Strategic experimentation with privately observed payoffs*

Jérôme Renault†, Eilon Solan‡, and Nicolas Vieille§

May 12, 2026

Abstract

We study a strategic experimentation game with exponential bandits, in which experiment outcomes are private. The equilibrium amount of experimentation is always weakly higher than with public outcomes. Moreover, in pure equilibria, experimentation is never significantly below the socially optimal level and may be strictly higher. We provide a tight bound on the extent of over-experimentation. The analysis rests on a novel encouragement effect.

Keywords: Strategic experimentation, exponential bandit, private payoffs.

1 Introduction

In many dynamic economic environments, decision makers face uncertainty about the profitability of different strategies and must balance exploration of new opportunities with exploitation of known ones.

*Renault acknowledges funding from ANITI, grant ANR-23-IACL-0002, and from the ANR under the Investments for the Future program, grant ANR-17-EURE-0010. Solan acknowledges the support of the Israel Science Foundation, Grant #211/22. Vieille thanks the HEC Foundation for support.

†Mathematics and Statistics Department, Toulouse School of Economics, Université Toulouse Capitole, France. E-mail: jerome.renault@tse-fr.eu.

‡School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: eilons@tauex.tau.ac.il.

§Department of Economics and Decision Sciences, HEC Paris, 1, rue de la Libération, 78 351 Jouy-en-Josas, France. E-mail: vieille@hec.fr.

This tension is central to strategic experimentation, in which multiple players learn through costly actions while observing one another's choices. Classical examples include firms investing in uncertain technologies, policymakers testing new policies, or investors navigating uncertain markets. Decision makers learn both from their own experience and from others' behavior, even when underlying information cannot be directly observed.

Beginning with Rothschild [24] in the single-agent case, and with Bolton and Harris [3] in strategic settings, strategic experimentation has been studied extensively in two-arm bandit settings, see Bergemann and Välimäki [2] for a survey. In these models, players repeatedly choose between a safe arm and a risky arm of unknown quality.

In many applications, experimentation outcomes are best thought of as private. However, the literature has largely assumed that both actions and experimentation outcomes are public, so that players share a common belief on the risky arm's type. Exceptions include Rosenberg et al. [23], Murto and Välimäki [19], Heidhues et al. [10], and Bonatti and Hörner [4]. As noted by Hörner and Skrzypacz [12], strategic experimentation with observed actions but unobserved outcomes remains largely unsolved. The present paper contributes to this literature.

We study a discrete-time version of the standard model of exponential bandits introduced in Keller et al. [15]. Two players alternate over time in choosing between two arms. The safe arm is costless and yields zero payoff. The risky arm is costly and generates positive payoffs at random times, only if its type is good. Actions are public, whereas payoffs are private.

In the benchmark case in which outcomes are public, a robust finding is that equilibrium experimentation is socially suboptimal in discrete time, see Heidhues et al. [10].

When outcomes are private, a player is unsure whether to interpret the experiments of other players as evidence that their previous experiments were successful, or as pure experimentation. This opens the way for potential belief manipulation: by continuing to experiment, player i may induce player j to believe that a success has occurred, thereby encouraging j to experiment. This mechanism resembles the *encouragement effect* of Bolton and Harris [3], which they defined as *the prospect of future experimentation by others encourages agents to increase current experimentation.* The difference here is that experimentation primarily serves as a signal of past success. As we show, this effect implies that the equilibrium amount of experimentation is always higher

when outcomes are private than when they are public. In equilibrium, though, players' inferences take into account the possibility that experimentation may be deceptive: equilibrium inferences are highly complex and involve beliefs of all orders.

In the one-player case and in the two-player public case, the exploration/exploitation trade-off between the option value of an additional experiment and the opportunity cost of that experiment is captured via a single belief cut-off p^*, which dictates the optimal behavior. When outcomes are private, we introduce a new *encouragement cut-off* $\widehat{p}$ that plays a comparable role, and that balances the cost of one extra experiment with the value of *two* experiments. As we show, players do not experiment when their belief is below $\widehat{p}$, and often experiment when it is above $\widehat{p}$. These findings allow us to bound the equilibrium amount of experimentation in pure equilibria: absent conclusive news, the amount of experimentation is at least socially optimal (up to two experiments), and at most twice the socially optimal level. Pure equilibria do not display under-experimentation, but may display over-experimentation.

The related literature is discussed in Section 7. The paper most closely related to ours is Heidhues et al. [10], which looks at a similar model but assumes that direct, costless communication between the players is possible at any point in time. Since players have no incentive to lie once they know that the risky arm is good, truthful communication is feasible, and a chief question is to what extent truthful disclosure should be postponed.[1]

The paper is organized as follows. Section 2 introduces the model. Section 3 discusses the benchmark case in which experiment outcomes are public, and Section 4 contains first comparative results on the public and private case. Our equilibrium concept is defined in Section 5, and our main results are stated in Section 6. Proofs are in the Appendix, Sections A through F.

[1] A further notable difference lies in the solution concept. Once successful, it is strictly dominant for a player to pull repeatedly the risky arm. Yet, the sequential equilibria in Heidhues et al. [10] share the feature that "a player who was made subjectively certain of the good state by an opponent's announcement of a success maintains this belief even in the face of the opponent's subsequent use of the safe arm", see p. 541 in Heidhues et al. [10]. Such a behavior in equilibrium may be deemed undesirable. We use a refinement of sequential equilibrium that excludes it.

2 Model

We consider a discrete-time strategic experimentation game, following Keller et al. [15]. There are two available arms: a safe arm S and a risky one R. The safe arm yields a constant payoff, which we normalize to zero. The risky arm is either good (G) or bad (B) and entails an opportunity cost $c > 0$. The type $\theta \in \{G, B\}$ of the risky arm is drawn at time $t = 0$ and remains fixed thereafter. We denote by $p_0 := \mathbf{P}(\theta = G) > 0$ the common prior.

In the bad state $\theta = B$, the risky arm yields zero payoff. If $\theta = G$, each pull of the risky arm yields a payoff of either 0 or m, with probabilities $1 - \lambda$ and $\lambda \in (0, 1)$, respectively.

Over time, two players alternate in choosing one of the two arms, with player 1 acting first. For convenience, a *period* consists of two consecutive choices, one by each player. That is, in each period $t \geq 1$, player 1 moves first, followed by player 2. Choices are *publicly* observed, but payoffs are *private*.[2] The common discount factor is $\delta \in (0, 1)$. This completes the model description.

If $\lambda m \leq c$, it is optimal to always pull the safe arm, even when the risky arm is known to be good ($\theta = G$). We rule out this trivial case and assume throughout that $g := \lambda m - c > 0$.

A player *experiments* when choosing R; the experiment is *successful* if it yields payoff m. In line with most of the strategic experimentation literature, news is conclusive: when a player with belief p experiments,[3] his belief jumps to 1 following a success and updates to

$$\phi(p) := \frac{p(1-\lambda)}{p(1-\lambda) + 1 - p} < p$$

following a failure.

Since $g > 0$, a successful player optimally chooses R in all subsequent periods. However, because only choices are observed, this success is not directly observed by the other player, who has to draw inferences from the choices of the successful player. The endogenous probabilistic inferences from such observational learning are the topic of the paper.

[2] Payoffs are private throughout the paper, except in Section 3.

[3] Unless specified, we mean the first-order belief of that player, identified with the belief assigned to $\theta = G$.

The assumption that payoffs are private is the primary departure from the existing literature. While either the alternating-move assumption or the two-player assumption can be relaxed without significantly affecting our qualitative results, doing so entails a substantial loss of transparency. By contrast, the assumption that time is discrete plays a critical role in ruling out cheap talk opportunities. Indeed, in continuous time players can switch arbitrarily fast between the arms, allowing them to encode information at no cost.

3 The case of public payoffs

In this section (only), we assume that both payoffs and choices are public. Since the results here are largely well-known, we provide only a brief discussion.

3.1 The one-player case and the social optimum

Assume there is a single player with discount factor δ. The optimal policy is unique up to ties and is a cut-off policy: in a given period, the optimal choice is to experiment if and only if the current belief exceeds some cut-off value p^*_δ. The value of p^*_δ is pinned down by the condition that when holding the belief p^*_δ, the agent is indifferent between choosing the safe arm forever, and experimenting one last time. This indifference condition yields $(1-\delta)\left(p^*_\delta \lambda m - c\right) + \delta p^*_\delta \lambda g = 0$, that is,

$$p^*_\delta = \frac{c(1-\delta)}{c(1-\delta) + g(1-\delta(1-\lambda))}. \tag{1}$$

It is thus optimal to pull the risky arm $N^*_\delta := \inf\{n \geq 0 : \phi^n(p_0) < p^*_\delta\}$ times before switching forever to the safe arm if all experiments failed. We write N^* and p^* for N^*_δ and p^*_δ when the discount factor is clear from the context.

Assume now that there are two agents, P1 and P2, the strategies of which are dictated by a social planner whose objective is to maximize joint payoffs. We note that the game is equivalent to a game in which P1 and P2 would act in odd and even stages respectively, players receive no payoff when inactive, and share a discount factor of $\sqrt{\delta}$. With this formulation, the problem of the social planner reduces to that of a single agent with a discount factor of $\sqrt{\delta}$. Accordingly, in this social optimum, agents

alternate in pulling the risky arm until the common belief falls below $p^{**} = p^{**}_{\delta} := p^{*}_{\sqrt{\delta}}$, and the optimal number of experiments is $N^{**} = N^{**}_{\delta} := N^{*}_{\sqrt{\delta}}$.

3.2 The two-player case

When payoffs are public, the experimentation game may be reinterpreted as a stochastic game with perfect information, where the state variable is the common belief over θ. Given this interpretation, a *Markov* strategy is a function $f : [0,1] \to \Delta(\{S,R\})$, with the understanding that $f(p)$ is the mixed move selected when the current belief is p. A *symmetric Markov equilibrium* is a Markov strategy f, such that the profile (f,f) is a subgame perfect equilibrium (henceforth, SPE).

Proposition 1 below summarizes the main results for that case. In this statement, the role of the assumption on p_0 is to rule out non-generic cases, and $N_{\rm e} \leq +\infty$ is the total (random) number of experiments over time. We recall that the n-th iterate $\phi^n(p_0)$ is the common belief after n failed experiments.

Proposition 1 (Public payoffs) *Assume that $\phi^n(p_0) \neq p^*$ for each n. The following properties hold:*

P1 *At any Nash equilibrium, $N_{\rm e} = N^*$ almost surely conditional on $\theta = B$.*

P2 *There is a pure* SPE*, in which players experiment N^* times in a row, and then stop experimenting if unsuccessful.*

P3 *There exists a unique symmetric Markov equilibrium, but a continuum of (non-symmetric)* SPE *payoffs if $p_0 > p^*$.*

If $\theta = G$, agents behave as when the state is $\theta = B$ until one experiment is successful. That is, $N_{\rm e} = +\infty$ if one of the first N^* experiments is successful, and $N_{\rm e} = N^*$ otherwise.

According to **P1**, which mirrors Proposition 1 in Heidhues et al. [10], the number of experiments is the same across all equilibria, and is inefficiently low since $p^{**} < p^*$ implies $N^* \leq N^{**}$. According to **P3** however, while the equilibrium number of experiments is uniquely pinned down, their timing is not.

Remark 1 *Define the* myopic *cutoff as* $p_{\text{myop}} := \frac{c}{\lambda m} = \frac{c}{c+g}$*, and note that for* $p > p_{\text{myop}}$*, the risky arm is more informative and produces a higher current reward than the safe arm.*

Let σ *be the Markov profile in which the active player experiments if and only if the current belief is above* p^**. We note that* σ *is not an SPE if* $p_0 \in (p^*, \min(p_{\text{myop}}, \phi^{-1}(p^*))$*. The reason is that given* σ*, player* 1 *faces the following trade-off at the initial period. If he chooses* R *as prescribed by* σ*, no one will experiment further, since* $\phi(p_0) < p^*$*. If he deviates to* S*, player* 2 *will experiment once, because the current belief will still be* p_0*. That is, deviating to* S *avoids paying the cost of experimenting and does not affect the amount of experimentation. Since* $p_0 < p_{\text{myop}}$*, deviating to* S *is profitable.*[4]

4 The private case: a first benchmark

4.1 Strategies

We introduce some terminology. *Full histories* specify entirely the play up to the current period, that is, the realized state θ, the sequence of past choices, and all experiment outcomes. *Public histories* list past choices and are therefore finite sequences of R's and S's. We denote by $H := \{R, S\}^{<\mathbb{N}}$ the set of public histories. Public histories play a leading role, and we reserve the symbol h to such histories. We note that P1 (resp., P2) is active at h if and only if the length $|h|$ of h is even (resp., odd). *Private histories* of player i specify in addition whether i's experiments were successful. A strategy σ^i of player i maps the set of i's private histories (where active) into the set $\Delta(\{R, S\})$ of mixed moves.

