# LEARNING TO DISAGREE IN A GAME OF EXPERIMENTATION

**By**

**Alessandro Bonatti and Johannes Hörner**

**March 2015**

**COWLES FOUNDATION DISCUSSION PAPER NO. 1991**

**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS**
**YALE UNIVERSITY**
**Box 208281**
**New Haven, Connecticut 06520-8281**

**http://cowles.econ.yale.edu/**

# Learning to Disagree in a Game of Experimentation*

Alessandro Bonatti† Johannes Hörner‡

March 2015

**Abstract**

We analyse strategic experimentation in which information arrives through fully revealing, publicly observable "breakdowns." With hidden actions, there exists a unique equilibrium that involves randomization over stopping times. This randomization induces belief disagreement on the equilibrium path. When actions are observable, the equilibrium is pure, and welfare improves. We analyse the role of policy interventions such as subsidies for experimentation and risk-sharing agreements. We show that the optimal risk-sharing agreement restores the first-best outcome, independent of the monitoring structure.

**Keywords:** Experimentation, free-riding, mixed strategies, monitoring, delay.

**JEL Codes:** C73, D83, O33.

*We would like to thank Francesc Dilmé, Daniel Gottlieb, Gabriele Gratton, Sven Rady, Tavneet Suri, Juuso Toikka and seminar participants at Bonn, Boston College, Glasgow, MIT, Montréal, Paris II, PSE, Toronto and SAET for their helpful comments, and we are grateful to Xiaosheng Mu for the excellent research assistance.

†MIT Sloan School of Management, 100 Main Street, Cambridge MA 02412, USA bonatti@mit.edu.

‡Yale University, 30 Hillhouse Ave., New Haven, CT 06520, USA, johannes.horner@yale.edu.

# 1 Introduction

We do not learn from observing others in the same way that we learn from our own experiences. There are at least two reasons for this difference. First, experimentation differs from classical conditioning: unlike Pavlov's dogs, we choose our actions to enable learning, whereas observational learning is largely passive. Second, our information concerning others is typically imperfect: often, we observe *how well* they do rather than *what* they do. This paper explains why these differences *must* lead to differences in beliefs, including higher-order beliefs, even provided perfectly aligned fundamentals and outcomes that are common knowledge. Belief disagreement leads to dispersion in actions, and the resulting lack of coordination has significant welfare consequences.

Both our premise and our conclusion are relevant: economic practices are often largely confidential (whether they pertain to sales or to distribution methods, consumer profiles, advertising strategies, lists of suppliers and clients, or manufacturing processes), especially when they involve projects rather than products; economic outcomes are instead easier to ascertain. Indeed, to the extent that they have been documented, there is substantial dispersion in such practices and the resulting performance across firms and industries.[1]

To explain how dispersion arises, we rely on a well-known strategic experimentation model. Players choose whether to experiment in the face of purely aggregate uncertainty. Formally, they continuously choose how much weight to assign to a risky action. Externalities are informational. Players observe only one another's outcomes, not their actions. We assume binary individual outcomes (a "breakdown" or not) and a common binary state of the world (good or bad). Occasional, publicly observable breakdowns occur when a player puts weight on the risky action and when the state is bad. Hence, whereas a breakdown reveals the state of the world to all players, the absence thereof causes objective and strategic uncertainty: inferences regarding the state interact with inferences regarding the actions of others. In the continuing absence of any breakdown, players grow increasingly optimistic about the state over time. As a result, they are tempted to delay their use of the risky arm to free-ride on the experimentation of others.

The game admits a unique symmetric mixed-strategy equilibrium. In particular, no pure-strategy Nash equilibrium exists.[2] In light of the literature, this is surprising because

[1] A long history of empirical literature has documented heterogeneity in the adoption rates for new technologies: Mansfield (1961) observes patterns of "slow imitation" for a small number of innovations; Coleman, Katz, and Menzel (1966) show distinct differences across physicians in the adoption of new medical technology; and more recent studies (Bloom and Van Reenen, 2007, Syverson, 2011) document the wide dispersion in managerial practices within an industry and relate it to persistent productivity differences.

[2] We show that with two players, no asymmetric mixed-strategy equilibrium exists either.

we do not assume discrete action sets: by definition, giving (say) equal weight to both the risky action and its safe alternative is a *pure* action in our framework. Because time is also continuous, mixing is caused not by discreteness but by the intrinsic nature of incentives. This relationship stands in contrast to the experimentation literature, discussed below, in which a "mixed strategy" is merely an interpretation of actions that are interior (*i.e.*, players assign positive weight to both arms) as opposed to extremal.

In equilibrium, mixing involves each player choosing *at random* a time before which he exclusively plays safe and after which he only plays risky. The distribution of switching times is continuously increasing over an interval, with an atom at the upper end. Despite being indifferent over an entire interval of such random times, players are unwilling to play an interior action during that interval (pure strategies involving such actions are strictly worse). Another way to appreciate the difference is that, unlike with conventional interior "mixed" strategies, players are uncertain of the aggregate amount of experimentation undertaken up to a given time.

Randomization over switching times drives the dispersion of beliefs. Not observing a breakdown can be explained in two ways: either the other players have not yet begun experimenting, or the state of the world is good. A player's own choice of action helps him sort through these competing explanations: the earlier he began experimenting himself, the more likely he is to believe that the lack of a breakdown can be attributed to the state of the world being good rather than to the other players waiting to experiment. In the absence of a breakdown, beliefs about the state remain private at all times: while there exists a finite time at which players commonly know that everyone is experimenting, they still do not know when everyone else began experimenting.

Why do players mix? Two forces are combined here. The first and familiar force is mentioned above: free-riding prevents players from adopting the same extremal pure strategy. If one's opponent is switching to the risky arm at a given time, then a player's best reply can involve experimenting immediately to avoid wasting time, as nothing will be learnt until then, or taking advantage of his experimentation by choosing to wait long enough to benefit from it. In standard experimentation models, this force drives the players' equilibrium choice of interior actions. Here, a second force compels players to choose extremal actions. Experimentation breeds experimentation: a player who deviates from an interior action to an action that places greater emphasis on the risky arm will see his choice confirmed by the absence of a breakdown; this observation makes him more optimistic about the state of the world. If he were indifferent between risky and safe if he had not deviated, his deviation would have led him to strictly prefer experimentation in the future. In a sense, payoffs are convex in the weight assigned to the risky action.

This equilibrium adds two sources of inefficiency (dispersion in practices and dispersion in beliefs) to the familiar cost of under-experimentation: given the overall amount of experimentation, the dispersion in timing is costly; holding his opponents' strategies fixed, a player would be strictly better off if he could determine when his opponents actually began experimenting.

This last statement extends to equilibrium analysis. As we show, players are better off in the (symmetric) Markov equilibrium in the game in which they can observe one another's actions. They also benefit from a mediator helping them to coordinate their play via private recommendations.

More radical policy interventions that modify the payoffs of the game can be further helpful. In particular, risk-sharing has several advantages over other types of interventions. By risk-sharing, we refer to a well-calibrated group-insurance scheme whereby a player who suffers a breakdown obtains partial compensation from the other players. First, such a scheme restores the first-best outcome in contrast to, for instance, externally funded subsidies that improve the amount of experimentation without solving the coordination problem. Second, the optimal scheme is robust to the specific monitoring structure: whether players observe one another's actions is irrelevant to the calibration of this scheme.

**Applications.** Our choice to adopt the simplest experimentation model is deliberate, as it makes the analysis as transparent as possible. One could certainly extend some of the analysis to richer settings.[3] However, we believe that even our simple model already captures the essential features of economically relevant applications and provides useful insights.

Bad-news learning processes naturally occur upon the introduction of a new technology that promises cost savings over the status quo but entails additional risks. Examples of such risky technologies that are especially relevant to our model include new drugs and medical devices; innovative processes such as hydraulic fracturing for oil production; and new crops in agriculture, such as hybrids and high-yielding varieties.

Although their technological and institutional details differ, these settings share the key features of our model: (i) *Privately observed adoption decisions:* there is growing evidence of significant barriers to information in several markets.[4] Factors such as geographical distance, social and information networks, and costly attention can hinder the observability of other agents' actions. (ii) *The potential for public catastrophic outcomes:* such breakdowns may

[3] Our earlier working paper provides an analysis of the case in which some learning occurs even in the absence of experimentation; Section 6 introduces payoff externalities in addition to informational externalities, and the conclusions discuss the case in which learning remains incomplete even after a breakdown.

[4] For example, Duflo, Kremer and Robinson (2006) document Kenyan farmers' limited knowledge of their neighbours' choice of crops.

include a malpractice suit following the failure of a new drug or medical equipment; an oil-drilling accident, such as a spill or a blowout;[5] or the loss of an entire harvest. (iii) *Weak market competition (pure informational externalities)*: drillers and farmers are largely price-takers, and doctors may not even be aware of other potential adopters of a new device.[6] Of course, when discussing our welfare implications, we must consider cross-market externalities, such as the environmental impact of an oil-drilling accident.

A growing body of empirical evidence is broadly consistent with our model's findings. In particular, Skinner and Staiger (2007) document U.S. state-level variation in the adoption rates for four technological innovations (hybrid corn, tractors, computers, and beta-blockers) and suggest informational barriers as a potential explanation. Consistent with the idea that barriers to information generate cross-sectional heterogeneity in the new technology adoption rate, Bandiera and Rasul (2006) relate farmers' decisions to adopt a new crop variety to the information available in their social networks, Conley and Udry (2010) collect direct data on social interconnections and document the importance of "information neighbourhoods" for the local diffusion of fertilizer and other chemicals, and Covert (2014) documents frictions in drilling companies' learning processes regarding the relationship between inputs and oil production.

With monitoring used as a design variable, our results in Section 6.1 help explain the information sharing observed in several industries. Indeed, in health care, industry associations and government agencies promote the sharing of information on best practices;[7] in oil drilling, regulations encourage sharing information regarding input choices for fracking operations (see Covert, 2014); and for an example in agriculture, BenYishay and Mobarak (2014) show how "extension agents" affect the flow of information and technology diffusion across villages in Malawi.

