuring of the current scientific system. They advocate a system that offers greater collegiality, freer sharing of information, and cooperation. In fact, our main findings that that sharing of information typically generates more experimentation and value is in line with their view.

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 that 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. 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 the 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, and they only share a prior belief about the feasibility of a high-value outcome.

Clearly, a researcher always stops experimenting if he obtains a breakthrough result, the best possible outcome of the project; but he may also choose to stop earlier, for a low outcome. Note that each additional draw can enhance a player's stopping value if it exceeds the value he already has. A researcher who has no successful experiments (i.e., no draws) so far, can run a successful experiment and receive either a low or high value draw; or a researcher who already has a low value draw can run a successful experiment and receive a high value draw. In addition, each successful experiment provides additional information about the project's potential value. This also happens for two reasons. If the researcher receives a high value draw, he learns that high value draws are possible. In this case the information about the future is not useful because the researcher immediately stops. If the researcher receives a low value draw, however, he becomes more pessimistic about the possibility of a future breakthrough. If the researcher becomes sufficiently pessimistic, he might simply stop.

Such decisions become more complex when the researcher faces the threat of being preempted by a competitor, in particular if the first mover receives recognition while the other gains nothing. The evaluation of a project's potential should then take into account also any outcomes that the competitor's activity might have produced. However, when truthful information sharing is not possible, a researcher does not know how many of the opponent's experiments produced results.[4] Obviously, this missing information could help the researcher obtain a sharper estimate of the unknown distribution. More importantly, it would resolve the higher order uncertainty about the opponent's beliefs about the true distribution, and the opponent's beliefs about his beliefs about the true distribution, and so forth. Resolving this higher order uncertainty would thus also help the researcher assess the threat of being preempted.

We construct perfect Bayesian equilibria in symmetric threshold strategies, supplemented in the case of no information sharing by an experimentation cutoff at some arbitrary date. 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. Then each player has to form beliefs regarding the draws his opponent has received. These beliefs and the player's own results determine in turn the player's belief about the feasibility of a high-value outcome. In addition, they determine the likelihood that the other player stops in the current or the next period. In general, since the players' beliefs are private, it can be rather 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 "position" of his opponent (i.e., how many low-value draws the other player has ob-

[4]Although he knows that all successful experiments must have produced low values. Otherwise, the opponent would have already stopped.

tained) still depend on that player's own position (i.e., the number of low-value draws the player has himself obtained). In fact, the players' beliefs are positively correlated: as we show in Lemma 1, for any strategy s, at any time t, a player's belief about the draws of his opponent stochastically increase in the number of his own draws. This eventually enables us to construct symmetric equilibria in strategies involving time-dependent thresholds and experimentation cutoffs. Each player experiments until he receives a high-value draw or he accumulates too many low-value results relative to the amount of time that has elapsed or simply until a certain time cutoff is reached. Note that time cutoffs occur naturally in many real-world settings, e.g., as a consequence of deadlines or of opportunity costs that increase as additional time is invested in a research project.

Our paper is related to two bodies of work. One deals with experimentation and learning when there is no threat of preemption. We have already mentioned the multi-armed bandit models (for a recent survey of this literature, see Hörner and Skrzypacz, 2016).[5] 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. The risky project is characterized by uncertainty about the 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. Here, the potential value of an observation is only realized when the player is first to disclose it (i.e., claims the safe option). The problem of information sharing in two-armed bandit frameworks is investigated by Akcigit and Liu (2015), who show that a researcher who finds that the risky arm is unprofitable has the incentive to keep his information secret, so as potentially to investigate the safe arm alone for some time. Like in our setting, each player has private beliefs about the position of his rival, that is, whether his rival has already switched to the safe arm. However, for any strategies, these beliefs are monotonic, a property not necessarily 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.[6]

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

[6]Private signals in patent races have been introduced by Reinganum (1982). Choi (1991) considers a

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 with a low-value outcome is not possible.[7] By contrast, the feature that researchers may *"publish their partial findings quickly, rather than dropping the bombshell of a completely solved problem on their surprised colleagues"* (Hagstrom, 1974, p.7) is an essential ingredient of our model. The possibility of preemption gives rise to different learning dynamics in our model: each player's beliefs regarding the position of his opponent are not only used to estimate the likelihood of achieving a high-value outcome, but also the probability of being preempted with a low-value result. Halac et al. (2016) investigate experimentation with uncertain outcomes and learning, but use a different model and address different questions. In their model, players engage in successive experiments that can either generate an innovation or yield nothing. Hence, as in Moscarini and Squintani (2010), preemption with a low-value outcome is not possible in their setting. The learning dynamics are therefore different from those in our model. Their paper focuses rather on the design of experimentation contest, i.e., the question of how to provide incentives for experimentation effort by designing prize-sharing schemes and information disclosure policies.[8]

The other related body of work deals with experimentation and preemption when there is no 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.[9] They find that private information about each

---

winner-take-all race in which participants have imperfect but symmetric information about the arrival rate of the R&D process.

[7]Related is also 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.

[8]When successes are immediately disclosed (a *public contest*), it is found that a winner-takes-it-all scheme incentivizes the agents best. However, without disclosure of innovations (a *hidden contest*), it is optimal to divide the prize equally among all agents achieving the innovation. In total, under mild conditions (the requirement that earlier successes are rewarded weakly more than later successes), the optimal contest is one in which the prize is shared equally and disclosure occurs when a certain number of successes is achieved.