Once a player is successful, his only sequentially rational continuation strategy is to choose repeatedly the risky arm. When constructing sequential equilibria, we will refrain from repeating that players choose R at such private histories. To simplify notations, we thus define a strategy profile as a map $\sigma : H \to \Delta(\{S, R\})$, with the understanding that $\sigma(h)$ is the mixed move at h of the active player if (i) the sequence of previous choices is h and (ii) all the past experiments of this active player failed.[5]

[4] If instead $p_0 \in [p_{\text{myop}}, \phi^{-1}(p^*)]$ (which requires $\phi^{-1}(p^*) \geq p_{\text{myop}}$), then σ is an SPE.

[5] The choice of this concise notation will raise two issues. The first will arise when specifying off-path beliefs in a sequential equilibrium: when checking the consistency of beliefs and strategies, we will have to allow for completely mixed strategies that play S with positive probability, even when

Thus, $\sigma(h)$ specifies behavior only at histories where the active player has not yet observed a success. For $h \in H$, we denote by $n_{\mathrm{e}}(h)$ the total number of experiments along h.

There is one case where the optimal arm choice is clear. Assume the active player holds the belief p, and chooses to experiment. His expected current reward is $p\lambda m - c$ and his expected continuation payoff is at most pg, irrespective of past play and of players' continuation strategies. If instead the active player chooses the safe arm, he can guarantee a non-negative payoff (e.g., by always choosing the safe arm). Thus, choosing the safe arm strictly dominates experimentation when $(1-\delta)(p\lambda m - c) + \delta pg < 0$, that is, when

$$p < \widetilde{p} := \frac{c(1-\delta)}{(1-\delta)\lambda m + \delta g}. \tag{2}$$

Note that $\widetilde{p} < p^*$. We denote by $\widetilde{N} = \widetilde{N}_\delta := \min\{n \geq 0 \colon \phi^n(p_0) < \widetilde{p}\}$ the number of failures needed to push the belief below $\widetilde{p}$.

4.2 Private vs. public payoffs: a first comparison

Our first main result states that the equilibrium amount of experimentation is *always* higher when payoffs are private rather than public.

Theorem 1 (Private payoffs) *Assume that $\phi^n(p_0) \neq p^*$ for each n. For any Nash equilibrium σ of the game with private payoffs, $\mathbf{P}_\sigma(N^* \leq N_{\mathrm{e}} < +\infty \mid \theta = B) = 1$.*

If $\theta = G$, the distribution over histories coincides with the distribution if $\theta = B$, as long as all experiments fail. On the equilibrium path, if one experiment is successful, the successful player repeats R forever, and the other player eventually sticks to R.

This result is driven by a form of encouragement. If player j sticks to the risky arm, player i will put some weight on the possibility that j was successful and will therefore be more optimistic than if j's failures were observed by i. As a result, i will experiment longer ($N_{\mathrm{e}} \geq N^*$).

This intuition leaves open the possibility that, by mutual encouragement, players might engage in an endless phase of experimentation, each player attributing the other

successful. The second issue will arise when discussing Nash equilibria, since Nash equilibria may play dominated continuation strategies off-path. We will deal with these issues when needed, without introducing additional notations.

player's insistence on R as evidence that he (the other player) was successful, and their own failures to bad luck. As the second part of Theorem 1 shows, this cannot be the case, and experimentation must eventually stop in the absence of conclusive evidence ($N_{\mathrm{e}} < +\infty$).

Proof Sketch. The formal proof is in Section B of the Appendix. Fix a Nash equilibrium σ.[6] The proof that $N_{\mathrm{e}} \geq N^*$ with probability 1 builds on the above intuition. Fix any on-path public history h. If players knew that they were both unsuccessful so far, they would share the belief $\phi^{n_{\mathrm{e}}(h)}(p_0)$. Since outcomes are private, the belief $p^i(h)$ of each player i at h accounts both for this worst-case scenario and for the possibility that j may have been successful, implying that $p^i(h) \geq \phi^{n_{\mathrm{e}}(h)}(p_0)$. The rest of the argument is quite similar to the proof for the public case.

To prove that $N_{\mathrm{e}} < +\infty$ if $\theta = B$, we first note that at equilibrium, players do not switch infinitely often between the two arms. The argument proceeds as follows. Assume there is an on-path history h that ends with S and such that $n_{\mathrm{e}}(h) > \widetilde{N}$. The player active at h infers that all experiments so far were failures. Therefore, he holds the belief $\phi^{n_{\mathrm{e}}(h)}(p_0) < \widetilde{p}$, and must play S. The same argument holds for the other player at the history hS. Hence $\sigma(h) = \sigma(hS) = S$, and the equilibrium sequence of choices following h must be S^∞. This implies that, with probability 1, the sequence of choices ends either with S^∞, or with R^∞.

Hence, with probability 1, one of the two arms is used finitely many times. Assume that conditional on B, there is a positive probability that S is pulled only finitely many times. On that event, each player i eventually assigns a probability arbitrarily close to 1 to the event that the other player, j, will always choose R in the future, independently of the experiments' outcomes. Consequently, player j's choices become increasingly non-informative from i's perspective. Eventually, the updating of $p^i(h)$ is mostly based on i's failures, so that $p^i(h) \to 0$, implying that i will eventually switch to S. ■

[6] Our proof allows for Nash equilibria in which a player may pick S after being successful.

5 The private case: conceptual issues

This section clarifies our solution concept, which is a refinement of sequential equilibrium (henceforth, SE). We start by discussing a simple Nash equilibrium σ_0. This discussion has two purposes. One is to provide simple insights into the construction of equilibria. The second is to motivate our concept. According to σ_0, deviations of player i to the safe arm are 'interpreted' by player j as evidence that player i was successful earlier, and 'trigger' an infinite sequence of experiments of j. That is, player j assigns probability 1 to player i having made a strictly dominated choice. While σ_0 is not part of a SE,[7] there are pure SEs that rely on such unreasonable beliefs. Following earlier literature, we will explicitly rule out such beliefs.

5.1 A simple pure Nash equilibrium

Constructing a pure Nash equilibrium is easy. Let $p_0 \geq p^*$ be given, and consider the infinite sequence of choices $h^*_\infty := (RR)^{N^*} S^\infty$. Define σ_0 as the strategy profile that follows the sequence h^*_∞ as long as past choices are consistent with it, and that selects R otherwise. That is, $\sigma_0(h) = S$ if and only if h is a prefix of h^*_∞ with length $|h| \geq 2N^*$.

Put differently, each player experiments for N^* periods under σ_0, then discloses his private information in period N^*+1. If P1 chooses R in period N^*+1, then P2 infers that $\theta = G$ and chooses R ever after. If P1 chooses S in period N^*+1, then P2 reports his information by choosing either S or R. An observable deviation from σ_0 triggers the other player to choose R forever.

During the experimentation phase, the fact that the other player keeps experimenting is uninformative, and each player's belief is updated only on the basis of his own experiments. Under σ_0, each player thus experiments as much as if he were alone, and only then learns the outcomes of N^* experiments of the other player. On the other hand, if a player deviates and experiments for a suboptimal number of periods, he additionally foregoes the opportunity to learn the outcomes of the other player's experiments. Hence, σ_0 is a Nash equilibrium.

According to the strategy profile σ_0, each player learns the other player's private information only once the experimentation phase is over. This positive externality can

[7] The main reason why σ_0 is not part of a SE is that a deviation to the safe arm cannot be interpreted as evidence of a success if the deviating player has not experimented before.

be used to incentivize players to experiment *more*. Specifically, given $n \geq 0$, consider the infinite sequence $h^*_\infty(n) := (RR)^{N^*+n} \cdot S^\infty$, and define σ_n as the strategy profile that follows $h^*_\infty(n)$ as long as the past sequence of choices is consistent with it, and selects R otherwise. According to σ_n, players are supposed to experiment for n additional periods once their belief falls below p^*, rather than to stop immediately. The rationale is that these n extra experiments will give them access *for free* to the outcome of N^*+n experiments (those performed by the other player), which they would never learn if they were to stop before.

For fixed $n \geq 0$, there are parameter values λ, δ, and p_0 under which σ_n is a Nash equilibrium. This implies that there exist λ and δ arbitrarily close to 0 and to 1, such that $N_{\rm e}/N^*$ is arbitrarily large, $N_{\rm e}/N^{**} > 2$ and the belief p_f held before the last experiment is such that p_f/p^* is arbitrarily close to $1/e$.[8] at equilibrium. Thus, the amount of experimentation may be significantly higher than in the public case.

5.2 Reasonable sequential equilibria

A sequential equilibrium is a pair (σ, π), where σ is a strategy profile and $\pi = (\pi^1, \pi^2)$ is a belief system, such that σ is sequentially rational given π, and π is consistent with σ. In our framework, a belief system of player i is a collection $(\pi^i(\cdot \mid h^i))_{h^i}$, which associates with each *private* history h^i of player i a probability distribution on the set of all full histories that are consistent with h^i.[9]

Given such a belief system, we denote by $p^i(h)$ the probability assigned by i to $\theta = G$, conditional on (i) the sequence of past choices is h, and (ii) player i was not successful along h. $p^i(h)$ is thus the first-order belief of i at h, when not successful.

The strategy profile σ_0 defined in Section 5.1 is not part of a sequential equilibrium but there exist *pure* sequential equilibria (σ, π) in which the choice by player i of the safe arm after an experiment is viewed by player j as evidence that i's experiment was successful. That is, player j's off-path belief assigns probability 1 to player i making a strictly dominated choice.[10] While such beliefs are not ruled out by the concept

[8] See Renault et al. [22], Proposition 2.

[9] Consistency means that there exists a sequence $(\sigma_n)_{n\in\mathbf{N}}$ of fully mixed strategies that converges to σ in the product topology, such that the belief systems $(\pi_n)_{n\in\mathbf{N}}$ deduced from $(\sigma_n)_{n\in\mathbf{N}}$ converge to π in the product topology: $\lim_{n\to+\infty} \pi^i_n(\cdot \mid h^i) = \pi^i(\cdot \mid h^i)$ for each player i and private history h^i.

[10] See Renault et al. [22] for details.

of sequential equilibrium, they are excluded by leading refinements of this concept. Accordingly, we restrict ourselves to *reasonable equilibria*, which we define next.[11] The relation of this concept to the existing equilibrium refinements is discussed in Section 7.

Definition 1 *A system π of beliefs is* reasonable *if the following holds for each history $h' \in H$ with active player i:*

- *If h' is of the form $h' = hS$, then $p^i(h') = \phi^{n_e(h)}(p_0)$.*
- *If h' is of the form $h' = hSaR$ with $a \in \{S, R\}$, then $p^i(h') = \phi^{n_e(hSa)}(p_0)$.*

The condition that $p^i(hS) = \phi^{n_e(h)}(p_0)$ captures the fact that player i infers from j's choice of S at h that past experiments of player j failed.

The additional condition on $p^i(hSaR)$ captures the further requirement that, after j has 'revealed' by choosing S that his experiments along h failed, his choice of the risky arm in the next period is considered to be non-informative.

As we now show, reasonable beliefs are uniquely determined by σ. For that reason, we will speak of reasonable SEs σ without specifying beliefs, with the understanding that the system of beliefs is the unique system of reasonable beliefs that is consistent with σ.

Lemma 1 *For every strategy profile σ there is a unique system of reasonable beliefs consistent with σ.*

Proof. Let σ be given. We prove uniqueness of the first-order beliefs $p^i(h)$ by induction.[12] Let $p^1(h)$ and $p^2(h)$ be given, and note that $p^1(h), p^2(h) > 0$. Indeed, no finite amount of evidence can rule out that $\theta = G$. We show that $p^i(ha)$ is uniquely defined by σ and the reasonable criterion, for each i and $a \in \{R, S\}$. For concreteness, assume that P1 is active at h. Since player 1 observes the outcome of his own experiments, one must have $p^1(hS) = p^1(h)$ and $p^1(hR) = \phi(p^1(h))$. On the other hand, $p^2(hS) = \phi^{n_e(h)}(p_0)$, since (p^1, p^2) is reasonable. Consider finally $p^2(hR)$. If the last move of P1 along h is S, the reasonability criterion implies that $p^2(hR) = \phi^{n_e(h)}(p_0)$.

[11] These are coined after Mas-Colell et al. [18], p. 468.

[12] The proof extends to the uniqueness of the belief systems π^i, at the cost of extra notation.

Assume instead that the last move of P1 is R. Since σ dictates R if this last experiment is a success, $p^2(hR)$ is uniquely deduced from $p^2(h)$ and σ using Bayes' rule.