Finally, our results in Section 6.2 highlight the benefits of risk-sharing agreements and warn against the ability of externally funded subsidies to reduce dispersion in adoption rates. These results may suggest an additional explanation for the low and heterogeneous take-up rates for agricultural subsidies, as documented, for example, by Carter, Laajaj and Yang

[5]See, for example, "The Downside of the Boom," *The New York Times*, November 22, 2014.

[6]The consequences of informational externalities for technology adoption in the context of agriculture are the object of a large body of empirical literature. However, most studies of social learning have assumed unconstrained information flows among potential adopters, *i.e.*, perfect monitoring of actions and outcomes, at least among subsets of agents. See Foster and Rosenzweig (2010) and Jack (2013) for an overview.

[7]The US Food and Drug Administration (FDA) is launching a Unique Device Identifier (UDI) system "to adequately identify medical devices through their distribution and use." The system provides information on outcomes and indirectly provides data on adoption rates through usage intensity (http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/UniqueDeviceIdentification/).

(2013). Surprisingly, although risk-sharing agreements in developing countries have been widely studied (*e.g.*, Townsend, 1994 and 1995), the link with technology adoption seems untested thus far.

**Related Literature.** Our motivation is closest to that of Murto and Välimäki (2011), emphasizing the interplay between experimentation and observational learning. In their model, players receive exogenous (random) private signals over time. In the present study, however, learning is endogenous, with players' actions influencing their private beliefs. Another major difference is that although signals are private in their model, actions (exit or not) are publicly observed. As a result, the dynamics that they identify are very different from ours, with waves of exits alternating with what they term "flow modes" until a collapse ends the game. In our game, however, unless a breakdown occurs, players' unobserved behaviour leads to smooth updating of their beliefs, except for the atom at the last switching time assigned a positive probability.

From a theoretical perspective, our paper is closest to the work of Keller and Rady (2015) and our earlier paper (Bonatti and Hörner, 2011). Our game differs from the former in that actions are not observed and differs from the latter in that the news is bad rather than good.[8] The resulting differences are significant: with good news, the equilibrium is not unique, and the symmetric equilibrium involves interior pure strategies, with experimentation dwindling but never ceasing altogether. With good news, an off-path deviation to greater experimentation leads to increased pessimism and hence less experimentation; thus, behaviour is "mean-reverting," and best replies are necessarily pure and possibly interior. This behaviour also explains why, in the context of good news, the Markov equilibrium with observable actions is actually worse than the symmetric equilibrium with unobservable actions, contrary to what we find with bad news. A comparison with the results of Keller and Rady is provided in Section 6.1.

As noted, the necessity of considering mixed strategies in our game should not be confused with the necessity of allowing pure actions that are not extremal appearing elsewhere. To restore existence in games of strategic experimentation without needing to confront the measure-theoretic difficulties raised by the modeling of independent randomization in continuous time, various authors (*e.g.*, Bolton and Harris 1999; Keller, Rady and Cripps, 2005; Keller and Rady, 2015) have redefined the space of actions available to a player at a given instant to be a convex set (that is, the set of pure strategies is sectionally convex). This redefinition is usually achieved by simple convexification, replacing the lotteries over $\{0, 1\}$ by the interval $[0, 1]$ (with the interpretation of players choosing how to allocate a unit resource),

[8]Another difference is that the game of Bonatti and Hörner (2011) features payoff externalities.

but is not always accomplished in this way: in Keller and Rady (2003), this redefinition involves players choosing two actions at every instant–a mean price and a mean variance. In these papers, this redefinition suffices to restore the existence of an equilibrium in pure (but not extremal) strategies.[9,10] In fact, there are special games for which privately randomizing over stopping times and using non-extremal pure strategies *are* equivalent. The question of whether to describe equilibrium strategies with a hazard rate or with a pure strategy taking value in a convexified set is then merely a matter of convenience. Examples include wars of attrition (for instance, Milgrom and Weber, 1985) or more recent versions of timing games allowing for additional learning (Murto and Välimäki, 2011; Rosenberg, Salomon and Vieille, 2013).

The following is a manifestation of the difference between the equilibria of the games considered in these papers (whether strategic experimentation or games of timing) and the approach in our study: in these equilibria, a player is indifferent regarding *all* his strategies (over the relevant time interval). In the unique equilibrium of our game, a player is indifferent over stopping times (over some interval), but he strictly prefers any of these stopping times to a strategy that uses an interior action over a set of times of positive measure: he is willing to mix, but not to play the pure strategy that specifies the expected value of the mixture.

We are not aware of another paper with a clear economic interpretation in which mixed strategies must be considered to ensure existence, despite convex action sets.[11] Our paper shows that such phenomena are both relevant to economic applications and amenable to mathematical analysis. (See Akcigit and Liu, 2014, Board and Meyer-ter-Vehn, 2014, for models in which mixed-strategy equilibria *might* exist.) Note that such equilibria might also arise when outcomes rather than actions are private, as in Rosenberg, Solan and Vieille (2007). Although this is not the case in their analysis, it is conceivable that equilibrium existence calls for such private strategies to be played in related environments. Relatedly, non-Markovian equilibria might also be required for games with incomplete (rather than

[9] Pure but non-extremal optimal policies contrast with the solution of decision-theoretic versions of bandit problems, which admit optimal solutions within the class of extremal policies. See Yushkevich (1988) or Presman and Sonin (1990).

[10] One should not confuse such a convexification with some clever application of Kuhn's theorem that would obviate mathematical difficulties. Kuhn's theorem also applies to continuous-time games (see Weizsäcker, 1974, or Shmaya and Solan, 2014), but the set of behavioural strategies (properly defined) is much larger than the set of pure strategies, even when action sets are convex.

[11] Of course, it is well known in optimization that sectional convexity is insufficient to guarantee the type of convexity in the policy space that is required for the existence of solutions of optimal control problems. *A fortiori*, the problem arises in games, and there are well-known examples of zero-sum games with sectionally convex action spaces for which the optimal policies cannot be found within the class of pure policies (see Karlin, 1959, and references therein).

imperfect) information and payoff externalities, see Décamps and Mariotti (2004).

Conditional on the state, breakdowns occur independently across players; thus, the heterogeneity in outcomes is obviously neither surprising nor original. Dispersion in beliefs would also be rather expected if players received idiosyncratic, private signals. In this context, however, breakdowns–or their absence–are common-knowledge events. Heterogeneity in beliefs despite public signals arises in only one other setting of which we are aware: repeated games in which players use private strategies (which is precisely what occurs in this situation). However, our game features incomplete information, not simply imperfect public monitoring. More important, in contrast to repeated games, equilibria in private strategies are not merely some of many possibilities: the equilibrium that we solve for is the unique equilibrium of the game.

Indeed, uniqueness of equilibrium is another surprising result, given the literature on strategic experimentation. By contrast, Keller, Rady and Cripps (2005), Bonatti and Hörner (2011) and Keller and Rady (2015) admit multiple equilibria, whereas Bolton and Harris (1999) do not attempt to characterize any equilibrium beyond the symmetric equilibrium.

# 2 The Model

## 2.1 Setup

Time is continuous, and the horizon is infinite. Players $i = 1, \ldots, I$ ($I \geq 2$) choose an action $u^i \in [0, 1]$ at all times.

There is a binary state of the world $\omega \in \{B, G\}$. Players assign a common prior probability $p^0 \in (0, 1)$ to the event $\{\omega = B\}$. Conditional on $\omega$, player $i$'s action controls the instantaneous intensity of a conditionally independent Poisson process $\{N_t^i : t \geq 0\}$. The process $N_t^i$ is interpreted as the number of lump-sum payoffs observed up to time $t$. That is, the action paths $u^i = (u_t^i)_{t=0}^{\infty}$, alongside $\omega$, define the instantaneous intensity of an inhomogeneous Poisson process with intensity $\lambda(t) := \lambda \mathbf{1}_{\{\omega=B\}}(1 - u_t^i)$, where $\lambda > 0$ and $\mathbf{1}_A$ is the indicator function of an event $A$. Note that this intensity is zero if $\omega = G$, independent of the actions chosen. When $u^i = 1$, player $i$ exclusively pulls the *safe arm*, as this choice prevents the occurrence of (costly) lump sums.[12] When player $i$ sets $u^i$ to 0, we state that he pulls the *risky arm* exclusively. Unless a player pulls the safe arm only, he might learn

[12] It might be desirable to allow for "background learning" such that learning is slowed when the safe arm is pulled but does not come to a halt. That is, we may assume that $\lambda(t) := \lambda \mathbf{1}_{\{\omega=B\}}(\bar{u}/I - u_t^i)$, where $\bar{u} > I$. Clearly, long-run beliefs are very different in that case. However, as we explain in Section 6.2, there is no discontinuity in payoffs or equilibrium policies.

about the state. Hence, we state that player $i$ experiments when $u^i < 1$.

Each lump sum entails a cost $h > 0$. That is, given an integrable function $u^i = (u^i_t)$ and the realization of the process $\{N^i_t : t \geq 0\}$, the realized cost of player $i$ is given by

$$\int_0^\infty re^{-rt}(h\mathrm{d}N^i_t + su^i_t\mathrm{d}t),$$

where $r, s > 0$. Note that this is a game of informational externalities only, as player $j \neq i$'s actions do not enter player $i$'s cost.

Throughout this setup, we assume that player $i$ observes the realization of the processes $\{N^i_t : i \in I\}$, the *breakdowns*, and can condition his action on it; however, he observes nothing else. In particular, player $i$ does not observe past values of $u^j_t$, $j \neq i$. That is, players observe outcomes but not actions.

We assume that $g := \lambda h > s$. Therefore, conditional on $\{\omega = B\}$, to minimize the expected cost, it is optimal to allocate the resource exclusively to the safe arm, that is, to set $u^i_t = 1\ \forall t$. Conditional on $\{\omega = G\}$, the risky arm is optimal, independent of other players' actions.