[9]A similar model has been introduced by Lippman and Mamer (1993). These authors, however, restrict their analysis to stationary strategies and construct Nash equilibria in time-invariant thresholds in which the players' beliefs about the value accumulated by their opponents (and therefore, about the likelihood of preemption) do not affect their decisions. In our paper, as in Hopenhayn and Squintani, players are allowed

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 multi-arm bandit, unlike in Hopenhayn and Squintani, where the players accumulate outcomes from a known distribution. Thus, in our problem, there are gains from sharing information regarding the draws the players obtain, in terms of learning about the unknown distribution, that are not present in the model of Hopenhayn and Squintani. Therefore, common learning can lead to more efficient outcomes than private learning in our research model with uncertain outputs, which is in sharp contrast to the results of Hopenhayn and Squintani. Bobtcheff at al. (2017) consider preemption in a model where two researchers privately have breakthroughs and decide how long to develop their ideas before disclosing them.[10] In their model, the returns to maturation are known with certainty, whereas the researchers' breakthrough times are random variables. By contrast, our paper focuses on situations in which the researchers face uncertainty about the distribution of potential returns and learn about both the project's potential value and the threat of being preempted. 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.[11]

To our knowledge, ours is the first paper that considers experimentation and preemption in the presence of uncertainty about research outcomes and learning about the underlying distribution.

The paper is organized as follows. In the next section, we present our model. In section 3, we analyze the single-player case. In section 4, we analyze the two-player case, under the assumption of information sharing. In section 5, we consider the case in which the two

to change their strategies over time, so that the constructed equilibria are in strategies with time-dependent thresholds.

[10]The term 'experimentation' in our paper encompasses the research activity studied in their paper.

[11]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).

players cannot observe one another's draws. We provide a comparison between the two information settings in Section 6, and conclude in section 7.

2 Model

Two players, 1 and 2, engage in a stopping game of successive experiments, taking place in discrete time periods. In each period $t \in \mathbb{Z}^+$, as long as the game continues, each player runs a new experiment. With probability $1-r$, $r \in (0,1)$, the experiment carried out by player $i \in \{1,2\}$ in period t is unsuccessful and fails to produce a valuable result. With probability r, the experiment is successful and the player receives a draw $x_t^i \in \{L, H\}$ where $0 < L < H$, interpreted as the value of the outcome for the experiment carried out in period t. 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

$$x_t^i \equiv L.$$

That is, whether an outcome of value H is at all possible is unknown to the players. At the beginning of the game, the distribution is chosen randomly (by *nature*) with probabilities p and $1-p$, respectively, in a manner unobservable to the players, and remains the same throughout the game. Conditional on the choice of distribution, the values x_t^i are independent across players and across periods.[12] 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

[12]Notice that this sampling procedure is equivalent to that in which both players receive draws $\tilde{x}_t^i \in \{0, L, H\}$ in every period with certainty, each of them from a distribution that attaches probability $\tilde{r}_0 = 1-r$ to $\tilde{x}_t^i = 0$, probability $\tilde{r}_L = r\,(1-pq)$ to $\tilde{x}_t^i = L$, and probability $\tilde{r}_H = r\,pq$ to $\tilde{x}_t^i = H$.

own draws.[13]

In 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).[14] 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.[15] 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.[16]

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

a. Player i's own past draws $x_\tau^i \in \{\emptyset, L, H,\}$, for $\tau = 0, 1, ..., t$, where $\emptyset$ denotes the occurrence of no draw;

b. Player j's past draws $x_\tau^j \in \{\emptyset, L, H,\}$, for $\tau = 0, 1, ..., t$, when their observation is possible;

c. Trivially, the two players' past decisions to continue, (c, c), for $\tau = 0, 1, ..., t-1$.

[13]In particular, we assume that each player i can observe the draw of his opponent, x_j^t, at the time of its occurrence, prior to his decision in period t; the analysis of the case in which a player can observe x_j^t after his decision in period t would require only a slight modification of our argument.

[14]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 non-trivial part of the value that the other player can claim.

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

[16]Our analysis extents 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.

A strategy of player i in period t indicates whether the player stops or continues in 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 an infinite sequence of time-t strategies,

$$\sigma^i \;=\; \{\sigma_t^i\}_{t=0}^{\infty}.$$

Equivalently, 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.

3 The Single-Player Problem

We start our analysis by examining the benchmark case in which experimentation is carried out by only one player. 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 so as 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 *multi-armed 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.[17] The solution to the multi-armed bandit problem is based on the *Gittins' index rule.* According to the rule, the player should remain in the game and

[17]See, e.g., Bertsekas (2001, Section 1.5). It helps to think that in each period t the player receives a draw with certainty; and this draw takes a value 0, L or H, respectively with probabilities $1-r$, $r(1-p_tq)$ and rp_tq, where $p_t \in (0,1)$ reflects the player's beliefs regarding the feasibility of H, at the beginning of period t.

keep experimenting with the stochastic arm, until its index falls below a certain threshold and then exit the game by choosing the sure arm.[18]

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.[19] Therefore, the expected value of staying in the game one more period, i.e., using the stochastic arm one more time, weakly decreases as the game progresses.[20] Hence, this is the *deteriorating case* of the multi-armed 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 its 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 get 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

[18]For reference, see Bertsekas (2001), vol. II, section 1.5. Our setting is a one-site version of the "Treasure Hunting" problem, analyzed in Example 1.5.1.

[19]Simply notice that $p(n)/p(n-1) < 1$, for all $n \in \mathbb{N}$.

[20]To be precise, we can redefine (or just re-interpret) p_n to be the probability that a *profitable* draw of H is feasible; then $p(n) = 0$, following the first draw of H.

$V(n+1)$. Finally, with probability $1-r$, he does not get 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, 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 stop otherwise. In particular, since the probability $p\,(n) \to 0$ as $n \to \infty$, player i will stop experimenting after receiving a finite number of L draws.

Finally, in our subsequent analysis of the impact of rivalry we have to adjust for the mere duplication of experiments in the setting with two players, as compared to the single-player case. For this, we slightly modify the single-player setting to allow the player to receive up to 2 draws in each period. 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

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

Our previous analysis implies that the single player will continue experimenting 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 next two sections, 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 single player benchmark developed in this section.