Existence can be shown along the following lines. Let $(\sigma_n)_{n\in\mathbf{N}}$ be a sequence of strategy profiles that converges to σ, with the property that the mixed action $\sigma_n(h) \in \Delta(\{S,R\})$ has full support for each $h \in H$. Given σ_n, players always play both arms with strictly positive probability unless successful. All public histories are on-path, hence the beliefs π_n induced by σ_n are uniquely defined and reasonable, and any accumulation point π of these beliefs is reasonable as well. Since σ_n repeats R once successful, it is not completely mixed. Yet, it can be approximated by a sequence $(\tau_m(n))_{m\in\mathbf{N}}$ of completely mixed profiles — both arms are chosen with positive probability, even after a success. The conclusion obtains, using a standard diagonal extraction argument. ■

Remark 2 *Let σ be pure. The proof of Lemma 1 shows that for each history h and each choice $a \in \{S,R\}$, the reasonable beliefs associated with σ are such that $p^i(ha) = 1$ or $p^i(ha) \leq p^i(h)$. That is, along any sequence of choices, on- or off-path, the players become gradually more pessimistic until, possibly, they become convinced that the other player was successful. This pattern may repeat over time (off-path).*

The proof also implies that for each history h and each player i, one has $p^i(h) = 1$ or $p^i(h) = \phi^n(p_0)$, for some $n \geq n^i_{\mathrm{e}}(h)$, where $n^i_{\mathrm{e}}(h)$ is the number of experiments of i along h.

Proposition 2 *Reasonable* SE *always exists.*

Proof. Consider a variant of the game, in which a player has no longer access to the safe arm S once successful. By Theorem 6.1 in Fudenberg and Levine [8], this variant has a sequential equilibrium (σ,π). The private histories of a player in the true game that are *not* private histories in the variant are histories along which that player was successful. Since we require that players choose R once successful, σ unambiguously induces a reasonable strategy profile in the true game, which we still denote σ. By Lemma 1, there is a unique system $\widetilde{\pi}$ of reasonable beliefs (in the true game) consistent with σ. One can verify that the restriction of $\widetilde{\pi}$ to the private histories of the variant game coincides with π, and that $(\sigma,\widetilde{\pi})$ is a reasonable SE in the true game. ■

While pure SEs always exist, *pure* reasonable SEs may fail to exist, see Section 6.2.

Example. For concreteness, we illustrate here how strategy profiles can be conveniently described as a function of the beliefs they induce. For some parameter values, the profile σ described next is a reasonable SE, such that $N_{\mathrm{e}} = 2 > N^* = 1$. This example also serves to illustrate the arguments we use later. Since this example is subsumed by Theorem 4, we provide no formal details.

According to σ, the active player experiments if no one has experimented so far or if the first experiment just took place. Apart from such histories, the active player experiments if and only if he is convinced that the other player was successful (or if he was successful).[13] The profile σ thus induces the play RRS^∞.

There is no circularity, because $p^i(h)$ depends only on the definition of σ at shorter histories. Yet finding a characterization of the histories h such that $\sigma(h) = R$ without using beliefs is involved, because the interpretation by player i of j's previous choices hinges on how player j previously interpreted earlier choices of i.

We identify conditions that p_0 must satisfy for σ to be a reasonable SE. On the one hand, p_0 must be high enough so that P1 and P2 are willing to experiment at the nodes $\emptyset$ and R respectively. At the initial node $h = \emptyset$, P1 should prefer the equilibrium path RRS^∞ over the path $SRRS^\infty$ which is followed if he deviates once. At the latter node $h = R$, P2 should prefer the continuation path RS^∞ over S^∞, taking into account that P2 will infer the outcome of P1's first experiment from P1's choice in the second stage. This learning externality lowers P2's value of experimenting at $h = R$, and yields a condition that is more demanding than $p_0 > p^*$.

On the other hand, consider P1's decision at the history $h = RR$. At h, P1 is required to play S and the equilibrium continuation path[14] is S^∞. By instead choosing R at h, P1 would thus convince P2 that his first experiment was successful ($p^2(h \cdot R) = 1$) and yield the continuation path $RR \cdot S^\infty$ after h. The requirement that the former should be preferred to the latter yields an upper bound on $p^1(h) = \phi(p_0)$.

Conversely, if these lower and upper bounds are consistent, then σ is a pure reasonable SE as soon as p_0 satisfy both bounds.

[13] That is, for each history h with active player i, $\sigma(h) = R$ if either (i) $h = \emptyset$, (ii) $h = S^k \cdot a$ for some $k \geq 0$ and $a \in \{S, R\}$, or (iii) $p^i(h) = 1$.

[14] If both experiments failed, as usual.

6 Main results

A complete characterization of reasonable SEs appears to be beyond reach, especially when allowing for non-pure equilibria. In fact, even describing specific equilibria is tricky (see, e.g., the proof of Theorem 4).

Instead, we introduce in Section 6.1 a belief threshold $\widehat{p}$, coined the encouragement cut-off, which we show is relevant in shaping equilibrium behavior. In particular, we relate the equilibrium terminal beliefs to $\widehat{p}$ and, for pure equilibria, bound the equilibrium amount of experimentation relative to the optimal amount N^{**}, see Theorem 3. In Section 6.1.4, we provide a characterization of the parameter values for which a natural profile is a reasonable SE and use it to prove that the bounds of Theorem 3 are tight. Finally, we briefly discuss existence issues of pure reasonable SEs and sketch non-pure equilibria.

6.1 The encouragement cut-off

An extra experiment of i may lead to additional *future* experiments of j. In this section, we introduce the *encouragement cut-off* $\widehat{p}$ that captures the simplest version of this positive externality.

6.1.1 The definition of $\widehat{p}$

Let h be a public history that ends with S. Assume that the active player, i, expects that choosing R will induce j to experiment once more, but that j will pick the safe arm in case i chooses S. Assume moreover that later choices are S.[15]

Choosing R yields a flow payoff of $(1-\delta)(p^i(h)\lambda m - c)$, and a per-stage continuation payoff of g if one of the two experiments is successful. If i's experiment is successful, this continuation payoff is received from the next period onward; if j's experiment is successful but i's experiment is not, it is received with a delay of one period, since it is the fact that j *repeats* R that will convey the good news to i. Consequently, i's continuation payoff is $\delta p^i(h)\left(\lambda + \delta(1-\lambda)\lambda\right) g$.

Facing such a situation, i prefers R if and only if $p^i(h) \geq \widehat{p}$, where

$$\widehat{p} := \frac{c(1-\delta)}{c(1-\delta) + g\big(1-\delta+\lambda\delta(1+\delta-\lambda\delta)\big)} < p^* : \tag{3}$$

[15] Except of course if one experiment, past or future, is successful.

for $p^i(h) \in (\widehat{p}, p^*)$, player i is willing to experiment only if this induces j to also experiment.

6.1.2 Encouragement and terminal beliefs

The definition of $\widehat{p}$ suggests that players will not stop experimenting until their belief falls below $\widehat{p}$, as long as they expect the other player to experiment. Theorem 2 below ties the terminal belief of the players to $\widehat{p}$.

Theorem 2 *For any reasonable SE σ, the following holds:*

- *If $p_0 > p^*$, final beliefs are at most $\widehat{p}$ with $\mathbf{P}_\sigma$-positive probability.*
- *For every h such that the belief of the active player satisfies $p^i(h) < \widehat{p}$, one has $\sigma(R \mid h) = 0$.*

In equilibrium, either some player is successful, in which case beliefs converge to 1, or both players remain unsuccessful, in which case experimentation eventually stops. According to Theorem 2, there is a positive probability that final beliefs are at most $\widehat{p}$. We note that these beliefs need not be less than $\widehat{p}$ with probability 1 conditional on $\theta = B$, see Section 6.2.

Theorem 2 also implies that as soon as the active player's belief drops below $\widehat{p}$, experimentation stops forever, unless the other player was successful. Indeed, the active player, i, must choose S at h. Since beliefs are reasonable, $p^j(hS) = \phi^{n_e(h)}(p_0) \leq p^i(h) < \widehat{p}$. Applying Theorem 2 again, this implies $\sigma(hS) = S$, etc.

For *pure* reasonable SEs, the contrast with the public case is even more explicit.

Corollary 1 *Assume that outcomes are* private *and $p_0 > p^*$. At any* pure *reasonable* SE*, final beliefs are below $\widehat{p}$ if $\theta = B$, and no player ever experiments once his belief is below $\widehat{p}$.*

Assume that outcomes are public*. At any Nash equilibrium, final beliefs are below p^* if $\theta = B$, and no player ever experiments once his belief is below p^*.*

Remark 3 *According to Corollary 1, in a pure reasonable SE, there is experimentation until a point where the encouragement effect no longer incentivizes experimentation. We stress that the conclusion requires some experimentation, that is, $p_0 > p^*$. If*

$p_0 < p^*$, there is no experimentation in any pure reasonable SE, see Lemma 7 in Section F of the Appendix.

Proof sketch of Theorem 2. The assumption that $p_0 > p^*$ ensures that $N_e \geq 1$, with probability 1. If final beliefs were always strictly above $\widehat{p}$, the number of experiments N_e would be bounded on-path. Hence, there would exist a 'last experiment', that is, some on-path history h ending with R, following which players choose S with probability 1 in all periods. For such h, the player who is active at hS is in the situation we discussed in Section 6.1.1 when defining $\widehat{p}$. Since his belief is above $\widehat{p}$, deviating to the risky arm is profitable.

The proof of the second statement is more involved. We argue by contradiction and assume that for some history h, the active player i holds a belief $p^i(h) < \widehat{p}$, yet $\sigma(R \mid h) > 0$. We claim that i's continuation payoff is strictly higher when choosing S at h.

Consider the longest history $\bar{h}$ of the form $h \cdot R^k$ along which j experiments with positive probability when active, and i experiments with probability 1 when active. Along such a history, the experiments of i are uninformative, hence the belief of j decreases to 0 as $k \to +\infty$, which shows that $\bar{h}$ is well-defined. For concreteness, assume that $\bar{h} = h \cdot R^{2\bar{n}+1}$, with $\bar{n} \geq 0$. By construction, assuming i chooses R at h, the continuation play is such that j experiments a random number of times $n \in [\![0, \bar{n}]\!]$, then stops, while i experiments for sure until that point. We will prove that, even if i knew in advance the realized value of n, he'd rather deviate to S at h, thus proving our claim. ■

6.1.3 Encouragement and optimality

The cut-off $\widehat{p}$ trades off the cost of one experiment with the individual marginal value of two consecutive experiments (the second one being observed with a delay). The optimal cut-off p^{**} trades off the cost of one experiment with the marginal collective value of that experiment. This suggests a close link between the two cut-offs, which we clarify below.

Lemma 2 *One has $\phi^2(\widehat{p}) < p^{**} < \widehat{p}$.*

Proof of Lemma 2. Suppose that the prior is such that the expected payoff of P1 is zero in the following scenario: both players experiment in period 1, then the experiment outcomes are publicly disclosed, and no further experimentation takes place from period 2 and on (unless one of the experiments in period 1 was successful). Denote by $\overline{p}$ the corresponding value of p_0.

In this scenario, the overall payoff of P2 is also zero, hence the social planner's payoff is zero, implying that $\overline{p} \geq p^{**}$. On the other hand, this scenario is more favorable to P1 than the scenario used to define $\widehat{p}$, since a success of P2 in period 1 would immediately be public. Hence, $\overline{p} < \widehat{p}$. This implies $p^{**} < \widehat{p}$.

Observe next that if the prior is such that $p_0 > \phi^{-1}(p^{**})$, a social planner would choose to experiment at least twice. Hence the expected payoff of P1 in the above scenario is positive, implying $\phi^{-1}(p^{**}) \geq \overline{p}$.

To complete the proof that $\widehat{p} < \phi^{-2}(p^{**})$, it suffices to show that $\widehat{p} < \phi^{-1}(\overline{p})$. Our argument is algebraic. The cut-off value $\overline{p}$ is given by $\overline{p} = \dfrac{c(1-\delta)}{c(1-\delta) + g(1-\delta(1-\lambda)^2)}$, so that $\phi^{-1}(\overline{p}) = \dfrac{c(1-\delta)}{c(1-\delta) + g(1-\lambda)(1-\delta(1-\lambda)^2)}$. Elementary algebra shows that $\widehat{p} < \phi^{-1}(\overline{p})$ is equivalent to the inequality $\lambda(1-\delta)^2 + \lambda^2\delta(3-\delta-\lambda) > 0$, which always holds. ■

Players experiment more when outcomes are private than when they are public. Theorem 3 provides bounds on the equilibrium amount of experimentation.

Theorem 3 *Assume $p_0 > p^*$. At any pure reasonable SE the equilibrium number of experiments satisfies*

$$N^{**} - 2 \leq N_{\mathrm{e}} \leq 2N^{**}, \text{ if } \theta = B.$$

The statement of Theorem 3 allows for the case where $N^{**} = N_{\mathrm{e}} + 2$, in which case there is under-experimentation at the equilibrium. Yet, if players are patient enough, then N^{**} is large if p_0 is not too close to p^{**}. The additive bound of -2 is therefore small relative to N^{**}. For this reason, we interpret Theorem 3 as stating that there cannot be (significant) under-experimentation at a pure reasonable SE.

The multiplicative factor 2 on the upper bound corresponds to the number of players. Allowing for an arbitrary number I of players would result in the upper bound IN^{**}.