Hence, player $i$'s problem reduces to a course of action up to the first arrival of a lump sum for *any* player, as it is strictly dominant to pull the safe arm thereafter. Let $\tau \in \mathbf{R}_+ \cup \{+\infty\}$ be the time of this first arrival. (Note that $\tau = +\infty$ if $\omega = G$.) Therefore, we can and do assume that the game ends at time $\tau$.

A terminal history $h^\tau$ specifies the stopped action paths $\{(u^i_t)^\tau_{t=0} : i = 1, \ldots, I\}$ up to time $\tau$. We can rewrite the cost for which we minimize the expectation as

$$\mathcal{C}^i(u^i) = \int_0^\tau \left(re^{-rt}su^i_t\mathrm{d}t + e^{-rt}rh\mathrm{d}N^i_t\right) + e^{-r\tau}s, \tag{1}$$

where the last term is the "terminal" cost equal to the expected cost over an infinite horizon conditional on $\{\omega = B\}$ under $u^i_t = 1\ \forall t$.

Some of the parameters are relevant only in combination. In particular, up to normalization, $g$ and $s$ enter only through the cost-benefit ratio $\gamma := (g - s)/s$, and by a standard change in the variable, the discount rate $r$ and intensity parameter $\lambda$ appear via the ratio $\mu := r/\lambda$ only.

## 2.2 Policies and Equilibrium

A deterministic (or "pure") policy for player $i$ is a measurable function $\pi^i : \mathbf{R}_+ \to [0, 1]$ that specifies player $i$'s action $u^i$ at time $t$ conditional on the event $\{t < \tau\}$.[13] Note that the

[13]The policy does not define behaviour after one's own deviation, an unnecessary specification given the information structure. In those rare instances in which we comment on behaviour after such off-path histories,

range of $\pi^i$ is convex (although the set of deterministic policies is not): we interpret $u_t^i$ as the share of $i$'s resources allocated to the safe arm. Let $\Pi^i$ denote the set of all deterministic policies. Of special importance are *stopping policies*, which are defined as follows. Given $t \geq 0$, let $\pi_t^i$ be the policy that sets $\pi_t^i(s) = 1$ for $s < t$ and $\pi_t^i(s) = 0$ for $s \geq t$. The set of stopping policies is denoted $\Pi_S^i$.

Ultimately, it is not sufficient to consider deterministic policies. Mixed policies must be introduced. We adopt the following definition of mixed policies based on Aumann (1964). A mixed policy is a measurable map $\phi^i : [0,1] \to \Pi^i$ such that for all $\beta^i \in [0,1]$, $\phi^i(\beta^i) \in \Pi^i$.[14] This definition can be interpreted as follows: player $i$ privately flips a "coin" at the beginning of the game, and its realization $\beta^i$ determines the deterministic policy that he then follows. Let $\Phi^i$ denote the set of (mixed) policies of player $i$.

Given $\phi^{-i} \in \Phi^{-i} := \times_{j \neq i} \Phi^j$, player $i$ minimizes

$$\mathcal{C}^{\phi^i} := \mathbf{E}_{p^0}^{\phi^i}\left[\mathcal{C}^i(u^i)\right]$$

over $\phi^i \in \Phi^i$.

Of particular interest are *stopping time* policies–"random" stopping policies. According to these policies, for some non-decreasing function $t^i : [0,1] \to \mathbf{R}_+$, $\phi^i(\beta^i) = \pi_{t^i(\beta^i)}^i$ (a.s.). Hence, in these policies, player $i$ randomizes over the time that he stops pulling the safe arm. Let $\Phi_S^i$ denote the set of stopping time policies of player $i$ (including $\Pi_S^i$). It is often more convenient to represent such policies using the distribution function $F^i : \mathbf{R}_+ \to [0,1]$, defined as $F^i(t) := \sup\{\beta^i \mid t^i(\beta^i) \leq t\}$; that is, $t^i$ is the quantile function of $F^i$.

Given that players do not observe one another's actions, there is no loss in considering Nash equilibria. Hence, an *equilibrium* is a vector $\phi^* \in \Phi := \times_i \Phi^i$ such that for all $i$ and for all $\beta^i \in [0,1]$, $\phi^{*i}(\beta^i)$ minimizes $\mathcal{C}^{\phi^i}$ over $\phi^i \in \Phi^i$, given $\phi^{*-i}$. Of particular interest are symmetric equilibria, which are equilibria in which $\phi^j = \phi^i$ for all $i, j$. However, our attention is not restricted to those equilibria.

# 3 Learning

Players face two sources of uncertainty. First, they do not know the state of the world. As time passes without lump sums occurring, they learn about the state. Second, players do

we use the word "strategy" instead.

[14] Let $\mathcal{B}_{[0,1]}$ (resp., $\mathcal{B}$) denote the $\sigma$-algebra of Borel sets of $[0,1]$ (resp., $\mathbf{R}_+$) with the Lebesgue measure. We endow the set of measurable functions from $(\mathbf{R}_+, \mathcal{B})$ to $([0,1], \mathcal{B}_{[0,1]})$ with the $\sigma$-algebra generated by sets of the form $\{f : f(s) \in A\}$ with $s \in \mathbf{R}_+$ and $A \in \mathcal{B}_{[0,1]}$. The notion that such a definition is equivalent to the use of "behavioural decision rules" follows from Weizsäcker (1974). See also Shmaya and Solan (2014) on the equivalence and Touzi and Vieille (2002) on mixed policies in timing games.

not know the specific deterministic policy selected by the other players–if indeed this policy was chosen at random. In this regard, time is also informative: because lump sums are more likely if others pull the risky arm, the absence of such lump sums is indicative of safe play. Both sources of uncertainty affect the choice of optimal action: if player $i$ knew that others were experimenting, then he might be tempted to "free-ride" on this experimentation and pull the safe arm unless he is very optimistic (about the risky arm being the correct one). Nonetheless, we argue here that it is unnecessary for player $i$'s belief to be "multidimensional." As we show below, a one-dimensional statistic suffices.

A second difficulty is that players' beliefs are private. A player who adopts a riskier policy becomes optimistic at a faster pace than if he had adopted a safer policy; indeed, if he pulled the safe arm exclusively, he would only learn from others. This statement implies that other players do not know player $i$'s beliefs. Those players have a belief about his belief, as in equilibrium, they know the distribution over deterministic policies that player $i$ is using.[15]

However, given player $i$'s policy, all these beliefs (including higher-order beliefs) are derived from a common source of information: time. Because the game ends with the first lump sum, there is only one information set corresponding to a given time $t$ (conditional on $i$'s policy throughout). Player $i$ faces no "uncertainty" regarding these conditional beliefs: he can perfectly forecast at time $t$ what his beliefs will be at any time $t' > t$, conditional on no lump sum occurring in the meantime. In particular, he can forecast the instantaneous probability with which a lump sum will occur on that date–but this forecast reflects the two sources of uncertainty that he faces. This *hazard rate* process is all that matters for determining best replies.

Formally, fix a player $i$ throughout. Define $p_t^i := \mathbf{P}_{p^0}^{\phi}[\omega = B \mid (u_s^i)_{s=0}^t]$ for $t < \tau$. As the conditioning clearly indicates, it is player $i$'s belief and his only, although we occasionally omit the superscript. Two properties of his belief process are important. First, we can express the probability of the event that no breakdown has occurred by time $t$, denoted $\varnothing_t$, in terms of $p^i$; by the martingale property of beliefs,

$$\mathbf{P}_{p^0}^{\phi}[\varnothing_t] \cdot p_t^i + (1 - \mathbf{P}_{p^0}^{\phi}[\varnothing_t]) \cdot 1 = p^0,$$

so that

$$\mathbf{P}_{p^0}^{\phi}[\varnothing_t] = \frac{1 - p^0}{1 - p_t^i}.$$

Second, we can derive the law of motion of the (deterministic) process $p_t^i$ taking into account the uncertainty regarding the realized policies $\pi^{-i}$ used by other players. Given that

[15]However, this second-order belief is not common knowledge because player $j$'s posterior belief regarding $i$'s adopted policy depends on $j$'s belief regarding the state of the world, which depends on his own policy (which is not known to others).

breakdowns follow an exponential distribution, the probability of no breakdown by time $t$, conditional on $\{\omega = B\}$ and $(u^i_s)^t_{s=0}$, is given by

$$\mathbf{P}^{\phi}_{p^0}[\varnothing_t \mid \omega = B] = \mathbf{E}^{\phi}_{p^0}\left[e^{-\int_0^t \lambda\left(I-\sum_j u^j_s\right)\mathrm{d}s} \mid (u^i_s)^t_{s=0}\right] = e^{-\int_0^t \lambda(I-u^i_s)\mathrm{d}s}\Pi_{j\neq i}\mathbf{E}^{\phi}_{p^0}\left[e^{\int_0^t \lambda u^j_s \mathrm{d}s}\right],$$

based on the independence of the players' policies. Hence,

$$\frac{p^i_t}{1-p^i_t}\Big/\frac{p^0}{1-p^0} = \frac{\mathbf{P}^{\phi}_{p^0}[\varnothing_t \mid \omega = B]}{\mathbf{P}^{\phi}_{p^0}[\varnothing_t \mid \omega = G]} = e^{-\int_0^t \lambda(I-u^i_s)\mathrm{d}s}\Pi_{j\neq i}\mathbf{E}^{\phi}_{p^0}\left[e^{\int_0^t \lambda u^j_s \mathrm{d}s}\right]. \tag{2}$$

Because the first term on the right-hand side is only a function of player $i$'s own action, uncertainty appears only via the second term. Hence, the log-likelihood ratio is given by

$$\ln \frac{\mathbf{P}^{\phi}_{p^0}[\varnothing_t \mid \omega = B]}{\mathbf{P}^{\phi}_{p^0}[\varnothing_t \mid \omega = G]} = \sum_{j\neq i} \ln \mathbf{E}^{\phi}_{p^0}\left[e^{\int_0^t \lambda(u^j_s - I)\mathrm{d}s}\right] - \lambda(I - u^i_t).$$

Indeed, player $i$'s belief (as measured by the log-likelihood ratio) is private, but his private information appears additively, as captured by the second term on the right-hand side. The contribution to *his* belief attributable to all other players' expected policies is common knowledge.