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. In our setting, this information is the direct consequence of our observability assumption; and it can stem, for example, from truthful communication between the players. Equivalently, this could be also the indirect consequence of perfect positive or negative correlation between the arrival times of the two players' draws.[21]

Thus, in every period $t \geq 0$, 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 in the next period, by either player, 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 if one or both players receive an H draw or 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; and this threshold is obtained by solving the Bellman equation describing each player's continuation problem.

At any time 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.[22] Thus, in the sequel, while analyzing the players' continuation and stopping incentives, we can restrict attention

[21] This alternative modeling approach was followed in an earlier version of the present work.

[22] In our setting, the continuation costs come only from the depreciation of the value that a player can claim in the current period, either because of time or because of preemption by the opponent. Therefore,

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

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

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 $V_{t+1}^i(n_t+1) = V_{t+1}^i(n_t+2) = L/2$ and $V_t^i(n_t) = V_{t+1}^i(n_t)$, so that player i's value function becomes

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

Hence,

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

there is no continuation cost for a player who cannot stop for a positive payoff.

that is, 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 i and the time t; it depends only the total number of draws n_t. Since the belief $p(n)$ 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 $V_t^i(n_t) > L$, 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.

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 $V_t^i(n_t, 0)$, is

$$V_t^i(n_t,0) = \max\{L,\ \delta\,[p^H(n_t)\frac{H}{2} + r^2(p(n_t)(1-q)^2 + (1-p(n_t)))V_{t+1}^i(n_t+2)$$
$$+\, r(1-r)(1-p(n_t)q)V_{t+1}^i(n_t+1,0) + r(1-r)(1-p(n_t)q)V_{t+1}^i(n_t+1) + (1-r)^2V_{t+1}^i(n_t,0)]\}$$

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

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

Hence,

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

that is, player i's continuation payoff 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 $V_t^i(n_t,0) < \delta L$, 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 $V_t^i(n_t,0)$ is decreasing in n_t; and 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 $V_t^i(n_t,0) > L$ we can define the threshold number of draws

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

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 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* $\sigma^* = \{\sigma_t^*\}_{t=0}^{\infty}$, 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.*

The equilibrium has a simple structure. The players 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. 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. As a consequence, the players decide to stop simultaneously in equilibrium.

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.

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 will experiment 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.[23]

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

5 Private Learning

We are now turning 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 in the 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 make the hypothesis that player j has received no draw of H either; otherwise, player i's decision will have no effect upon his payoff.[24] 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.[25]

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.[26] 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.[27] 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 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*

[24]Obviously, player j will stop if he has received a draw of H, for a zero payoff for player i.

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

[26]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 $n_t^i = 1$.

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

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 $h_t(n^j, s^j) \le \binom{t+1}{n^j}$ 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 $p_t(\,\cdot, n^i, s^j)$.

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 $N_\tau^j \le \hat{N}_\tau^j$, for all $\tau < t$. Then, for all n_t^i, the distribution $p_t(\,\cdot, n_t^i, \hat{s}^j)$ describing player i's beliefs about n_t^j at time t, conditionally on player j having received no draw of H, first-order stochastically dominates the distribution $p_t(\,\cdot, n_t^i, s^j)$.*

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, in any period $t+1$, with thresholds $N_{t-1}^j > N_t^j \ge N_{t+1}^j$, player i updates his beliefs about n_{t+1}^j in a manner that can make H more likely to be feasible and stopping by player j less likely to occur in that period. 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 consider equilibria in which after some period T the two players stop experimenting independently of their histories and beliefs. We say that an equilibrium involves an *experimentation cutoff at time T*, if the stopping thresholds in all periods $t \ge T$ are $N_t = 0$. Experimentation cutoffs arise naturally in several real-world settings; for example, after a period of time, a researcher may need to submit results or the aim of a research project may become obsolete.[28] In our setting, such cutoffs serve as a starting point for a backwards induction argument with which the players can calculate their best-response strategies.

In addition to an experimentation cutoff, we need a technical condition to guarantee the existence of an equilibrium in non-trivial symmetric strategies.[29] Condition 1, which we present in Appendix A, 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:

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

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 (and surely stopping) 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; and 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 $\{s_\tau^j\}_{\tau>t}$ to $\{\hat{s}_\tau^j\}_{\tau>t}$. Then player i's calculations about the benefits of further experimentation should not only involve 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

[28]The role of deadlines in experimentation activity is analyzed, for instance, by Bonatti and Hörner (2011) in a model of collaboration among researchers.

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

may not exist even for short time horizons. For example, when $\delta = 0.9$, $p = 0.8$, $q = 0.9$, $H = 8$, $L = 1$, with a time cutoff at $T = 1$, each player's strategy reduces to deciding whether to stop or to continue with one draw of L at the end of period $t = 0$. 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.[30]

Lemma 3 *Under Condition 1, for any threshold strategy σ^j involving an experimentation cutoff in period T, the best response of player i is characterized by stopping thresholds $\{N_t^i\}_{t=0}^{T-1}$ and an experimentation cutoff in period T.*

The 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; and 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.[31] 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 $\{N_t^j\}_{t=0}^{T-1}$ and an experimentation cutoff in period T. 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 at period t, when he has obtained n_t^i draws of L, is

[30] In this example, it is interesting to notice that the probability 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.

[31] 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.

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

defined recursively by equations (8)–(13) 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.

A strategy σ with thresholds $\{N_t\}_{t=0}^{T-1}$ and experimentation cutoff in period T 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.[32]

Proposition 2 *For any period $T \in \mathbb{Z}^+$ such that Condition 1 holds, there is a symmetric perfect Bayesian equilibrium in strategies with stopping thresholds $\{N_t\}_{t=0}^{T-1}$ and experimentation cutoffs in period T.*

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.[33] 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.