Proof of Theorem 3. Let σ be a pure reasonable SE. Set $\widehat{N} := \inf\{n \geq 1 \colon \phi^n(p_0) < \widehat{p}\}$. By Lemma 2, one has $N^{**} - 2 \leq \widehat{N} \leq N^{**}$. By Theorem 2, the final belief $\phi^{N_{\rm e}}(p_0)$ is at most $\widehat{p}$ in the absence of a success, which implies $N_{\rm e} \geq \widehat{N} \geq N^{**} - 2$.

Suppose now that $\theta = B$. For any on-path history h with active player i one has $p^i(h) \leq \phi^{n^i_{\rm e}(h)}(p_0)$, see Remark 2. This implies that $p^i(h) < \widehat{p}$ as soon as $n^i_{\rm e}(h) \geq \widehat{N}$. The second part of Theorem 2 implies that $\sigma(h) = S$. Hence the total number of experiments is at most $N_{\rm e} \leq 2\widehat{N} \leq 2N^{**}$. ■

6.1.4 Encouragement and over-experimentation

A limitation of Theorem 3 is that it applies only to pure reasonable SE, which may fail to exist. It is therefore critical to provide conditions on primitives under which pure reasonable SEs do exist.

When outcomes are public, the strategy profile that experiments whenever the current belief is above p^* is an SPE in the public case for a continuum of values of $p_0 > p^*$, if and only if $\phi^{-1}(p^*) > p_{\rm myop}$, see Remark 1. For private outcomes, we address the existence issue by asking whether the profile in which the active player experiments whenever his current belief is above $\widehat{p}$ is a reasonable SE. Theorem 4 below characterizes the parameter values for which this property holds. We next exploit such parameter values to show in Theorem 5 that the bound $N_{\rm e} \leq 2N^{**}$ on the equilibrium amount of experimentation is tight.

The statement of Theorem 4 involves a new cut-off p^*_n, which is obtained as follows. Fix a history h, with active player i. Assume that player j just experimented exactly n times consecutively along h and is about to switch (forever) to S if unsuccessful, independently of i's choices. At h, the informational value of choosing R is reduced for i, since this experiment is irrelevant in the event where j was successful. Choosing R at h is optimal only if

$$p^i(h) \geq p^*_n := \frac{c(1-\delta)}{c(1-\delta) + g(1 - \delta + \delta\lambda(1-\lambda)^n)}. \tag{4}$$

We recall that $\widehat{N} := \inf\{n \geq 1 \colon \phi^n(p_0) < \widehat{p}\}$.

Theorem 4 *Let σ be the strategy profile defined by $\sigma(h) = R$ if and only if the belief of the active player satisfies $p^i(h) \geq \widehat{p}$. Then σ is a reasonable* SE *if and only if $\phi^{\widehat{N}-1}(p_0) \geq p^*_{\widehat{N}}$.*

At this equilibrium we have $N_{\rm e} = 2\widehat{N}$. For instance, if $p_1^* \leq p_0 < \varphi^{-1}(\widehat{p})$, the theorem applies and we have $\widehat{N} = N^* = 1$ and $N_{\rm e} = 2$. More generally, given $n \geq 1$, the inequality $\varphi^{-1}(\widehat{p}) \geq p_n^*$ is necessary and sufficient for the existence of p_0 such that σ is a pure reasonable SE with $N_{\rm e} = 2n$. This condition is restrictive and is not satisfied when players are very patient, unless λ is close to 1

Proof sketch. Consider the public history $h = (RR)^{\widehat{N}-1} \cdot R$, at which P2 is active and holds the belief $p^2(h) = \phi^{\widehat{N}-1}(p_0)$. At h P2 anticipates that he is about to learn whether P1 was successful in the first $\widehat{N}$ periods and is thus in the situation described above. Thus, deviating to S is optimal unless $p^2(h) \geq p^*_{\widehat{N}}$. This proves the necessity part. The proof of the sufficiency part is much more involved and appears in Section D of the Appendix. It requires to prove that, for each history, the payoff on the equilibrium continuation path exceeds the payoff on the continuation path induced after a one-step deviation. ■

Theorem 5 *For every $\alpha < 2$, there exist $\delta, \lambda, p_0 > p^*$, and a pure reasonable SE such that $N_{\rm e} \geq \alpha N^{**}$.*

The proof of Theorem 5 is algebraic and relies on Theorem 4. Given n, we identify parameter values such that $\phi^{n-1}(p_0) \geq p_n^*$ and $\widehat{N} = n$, and we show that the ratio n/N^{**} is arbitrarily close to 1. Since the total number of experiments is $N_{\rm e} = 2n$, the result follows, see Section E of the Appendix.

According to the strategy defined in Theorem 4, each player experiments as long as his belief exceeds $\widehat{p}$, thus for exactly $n := \widehat{N}$ periods. The equilibrium path is (absent a success) $(RR)^n S^\infty$. How come P2 is willing to do the *last* experiment at $h = (RR)^{n-1} \cdot R$, given that n is potentially large? At this point, P2 anticipates that, irrespective of his choice, he will get to know the private information of P2 immediately and it seems that deviating to S might be profitable.

Choosing R is optimal at h if and only if $p^2(h) \geq p_n^*$, by definition of the latter cut-off. Since $p^2(h) = \phi^{n-1}(p_0)$, this amounts to $\phi^n(p_0) \geq \phi(p_n^*)$, or, equivalently,

$$\phi^{N^{**}}(p_0) \geq \phi^{N^{**}-n+1}(p_n^*). \tag{5}$$

Observe that $p_n^* \leq p_{myop}$ and $\phi^{N^{**}}(p_0) \geq \phi(p^{**})$. The reason for the former inequality is that p_n^* is dictated by the option value of experimenting when a player is about to

learn the private information of the other, which consists of n experiments; this option value decreases and eventually vanishes as $n \to +\infty$, see Eq. 4. The reason for the latter is that $\phi^{N^{**}-1}(p_0) \geq p^{**}$ by the definition of N^{**}.

Substituting these two inequalities into (5), it follows that choosing R is optimal at h when $\phi^{N^{**}-n}(p_{myop}) \leq p^{**}$. Choosing any $\alpha < 1$, and $n = \alpha N^{**}$, this last inequality holds when p_0 is close enough to 1, since $N^{**} - n$ is then arbitrarily large.

6.2 Pure vs. mixed reasonable SEs

For all parameter values, pure SEs do exist, see Renault et al. [22]. However, for some parameter values, such SEs must rely on non-reasonable beliefs, as pure reasonable SEs do not always exist.

Proposition 3 *If $\phi^{-1}(\widehat{p}) < p^*$, there is no pure reasonable* SE *for $p_0 \in (\phi^{-1}(\widehat{p}), p_1^*)$.*

Proposition 3, together with Remark 1, clarifies the existence issue of a pure reasonable SE for p_0 close to p^*. In particular, for $p_0 = p^*$, there is a pure reasonable SE if and only if $\phi^{-1}(\widehat{p}) \geq p^*$. When this inequality does not hold, players must resort to randomization, see below. Proposition 3 implies that there is no pure reasonable SE when players are very patient, unless if λ is close to one.

Proof Sketch. Let σ be a pure reasonable SE. The proof relies on the claim that the continuation path induced by σ after a history h is S^∞, as soon as $p^1(h), p^2(h) < p^*$. That is, the active player does not experiment unless at least one of the players has a belief above p^*.

To see the logic of this claim, we assume for simplicity that $p_0 < p^*$ and we argue that $\sigma(hS) = S$ for each h.[16] Assume instead that $\sigma(hS) = R$ for some h. Among such h's, pick one for which the number of experiments is maximal. This implies that the continuation path following hS is either R^∞, or $R^k S^\infty$ for some $k \geq 1$. In the former case, j's experiments are uninformative from i's perspective, hence i's beliefs eventually fall below $\widetilde{p}$, and deviating to S is then profitable. In the latter case, the player who experiments last holds a belief below p^* and his choice does not affect the continuation play. Hence, he would rather not experiment. The general claim follows a similar logic, but is more involved, see Lemma 8 in Section F.

[16] In particular, no player ever experiments along the equilibrium path induced by σ.

Since $\phi(p_0) < p^*$, this claim implies that the equilibrium sequence of choices is S^∞, following both histories RSS and $RS{\cdot}RR{\cdot}S$. Therefore $\sigma(RS) = S$, and hence $p^2(RS{\cdot}R) = 1$. Moreover, the assumption $p_0 < p_1^*$ implies that the sequence of choices following RS must be S^∞ as well, see Lemma 9. At $h = RS$, P1 is therefore in the situation used to define $\widehat{p}$ in Section 4.2. Since $p^1(h) = \phi(p_0) > \widehat{p}$, deviating to R (and then S) yields a higher continuation payoff than the equilibrium choice S. This is the desired contradiction. ■

Intuitively, the previous argument can be summarized as follows. If $\sigma(RS) = S$, P2 interprets the history RSR as evidence of a success, and therefore experiments, which induces P1 to deviate to R at RS because $p^1(RS) > \widehat{p}$. If instead $\sigma(RS) = R$, P1's second experiment is non-informative. Then both players' beliefs at RSR are below p^*. Assuming no player experiments afterwards, this induces P1 to deviate to S at RS.

Mixing is therefore required. At equilibrium, P2 randomizes at the history RSR in such a way that P1 is indifferent between S and R at the history RS. In turn, P1 randomizes at RS in such a way that P2's belief at RSR makes P2 indifferent at RSR between the two arms. Except for these two histories, the equilibrium behavior is pure.[17] Final beliefs are either $\phi(p_0), \phi^2(p_0)$ or $\phi^3(p_0)$. Since $p_0 > \phi^{-1}(\widehat{p})$, terminal beliefs in a mixed equilibrium may thus exceed $\widehat{p}$. At this equilibrium, the support of the equilibrium amount of experimentation N_{e} is $\{1, 2, 3\}$, while $N^* = 1$ and $N^{**} = 2$.

7 Related literature

The literature on strategic experimentation with exponential bandits is by now quite large, see, e.g., Das et al. [7], Keller et al. [13], Keller and Rady [14], Klein and Rady [16], or Marlats and Ménager [17]. However, there are only a few papers that relax the assumption that both actions/efforts and outcomes are observed. Bonatti and Hörner [4] focus on the hidden actions/observed payoffs case, and develop a model where agents collectively work on a project with uncertain prospects, facing a moral hazard problem due to the inability to monitor each other's actions. They find that free-riding behavior among team members leads to reduced effort and procrastination, and that simply improving monitoring does not necessarily lead to better outcomes.

[17] See Renault et al. [22].

Rosenberg et al. [23] analyze a discrete-time, two-arm bandit problem with hidden outcomes, but assume that the choice of the safe arm is irreversible. Murto and Välimäki [19] analyze information aggregation in a game with irreversible exit, with private payoffs and public actions.

Heidhues et al. [10] consider a model that is essentially identical to ours, but allow for cheap talk communication between the players. They show that the socially optimal level of experimentation can be achieved when initial beliefs are sufficiently optimistic. They also find pure sequential equilibria that exhibit over-experimentation. However, their sequential equilibria rely on 'non-reasonable' beliefs, which we rule out here. On p. 544, they write "*What kind of equilibria exist absent communication remains [...] for future research.*" In that sense, our paper is a follow-up to theirs, where we restrict ourselves to reasonable equilibria.

The issue of over-experimentation relative to the socially efficient level also arises in Halac et al. [9], who examine how to design dynamic contests that promote innovation when there is uncertainty about the feasibility of the innovation, and show the optimality of hiding information about successes, under certain conditions.

The idea that some sequential equilibria are unreasonable because they rely on non-credible beliefs has led to many refinements, see the surveys Hillas and Kohlberg [11] and Van Damme [26]. While most concepts were defined for finite extensive games and/or normal form games, and hence do not apply as such, the idea that underlies our notion of reasonable beliefs is implied by most existing refinements. For instance, any quasi-perfect equilibrium (Van Damme [25]) corresponds to a reasonable equilibrium.[18] A fortiori, any proper equilibrium (Myerson [20]) of the normal form of the (truncation) of the experimentation game corresponds to a reasonable equilibrium in the extensive form. The restriction to reasonable beliefs is also reminiscent of refinements defined for signaling games (Banks and Sobel [1], Cho and Kreps [6]), and is implied by the *intuitive criterion* as defined in Cho [5].

[18]Formally, this applies to finite-horizon truncations of our game.

A Proof of Proposition 1

Proof of P1. Given a Nash equilibrium (NE) σ, we denote by $\mathbf{P}_\sigma(\cdot \mid B)$ the distribution over plays conditional on $\{\theta = B\}$. Recall that $n_{\mathrm{e}}(h)$ is the total number of experiments along h, for each history h. Note that n_{e} is defined over the set of (finite) public histories while the total number of experiments N_{e} is defined over (infinite) plays.

Since the active player chooses S if the common belief is below $\widetilde{p}$, on-path N_{e} is bounded by $\widetilde{N} := \min\{n \geq 0 : \phi^n(p_0) < \widetilde{p}\}$. This implies the existence of an on-path history h such that the player active at h experiments with positive probability and $n_{\mathrm{e}}(h) + 1 = \max N_{\mathrm{e}}$. Since that player is the *last* one to experiment, his belief at h is at least p^*, hence $n_{\mathrm{e}}(h) < N^*$, and, therefore, $N_{\mathrm{e}} \leq N^*$, with probability 1.