Because this log-likelihood ratio is differentiable with respect to $t$, we define

$$\nu^{-i}_t := \sum_{j\neq i} \frac{1}{\lambda}\frac{\partial}{\partial t} \ln \mathbf{E}^{\phi}_{p^0}[e^{\int_0^t \lambda u^j_s \mathrm{d}s}]. \tag{3}$$

Note that $\nu^{-i}_t \in [0, I-1]$ because $u^j_s \in [0,1]$, all $s \leq t, j \neq i$.

This can be interpreted as the expected contribution from the other players' experimentation to the hazard rate of player $i$'s belief. Player $j \neq i$'s experimentation affects player $i$'s belief revision at time $t$, and it is not simply a matter of whether player $j$ is playing safe at that time. The entire path of player $j$'s actions affects player $i$'s belief regarding the state of the world at time $t$ and, hence, how much this belief must be revised if no breakdown occurs in the next instant.

It follows from (2) that $p^i$ is also differentiable and that it solves the differential equation

$$\dot{p}^i_t = -\lambda p^i_t(1-p^i_t)(I - u^i_t - \nu^{-i}_t), \quad p^i_0 = p^0. \tag{4}$$

Because the function $\nu^{-i}$ plays an important role in the analysis, it is important to develop some intuition for it. Suppose that players use stopping time policies such that $\phi \in \Phi_S$. Hence, players switch from the safe arm to the risky arm at time $t$ according to some distribution function $F^j : \mathbf{R}_+ \to [0,1]$. Write $\bar{F}^j = 1 - F^j$ for the complementary distribution function. This approach also allows us to provide an alternative, perhaps more expressive, formula for $\nu^{-i}$. By definition,

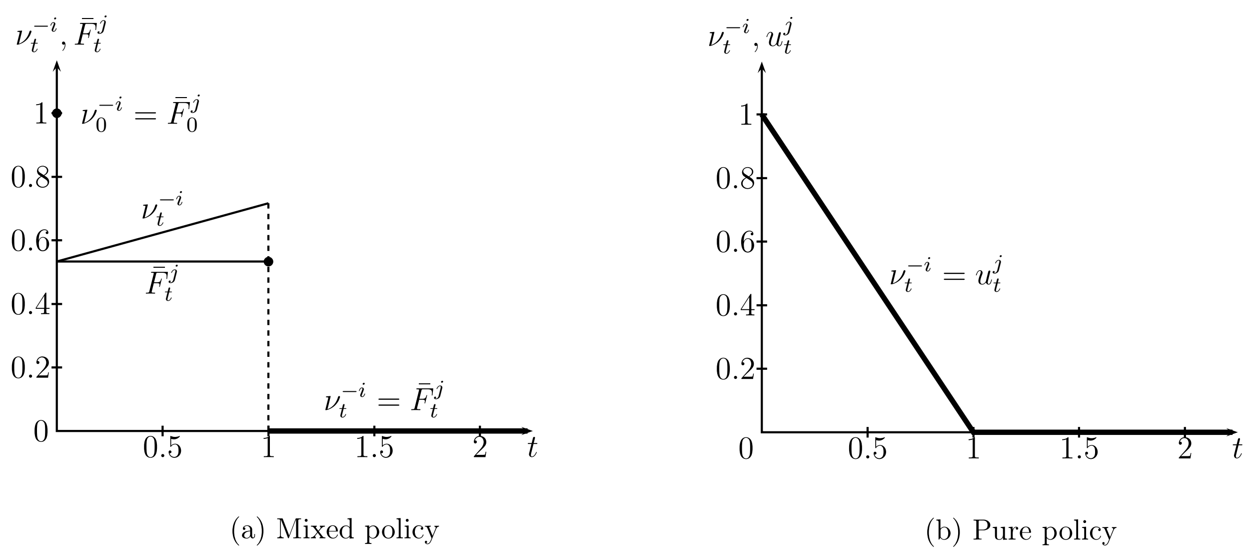
$$\nu_t^{-i} = \frac{1}{\lambda}\sum_{j\neq i}\frac{\partial}{\partial t}\ln \mathbf{E}_{p^0}^{\phi}[e^{\int_0^t \lambda \mathbf{1}_{\{s\leq t^i(\beta^i)\}}\mathrm{d}s}] = \frac{1}{\lambda}\sum_{j\neq i}\frac{\partial}{\partial t}\ln\left[e^{\lambda t}(1-F_t^j) + \int_0^t e^{\lambda s}\mathrm{d}F_s^j\right],$$

such that, explicitly,

$$\nu_t^{-i} = \sum_{j\neq i}\frac{e^{\lambda t}\bar{F}_t^j}{\bar{F}_0^j + \int_0^t \lambda e^{\lambda s}\bar{F}_s^j \mathrm{d}s} \geq \sum_{j\neq i}\frac{e^{\lambda t}\bar{F}_t^j}{1+\int_0^t \lambda e^{\lambda s}\mathrm{d}s} = \sum_{j\neq i}\bar{F}_t^j. \tag{5}$$

It follows that $\nu_t^{-i}$ is a function that begins at $I-1$, remains there as long as $F^j(t)=0$ for all $j\neq i$, discontinuously decreases when $F^j$ discontinuously increases for some $j\neq i$, and continuously increases when $t\notin\cup_{j\neq i}\operatorname{supp}F^j$, strictly, unless it is equal to 0, which occurs when $F^j(t)=1$ for all $j\neq i$.[16] It always exceeds the total probability of others not having stopped pulling their risky arm, as the likelihood of a breakdown is lowest when they have not yet done so, and player $i$ must entertain the possibility that they already have begun pulling the risky arm.

Figure 1 illustrates this scenario for the case of two players from the perspective of player $i$: in the case of some (arbitrary) pure policy (when player $j$ selects a deterministic policy $u^j$), the hazard rate $\nu^{-i}$ coincides with it; in the case of a mixed policy, $\nu^{-i}$ and $\bar{F}^j$ coincide (at least) at the initial instant and once they reach 0.




(a) Mixed policy

(b) Pure policy

Figure 1: Hazard rate $\nu^{-i}$ compared with $\bar{F}^j, u^j$, $I=2$.

[16] Given a distribution $G$, we write $\operatorname{supp}G$ for the set of points of increases in $G$.

# 4 Best Replies

This section elucidates the structure of the best-reply function of a fixed player $i$, taking the behaviour of others as given, as summarized by the hazard rate $\nu^{-i}$. This procedure involves five steps. First, we explain why this hazard rate is indeed a summary statistic for the best-reply problem. Second, we show that any best reply is necessarily a stopping time policy. Third, we derive the unique cooperative solution as an immediate by-product, in which $\nu^{-i} = I-1$. Fourth, in the case of two players, we solve for the best-reply function and show how its structure–first increasing and then decreasing to 0–eliminates the possibility of the existence of an equilibrium in pure policies. Fifth, we explain why this non-existence extends to the case of more than two players.

## 4.1 The Certainty-Equivalent Problem

The optimization problem faced by player $i$ satisfies certainty equivalence: *the optimal action $u_t^i$ is exactly the same as it would be if all unknowns were known and if their values equaled their best estimates (the conditional expectations), given by $(p^i, \nu^{-i})$.* Furthermore, a separation principle holds: optimal estimation and optimal control can be decoupled. As clearly illustrated in the definition of $(p^i, \nu^{-i})$, the choice of $u^i$ does not affect this estimate.

As an illustration of this, according to the law of iterated expectations, we may now rewrite the problem of minimizing (1) as, equivalently, minimizing

$$\int_{t\geq 0} e^{-rt}\left(rp_t^i g\left(1-u_t^i\right) + ru_t^i s + \lambda p_t^i (I - u_t^i - \nu_t^{-i})s\right)\frac{1-p^0}{1-p_t^i}\mathrm{d}t \tag{6}$$

over measurable policies $\pi^i : \mathbf{R}_+ \to [0,1]$, subject to (4), *e.g.*,

$$\dot{p}_t^i = -\lambda p_t^i(1-p_t^i)(I - u_t^i - \nu_t^{-i}), \quad p_0^i = p^0.$$

This is the program $\mathcal{P}$.[17] Here, the function $\nu^{-i} : \mathbf{R}_+ \to [0, I-1]$ is treated as an exogenous (measurable) function. We omit it as an explicit argument of $\mathcal{P}$. By the Filippov-Cesari theorem (see Cesari, 1983), a solution exists, that is to say, the infimum is achieved. We will examine the necessary conditions given by Pontryagin's maximum principle.

The interpretation of the objective is as follows. As explained above, $(1-p^0)/(1-p_t^i)$ is the probability of reaching time $t$ without a breakdown. At that time, if player $i$ invests $u_t^i$ in the safe arm, then the instantaneous probability that he suffers a breakdown is $(1-u_t^i)\lambda p_t^i \mathrm{d}t$, with the expected cost $rh$. If any of the players has a breakdown (which occurs with probability

[17] With a slight abuse: the program $\mathcal{P}$ examined in the appendix is a slight modification of it.

$\lambda p_t^i(I - u_t^i - \nu_t^{-i})\mathrm{d}t)$, then player $i$ switches to the safe arm, yielding the net present cost $s$. As was the case for learning dynamics, the pair $(p^i, \nu^{-i})$ also summarizes all the information that matters for computing payoffs.

## 4.2 Stopping Time Policies

Here, we show that any best reply must be within the class of stopping time policies.