[32]Although we allow for arbitrarily long experimentation until the cutoff T, there exist equilibria in strategies without cutoffs. For a simple example, suppose that $q = 1$, so that the first draw reveals perfectly whether H is feasible; in this case, there is an equilibrium with an infinite sequence of thresholds, $N_t^i \equiv 1$, for all $t \in \mathbb{Z}^+$, allowing the two players' experimentation to exceed any cutoff time with positive probability.

[33]It is Condition 1 which ensures that the increase in N_t^j is more positive for a player who chooses to continue rather than to stop, also when one accounts for a potential change in the other player's continuation strategy in period $t + 1$.

6 Comparison of Common and Private Learning

Clearly, both common and private learning generate less experimentation than the single agent case where there is no concern for preemption. The more important and policy relevant comparison is the one between the common and private learning settings. The intuition that private learning would soften competition and lead to longer experimentation and higher values turns out to be incomplete. This is because under private learning players are not able to coordinate on the information they obtain during experimentation. As the two examples below indicate, this coordination failure can result in either shorter or longer experimentation horizons.

Example 1 *Suppose that* $\delta = 0.9$*,* $H = 8$*,* $L = 1$*,* $r = 0.75$*,* $p = 0.6$*,* $q = 0.5$*; and that there is a cutoff time* $T = 4$*. In this setting, a single player will keep experimenting until he receives* 6 *draws of* L*. With two players and common learning, the players will stop simultaneously after receiving a total of* $N = 3$ *draws of* L*, for expected experimentation time of* 1.811 *periods and value of* 4.568.[34] *With private learning, there are multiple equilibria. The one resulting in the most positive outcome is in thresholds* $(2, 1, 2, 1, 0)$*, with expected experimentation time of* 1.641 *periods and value of* 4.406*. In both cases, the two players never stop in the first period; in the second period, however, under private learning, a player would stop with a single draw even if his opponent adopted a threshold of* 2 *in that period.*[35] *That is, each player's beliefs about the draws of his opponent make him stop even when he would prefer to continue, if he knew the actual number of the draws obtained.*

Example 2 *Similar to the previous example, suppose that* $\delta = 0.9$*,* $H = 8$*,* $L = 1$*,* $r = 0.75$*,* $p = 0.6$*,* $q = 0.6$*; and that there is a cutoff time* $T = 4$*. Because of the higher value of* q*, a single player will now experiment until he receives* 5 *draws of* L*. Under common learning, the two players will stop simultaneously after receiving a total of* $N = 2$ *draws of* L*, for expected experimentation time of* 1.340 *periods and value of* 4.515*. Under private*

[34] For a more direct comparison, we have "truncated" the players' strategies under common learning, incorporating a cutoff time $T = 4$, utilizing the fact that immediate stopping by the two players constitutes an equilibrium in any continuation game.

[35] In particular, in any strategy profile involving thresholds of 2 in the first two periods, each player is better-off deviating to stopping with one draw in the second period.

learning, again there are multiple equilibria, with $(2,1,2,1,0)$ *resulting in the most positive outcome, expected experimentation time of* 1.588 *periods and value of* 4.68*. Contrary to the previous example, here, the players' uncertainty about the draws of their opponent helps them experiment longer, by preventing them from stopping when they both receive a draw in the first period.*

Our simulations indicate that common learning can generate more experimentation and higher value than private learning for a relatively large set of parameters. In fact, it is the opposite conclusion, in favor of private learning, that appears to be true only in limited cases. For instance, in the graphs below, corresponding to the parameters of Examples 1-2, private learning produces better expected outcomes only for values of $q \approx 0.6$.

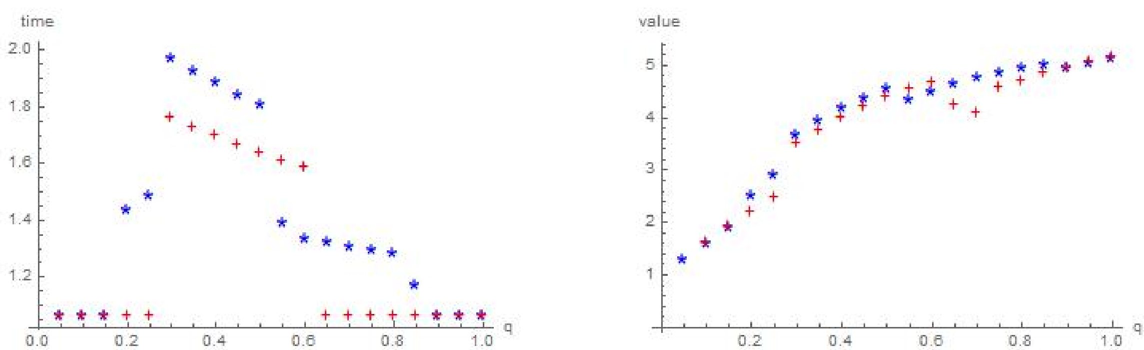

Figure 1: Experimentation length and value for common ($*$) and private ($+$) learning, when $\delta = 0.9$, $H = 8$, $L = 1$, $r = 0.75$, $p = 0.6$, for a cutoff time $T = 4$.

Figure 1 also demonstrates the effect of the parameter q upon the players' experimentation strategies. 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; and 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 equilibria 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.[36] Finally, although higher values of q correspond to more efficient experimentation, that is, to experiments more likely to result in H, also more informative for the players, the players' expected payoff is not monotonically increasing in q. For both common and private learning, since increasing q eventually results in equilibrium strategies that involve less experimentation, the players' expected payoff may also be lower when q is higher.

From the remaining parameters, p and H determine the initial position of the two players without affecting the experimentation dynamics. On the other hand, as the next example shows the probability r, which reflects how often experiments are successful, impacts the length of experimentation under the two regimes.