If, on the other hand, $\mathbf{P}_\sigma(N_{\mathrm{e}} < N^* \mid B) > 0$, then for each $\varepsilon > 0$ there exists an on-path history h, such that $n_{\mathrm{e}}(h) < N^*$ and such that the probability that someone will ever experiment after h is at most ε. At h, the expected continuation payoff of the active player is at most εg. On the other hand, since $n_{\mathrm{e}}(h) < N^*$, the continuation payoff from deviating to the one-player optimal strategy is bounded away from zero. Hence, for $\varepsilon > 0$ sufficiently small, there is a profitable deviation from σ. This concludes the proof of **P1**.

Proof of P2. Assume for simplicity that $p_0 < p_{\mathrm{myop}}$. We define a pure profile σ inductively. Given h, set $k(h) := N^* - n_{\mathrm{e}}(h)$, which we interpret as a remaining budget of experiments. Accordingly, we set $\sigma(h) = S$ whenever $k(h) \leq 0$. Let now h be such that $k(h) > 0$ and denote by i the player active at h, and by j the other player. We set $\sigma(h) = R$ if j's previous choice is consistent with σ, or if $k(h)$ is even. Otherwise, we set $\sigma(h) = S$.

We argue that i has no one-step profitable deviation. If $\sigma(h) = R$, then the continuation sequence of choices under σ is $R^{k(h)} := \underbrace{RR\cdots R}_{k(h) \text{ times}}$, at which point players switch to S (unless successful). If instead i deviates to S, the continuation path is $SS{\cdot}R^{k(h)}$ if $k(h)$ is odd, and $SR^{k(h)}$ if $k(h)$ is even. In each case, the deviation does not modify the number of experiments of each player, and only delays by one period the payoffs of i.

If $\sigma(h) = S$, then the continuation path after h is $SR{\cdot}R^{k(h)-1}$. If instead i deviates

to R, then the resulting sequence of choices is $RS \cdot R^{k(h)-1}$. While the total number of experiments is the same, the cost of experimentation has shifted to i, making the deviation not profitable.

Proof of P3. We prove the existence of a continuum of SPE payoffs when $p_0 \in (p^*, \min(p_{\text{myop}}, \phi^{-1}(p^*)))$.

By **P1**, there is exactly one experiment ($N^* = 1$) in any SPE. Since $p_0 < p_{\text{myop}}$, each player prefers that the cost of experimentation is borne by the other player. Specifically, if no player has experimented so far, the active player prefers the continuation path $SR \cdot S^\infty$ to $RS \cdot S^\infty$.

As a result, there are exactly two pure SPEs: (i) the SPE σ_1 described in the proof of **P2** (whenever active, player 1 plays S if R has been pulled at least once, and R otherwise, while player 2 always plays S) and (ii) an analogous SPE σ_2, obtained by exchanging the roles of the players.

There is an additional mixed SPE σ^*, in which the active player experiments with probability α as long as R was never chosen. The value of α is such that the benefit of potentially having the other player carry the experiment cost compensates for the delay that is incurred, should the active player eventually be the one to experiment.

Consider finally the strategy profile σ_y that coincides with σ^*, except for the fact that P1 chooses R with probability y rather than α at the initial node. The profile σ_y inherits its SPE property from σ^*. Note that the equilibrium payoff of player 2 is increasing in y. The result follows.

We turn to the uniqueness of symmetric Markov equilibria. We interpret the game as a stochastic game, whose state space is the set $P := \{1\} \cup \{\phi^n(p_0), n \geq 0\}$ of the beliefs that are consistent with Bayesian updating and the prior p_0. The existence of a symmetric Markov equilibrium follows from the fact that every symmetric stochastic game with countably many states and finite sets of actions has a symmetric equilibrium, see, e.g., [21].

Assume $f : [0,1] \to [0,1]$ is a symmetric Markov equilibrium, and denote by $\gamma_1(p)$

and $\gamma_2(p)$ the payoffs[19] of the two players when the prior is $p_0 = p$. Note that

$$\gamma_1(p) \geq \delta\gamma_2(p), \text{ with equality if } f(p) < 1. \tag{6}$$

Indeed, since f is symmetric, the RHS $\delta\gamma_2(p)$ is the expected payoff of P1 when deviating to S in the first period and playing according to f afterwards, so the inequality (6) follows from the equilibrium property.

We first claim that $f(p) > 0$ if and only if $p > p^*$. Indeed, if $f(p_0) = 0$ for some $p_0 > p^*$, then no player ever experiments when the prior is p_0, in contradiction with **P1**. Note next that if $f(p_0) > 0$ for some $p_0 < p^*$, then $N_{\mathrm{e}} \geq 1$ when the prior is p_0, again in contradiction with **P1**. It remains to show that $f(p^*) = 0$. This follows from (6) since (a) if $f(p^*) > 0$, one has $\gamma_1(p^*) = 0$, and (b) $\gamma_2(p^*) > 0$. Indeed, the equality $\gamma_1(p^*) = 0$ holds since the overall payoff of P1 when experimenting at p^* is zero; $\gamma_2(p^*) > 0$ since P2 benefits from the fact that P1 experiments with positive probability.

We now prove the uniqueness claim by induction. Assume that for some $n \geq 0$, $f(p)$ is uniquely defined for every $p \in [0, \phi^{-n}(p^*)]$,[20] and let $p_0 = p \in [\phi^{-n}(p^*), \phi^{-(n+1)}(p^*)]$.

The continuation payoff of P1 when choosing R in the first period is $\delta\gamma_2(\phi(p))$ if unsuccessful, and g if successful. Since $f(p) > 0$ and $\gamma_2(\phi(p))$ is uniquely defined, the equilibrium payoff of P1 is uniquely defined and is given by

$$\gamma_1(p) = (1-\delta)\big(p\lambda m - c\big) + \delta\left(p\lambda g + (1-p\lambda)\gamma_2(\phi(p))\right). \tag{7}$$

After P1's first move, P2 is in the position of the first player. Hence, denoting $\widetilde{p}$ the (random) belief of P2 after the first move of P1, one has

$$\gamma_2(p) = \mathbf{E}\left[\gamma_1(\widetilde{p})\right] = f(p)\left(p\lambda g + (1-p\lambda)\gamma_1(\phi(p))\right) + (1-f(p))\gamma_1(p). \tag{8}$$

We claim that there is at most one equilibrium[21] such that $f(p) < 1$. Assume that $f(p) < 1$. Then P1 is indifferent at p between S and R, so that $\gamma_1(p) = \delta\gamma_2(p) < \gamma_2(p)$. Therefore, the right-hand side of (8) is (i) not constant in $f(p)$, since this would imply

[19] Note that $\gamma_1(p)$ is typically different than $\gamma_2(p)$. Indeed, $\gamma_2(p)$ is player 2's payoff when player 1 makes a choice, while since the equilibrium is symmetric, $\gamma_1(p)$ is player 2's payoff when she (player 2) makes a choice.

[20] This holds for $n = 0$.

[21] Starting from $p_0 = p$.

$\gamma_1(p) = \gamma_2(p)$, and (ii) increasing in $f(p)$, since $\gamma_2(p) > \gamma_1(p)$. This implies that there is *at most* one value of $f(p) \in (0,1)$ such that both (7) and (8) hold, which proves our claim.

Moreover, we claim that if such a value $f(p)$ exists, then there cannot be another equilibrium such that $f(p) = 1$. Assume such an additional equilibrium exists, with equilibrium payoffs $\widetilde{\gamma}_1(p), \widetilde{\gamma}_2(p)$. Eq. (7) still holds, hence $\widetilde{\gamma}_1(p) = \gamma_1(p)$. Moreover, one must have $\widetilde{\gamma}_2(p) > \gamma_2(p)$, since (8) is increasing in f. Since $\gamma_1(p) = \delta\gamma_2(p)$, this implies $\widetilde{\gamma}_1(p) < \delta\widetilde{\gamma}_2(p)$: in the additional equilibrium, P1 is better off deviating to the safe arm – a contradiction. This concludes the proof of the uniqueness claim.

B Proof of Theorem 1

Fix a Nash equilibrium σ, and denote by $N_{\rm s} \in \mathbf{N} \cup \{+\infty\}$ the total number of times S is used along the play. We first prove that either $N_{\rm s} < +\infty$ or $N_{\rm e} < +\infty$: players eventually settle on the same arm.

Claim 1 *One has* $\mathbf{P}_\sigma(N_{\rm s} = +\infty \text{ and } N_{\rm e} = +\infty) = 0$.

Proof. Assume there is an on-path history of the form $h = \overline{h}S$ such that $n_{\rm e}(h) > \widetilde{N}$, and let i be the player active at h. Then $p^i(h) = \phi^{n_{\rm e}(h)}(p_0) < \phi^{\widetilde{N}}(p_0) \leq \widetilde{p}$, and hence $\sigma(h) = S$. By the same argument, $\sigma(hS^k) = S$ for each $k \geq 1$. This implies that on the event $N_{\rm e} > \widetilde{N}$, one has either $N_{\rm s} < +\infty$, or $N_{\rm e} < +\infty$, $\mathbf{P}_\sigma$-a.s. The result follows. ■

We next prove that $\mathbf{P}_\sigma(hR^\infty \mid B) = 0$, for every history $h \in H$. Summing over the countable set H of finite histories will imply that $\mathbf{P}_\sigma\left(N_{\rm s} < +\infty \text{ and } N_{\rm e} = +\infty \mid B\right) = 0$. Combined with Claim 1, this implies that $N_{\rm e} < \infty$ a.s. when $\theta = B$.

Claim 2 *One has* $\mathbf{P}_\sigma(hR^\infty \mid B) = 0$ *for each* $h \in H$.

Proof. We argue by contradiction. Let $h \in H$ such that $\mathbf{P}_\sigma(hR^\infty \mid B) > 0$. For $n \geq 0$, let h_n denote the restriction of hR^∞ to the first n periods, so that $|h_n| = 2n$. Denote by h_n^1 the *private* history of P1 along which the sequence of choices is as in h_n and all experiments of P1 failed. Then $p^1(h_n) = \mathbf{P}_\sigma(\theta = G \mid h_n^1)$.

By Bayes' rule,

$$\frac{p^1(h_n)}{1-p^1(h_n)} = \frac{\mathbf{P}_\sigma\left(\theta = G \mid h_n^1\right)}{\mathbf{P}_\sigma\left(\theta = B \mid h_n^1\right)} = \frac{p_0}{1-p_0} \times \frac{\mathbf{P}_\sigma\left(h_n^1 \mid \theta = G\right)}{\mathbf{P}_\sigma\left(h_n^1 \mid \theta = B\right)}. \tag{9}$$

Plainly, $\mathbf{P}_\sigma\left(h_n^1 \mid \theta = B\right) = \mathbf{P}_\sigma(h_n \mid B)$, and therefore converges to $\mathbf{P}_\sigma(hR^\infty \mid B) > 0$. Hence the denominator on the RHS of (9) has a non-zero limit.

Since the probability of a failure is $1-\lambda$ if $\theta = G$, we have $\mathbf{P}_\sigma\left(h_n^1 \mid \theta = G\right) \leq (1-\lambda)^{n_{\mathrm{e}}^1(h_n)}$. Since $n_{\mathrm{e}}^1(h_n) \to +\infty$, Eq. (9) implies that $\lim_{n\to+\infty} p^1(h_n) = 0$. In particular, $p^1(h_n) < \widetilde{p}$ for all n large enough, so that $\sigma(h_n) = S$ for all such n's. This contradicts the assumption that $\mathbf{P}_\sigma(hR^\infty \mid B) > 0$. ■

C Proof of Theorem 2

C.1 Proof of the first statement

We argue by contradiction, so let σ be a reasonable SE such that $\mathbf{P}_\sigma$-a.s., the limit belief of each player i satisfies $p_\infty^i > \widehat{p}$. Since the sequence of beliefs of player i forms a bounded martingale (w.r.t. to the filtration of his private histories), $p^i(h) > \widehat{p}$ for every on-path public history h.

Let $\widetilde{h}$ such that $\mathbf{P}_\sigma(\widetilde{h} \mid \theta = B) > 0$. By Theorem 1, players eventually switch to S forever if $\theta = B$, hence there is an on-path extension of $\widetilde{h}$ of the form hSS. Since σ is a reasonable SE, one has $\phi^{n_{\mathrm{e}}(h)}(p_0) = p^i(hSS) > \widehat{p} \geq \widetilde{p} \geq \phi^{\widetilde{N}}(p_0)$, and since $n_{\mathrm{e}}(\widetilde{h}) \leq n_{\mathrm{e}}(h)$, it follows that $n_{\mathrm{e}}(\widetilde{h}) \leq \widetilde{N}$.