**Lemma 1** *If $\pi^i \in \Pi^i$ solves $\mathcal{P}$, then $\pi^i \in \Pi_S^i$.*

Informally, Lemma 1 states that if a player begins experimenting, he should do so indefinitely (*i.e.*, until a breakdown occurs), and conversely, if he plays safe, he must have played safe at all earlier times. To gain further intuition, consider the *arbitrage equation* of player $i$, which describes the trade-off between *backloading* and *frontloading* experimentation. This equation is silent regarding the optimal amount of experimentation at time $t$ and suggests only the optimal *timing* of a fixed amount of experimentation. The marginal value of backloading experimentation is given by

$$r\left(p_t^i g - s\right) + \lambda p_t^i\left(I - u_t^i - \nu_t^{-i}\right)(g - s) - \lambda p_t^i (g - s)\left(1 - u_t^i\right). \tag{7}$$

The first term is the time-preference effect of delaying the expected flow cost $p_t g$ and anticipating the cost $s$. The second term pertains to the event of a breakdown (at rate $\lambda p_t^i\left(I - u_t^i - \nu_t^{-i}\right)$): if so, safe play would occur at $t + \mathrm{d}t$ regardless of the player's earlier action; in that event, pulling the safe arm more at $t$ yields marginal savings of $g - s$. Finally, the third term considers the effect of the player's action on the likelihood of a breakdown: by frontloading safe play, the player reduces (at a rate $\lambda p_t^i$) the arrival of a breakdown, in which case he would switch from the current action $u_t^i$ to $u^i = 1$; because this scenario can occur only in the bad state, this action yields a loss $g - s$.

Note that the sum of the last two terms is non-negative. Hence, equation (7) implies that backloading is profitable when $p$ is sufficiently large. Conversely, if a player were certain that the state is good, discounting would suggest frontloading the risky action. Lemma 1 then establishes that over the relevant range of beliefs (*i.e.*, for $p^i \geq p^*$; see Lemma 2), the marginal value of backloading is positive.

Finally, note that Lemma 1 does not imply that the solution to $\mathcal{P}$ is unique; rather, it implies that all deterministic solutions are in $\Pi_S^i$. Furthermore, one can determine the bounds on how early or late a player is willing to switch to the risky arm. Next, we provide such bounds in terms of player $i$'s beliefs. We define

$$\frac{p^*}{1 - p^*} := \frac{\mu + I}{\mu + I - 1}\frac{1}{\gamma},$$

as well as

$$\frac{p^{**}}{1-p^{**}} := \frac{\mu+1}{\mu}\frac{1}{\gamma},$$

where we recall that $\gamma = (g-s)/s$ and $\mu = r/\lambda$. As is immediately observed from the formula, $p^{**} > p^{*}$, given $I \geq 2$. The next result establishes that once a player becomes sufficiently optimistic (specifically, when $p_t^i < p^*$), he allocates his entire resource to the risky arm.

**Lemma 2** *If $\pi^i$ solves $\mathcal{P}$, then $u_t^i = 0$ for all $t$ such that $p_t^i < p^*$. Conversely, if $u_t^i = 0$ yet $p_t^i > p^*$, then $\nu_t^{-i} > 0$.*

The belief $p^*$ is the threshold value at which a player is willing to experiment if all other players are pulling the risky arm thereafter. Lemma 2 establishes a lower bound on experimentation because, intuitively, the temptation to free-ride is strongest when all other players are pulling the risky arm. Hence, if $p^0 < p^*$, we have finished: in the unique equilibrium, all players choose $u_t^i = 0$ at all times. In what follows, we assume that $p^0 \geq p^*$.

The next result establishes that at least one player must assign positive probability to switching to the risky arm no later than when his belief reaches $p^{**}$.

**Lemma 3** *If $\nu_t^{-i} = I - 1$ and player $i$'s best reply is $\bar{F}^i(t) = 1$, then $p_t^i \geq p^{**}$.*

It can be shown (previewing the cooperative solution characterized below) that the upper bound on the amount of experimentation $p^{**}$ coincides with the threshold belief for the single-agent problem. This theme is familiar in exponential-bandit models (Keller, Rady and Cripps, 2005, and Bonatti and Hörner, 2011) in which players are not willing to experiment more than in the single-player case. However, in contrast to good-news models, in which equilibrium beliefs "reach" the stopping region, the lower bound $p^*$ depends on the number of players because of the amount of information generated by $I-1$ players pulling the risky arm.

## 4.3 Cooperative Solution

We briefly mention the cooperative solution, which is easily derived from the previous result. Assume that players perfectly observe one another's actions (an innocuous assumption because the optimum involves pure policies) and choose them so as to minimize the sum of their costs. We define

$$\frac{p^{FB}}{1-p^{FB}} := \frac{\mu+I}{\mu}\frac{1}{\gamma}.$$

Note that $p^{FB}$ is larger than $p^{**}$, given $I \geq 2$. In fact, it coincides with $p^{**}$ and hence also $p^*$ when $I = 1$ is inserted into the formulas.

Given a pair $(p, u)$ such that $p$ is the belief path generated by $u := \sum_i u^i$, given $p^0$, along the history with no breakdown, the action path $(u_t)_t$ is *measurable with respect to the belief path* $(p_t)_t$ if $p_t = p_{t'} \Rightarrow u_t = u_{t'}$ for all $t, t'$. We write $u(p)$ for the value of $u$ of belief $p \leq p^0$, which is then well defined. The cooperative solution given in the next lemma is measurable with respect to its belief path.

Unsurprisingly, the optimal policy involves having all players employ the safe arm until belief $p^{FB}$ is reached and then switching to the risky arm. This approach is stated below (see also Keller and Rady 2015, Proposition 1). The next lemma also establishes that total costs decrease in the intensity with which the risky arm is pulled, as long as the ranking of intensities holds pointwise in the beliefs.

**Lemma 4** *The cooperative solution $u^{FB}$ is given by $u_t^{FB} = I$ for all $t$ such that $p_t \geq p^{FB}$ and $u_t^{FB} = 0$ otherwise. Furthermore, let $p', p'' : \mathbf{R}_+ \to \mathbf{R}$ be two feasible paths such that the corresponding action paths $u', u''$ are measurable with respect to their belief path, with $u^{FB}(p) \leq u'(p) \leq u''(p)$ for all $p \leq p^0$. The cost is then weakly lower under $p'$ than under $p''$ and strictly lower when $u'(p'_t) < u''(p'_t)$ for a set of times $t$ of positive measure.*

## 4.4 Best Replies with Two Players

To understand why equilibrium is necessarily in mixed policies (unless $p^0 \notin (p^*, p^{**})$), it is useful to derive the best-reply correspondence in the special case of two players. Suppose that player $j \neq i$ switches (with probability one) to the risky arm at time $t^j$. We may distinguish player $i$'s cost according to whether he switches to the safe arm first or second.

If player $i$ decides to go second, he must do so when his private belief reaches the threshold $p^*$ (or immediately if this belief has been reached by the time $j$ switches). Hence, if going second is best, then player $i$'s best reply must be

$$t^i = t^j + \lambda^{-1} \ln \left( \frac{p^0}{1 - p^0} \Big/ \frac{p^*}{1 - p^*} \right).$$

The fixed delay is equal to the time required for beliefs to reach the threshold $p^*$ based on player $j$'s experimentation alone.

If $i$ decides to go first, then player $i$ will begin experimenting immediately, conditional on wanting to preempt player $j$. Intuitively, player $i$ will not learn before time $t^j$ unless he experiments. If player $i$ is not willing to wait until then, he should begin immediately. If delaying experimentation is not as costly as choosing the risky arm while still being

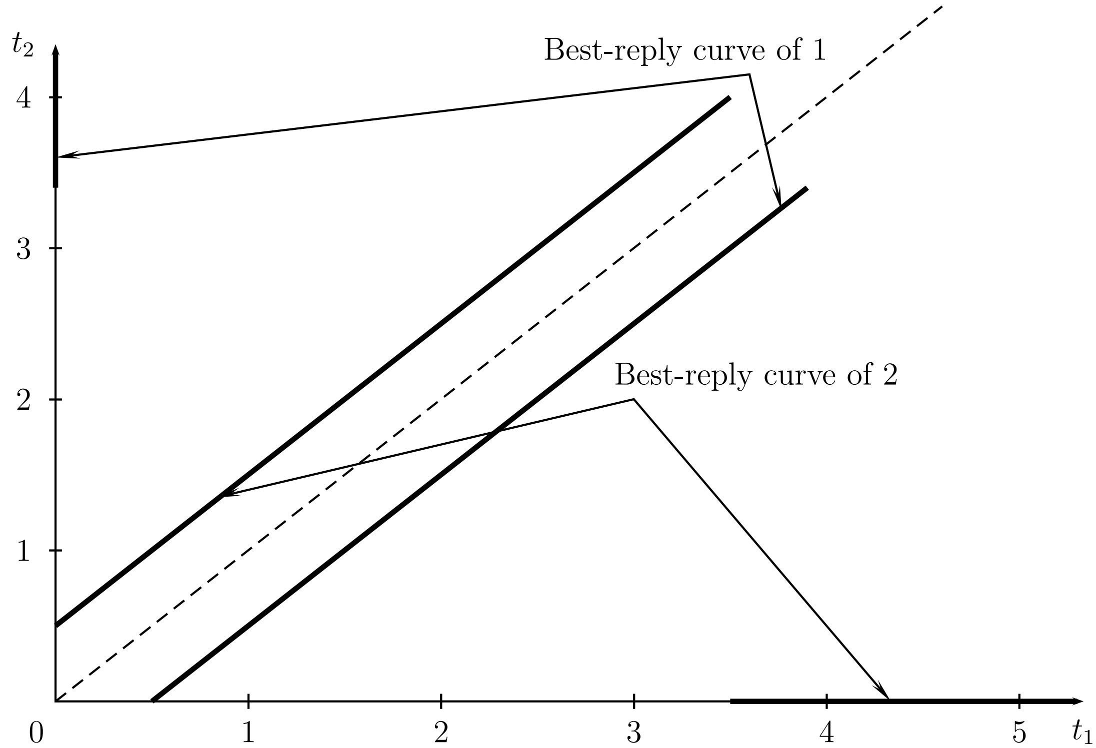



Figure 2: Best-reply curves with two players for $(r, \lambda, \gamma, p^0) = (1/10, 1, 4, 1/2)$.

pessimistic, then player $i$ will choose to "freeze" beliefs until $t^j$. This scenario is intuitive: the incentive to wait and "be second" only grows as time passes: if a player were to delay his decision to switch to the risky arm, then his incentive to move second would only grow, as he would have to wait for less time before benefiting from the other player's experimentation. Hence, if moving first is the preferred course of action, then moving immediately is best.