Example 3 *Suppose that* $\delta = 0.9$*,* $H = 8$*,* $L = 1$*,* $p = 0.6$*,* $q = 0.6$*; that* $r = 0.5, 0.6$ *or* 0.7*; and that there is a cutoff time* $T = 4$*. Under all three values of* r*, a single player would experiment until he receives* 5 *draws of* L*; two players under common learning would experiment until they receive* $N = 2$ *draws of* L *in total; and with two players under private learning, there are multiple symmetric equilibria, the most efficient of which is* $(2, 1, 2, 1, 0)$*. Therefore, the above change in the probability* r *does not affect the players' equilibrium strategies in each setting. However, it turns out that it affects the length of experimentation and the payoff comparison between the two modes of learning. In particular, as* r *increases, the probability of stopping in the first period increases under common learning but remains zero under private learning. Thus, although common learning is more efficient when* $r = 0.5$ *or* 0.6*, private learning becomes more efficient when* $r = 0.7$*. In the latter case, again, as in Example 2, the two players benefit from the absence of information exchange.*

[36] 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.

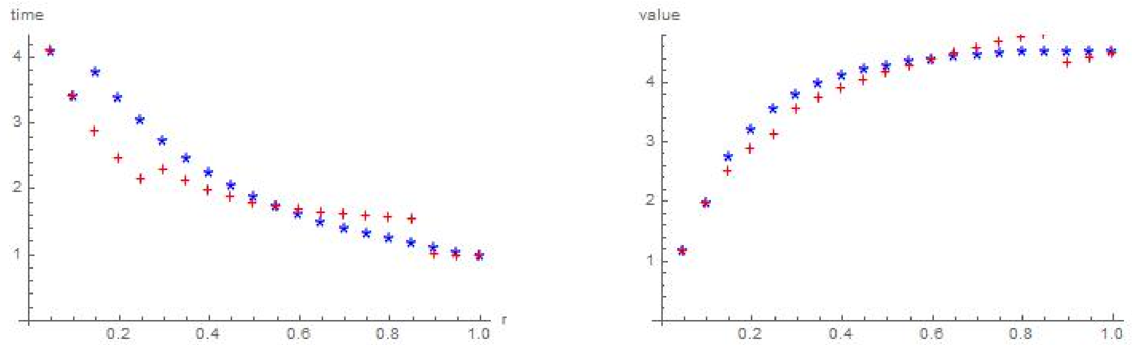

Figure 2: Experimentation length and value for common ($*$) and private ($+$) learning, when $\delta = 0.9$, $H = 8$, $L = 1$, $p = 0.6$ $q = 0.6$, for a cutoff time $T = 4$.

The setting of the last example allows us to consider the way 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 rise 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; and 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 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 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 keep experimenting, trying to obtain a high-value outcome, until their beliefs about its feasibility become too pessimistic. Because of the possibility of preemption, experimentation lasts 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 do not only depend 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 existence of equilibria in strategies involving non-monotone time-variant thresholds and experimentation cutoffs.

Information sharing is an important variable that can be influenced by policy and our paper sheds light on which information sharing regime, public or private, generates longer experimentation horizons and values for the 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 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 would push the player to prematurely stop experimenting.

Our model can be extended in at least two directions although both extensions would introduce substantial added complexity. One extension would be to introduce a richer range of potential payoffs with values beyond L and H. The issue of tractability of beliefs should then become critical. Another extension would involve relaxing the severity of preemption by allowing a player to continue experimentation, if his rival's exit does not exhaust all potential rewards. In this case, one has to consider strategies without cutoffs, leading to different forms of equilibria.

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

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

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

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

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

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

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

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

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

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

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

[13] Halac, M., Kartik, N. and Liu, Q. (2016), Contests for experimentation, *Journal of Political Economy*, forthcoming

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

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

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

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

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

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

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

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

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

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

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

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

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

[27] 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

Appendix A: Complete 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 the cutoff time* 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* $N_\tau = \tau-(t-N)+1$*, for* $\tau \geq t-N$*.*

Using the expression for $p_t(n, n_t^i, s^j)$ in Lemma 1, with $n_t^i = 1$ and $s^j = \underline{s}$, the inequality in Condition 1 becomes

$$r^N\left[p(1-q)^{N+1}+(1-p)\right]\;\left[p(2N)\left[1-(1-rq)^2\right] \;-\; \frac{1-\delta}{\delta}\,\frac{L}{H-L}\right] \;+$$
$$\sum_{n<N} \binom{N}{n}\, r^n\,(1-r)^{N-n}\left[p(1-q)^{n+1}+(1-p)\right]\,p(n+1)\left[(1-rq)^{2(T-t)} \;-\; (1-rq)^2\right] \;\geq\; 0,$$

for all $N \leq t+1$, for all $t < T$, which is easier to check.

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 Lemma 1:

It is clear that 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 shall 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 (strong) induction on the number of draws.

First, in period t, for any $t \in \mathbb{Z}^+$, 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 $V_{t+1}^i(n_t+1) = V_{t+1}^i(n_t+2) = L/2$ and $V_t^i(n_t) = V_{t+1}^i(n_t)$, so that player i's continuation payoff becomes

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

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

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

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

Now, let $n_t < N_1 - 2$, if feasible, and suppose that $V_{t+1}^i(n) \geq L$, 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

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

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

Second, suppose that player j has obtained no draw up to period t, for $t \in \mathbb{Z}^+$. As argued in the text, if $H/L < (3-2rq)/(2-rq)$, we have $V_t^i(n_t, 0) < \delta\, L$, 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 $V_t^i(n_t, 0) \geq L$, 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.[37] 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 $n_t \geq N_2$.

$\square$

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

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 be stopping simultaneously; otherwise, the preempted player would be able to profit by deviating from his strategy 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 his 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.