Hence, N_{e} is bounded. This implies the existence of some history h, with active player i, such that $\mathbf{P}_\sigma(hR \mid \theta = B) > 0$ and σ induces S^∞ after hR. At the history $h{\cdot}RS$, player i is in the situation used to define $\widehat{p}$. Since by assumption $p^i(h{\cdot}RS) > \widehat{p}$, playing R and then S is a profitable deviation of player i from σ.

C.2 Proof of the second statement

We argue by contradiction and assume that there is a reasonable SE σ such that $p^i(h) < \widehat{p}$ and $\sigma(R \mid h) > 0$ for some h with active player i. We denote by $\underline{p}$ the infimum of $p^i(h)$ over such histories and players. Below, we fix such a history h (and active player i) such that $\phi(p^i(h)) < \underline{p}$. This implies that the continuation play induced by σ after hR is of the form $R^k S^\infty$ for some $k \geq 0$.

We denote by $q^i(h)$ the probability that j was successful, conditional on $\theta = G$, if the past sequence of choices is h and i was not successful. Then, $p^i(h)q^i(h)$ is the probability that i assigns at h to the event that j was successful.

Claim 3 *The continuation payoff of player i at h when choosing S is at least*

$$\delta p^i(h)g \times q^i(h), \tag{10}$$

Proof. If j's last choice along h is S, then $q^i(h) = 0$ and the claim holds. Assume then that j's last choice along h is R. This implies that $p^j(hS) \leq \phi(p^i(h)) < \underline{p}$, and hence $\sigma(hS) = S$. Thus, player j chooses R at hS if and only if he was successful along h, which has probability $p^i(h)q^i(h)$. Hence, the continuation play induced by σ after $h{\cdot}SR$ is R^∞, and i's expected continuation payoff is g. Since i's expected continuation payoff after $h{\cdot}SS$ is nonnegative, this concludes the proof of the claim. ■

We will prove that the expected continuation payoff of i when choosing R at h is strictly lower than (10). For convenience, we assume w.l.o.g. that i assigns probability 1 to R at h.[22]

As in the main text, we denote $\omega_i := \inf\{n \geq 1 : \sigma(S \mid h{\cdot}(RR)^n) > 0\}$ and $\omega_j := \inf\{n \geq 0 : \sigma(S \mid h{\cdot}(RR)^n{\cdot}R) = 1\}$.

Claim 4 *One has $\omega_i < +\infty$ or $\omega_j < +\infty$ (or both).*

That is, following hR, player i will eventually choose the safe arm with positive probability in the event that player j keeps experimenting, or player j will eventually pick the safe arm for sure.

Proof of Claim 4. Assume that $\omega_i = +\infty$. Then, the experiments of player i following $h \cdot RR$ are uninformative as long as player j keeps choosing R, so that $p^j(h{\cdot}(RR)^n{\cdot}R) = \phi^n(p^j(hR))$ for each $n \geq 0$. Hence $p^j(h{\cdot}(RR)^n{\cdot}R) < \widetilde{p}$ for n large enough, which implies that $\sigma(h{\cdot}(RR)^n{\cdot}R) = S$ for n large and therefore, $\omega_j < +\infty$. ■

Set $\omega := \inf(\omega_i, \omega_j)$.

Case 1: $\omega = \omega_j < \omega_i$.

[22]Formally, let $\widetilde{\sigma}_i$ be the strategy of i that coincides with σ_i, except that $\widetilde{\sigma}_i(h) = R$. As $\sigma_i(R \mid h) > 0$, and as σ_i is sequentially rational at h against σ_j, the expected continuation payoff of player i when facing σ_j is the same, whether he uses σ_i or $\widetilde{\sigma}_i$.

In this case, when starting from h, player i chooses R until the first random time where player j chooses S, which occurs after at most ω periods.

For $k \in [\![0, \omega]\!]$, denote by π_k the probability that j first chooses S in the k-th period that follows h conditional on $\theta = B$.[23]

For $k \in [\![0, \omega]\!]$, let $\sigma_j(k)$ be the strategy of j that (i) chooses the risky arm k times when starting from h, then switches to S, and (ii) coincides with σ_j elsewhere. We note that the distribution of continuation plays after h induced by the behavior strategy profile σ is a mixture of the distributions induced by $(\sigma_i, \sigma_j(k))$, with weights π_k, $k \in [\![0, \omega]\!]$.

We denote by $\gamma^i(k)$ the continuation payoff of player i at h when facing $\sigma_j(k)$, so that the expected continuation payoff of player i at h is equal to $\sum_{k=0}^{\omega} \pi_k \gamma^i(k)$.

Claim 5 *For each $k \in [\![0, \omega]\!]$, $\gamma^i(k) < \delta g p^i(h) q^i(h)$.*

Since $\delta g p^i(h) q^i(h)$ is a lower bound on i's continuation payoff when choosing S at h, Claim 5 implies that i's continuation payoff is strictly lower when playing $R = \sigma(h)$ than when playing S, a contradiction.

The proof will make use of the following observation. If $p^i(h) < \widehat{p}$, choosing R *once* to encourage *one* additional experiment is not worth it. Deceiving j for an extended duration is *a fortiori* not worth it either. Indeed, consider the case where i expects that choosing R for k times in a row will induce j to experiment k times as well. Assume that $k = 2$ for simplicity. At the history $h \cdot RR$, the belief of i is $\phi(p^i(h))$, because j's most recent experiment is uninformative. Moreover, the marginal benefit from experimenting once is lower than when at h, because choosing S will give access to the outcome of j's experiment. These two observations imply that i would rather choose S when at $h \cdot RR$.

Proof of Claim 5. In each period $t \in [\![0, k]\!]$ after h the flow reward of i is $p^i(h)\lambda m - c$, and the expected continuation payoff reward from period $k+1$ onward is g if some player was successful, and 0 otherwise. Thus, the overall continuation payoff

[23] Thus, $\pi_0 = \sigma(S \mid h \cdot R)$, $\pi_1 = \sigma(S \mid h \cdot (RR) \cdot R) \times \sigma(R \mid h \cdot R)$, etc.

$\gamma^i(k)$ is given by

$$\begin{aligned}&\left(p^i(h)\lambda m - c\right)\left(1-\delta^{k+1}\right) \\ &+\delta^{k+1}p^i(h)g\left\{\left(1-(1-\lambda)^{k+1}\right)+(1-\lambda)^{k+1}\left(q^i(h)+(1-q^i(h))\times\left(1-(1-\lambda)^k\right)\right)\right\}.\end{aligned} \tag{11}$$

The first term between braces is the probability that i is successful, and the second term is the probability that j was successful (but not i), either along h or after h.

We claim that the expression in (11) is strictly lower than (10). This follows from three observations.

1. The difference between (10) and (11) is increasing in $q^i(h)$, so we only need to prove that the inequality holds when $q^i(h)=0$, that is,

$$\begin{aligned}&\left(p^i(h)\lambda m - c\right)\left(1-\delta^{k+1}\right) \\ &+\delta^{k+1}p^i(h)g\left\{\left(1-(1-\lambda)^{k+1}\right)+(1-\lambda)^{k+1}\left(1-(1-\lambda)^k\right)\right\}<0.\end{aligned} \tag{12}$$

2. The choice of player i in period k does not affect j's subsequent choice. Since i's belief at $h\cdot(RR)^k$ is $\phi^k(p^i(h))<p^*$, the continuation payoff in (12) is lower than if i chooses S in that period. The latter payoff is given by

$$\left(p^i(h)\lambda m - c\right)\left(1-\delta^{k}\right)+\delta^{k}p^i(h)g\left\{\left(1-(1-\lambda)^{k}\right)+\delta(1-\lambda)^{k}\left(1-(1-\lambda)^k\right)\right\}. \tag{13}$$

3. The quantity in (13) is i's payoff when experimenting k times induces j to experiment k times as well. Since $p^i(h)<\widehat{p}$, this payoff is *strictly* negative.

■

Case 2: $\omega_j \geq \omega = \omega_i \geq 1$.

This is a variant of the first case. When starting from h, player i chooses R until the ω-th period after h, or until the first random time where player j chooses S, whichever occurs first. As in the first case, the continuation payoff of i given σ is a convex combination of the continuation payoffs $\gamma^i(k)$ induced by $(\sigma_i,\sigma_j(k))$, $k\in[\![0,\omega]\!]$.

For $k<\omega$, the expression of $\gamma^i(k)$ is still given by (11), and $\gamma^i(k)<\delta g p^i(h)q^i(h)$, as in Claim 5. Hence the result will follow if we prove that $\gamma^i(\omega)<\delta g p^i(h)q^i(h)$ as well.

Under $(\sigma_i, \sigma_j(\omega))$, players choose R for ω periods, then i switches to S. Hence $\gamma^i(\omega)$ is given by

$$\begin{aligned}&\left(p^i(h)\lambda m - c\right)(1-\delta^\omega) \\ &+ \delta^\omega p^i(h) g\left\{(1-(1-\lambda)^\omega) + \delta(1-\lambda)^\omega\left(q^i(h) + (1-q^i(h)) \times (1-(1-\lambda)^\omega)\right)\right\}.\end{aligned} \tag{14}$$

As above, the difference between (10) and (14) is increasing in $q^i(h)$, so we need only prove that

$$\left(p^i(h)\lambda m - c\right)(1-\delta^\omega) + \delta^\omega p^i(h) g\left(1-(1-\lambda)^\omega\right) + \delta^{\omega+1} p^i(h) g(1-\lambda)^\omega\left(1-(1-\lambda)^\omega\right) < 0. \tag{15}$$

The result follows by noting that the left-hand side of (15) coincides with (13).

D Proof of Theorem 4

We recall that $\sigma(h) = R$ if and only if $p^i(h) \geq \widehat{p}$, where i is the player active at h. We prove the sufficiency part. Setting $n := \widehat{N}$, we assume that $\phi^{n-1}(p_0) \geq p_n^*$, and use the one-shot deviation principle.

For $i = 1, 2$ and $h \in H$, denote by $d_i(h)$ the number of stages in which i experimented and, at some later stage in h, chose S, thus *disclosing* to j that these experiments failed. Recall that $n_{\mathrm{e}}^i(h)$ is the number of experiments of i along h. The difference $u_i(h) := n_{\mathrm{e}}^i(h) - d_i(h)$ is the number of *un*-disclosed outcomes. That is, the last $u_i(h)$ choices of i along h were R. We will write n_i, d_i, u_i unless there is a potential ambiguity.

It follows from the definition of σ that $p^i(h)$ is either equal to $\phi^{n_i+d_j}(p_0)$ (if the last u_j choices of j are non-revealing), or to 1 (if the last u_j choices of j indicate a success of j).

We will use the following observation:

$$\widehat{p} \leq p^i(h) < 1 \Rightarrow n_i + d_j < n \Rightarrow p^j(h) < 1, \tag{16}$$

for each $h \in H$ where i is the active player. The first implication holds since $p^i(h)$ is either 1 or $\phi^{n_i+d_j}(p_0)$. The second implication holds since as long as $n_i + d_j < n$, the strategy profile σ instructs i to choose R, and hence the u_i undisclosed experiments of i (if any) are not informative to j.

Fix $h \in H$, and let i be the active player at h. Given two infinite plays h' and h'' in $\{S,R\}^{\mathbf{N}}$, we write $h' \succeq_h h''$ if i (weakly) prefers the continuation play h' to h''. That is, consider the two strategy profiles σ' and σ'' that coincide with σ up to h, and then follow h' and h'', respectively. We say that $h' \succeq_h h''$ if i's expected continuation payoff, computed using i's belief at h, is higher under σ' than under σ''.

Lemma 3 *Let h be given.*

1. *For each $k \geq 1$,*

$$(RR)^k S^\infty \succeq_h S(RR)^k S^\infty \text{ if and only if } \phi^{k-1}\left(p^i(h)\right) \geq p^*.$$

2. *For each $k \geq 1$,*

$$(RR)^k S^\infty \succeq_h (RR)^{k-1} S^\infty \text{ if and only if } \phi^{k-1}\left(p^i(h)\right) \geq \widehat{p}_{u_j+k-1}.$$

3. *For each $k \geq 1$,*

$$(RR)^k RSS^\infty \succeq_h (RR)^k S^\infty \text{ if and only if } \phi^{k}\left(p^i(h)\right) \geq p^*_{u_j+k}.$$

Proof. For each claim, the result is obvious if $p^i(h) = 1$. We thus assume that $p^i(h) < 1$.

Proof of 1. The expected continuation payoff under $(RR)^k S^\infty$ is

$$(1-\delta^k)\left(p^i(h)\lambda m - c\right) + \delta^k p^i(h) g\left\{\left(1-(1-\lambda)^k\right) + \delta(1-\lambda)^k\left(1-(1-\lambda)^{k+u_j}\right)\right\}.$$

The expected payoff under $S(RR)^k S^\infty$ is

$$\delta\left\{(1-\delta^k)\left(p^i(h)\lambda m - c\right) + \delta^k p^i(h) g\left\{\left(1-(1-\lambda)^k\right) + (1-\lambda)^k\left(1-(1-\lambda)^{k+u_j}\right)\right\}\right\}.$$

Comparison of the two shows that $(RR)^k S^\infty \succeq_h S(RR)^k S^\infty$ if and only if

$$(1-\delta^k)\left(p^i(h)\lambda m - c\right) + \delta^k p(h) g\left\{\left(1-(1-\lambda)^k\right)\right\} \geq 0. \tag{17}$$

The LHS of (17) is the expected payoff of a single agent holding a prior of $p^i(h)$, who experiments for k periods before switching to S. This proves the first claim.