What remains to be determined is when player $i$ prefers to go first or second. This preference depends on when player $j$ switches. Unsurprisingly, the larger $t^j$ is, the more tempting it is to go first. Intuitively, if $t^j$ is very high, then the cost of waiting until player $j$'s actions take beliefs to the threshold causes an overly costly delay in learning. Conversely, when player $j$ is expected to switch to the risky arm soon, the benefits of free-riding on his experimentation when beliefs are most pessimistic outweigh the cost of delay. This scenario is summarized by the following lemma.

**Lemma 5** *The best-reply correspondence $t^i : \mathbf{R}_+ \rightrightarrows \mathbf{R}_+$ is given by, for some $\hat{t} \in \mathbf{R}_+$,*

$$t^i(t^j) = \begin{cases} t^j + \lambda^{-1} \ln \left( \frac{p^0}{1-p^0} \Big/ \frac{p^*}{1-p^*} \right) & \textit{if } t^j < \hat{t}, \\ \left\{ 0, \hat{t} + \lambda^{-1} \ln \left( \frac{p^0}{1-p^0} \Big/ \frac{p^*}{1-p^*} \right) \right\} & \textit{if } t^j = \hat{t}, \\ 0 & \textit{if } t^j > \hat{t}. \end{cases}$$

Consequently, player $i$'s best-reply curve shifts downward at $\hat{t}$. The two best-reply curves do not cross, and no pure-policy equilibrium exists. Figure 2 provides an illustration.

If player $j$ uses *precisely* $\hat{t}$, then player $i$ has two best replies. However, consider Figure 2: for each of his choices, player $j$'s best reply would be much smaller than $\hat{p}$ (indeed, it would be 0 if $i$ used the larger best reply and $\lambda^{-1} \ln \left( \frac{p^0}{1-p^0} \Big/ \frac{p^*}{1-p^*} \right)$ if he uses 0). Unsurprisingly, regardless of how player $i$ randomizes between these two choices, player $j$'s best reply is strictly lower than $\hat{t}$. We immediately obtain the following result.

**Lemma 6** *Suppose that $I = 2$ and $p^0 > p^*$. There exists no equilibrium in which either player uses a pure policy.*

### 4.5 More Than Two Players

Can an equilibrium in pure policies exist when $I > 2$? Deriving best-reply curves is no longer an easy task. However, a pure-policy equilibrium cannot exist based on the following simple argument. Suppose that such an equilibrium exists, and let $t^i$ denote the time at which player $i$ switches to the risky arm. Without loss of generality, suppose that $t^1 \geq t^2 \geq \cdots$. Suppose first that $p(t^3) > p^*$. Consider the game starting at time $t^3$ and the corresponding initial belief $p(t^3)$. This game involves only two players, players 1 and 2 (assuming indeed that $t^3$ is optimal for player 3). A necessary condition for the policy profile to be an equilibrium is that players 1 and 2 play mutual best replies in this game–yet the two-player game admits no pure-policy equilibrium. If instead $p(t^3) = p^*$, then given Lemma 3, because $p(t^I) \geq p^{**}$, there exists $j$ such that $p(t^j) > p^* = p(t^{j-1}) = \cdots = p(t^1)$. As in the two-player case, past time $t^j$, any player $i = 1, \ldots, j-1$ would gain from unilaterally deviating to the risky arm immediately.[18]

## 5 Main Results

### 5.1 Symmetric Equilibrium

We now turn to the equilibrium analysis. Recall that we assume throughout that $p^0 > p^*$. Given $F^{-i}$ and, hence, given $\nu^{-i}$, each time $\tau \in \operatorname{supp} F^i$ is such that the stopping policy $\pi^i_\tau$ is a solution to $\mathcal{P}$. Furthermore, it holds that, given any $\tau \in \operatorname{supp} F^i$, $p_\tau \geq p^*$. We let $\bar{\tau}^i := \max\{\tau \in \mathbf{R}_+ : \tau \in \operatorname{supp} F^i\}$.

[18] For any number of players $I$, the proof of Lemma 5 establishes that if players $-i$ switch at $t = T$, then player $i$ wants to switch at a different time.

First, we focus on symmetric equilibria and accordingly write $F$, $\bar{\tau}$ for $F^i$, $\bar{\tau}^i$, unless we emphasize a given player's perspective. The next result derives the unique symmetric equilibrium of the game.

**Theorem 1** *There exists a unique symmetric equilibrium. If $p^0 \geq p^{**}$, then the equilibrium is pure and involves $F^i(t) = 0$ at all times.*

*If $p^0 \in (p^*, p^{**})$, the equilibrium involves mixed policies. Specifically, player $i$ chooses a stopping policy $\pi_t$ among the set $[0, \bar{\tau}]$, with $\bar{\tau} > 0$ and $p_{\bar{\tau}} = p^*$; this distribution is positive and continuous over $(0, \bar{\tau})$ and has an atom at times $t = 0, \bar{\tau}$.*

The equilibrium distribution function can be solved in closed form. Namely, let

$$A := \left(1 - \gamma \frac{\mu}{1+\mu} \frac{p^0}{1-p^0}\right)^{-1}.$$

We obtain, normalizing $\lambda$ to 1,

$$\bar{F}(t) = \left(\frac{A - e^{\mu t}}{A - 1}\right)^{\frac{1}{I-1}} \left(1 - \frac{\mu}{(I-1)(Ae^{-\mu t} - 1)}\right),$$

and

$$\bar{\tau} = \frac{1}{\mu} \ln \left( \frac{I-1}{(I+\mu-1)\left(1+\mu-\gamma\mu\frac{p^0}{1-p^0}\right)} \right).$$

Figure 3 illustrates the equilibrium distribution (left panel) and compares the complementary distribution function $\bar{F}^i_t$ with the hazard rate $\nu^{-i}_t$ (right panel). Recall the two sources of uncertainty that each player faces. Over time, players learn from their own experience and from that of others. In particular, as time passes, a player assigns growing weight to the event in which this opponent has already switched to the risky arm, conditional on which learning occurs faster. Moreover, the contribution of other another player's experimentation to player $i$'s learning ($\nu^{-i}$) is always smaller than the survival rate of the opponent's distribution because, as time passes and no breakdown occurs, player $i$ also assigns growing weight to subsequent realizations of his opponent's switching time, which slows this learning process.

The maximum range of stopping times in the symmetric equilibrium has a natural interpretation: the "earliest" that a player may switch to the risky arm is when his belief is $p^{**}$: this is the belief for which he would switch if he were on his own (cf. Section 4.3). The latest he might switch is when his belief reaches $p^*$: this would be his uniquely optimal belief if all others were always experimenting. Because his opponents' behaviour lies somewhere between these two extremes, so does his set of best replies.

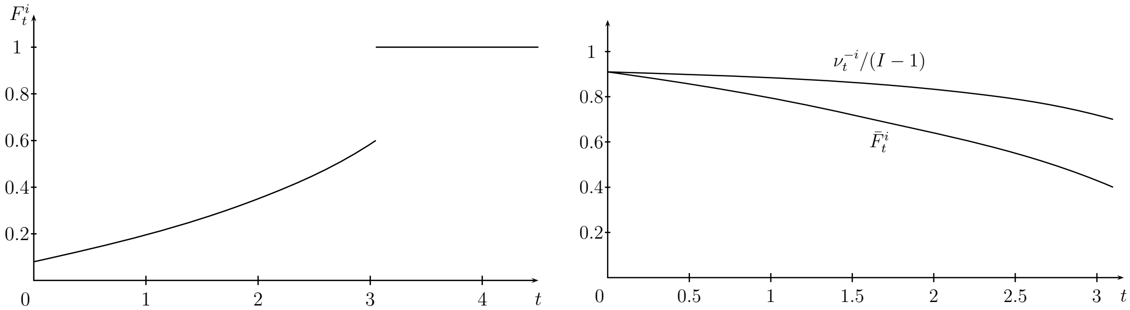



Figure 3: Equilibrium distributions $F_t^i$ (left), $\bar{F}_t^i$ and hazard rate $\nu_t^{-i}$ (right) for $(r, \lambda, \gamma, I, p^0) = (1/10, 1, 4, 3, 1/2)$.

In contrast to typical mixed-strategy equilibria of normal-form games, player $j$ does not *need* to randomize over stopping policies to make player $i$ indifferent over all stopping times in the relevant time interval. He could just as easily play a deterministic policy $u^j = \nu^{-i}$. The reason that players randomize is that player $j$ is not willing to play such a deterministic but interior policy. Randomizing over stopping policies is the unique cost-minimizing way to make player $i$ indifferent over stopping times. This fact has rich implications for the dispersion of equilibrium beliefs.

Figure 4 illustrates belief paths as a function of time and behaviour. For instance, $p^c$ is the belief path of player $i$ who chooses the latest possible equilibrium stopping time, $\bar{\tau}$. The solid line indicates when he pulls the safe arm; the dashed line indicates when he has already switched to the risky arm. Trajectories $p^a$ and $p^b$ correspond to earlier switching stopping times. Once the player begins pulling the risky arm, his belief decreases faster, reinforcing his preference for the risky arm (absent any breakdown). Trajectories $p^d$ and $p^e$ are "counterfactual" trajectories in which player $i$ is more pessimistic than is possible. He then has an incentive to pull the safe arm longer than in equilibrium and to switch once his belief reaches $p^*$, which is later than the latest possible equilibrium time at which his opponents might have switched.

This elucidates the "off-path" behaviour of player $i$. After an arbitrary history $(u_s^i)_{s=0}^t$ (along which he might have deviated from the prescribed behaviour), Lemmas 1–2 remain valid: player $i$'s optimal policy is a stopping policy (from time $t$ onward) that prescribes stopping no later than the first time his belief reaches $p^*$.