$\square$

Proof of Corollary 2:

Comparing the inequalities in (4) and (7), defining the thresholds N^* and N_2, we find that a player's gain from continuing experimenting for exactly one more period is larger when he is alone.

$\square$

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

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

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

$$\frac{dp_t}{dn^i_{t,}}(n^j_t, n^i_t, s^j) \gtreqless 0 \iff \sum_{n=0}^{t} h_t(n, s^j)\ r^n\,(1-r)^{t-n}\,[(1-q)^{n^j_t} - (1-q)^n] \lesseqgtr 0,$$

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

$$\frac{dp_t}{dn^i_{t,}}(n^j_t, n^i_t, s^j) \gtreqless 0 \iff n^j_t \gtreqless \bar{n}^j_t$$

Let $\tilde{n}^i_t > n^i_t$. To show that

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

as required for first-order stochastic dominance, notice that

$$p_t(n^j_t, \tilde{n}^i_t, s^j) \gtreqless p_t(n^j_t, n^i_t, s^j) \iff n^j_t \gtreqless \bar{n}^j_t.$$

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

$$\sum_{n^j_t=0}^{n} [\,p_t(n^j_t, n^i_t, s^j) - p_t(n^j_t, \tilde{n}^i_t, s^j)\,] \ = \ -\sum_{n^j_t=n+1}^{t} [\,p_t(n^j_t, n^i_t, s^j) - p_t(n^j_t, \tilde{n}^i_t, s^j)\,]$$

so that again the sum is positive, as required.

$\square$

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 manner, 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 $M \leq t+1$, we have

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

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 $M \leq t+1$,

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

as required for for the result, if and only if

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

or, after canceling equal terms, if and only if

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

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 $m, n \leq t+1$ such that $m \leq M < n$.

Notice that for all strategies s with thresholds $\{N_\tau\}_{\tau=0}^{t-1}$ and any time $t_0 < t$, we have

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

where $h'_{t_0}[l,(N_\tau)_{\tau=0}^{t_0})$ 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

$$\begin{aligned}
&\sum_{k=0}^{n_t^j} \; h'_{t_0}[k,(\hat{N}_\tau^j)_{\tau=0}^{t_0}] \; h_{t-1-t_0}[n_t^j - k,(\hat{N}_\tau^j - k)_{\tau=t_0+1}^{t-1}] \\
&\qquad\qquad \times \sum_{l=0}^{n} \; h'_{t_0}[l,(N_\tau^j)_{\tau=0}^{t_0}] \; h_{t-1-t_0}[n-l,(N_\tau^j - l)_{\tau=t_0+1}^{t-1}] \;\; - \\
&\sum_{k=0}^{n_t^j} \; h'_{t_0}[k,(N_\tau^j)_{\tau=0}^{t_0}] \; h_{t-1-t_0}[n_t^j - k,(N_\tau^j - k)_{\tau=t_0+1}^{t-1}] \\
&\qquad\qquad \times \sum_{l=0}^{n} \; h'_{t_0}[l,(\hat{N}_\tau^j)_{\tau=0}^{t_0}] \; h_{t-1-t_0}[n-l,(\hat{N}_\tau^j - l)_{\tau=t_0+1}^{t-1}] \;\; \leq \;\; 0
\end{aligned}$$

Since $\hat{N}_\tau^j = N_\tau^j$, for all $\tau > t_0$, this reduces to showing that

$$\begin{aligned}
&\sum_{k=0}^{m} \sum_{l=0}^{n} \; h_{t-1-t_0}[m-k,(N_\tau^j - k)_{\tau=t_0+1}^{t-1}] \; h_{t-1-t_0}[m-l,(N_\tau^j - l)_{\tau=t_0+1}^{t-1}] \\
&\qquad\qquad\qquad \times \left[\begin{array}{l} h'_{t_0}[k,(\hat{N}_\tau^j)_{\tau=0}^{t_0}] \; h'_{t_0}[l,(N_\tau^j)_{\tau=0}^{t_0}] \; - \\ h'_{t_0}[k,(N_\tau^j)_{\tau=0}^{t_0}] \; h'_{t_0}[l,(\hat{N}_\tau^j)_{\tau=0}^{t_0}] \end{array} \right] \;\; \leq \;\; 0
\end{aligned}$$

or, after again canceling equal terms, that

$$\sum_{k=0}^{m} \sum_{l=m+1}^{n} h_{t-1-t_0}[m-k,(N_\tau^j-k)_{\tau=t_0+1}^{t-1}]\ h_{t-1-t_0}[m-l,(N_\tau^j-l)_{\tau=t_0+1}^{t-1}]$$

$$\times \begin{bmatrix} h'_{t_0}[k,(\hat{N}_\tau^j)_{\tau=0}^{t_0}]\ h'_{t_0}[l,(N_\tau^j)_{\tau=0}^{t_0}] - \\ h'_{t_0}[k,(N_\tau^j)_{\tau=0}^{t_0}]\ h'_{t_0}[l,(\hat{N}_\tau^j)_{\tau=0}^{t_0}] \end{bmatrix} \leq 0$$

for all $m, n \leq t+1$ such that $m \leq M < n$.

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

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

$\square$

Proof of Lemma 3:

We argue by means of backwards induction, in periods $T, T-1, \ldots, 1, 0$, 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.

Suppose that player j's strategy s^j is such that he stops in periods $t < T$ if and only if $n_t^j \geq N_t^j$, for some sequence of thresholds $\{N_t^j\}_{t=0}^{T-1}$; and that he never continues experimenting beyond period T, where T is arbitrary, fixed.