Proof of 2. The continuation payoffs along the two plays $(RR)^k S^\infty$ and $(RR)^{k-1} S^\infty$ only differ in the event where player i has no success along $(RR)^{k-1}$. At that point,

player i's belief is $\phi^{k-1}(p^i(h))$, and player j has experimented $u_j + k - 1$ times in sequence. Hence player i prefers the continuation play $RR\cdot S^\infty$ to S^∞ if and only if $\phi^{k-1}\left(p^i(h)\right) \geq \widehat{p}_{u_j+k-1}$, as claimed.

Proof of 3. The continuation payoffs along the two plays $(RR)^k\cdot RS\cdot S^\infty$ and $(RR)^k S^\infty$ only differ in the event where player i has no success along $(RR)^k$. At that point, player i's belief is $\phi^k(p^i(h))$, and player j has experimented $u_j + k$ in sequence. Hence, player i prefers the continuation play $RS \cdot S^\infty$ to S^∞ if and only if $\phi^k\left(p^i(h)\right) \geq p^*_{u_j+k}$, as claimed. ■

Let $h \in H$ be arbitrary, with active player i. We claim that i has no one-step profitable deviation at h. This is clear if $p^i(h) = 1$, so we assume that $p^i(h) < 1$ and recall that $p^i(h) = \phi^{n_i+d_j}(p_0)$.

Assume first that $p^i(h) < \widehat{p}$, so that $\sigma(h) = S$. Either player j was not successful in the past, in which case $p^j(hS) \leq p^i(h) < \widehat{p}$ and the continuation play is S^∞; or player j was successful and chooses R. Hence $p^i(h\cdot SR) = 1$ and i's continuation payoff is δg.

If instead i deviates to R, the continuation play depends on j's beliefs. If $p^j(hR) \geq \widehat{p}$, then $\sigma(hR) = R$ and hence $p^i(h\cdot RR) = \phi(p^i(h)) < \widehat{p}$, which implies that $\sigma(h\cdot RR) = S$. Thus, deviating to R triggers one additional experiment by j. Since $p^i(h) < \widehat{p}$, the deviation is not profitable. If instead $p^j(hR) < \widehat{p}$, the continuation play after hR is the same as after hS and hinges on whether j was successful in the past. Since $p^i(h) < p^*$, deviating to R is not profitable in that case either.

We now assume that $p^i(h) \geq \widehat{p}$. The continuation play induced by σ after h is $R^t S^\infty$ for some $t \geq 1$, and it is $SR^q S^\infty$ in case player i deviates to S, for some $q \geq 0$. The exact values of t and q depend on h, as follows.

Equilibrium continuation: The parity of t depends on who stops experimenting. By (16), $p^i(h), p^j(h) < 1$. This implies that the first player to stop is player j if $n_i + d_j < n_j + d_i$, and is player i if $n_i + d_j \geq n_j + d_i$:

- If $n_i + d_j < n_j + d_i$, the continuation play is $(RR)^k \cdot RS \cdot S^\infty$, with $k = n - (n_j + d_i)$.
- If $n_i + d_j \geq n_j + d_i$, the continuation play is $(RR)^k \cdot S^\infty$, with $k = n - (n_i + d_j)$.

Deviation continuation: Player j interprets the choice of S by i as evidence that i was not successful, hence $p^j(hS) = \phi^{n_1+n_2}(p_0) \leq p^i(h)$. Therefore, the first player

to stop experimenting is player j and he will experiment $\max(n-(n_1+n_2),0)$ times: if player i deviates to S at h, the continuation play (deviation included) is

- S^∞, if $n_i+n_j \geq n$;
- $SR\cdot(RR)^{k-1}\cdot RS\cdot S^\infty$, if $n_i+n_j<n$, with $n_i+n_j+k=n$.

We will prove in each case that the deviation to S is not profitable, using the next lemma.

Lemma 4 *One has:*

Q1. *If $n_i+d_j<n_j+d_i$, then $(RR)^{n-(n_j+d_i)}\cdot RS\cdot S^\infty \succeq_h (RR)^k\cdot S^\infty$ for each $0\leq k\leq n-(n_j+d_i)$.*

Q2. *If $n_i+d_j\geq n_j+d_i$, then $(RR)^{n-(n_i+d_j)}\cdot S^\infty \succeq_h (RR)^{k-1}\cdot S^\infty$ for each $1\leq k\leq n-(n_i+d_j)$.*

Q3. *If $n_i+n_j<n$, then $(RR)^{n-(n_i+n_j)}\cdot S^\infty \succeq_h SR\cdot(RR)^{n-(n_i+n_j)-1}\cdot RS\cdot S^\infty$.*

Proof. We start with **Q1**. Since the claim is empty if $n_j+d_i=n$, we assume $n_j+d_i<n$.

Since the sequences (p_k^*) and $(\phi^k(p))$ are increasing and decreasing, respectively, and since $\phi^{n-1}(p)\geq p_n^*$, Lemma 3(3) implies that $(RR)^{n-(n_j+d_i)}\cdot RS\cdot S^\infty \succeq_h (RR)^{n-(n_j+d_i)}\cdot S^\infty$.

We argue below that, moreover, $(RR)^k\cdot S^\infty \succeq_h (RR)^{k-1}\cdot S^\infty$ holds for each $1\leq k\leq n-(n_j+d_i)$. This will imply **Q1**.

By Lemma 3(2), it suffices to check that $\phi^{k-1}(p^i(h))\geq \widehat{p}_{u_j+k-1}$ for each $1\leq k\leq n-(n_j+d_i)$. The LHS of this inequality is decreasing in k, and the RHS is increasing in k. Hence, it suffices to check that it holds for $k=n-(n_j+d_i)$.

For $k=n-(n_j+d_i)$, we have $\phi^{k-1}(p^i(h))=\phi^{n-(n_j+d_i)-1}\left(\phi^{n_i+d_j}(p_0)\right)\geq \phi^{n-2}(p_0)\geq \phi^{n-1}(p_0)$, and $\widehat{p}_{u_j+n-(n_j+d_i)-1}=\widehat{p}_{n-(d_i+d_j)-1}\leq \widehat{p}_n\leq p_n^*$, hence the desired inequality follows from the assumption $\phi^{n-1}(p_0)\geq p_n^*$.

Proof of Q2. We proceed as in the second part of **Q1** and apply Lemma 3(2) to show that $(RR)^k\cdot S^\infty \succeq_h (RR)^{k-1}\cdot S^\infty$ for each $1\leq k\leq n-(n_i+d_j)$. As before, the necessary and sufficient condition from Lemma 3(2) is most demanding when k is highest. So we only need to check that the condition is satisfied when $k=n-(n_i+d_j)$. For this k,

the condition reduces to $\phi^{n-1}(p_0) \geq \widehat{p}_{u_j+n-(n_i+d_j)-1}$. Since $u_j + n - (n_i + d_j) - 1 \leq n$, and since $\widehat{p}_n \leq p_n^*$, the inequality does indeed hold.

Proof of Q3. Thanks to Lemma 3(1), it suffices to check that $\phi^{n-(n_i+n_j)-1}(p^i(h)) \geq p^*$, which holds since $n_i + n_j \geq 0$ and $p^* \leq p_n^*$. ■

We now prove that deviating to S is not profitable.

Case 1: $n_i + d_j < n_j + d_i$ and $n_i + n_j \geq n$.

We need to prove that $(RR)^{n-(n_j+d_i)} \cdot RS \cdot S^\infty \succeq_h S^\infty$, which follows from **Q1** with $k = 0$.

Case 2: $n_i + d_j < n_j + d_i$ **and** $n_i + n_j < n$.

We need to prove that $(RR)^{n-(n_j+d_i)} \cdot RS \cdot S^\infty \succeq_h SR \cdot (RR)^{n-(n_i+n_j)-1} \cdot RS \cdot S^\infty$, which follows by applying **Q1** with $k = n - (n_i + n_j)$ and **Q3**.

Case 3: $n_i + d_j \geq n_j + d_i$ and $n_i + n_j \geq n$.

We need to prove that $(RR)^{n-(n_i+d_j)} \cdot S^\infty \succeq_h S^\infty$, which follows from **Q2** with $k = 1$.

Case 4: $n_i + d_j \geq n_j + d_i$ and $n_i + n_j < n$.

We need to prove that $(RR)^{n-(n_i+d_j)} \cdot S^\infty \succeq_h S \cdot (RR)^{n-(n_i+n_j)} \cdot S^\infty$. This follows by applying **Q2** with $k = n - (n_i + n_j)$ and **Q3**.

E Proof of Theorem 5

Theorem 5 will follow from Theorem 4 and simple algebra.

Lemma 5 *Set $\delta = \frac{1}{2}$ and $\lambda = \frac{1}{2n}$, with $n \geq 1$. Then $\phi(p_n^*) < \widehat{p}$.*

The conclusion implies that for such δ, λ, the interval $I := (\max(\phi(\widehat{p}), \phi(p_n^*), \widehat{p})$ is non-empty. For each $p_0 \in \phi^{-n}(I)$, one has $n := \inf\{k \geq 0 : \phi^k(p_0) < \widehat{p}\}$ and $\phi^{n-1}(p_0) \geq p_n^*$, which implies that for such p_0, the strategy analyzed in Theorem 4 is a reasonable SE with $N_{\mathrm{e}} = 2n$ if $\theta = B$.

Proof. By Eqs. (3) and (4), and since $\phi(p) = \frac{(1-\lambda)p}{1-\lambda p}$, the condition $\phi(p_n^*) < \widehat{p}$ reduces to

$$(1-\lambda)\big(1 - \delta + \delta\lambda(1 + \delta - \delta\lambda)\big) < (1-\delta) + \lambda\delta(1-\lambda)^n. \tag{18}$$

For $\delta = \frac{1}{2}$, the inequality (18) simplifies to $(1-\lambda)\left(2+3\lambda-\lambda^2\right) < 2+2(1-\lambda)^n\lambda$, or equivalently,

$$1-4\lambda+\lambda^2 < 2(1-\lambda)^n. \tag{19}$$

Since the LHS of (19) is at most $1-3\lambda$, and the RHS is at least $2(1-n\lambda)$, the inequality is satisfied whenever $1-3\lambda < 2(1-n\lambda)$, or $\lambda < \frac{1}{2n-3}$, and, in particular, when $\lambda = \frac{1}{2n}$. ■

Lemma 6 *Let $\eta > 0$ be given. For all n large enough, the cut-offs $\widehat{p}$ and $p^*_{\sqrt{\delta}}$ associated with $\delta = \frac{1}{2}$ and $\lambda = \frac{1}{2n}$ satisfy $\phi^{\lfloor \eta n \rfloor}(\widehat{p}) < p^*_{\sqrt{\delta}}$.*

Proof. Substituting $\delta = \frac{1}{2}$ and $\lambda = \frac{1}{2n}$ in Eq. (1), we obtain

$$p^*_{\sqrt{\delta}} = \frac{c\left(\sqrt{2}-1\right)}{c\left(\sqrt{2}-1\right)+g\left(\sqrt{2}-(1-\frac{1}{2n})\right)},$$

so that $\lim_{n\to+\infty} p^*_{\sqrt{\delta}} = \frac{c}{c+g}$.

On the other hand, $\phi^k(p) = \frac{(1-\lambda)^k p}{1-(1-(1-\lambda)^k)p}$ for each $p \in (0,1)$ and $k \in \mathbb{N}$. With $\delta = \frac{1}{2}$ and $\lambda = \frac{1}{2n}$, this yields

$$\phi^k(\widehat{p}) = \frac{c\left(2(1-\frac{1}{2n})^k\right)}{g\left(2+\frac{3}{2n}-\frac{1}{4n^2}\right)+c\left(2(1-\frac{1}{2n})^k\right)}.$$

Substituting there $k = \lfloor \eta n \rfloor$, we get

$$\lim_{n\to+\infty} \phi^{\lfloor \eta n \rfloor}(\widehat{p}) = \frac{ce^{-\eta/2}}{g+ce^{-\eta/2}} < \frac{1}{1+g}.$$

It follows that $\phi^{\lfloor \eta n \rfloor}(\widehat{p}) < p^*_{\sqrt{\delta}}$ for every n large enough, as desired. ■

Proof of Theorem 5. As noted after the statement of Lemma 5, with $\delta = \frac{1}{2}$, $\lambda = \frac{1}{2n}$, and $p_0 \in \left(\phi^{-n+1}(\widehat{p}), \phi^{-n}(\widehat{p})\right]$, there is a pure SE with $N_{\mathrm{e}} = 2n$.