Finally, the necessity to randomize is a robust phenomenon: as Figure 2 clearly indicates, given that best-reply curves vary continuously with the parameters, the non-existence of pure-policy equilibrium is robust to perturbations in parameters, regardless of whether symmetry is preserved. Furthermore, it is not necessary to consider that when the safe

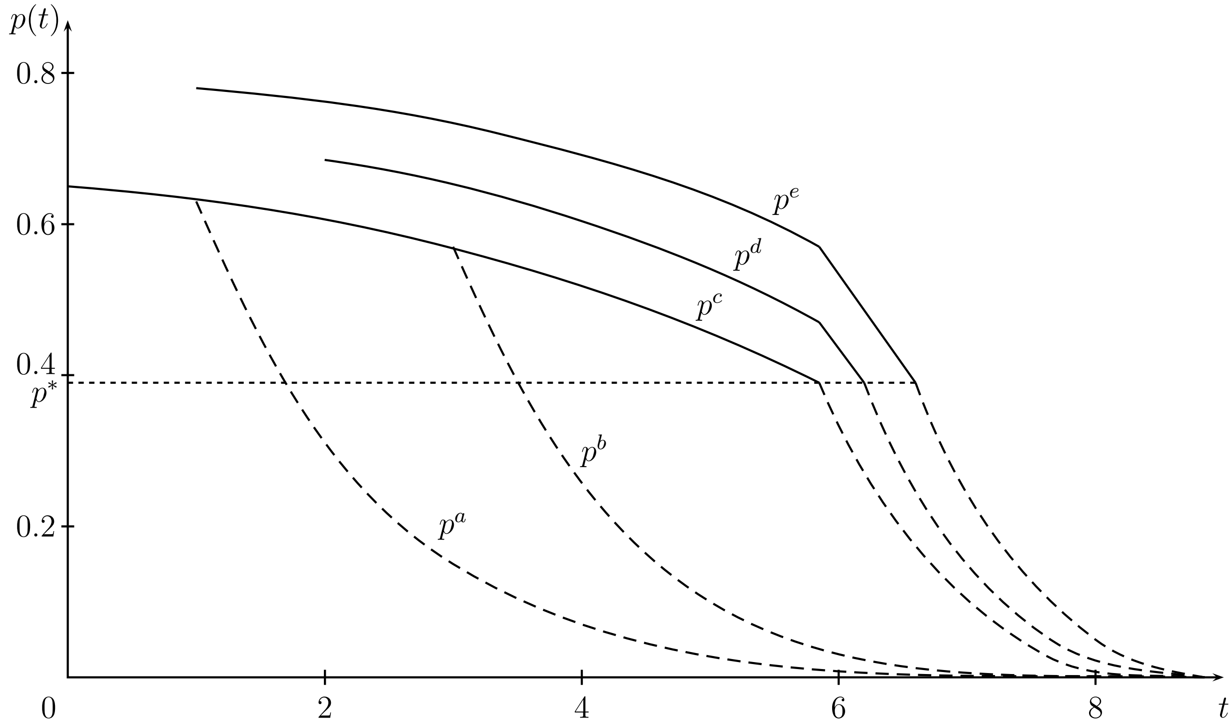



Figure 4: Belief trajectories for $(\gamma, p^0, I, \mu) = (3, 13/20, 2, 1/8)$.

arm is pulled, no learning occurs. Our results generalize to the case containing background learning. In that case, even if the initial belief is above $p^{**}$, players use stopping time policies in the unique symmetric equilibrium, which is mixed. However, the smallest stopping policy with the support of randomization has a stopping time that is strictly positive and that precisely corresponds to the time when their belief reaches $p^{**}$. Background learning is further discussed in Section 6.2 below.

**Derivation of the equilibrium distribution.** Fix the other players' behaviour in terms of $\nu_t^{-i}$, and consider player $i$'s stopping time. The first-order effect of playing safe longer (differentiating (6)) is given by

$$\frac{e^{-\mu T}}{1-p_T}\left(\mu s - p_T\left(\mu g + s\right)\right) + \int_T^\infty e^{-\mu t}\left(\mu g + (I - \nu_t^{-i})s\right)\frac{p_t}{1-p_t}\mathrm{d}t.$$

The first term (which is negative) captures the myopic benefit (cost reduction) of playing safe longer. The second term is instead the added cost of slower learning, which is captured by a higher hazard rate of a breakdown at all future times.

Pointwise indifference requires the marginal cost of playing safe longer to be nil over the entire support. Thus, we turn to the second-order effect of playing safe, which is given by the sum of the following four terms,

$$-\frac{e^{-\mu T}\mu}{1-p_T}\left(\mu s - p_T\left(\mu g + s\right)\right) - (I-\nu_T^{-i}-1)\frac{e^{-\mu T}p_T}{1-p_T}\left(\mu s - \left(\mu g + s\right)\right)$$
$$-\frac{e^{-\mu T}p_T}{1-p_T}\left(\mu g + (I-\nu_T^{-i})s\right) + \int_T^\infty \frac{e^{-\mu t}p_t}{1-p_t}\left(\mu g + (I-\nu_t^{-i})s\right)\mathrm{d}t.$$

The second-order effect is given by (a) the delayed myopic benefit, (b) the lower myopic benefit (note that $(I-\nu_T^{-i}-1)p_T/(1-p_T)$ is the derivative of the hazard rate), (c) the postponed cost of diminished learning, and (d) the higher marginal cost of diminished learning (because the hazard rate is exponential in $u^i$). Because the first-order condition must hold pointwise, the last term is equal to the myopic (first-order) benefit of delaying switching.

These four terms can be combined into an expression characterizing the equilibrium $\nu^{-i}$ as a function of the belief $p$,

$$\frac{p}{1-p}(g-s)\left(I+\mu-\nu^{-i}-1\right) - s(\mu+1). \tag{8}$$

Note that these beliefs are those of the most pessimistic type, *i.e.*, the player who has not yet switched to the risky arm.

Next, we use the state equation to derive $\nu_t^{-i}$ as a function of time alone. We then derive the equilibrium distribution from the definition of $\nu^{-i}$ as

$$\frac{\nu_t^{-i}}{I-1} = \frac{(1-F_t)e^{\lambda t}}{(1-F_t)e^{\lambda t} + \int_0^t e^{\lambda s}\mathrm{d}F_s},$$

which yields a differential equation for $F_t$, resulting in

$$F_t = 1 - \frac{\nu_t^{-i}}{I-1}e^{\int_0^t\left(\frac{\nu_s^{-i}}{I-1}-1\right)\mathrm{d}s},$$

and we then plug the formula for $\nu_t^{-i}$.

## 5.2 Uniqueness

As discussed above, no equilibrium in pure (deterministic) policies exists, regardless of symmetry. Thus, asymmetric equilibria in mixed policies is possible. As Figure 2 clarifies, our game is not supermodular: in particular, best-reply curves are not monotone, which makes it difficult to establish uniqueness. This implies the failure of standard methods to prove uniqueness.[19] Moreover, the different methods and tricks described in Karlin (1959) do not

[19]See Vives (1999) for an excellent discussion. Because best-reply curves are downward sloping, existing arguments based on supermodular games are ineffective; the best-reply function is not a contraction either (otherwise, the equilibrium would be pure), and the fact that the equilibrium is mixed implies that the Gale-Nikaido theorem or the Poincaré-Hopf theorem cannot work either or rather that one should work with the mixed-strategy space directly and possibly use an infinite-dimensional extension of those.

appear to be effective. In the case of two players, it can be shown that no such equilibrium exists.[20] Our proof carries no philosophical charm and is based on particular features of the payoff function.

**Theorem 2** *Assume that $I = 2$. The equilibrium is then unique (and thus equal to the mixed equilibrium of Theorem 1).*

Uniqueness contrasts with the multiplicity that is prevalent in games with strategic experimentation, not only when actions are observable (Keller, Rady and Cripps, 2005; Keller and Rady, 2015) but also when they are not (Bonatti and Hörner, 2011). Because of the pervasive free-riding incentives, asymmetric equilibria typically exist when players alternate (finitely or infinitely often) between experimenting and taking advantage of the opponent's experimentation–leading to the existence of additional asymmetric equilibria. By contrast, in our game, free-riding finds its expression in how early a player is willing to begin experimenting; the earlier the opponent begins experimenting, the later the player finds it optimal to do so. However, the *ordering* of actions is unambiguous: for a given total amount of experimentation, it is always optimal to use a stopping policy, pulling the safe arm if and only if a threshold time has not yet been reached.[21] It is impossible to determine a player's incentives to use a policy that would involve pulling the risky arm before the safe arm, precluding any type of alternation in the experimentation that players conduct.

### 5.3 Comparative Statics

As the number of players increases, the free-rider problem worsens in terms of both the timing and the amount of experimentation.

In computing beliefs, we encounter a difficulty: the belief that player $i$ holds at a given time is not uniquely determined in the mixed equilibrium; the earlier a player stops, the lower is his belief at a given time $t$, provided that no breakdown has occurred. We are thus led to adopt the perspective of an *outside observer* who observes nothing at all: conditional on a given time $t$ being reached without a breakdown under either informational assumption, what probability does he attach to the event $\{\omega = B\}$? In the observable case, this belief coincides with that of any player on path. In the unobservable case, it is some weighted average of a player's belief where the weight reflects the probability attached by this observer to a player

[20]For more than two players, uniqueness is an open problem.

[21]This is also the key reason why the equilibrium must be in mixed policies and not in pure policies with non-extremal actions: for a given amount of experimentation, players have strict incentives to backload, eliminating the possibility of pulling arms with non-extremal intensity over any interval of time.

switching to risky play at a given time, suitably updated given that time $t$ is reached without a breakdown. Formally, we compute

$$p_t = \mathbf{P}^{\phi}_{p^0}[\omega = B \mid \varnothing_t],$$

where, unlike in Section 3, we do not condition on any particular player's action path. It follows that the outside observer's belief satisfies

$$\dot{p}_t = -p_t(1 - p_t)I(1 - \nu^i_t).$$

For the purpose of the next proposition, we index distributions and stopping times, among others, by the number of players $I \geq 1$.