Clearly, against such a strategy, player i will not continue experimenting beyond period T either, establishing his experimentation cutoff in that period.[38]

Moving backwards, suppose that player i has obtained $n_{T-1}^i > 0$ draws of L by the end of period $T-1$.[39] Then player i's expected payoff at the continuation game starting (and ending) in period T, conditionally on player j having obtained n_{T-1}^j draws of L and on the game reaching period T, is

$$U_T(n_{T-1}^i \,|\, n_{T-1}^j, s^j) \;=\; \begin{cases} (1/2)\,\delta\,[p^H(n_{T-1}^i + n_{T-1}^j)(H-L) + L], & n_{T-1}^j > 0; \\[2ex] (1/2)\,\delta\,[p^H(n_{T-1}^i)(H-L) + L] + \\ (1/2)\,\delta\,[1 - r\,p(n_{T-1}^i)\,q]\,(1-r)\,L, & n_{T-1}^j = 0. \end{cases} \tag{8}$$

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

$$\Delta V_{T-1}(n_{T-1}^i \,|\, n_{T-1}^j, s^j) \;=\; \begin{cases} -L/2, & n_{T-1}^j \geq N_{T-1}^j; \\[2ex] U_T(n_{T-1}^i \,|\, n_{T-1}^j, s^j) \,-\, L, & n_{T-1}^j < N_{T-1}^j. \end{cases}$$

[38]Notice that player i's incentive to stop experimenting in period T is weak, when he has received no draw up to that period; and strict, otherwise.

[39]When $n_{T-1}^i = 0$, player i has an incentive to continue into period T, independently of h_{T-1}^j.

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

$$\Delta V_{T-1}(n^i_{T-1}\,|\,s^j) \;=\; \sum_{n^j_{T-1}=0}^{T-1} p_{T-1}(n^j, n^i_{T-1}, s^j)\,\Delta V_{T-1}(n^i_{T-1}\,|\,n^j_{T-1}, s^j), \tag{9}$$

Under Condition 1, the function $\Delta V_{T-1}(\,\cdot\,|\,\cdot, s^j)$ is decreasing in n^j_{T-1}.[40] In addition,

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

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

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

[40] 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}.

with the second inequality being obtained from the fact that the 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 $T-1$,

$$V^*_{T-1}(n^i_{T-1}\,|\,s^j) \;=\; \max\{\,V^c_{T-1}(n^i_{T-1}\,|\,s^j),\, V^s_{T-1}(n^i_{T-1}\,|\,s^j)\,\} \tag{10}$$

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

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

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

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

$$
\begin{aligned}
U^*_{t+1}(n^i_t \,|\, s^j) &= U_{t+1}[n^i_t \,|\, s^j, (N^i_\tau)_{\tau=t+1}^{T-1}] \\
&= \hat{p}_H(n^i_t \,|\, s^j)\ (1/2)\, H \\
&\quad + [1 - \hat{p}^H_t(n^i_t \,|\, s^j)]\ \hat{p}^L_t(n^i_t \,|\, s^j)\ V^*_{t+1}(n^i_t + 1 \,|\, s^j) \\
&\quad + [1 - \hat{p}^H_t(n^i_t \,|\, s^j)]\ [1 - \hat{p}^L_t(n^i_t \,|\, s^j)]\ V^*_{t+1}(n^i_t \,|\, s^j)
\end{aligned} \tag{11}
$$

where

$$
\hat{p}^H_t(n^i_t \,|\, s^j) \;=\; \sum_{n^j_t=0}^{t+1} \hat{p}_t(n^j_t, n^i_t, s^j)\, p_H(n^j_t + n^i_t)
$$

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,

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

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

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

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.[41]

[41] In particular, $h'_t(n^j_t, s^j) \le \binom{t+1}{n^j_t}$ 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.[42]

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

$$\begin{aligned}
\Delta V_t(n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}) &= P[n^j_t < N^j_t \,|\, n^i_t, s^j]\; [\, U_{t+1}[n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}] - L\,] \\
&\quad + P[n^j_t \geq N^j_t \,|\, n^i_t, s^j]\; (-L/2) \qquad (12) \\
&= P[n^j_t < N^j_t \,|\, n^i_t, s^j]\; [\, U_{t+1}[n^i_t \,|\, s^j, (N^i_\tau)^{T-1}_{\tau=t+1}] - L/2\,] \;-\; L/2
\end{aligned}$$

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 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

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

and

[42] Simply, let $U(x) = \alpha(x)(H/2) + [1-\alpha(x)][\beta(x)V(x+1) + (1-\beta(x))V(x)]$, for $x \geq 0$, where $\alpha(\cdot), \beta(\cdot)$ are respectively decreasing and increasing probabilities, $V(x)$ is decreasing, and $V(x) \leq H/2$, for all $x \geq 0$. Then it is straightforward to show that $U'(x) \leq 0$, being the sum of non-positive terms.

$$\begin{aligned} V_t^*(n_t^i \,|\, s^j) &= V_t[n_t^i \,|\, s^j, (N_\tau^i)_{\tau=t+1}^{T-1}] \\ &= \max\{\, V_t^c[n_t^i \,|\, s^j, (N_\tau^i)_{\tau=t+1}^{T-1}],\ V_t^s[n_t^i \,|\, s^j, (N_\tau^i)_{\tau=t+1}^{T-1}] \,\} \end{aligned} \tag{13}$$

are decreasing in n_t^i, completing the induction.

$\square$

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 addition, throughout our argument, we make a hypothesis that the two players derive their beliefs regarding the private histories of their opponent in an ex ante symmetric manner.[43] In particular, each player attaches positive probability to the event that his opponent has received the same number of draws as he has. This hypothesis is validated at the end of the argument, by the symmetric character of the constructed equilibrium.