Given $\eta > 0$, for $p_0 \in \left(\phi^{-n+1}(\widehat{p}), \phi^{-n}(\widehat{p})\right]$ one has $\phi^{n+\lfloor \eta n \rfloor}(p_0) < \phi^{\lfloor \eta n \rfloor}(\widehat{p})$, which is less than $p^*_{\sqrt{\delta}}$ for all n large enough. This implies that $N^{**} \le (1+\eta)n$.

For such an SE, $\dfrac{N_{\mathrm{e}}}{N^{**}} \ge \dfrac{2}{1+\eta}$. Since $\eta > 0$ is arbitrary, this implies the result. ■

F Proofs for Proposition 3

The following three lemmas are used in the proof of Proposition 3.

Lemma 7 *Let σ be a pure reasonable* SE. *If $p_0 < p^*$, then the path induced by σ is S^∞.*

Proof. We first show that $\sigma(hS) = S$ whenever $p^i(hS) < 1$ for each player i. We argue by contradiction and assume that the set

$$H_0 := \{h : p^1(hS) < 1,\ p^2(hS) < 1,\ \text{and } \sigma(hS) = R\}$$

is non-empty.

Let $h \in H_0$ be arbitrary and i be the player active at hS. Since σ is a pure reasonable SE, one has $p^i(hS) = \phi^{n_{\rm e}(h)}(p_0)$. On the other hand, $p^i(h) \geq \widetilde{p}$, since $\sigma(hS) = R$. Therefore, $\max_{h\in H_0} n_{\rm e}(h) \leq \widetilde{N}$.

Let $h \in H_0$ be such that $n_{\rm e}(h) = \max_{h'\in H_0} n_{\rm e}(h')$. We prove that player i has a profitable one-step deviation at h, which will yield the desired contradiction.

Since h maximizes $n_{\rm e}(\cdot)$ over H_0, the continuation play induced by σ after hS is either equal to R^kS^∞ for some $k \geq 1$, or to R^∞. We rule out both cases in turn.

Case 1: The continuation play is R^∞.

In this case, $p^i(hS{\cdot}(RR)^k) = \phi^{n_{\rm e}(h)+k}(p_0)$ for each k, and hence $p^i(hS{\cdot}(RR)^k) < \widetilde{p}$ for all k large enough, contradicting the assumption $\sigma(hS{\cdot}(RR)^k) = R$.

Case 2: The continuation play is R^kS^∞, for some $k \geq 1$.

Assume first that $k = 1$, and let us place ourselves at the history hS. Since player j just played S, player i expects his experiment $R = \sigma(h)$ at hS to be the last. Since $p^i(hS) \leq p_0 < p^*$, the equilibrium continuation payoff of i at hS is (strictly) negative. On the other hand the optimal continuation payoff when deviating to S is non-negative, irrespective of the continuation strategy of player j — a contradiction.

Assume now that $k > 1$. Set $\widehat{h} := hS{\cdot}R^{k-1}$, and let i be the player active at $\widehat{h}$. Since $p^1(hS), p^2(hS) \leq p_0$, and since σ induces the sequence of choices R^{k-1} after hS, it follows that $p^i(\widehat{h}) \leq p^i(hS) \leq p_0 < p^*$. By the maximality property of h and since $n_{\rm e}(\widehat{h}) > n_{\rm e}(h)$, the continuation play induced by σ after $\widehat{h}a$ is S^∞ for each $a \in \{R, S\}$. Since $p^i(\widehat{h}) < p^*$, player i's continuation payoff at $\widehat{h}$ is higher when deviating to S than with $R = \sigma(\widehat{h})$.

We now conclude by showing that $N_{\rm e} = 0$ under σ. Assume instead that $N_{\rm e} \geq 1$. The first part of the proof implies that the play induced by σ is of the form R^kS^∞ for

some $k \geq 1$. A contradiction is obtained by repeating the arguments in Case 2. This concludes the proof of Lemma 7. ■

Lemma 8 *Let σ be a pure reasonable* SE. *Let h be a history such that $p^i(h) < p^*$ for each i. Then the continuation play following h is S^∞.*

Proof. Define A to be the interval of beliefs $p \in [0,1]$ such that there exist a prior $p_0 \in [0,1]$, a pure reasonable SE σ, and a history h satisfying $p^1(h) < p$, $p^2(h) < p$, and $\sigma(h) = R$. Note that $\inf A \geq \widetilde{p} > 0$. It follows that $\phi(p) < \inf A$ whenever $p \in A$ is sufficiently close to $\inf A$.

We need to prove that $\inf A \geq p^*$. We argue by contradiction, and assume that $\inf A < p^*$. This implies the existence of $p < p^*$ such that $p \in A$ and $\phi(p) < \inf A$. We fix such a belief p, and let p_0 and σ be given as in the definition of the set A. By assumption, the set $H_1 = \{h\colon p^1(h) < p, p^2(h) < p, \sigma(h) = R\}$ is not empty.

The main part of the proof consists in showing that $h \in H_1$ implies $hS \in H_1$. The argument follows closely the proof of Lemma 7.

Let $h \in H_1$ be arbitrary. Since $h \in H_1$, the continuation play induced by σ after h starts with R and ends with S^∞. It can thus be written as $\overline{h}RS^\infty$, where $\overline{h}$ is either empty or starts with R.

We denote by i the player active at $h\overline{h}$ and note that $\sigma(h\overline{h}) = R$.

Part 1. We prove that $\overline{h} = \emptyset$. Assume to the contrary that $\overline{h} \neq \emptyset$, so that $\overline{h}$ starts with R. Then, the beliefs of both players j following $h\overline{h}S$ are such that $p^j(h\overline{h}S) \leq \phi(p^j(h)) < \phi(p)$.[24] Since $\phi(p) < \inf A$, one has $\sigma(h\overline{h}S) = S$. Moreover, by the definition of $\overline{h}$, $\sigma(h\overline{h}R) = S$. Hence, the continuation play induced by σ after $h\overline{h}a$ is S^∞ for each $a \in \{R, S\}$. Since $p^i(h\overline{h}) < p^*$, player i's continuation payoff at $h\overline{h}$ is higher when choosing S. This contradicts $\sigma(h\overline{h}) = R$, and hence $\overline{h} = \emptyset$.

Part 2. We prove that $\sigma(hS) = R$, which implies that $hS \in H_1$. Assume to the contrary that $\sigma(hS) = S$. Since $\overline{h} = \emptyset$, the continuation play induced by σ after hR is S^∞. On the other hand, $\sigma(hS) = S$ implies that at h, i is expecting that the outcomes of j's most recent experiments (if any) will be immediately disclosed if i chooses S. Since $p^1(h) < p^*$, player i is better off deviating to S at h, contradicting $\sigma(h) = R$.

We have thus proved that $hS \in H_1$ whenever $h \in H_1$. Let any $h \in H_1$ be given. Thus, $hS \in H_1$ as well, which implies in turn that $hSS \in H_1$. At the history hSS, the

[24] The first inequality holds since $n_{\mathrm{e}}(\overline{h}) \geq 1$, and the second since $h \in H_1$.

most recent choice of each agent was S, and hence the continuation game is 'isomorphic' to the initial game with prior $p_0' := p^1(hSS) = p^2(hSS)$. That is, the profile induced by σ following hSS is a sequential equilibrium of the entire game, in which the prior belief is p_0'. It then follows from Lemma 7 that the play path induced by σ after hSS is S^∞. In particular, $\sigma(hSS) = S$, which contradicts $hSS \in H_1$. ■

Lemma 9 *Let σ be a pure reasonable* SE. *If $p_0 < p_1^*$, then the path induced by σ after RS is S^∞.*

Proof. We use the inequality $p_0 < \min(p_2^*, \phi^{-1}(p^*))$, which follows from $p_0 < p_1^*$.

We note that $p^1((RS)^n) = \phi^n(p_0)$ for each n, hence $N := \min\{n \geq 1 : \sigma((RS)^n) = S\}$ is well defined, finite, and at least 1. Since $p^i((RS)^N \cdot S) \leq \phi(p_0) < p^*$ for $i = 1, 2$, by Lemma 8 the continuation play induced by σ after $(RS)^N \cdot S$ is S^∞. Hence, we only need to prove that $N = 1$. This is implied by the following two contradictory claims.

Claim: If $N \geq 2$, one has $\sigma((RS)^{N-1}R) = R$.

Assume that $N \geq 2$ and $\sigma((RS)^{N-1}R) = S$, so that the equilibrium continuation play at $(RS)^{N-1}$ is RS^∞. Since $p^1((RS)^{N-1}) \leq \phi(p_0) < p^*$, the induced payoff for P1 is negative — a contradiction.

Claim: If $N \geq 2$, one has $\sigma((RS)^{N-1}R) = S$.

Assume that $N \geq 2$ and that $\sigma((RS)^{N-1}R) = R$. By the definition of N, one has $p^2((RS)^{N-1}R) = p_0$. This implies that $p^i((RS)^{N-1} \cdot RR) < p^*$ for both players, hence the continuation play induced by σ following $(RS)^{N-1} \cdot RR$ is S^∞, by Lemma 8. On the other hand, the continuation play in the event where P2 deviates to S at $(RS)^{N-1} \cdot R$ is also S^∞. Since the belief of P2 at $(RS)^{N-1} \cdot R$ is below p_2^*, deviating to S is profitable. ■

References

[1] Jeffrey S. Banks and Joel Sobel. Equilibrium selection in signaling games. *Econometrica*, 55(3):647–661, 1987.

[2] Dirk Bergemann and Juuso Välimäki. *Bandit problems*. Palgrave McMillan Ltd., 2008.

[3] Patrick Bolton and Christopher Harris. Strategic experimentation. *Econometrica*, 67(2):349–374, 1999.

[4] Alessandro Bonatti and Johannes Hörner. Collaborating. *American Economic Review*, 101(2):632–663, 2011.

[5] In-Koo Cho. A refinement of sequential equilibrium. *Econometrica*, 55(6):1367–1389, 1987.

[6] In-Koo Cho and David M. Kreps. Signaling games and stable equilibria. *The Quarterly Journal of Economics*, 102(2):179–221, 1987.

[7] Kaustav Das, Nicolas Klein, and Katharina Schmid. Strategic experimentation with asymmetric players. *Economic Theory*, 69:1147–1175, 2020.

[8] Drew Fudenberg and David Levine. Subgame-perfect equilibria of finite-and infinite-horizon games. *Journal of Economic Theory*, 31(2):251–268, 1983.

[9] Marina Halac, Navin Kartik, and Qingmin Liu. Contests for experimentation. *Journal of Political Economy*, 125(5):1523–1569, 2017.

[10] Paul Heidhues, Sven Rady, and Philipp Strack. Strategic experimentation with private payoffs. *Journal of Economic Theory*, 159:531–551, 2015.

[11] John Hillas and Elon Kohlberg. Foundations of strategic equilibrium. *Handbook of Game Theory with Economic Applications*, 3:1597–1663, 2002.

[12] Johannes Hörner and Andrzej Skrzypacz. Learning, experimentation and information design. *Advances in Economics and Econometrics*, 1:63–98, 2017.

[13] Godfrey Keller and Sven Rady. Strategic experimentation with poisson bandits. *Theoretical Economics*, 5(2):275–311, 2010.

[14] Godfrey Keller and Sven Rady. Breakdowns. *Theoretical Economics*, 10(1):175–202, 2015.

[15] Godfrey Keller, Sven Rady, and Martin Cripps. Strategic experimentation with exponential bandits. *Econometrica*, 73(1):39–68, 2005.

[16] Nicolas Klein and Sven Rady. Negatively correlated bandits. *The Review of Economic Studies*, 78(2):693–732, 2011.

[17] Chantal Marlats and Lucie Ménager. Strategic observation with exponential bandits. *Journal of Economic Theory*, 193:105232, 2021.

[18] Andreu Mas-Colell, Michael D. Whinston, and Jerry R. Green. *Microeconomic theory*. Oxford University Press, New York, 1995.

[19] Pauli Murto and Juuso Välimäki. Learning and information aggregation in an exit game. *The Review of Economic Studies*, 78(4):1426–1461, 2011.

[20] Roger B. Myerson. Refinements of the nash equilibrium concept. *International Journal of Game Theory*, 7:73–80, 1978.

[21] Thiruvenkatachari Parthasarathy. Discounted, positive, and noncooperative stochastic games. *International Journal of Game Theory*, 2:25–37, 1973.

[22] Jérôme Renault, Eilon Solan, and Nicolas Vieille. Strategic experimentation with private payoffs. *arXiv preprint arXiv:2512.06180*, 2025.

[23] Dinah Rosenberg, Eilon Solan, and Nicolas Vieille. Social learning in one-arm bandit problems. *Econometrica*, 75(6):1591–1611, 2007.

[24] Michael Rothschild. A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9(2):185–202, 1974.

[25] Eric Van Damme. A relation between perfect equilibria in extensive form games and proper equilibria in normal form games. *International Journal of Game Theory*, 13:1–13, 1984.

[26] Eric Van Damme. Strategic equilibrium. *Handbook of Game Theory with Economic Applications*, 3:1521–1596, 2002.