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 (8) and (9) in the proof of Lemma 3,[44]

$$\Delta V_{T-1}(n^i_{T-1}\,|\,s',\,N^j_{T-1}\,) \;=\; \sum_{n^j_{T-1}=0}^{T-1} p_{T-1}(n^j, n^i_{T-1}, s')\,\Delta V_{T-1}(n^i_{T-1},\,|\,n^j_{T-1}, s',\,N^j_{T-1}\,),$$

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 $N_{T-1} = 1$.[45]

Otherwise, 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}.

[43] In other words, in each period t, the mapping associating a player's number of L draws to his beliefs about his opponent's draws is the same for both players. Since this mapping is parametrized by the opponent's strategy and the time t, this symmetry is a consequence of the two players using identical strategies.

[44] 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}.

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

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

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

with $BR^i_{T-1}(s^j)=1$ when the set is empty, is an increasing function of the threshold N^j_{T-1} in the strategy s^j.[46]

The set $\{1,2,...,T+1\}$ is a lattice with respect to the order $\geq$, complete because of finiteness. Therefore, since the function $BR^i_{T-1}(\,\cdot\,|\,s')$ is increasing in the variable N^j_{T-1}, it has at least one fixed point. Hence, for each symmetric strategy $s'=\{N_t\}_{t=0}^{T-2}$ prior to period $T-1$, we can define the players' common threshold at time $T-1$ as the maximal fixed point of $BR^i_{T-1}(\,\cdot\,|\,s')$.[47]

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

[47]Equivalently, one can set

$$N_{T-1}\;=\;\max\{\,N=2,3\ldots,T:\;\Delta V_{T-1}[N-1\,|\,(N^j_t)_{t=0}^{T-2},N]\;>\;0\,\}\;+\;1,$$

where $\{N^j_t\}_{t=0}^{T-2},N$ are the thresholds characterizing the strategy of each player's opponent, with $N_{T-1}=1$, when the set is empty. To validate this definition, notice that for all $n^i_{T-1}<N_{T-1}$,

$$\Delta V_{T-1}[n^i_{T-1}\,|\,(N^j_t)_{t=0}^{T-2},N_{T-1}]\;\geq\;\Delta V_{T-1}[N_{T-1}-1\,|\,(N^j_t)_{t=0}^{T-2},N_{T-1}]\;>\;0,$$

since the function ΔV_{T-1} is decreasing in n^i_{T-1}; and that for all $n^i_{T-1}\geq N_{T-1}$,

$$\Delta V_{T-1}[n^i_{T-1}\,|\,(N^j_t)_{t=0}^{T-2},N_{T-1}]\;\leq\;\Delta V_{T-1}[n^i_{T-1}\,|\,(N^j_t)_{t=0}^{T-2},n^i_{T-1}+1]\;\leq\;0,$$

since the function ΔV_{T-1} is increasing in N^j_{T-1} and the threshold N_{T-1} has been defined as the maximal element in the set. Therefore, each player is willing to continue at the end of period $T-1$ if and only if he has $n^i_{T-1}<N_{T-1}$ draws of L, as required for a symmetric equilibrium in the continuation game starting in period $T-1$.

Moving backwards to periods $t = T-2, T-3, \ldots, 1, 0$, suppose that for each symmetric strategy profile with stopping thresholds $\{N_\tau\}_{\tau=0}^{t}$ up to the end of period t, there is a symmetric equilibrium $s''[(N_\tau)_{\tau=0}^{t}]$ for the continuation game starting in period $t+1$, with thresholds that depend on $\{N_\tau\}_{\tau=0}^{t}$. (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_t^i 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

$$\begin{aligned} \Delta V_t[(n_t^i \,|\, s(s', M)] &= P(n_t^j \geq M \,|\, n_t^i, s')(-L/2) \\ &+ P(n_t^j < M \,|\, n_t^i, s') \; [\, U_{t+1}[n_t^i \,|\, s(s', M)] - L\,], \end{aligned}$$

where $U_{t+1}(n_t^i \,|\, s(s', M))$, defined recursively by equations (8)–(13) 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 $U_{t+1}(n_t^i \,|\, s(s', M))$ is achieved with player i also following the continuation strategy $s''(s', M)$.

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

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

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

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

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

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

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 the cutoff T.

Therefore, we have

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

or, after some rearrangement of the terms,

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

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

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

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

$$\begin{aligned}& p_t(N, n_t^i, s')\ \left[p(2N)\,[1-(1-rq)^2] \;-\; \tfrac{1-\delta}{\delta}\ \tfrac{L}{H-L}\right] \\ + & \sum_{n_t^j=0}^{N-1} p_t(n_t^j, n_t^i, s')\ p(n_t^j)\,[(1-rq)^{2(T-t)} \;-\; (1-rq)^2] \;\;\geq\;\; 0\end{aligned}$$

After noticing that 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 that it achieves its minimal value for $n_t^i = 1$, it follows that for the inequality to hold it is sufficient that

$$p_t(N,1,s')\ \left[p(2N)\,[1-(1-rq)^2]\ -\ \tfrac{1-\delta}{\delta}\ \tfrac{L}{H-L}\right]$$
$$+\ \sum_{n_t^j=0}^{N-1} p_t(n_t^j,1,s')\ p(n_t^j)\,[(1-rq)^{2(T-t)}\ -\ (1-rq)^2]\ \geq\ 0,$$

which 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 $\Delta V_t[n_t^i \mid s(s', N_t^j)]$ 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,

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

with $BR_t^i(N_t^j \mid s') = 1$ 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 $\{1, 2, ..., t+2\}$ of possible thresholds in period t is a lattice with respect to the order $\geq$, complete because of finiteness. Therefore, since the function $BR_t^i(\,\cdot \mid s')$ is increasing in N_t^j, 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 $BR_{T-1}^i(\,\cdot \mid s')$; 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_0 for the first period of the game, with the impled strategy $[N_0, s''(N_0)]$ forming a symmetric perfect Bayesian equilibrium for the entire game.

$\square$