**Proposition 1**

1. *The distributions $F^I$ are ranked by stochastic dominance: $F^I_t$ decreases in $I$, for all $t$.*

2. *For an outside observer, $\nu^{I'}_t \geq \nu^I_t$ for all $I' > I$, with strict inequality for all $t < \bar{\tau}_{I'}$.*

3. *For $I' > I$, the belief path $p^{I'}_t$ crosses $p^I_t$ once (from above).*

4. *For all $I > 1$, total individual costs in the symmetric equilibrium are given by*

$$p^0(g\mu + s) - \mu s. \tag{9}$$

In summary, as the number of players increases, the "mixing phase" lasts longer (until $\bar{\tau}_I$) and drives beliefs to a lower threshold $p^*$. The worsening free-riding problem implies full dissipation of the positive informational externalities generated by an additional player. Moreover, adding one player does not modify the upper bound on experimentation $p^{**}$. Thus, although the social planner would like to experiment even under more pessimistic prior beliefs, there can be no experimentation in equilibrium if the prior is below a constant threshold.

The distributions of stopping times $F_t$ with different $I$ are ranked by first-order stochastic dominance: a larger number of players increases the likelihood of later stopping times. Furthermore, the expected hazard rate from the outside observer's perspective $I - \nu^I_t$ is decreasing in $I$ as long as $p^*$ has not been reached. The outside observer's beliefs facing $I' > I$ players eventually overtake the beliefs that he would hold with $I$ players. In Figure 5, we illustrate the hazard rate $I - \nu^I_t$ and the belief paths for $I = 2, 4$.

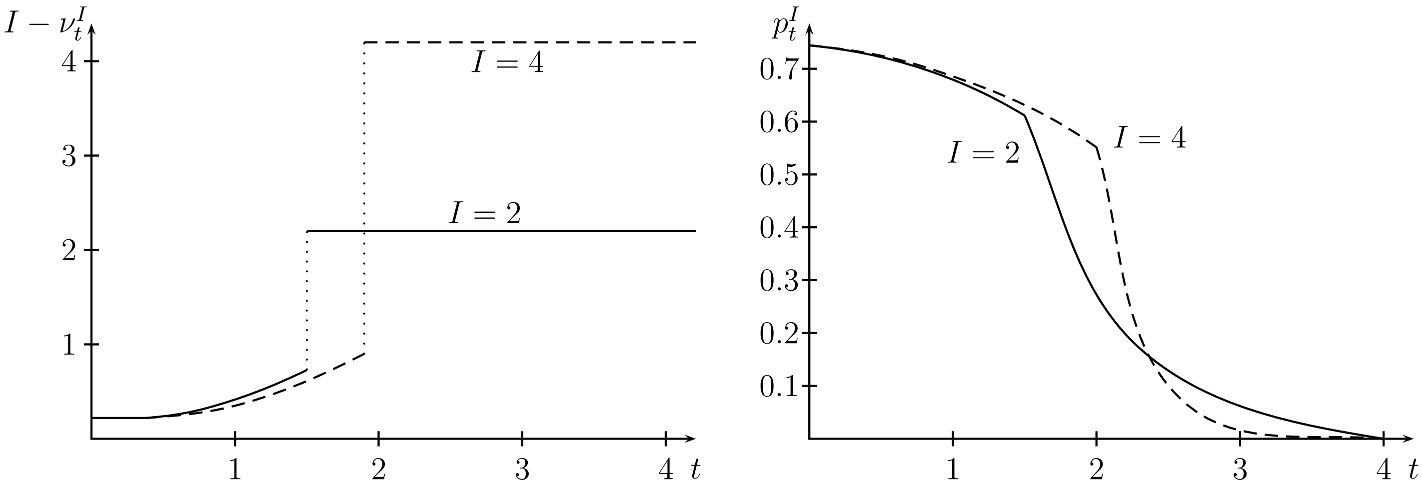



Figure 5: Hazard rate and belief paths for $(\mu, \gamma, p^0) = (1/4, 1, 1/2)$.

# 6 Remedies: Information and Subsidies

Given the inefficiencies of the equilibrium outcome, it is natural to ask what interventions could possibly be helpful. The first possibility is to provide more information to the players. For instance, does better monitoring help in this scenario, in contrast to good-news models? If an intermediary could provide private recommendations to the players, what would he do? These questions are examined in the first subsection. In the second, we consider the scope for transfers (specifically, subsidies and cross-insurance) in improving the outcome.

## 6.1 The Role of Information

### 6.1.1 Perfect Monitoring

We begin by recalling Keller and Rady's result regarding symmetric Markov equilibria in the game with observable actions. Players are restricted to Markov policies $u^i : [0,1] \to [0,1]$ with the left limit $p_{t-}$ of the common posterior belief as the state variable. Policies are required to be left-continuous and piecewise Lipschitz. We define $\bar{p}$ as (the unique solution of)

$$\frac{\bar{p}}{1-\bar{p}} := \frac{p^*}{1-p^*} \exp\left(-\frac{1+\mu}{\mu} - W_{-1}\left(-\gamma \frac{p^*}{1-p^*} e^{-\frac{1+\mu}{\mu}}\right)\right),$$

where $W_{-1}$ is the (negative branch of the) Lambert function. We also define $u^o : [0,1] \to [0,1]$ as

$$u^o(p) = \begin{cases} 1 & \text{if } p \geq \bar{p}, \\ \frac{I+\mu-1}{I-1} - \frac{\mu(\ln(p/(1-p))-\ln(p^*/(1-p^*)))+1}{(I-1)(\gamma p/(1-p)-1)} & \text{if } p \in [p^*, \bar{p}), \\ 0 & \text{if } p < p^*. \end{cases}$$

**Theorem 3 (Keller and Rady, 2015)** *The unique symmetric Markov equilibrium is given by $u^o$.*

It is worth emphasizing that this is not the unique Markov equilibrium: asymmetric Markov equilibria exist, and the ranking in terms of welfare can go either way. (See Section 3.3 of Keller and Rady, 2015.) Theorem 3 is established by Keller and Rady (2015), although they do not provide the closed-form expression for the policy, and the reader is referred to their paper for a proof of this result.

Under observable actions, players benefit from a larger group, although not at the same rate as the social planner. Furthermore, as the number of players (and hence the value of information) grows, the first-best policy eventually involves immediate full experimentation. Each player's cost then converges to its level under complete information: the arrival rate of a breakdown grows, conditional on the bad state, and the probability of suffering a breakdown is inversely proportional to $I$. This relationship cannot be found when actions are not observable because experimentation does not even begin unless $p^0 < p^{**}$. Even under observable actions, the threshold belief $\bar{p}$ for experimentation to begin is increasing in $I$ but converges to a finite value. Furthermore, the duration of the mixing phase does not decline as the number of players grows. Therefore, the cost under complete information is not attainable if $p^0 > p^*$.

Next, we compare the total amount of experimentation up to some $t$ under both observable and unobservable actions. We write $p_t^o$, $p_t^n$ and $p_t^{FB}$ for these beliefs, depending on whether we consider the observable, unobservable or cooperative case, respectively. We can show stronger results than the ranking of the belief paths. For the unobservable case, let $\nu(p) := \nu_{t(p)}$, where $t(p)$ denotes the time at which the outside observer's belief reaches a value of $p$. Formally, we can show that $\nu(p)$ is ranked across the three cases.

**Proposition 2** *The following inequalities hold for all $p$*

$$\nu^n(p) \geq \nu^o(p) \geq \nu^{FB}(p).$$

*The second inequality is strict when $p < p^{FB}$, and the first is strict when $p^o < \bar{p}$. In particular, $\bar{p} > p^{**}$.*

Therefore, for all $t$,

$$p_t^n \geq p_t^o \geq p_t^{FB},$$

with strict inequalities as described in the previous proposition. Furthermore, Lemma 4 implies the ranking of the symmetric equilibrium costs $\mathcal{C}(p^0)$.

**Corollary 1** *The following inequalities hold for all $p = p^0$*

$$\mathcal{C}^n(p) \geq \mathcal{C}^o(p) \geq \mathcal{C}^{FB}(p).$$

*Both inequalities are strict when $p > p^*$ and $\mu > 0$.*

Hence, monitoring is helpful in our context, although it is not helpful with good news.[22] The basic intuition is easy to grasp: when actions are observable, a player's incentive to deviate is related not only to the direct cost or benefit from this deviation but also to the indirect cost or benefit in terms of the change in actions by other players. By deviating to the risky arm, a player accelerates the common learning that, in the absence of news, leads to greater optimism and more experimentation by others; this outcome is good because players do not experiment enough. By contrast, with good news, experimentation by a player leads to greater pessimism in the absence of news and hence depresses experimentation provision.

### 6.1.2 Coordination

Forcing players to disclose their actions might not be easy to achieve in practice. A less drastic intervention might involve introducing a disinterested intermediary who makes private but correlated recommendations to each player regarding when they should begin experimenting. Mediation is a particularly weak form of intervention; it is self-enforcing and costless. Its formal implementation requires no more than a private correlation device, but in practice, this mediation is undertaken by trade associations, political representatives, or any institution commonly involved in the social dialogue.

Clearly, the optimal correlation scheme should be in the extensive form: there is no benefit in telling a player when to switch before the intermediary intends for him to do so, as telling him the specific stopping time in advance only makes it more difficult to induce compliance with the recommendation, giving him more information than needed. However,

[22] See Bonatti and Hörner (2011) and related games with incomplete information. Holmström (1999) is perhaps the most famous example, although arguably the mechanism through which the lack of observability operates is very different.

as we explain, *even* in the normal form (by telling each player privately at the beginning of the game when he should stop playing safe), such correlation is helpful.[23]

Specifically, for the case of two players, consider the joint distribution over switching times in our symmetric equilibrium, $F(t_1, t_2) = F(t_1)F(t_2)$. We construct a new distribution by slightly perturbing the independent randomization according to a bivariate FGM copula.[24] Let $\rho$ denote the correlation parameter of the joint distribution $F$.

We modify our equilibrium marginal distribution to introduce a small amount of correlation and preserve incentives. At an abstract level, the incentive-compatibility constraint for obeying the recommendation to switch at time $t$ is a functional equation that is linear in the distribution $F$. We can then write this constraint as the combination of two *linear* operators $K_0$ and $K_1$. In particular, we have

$$K_0(F) + \rho K_1(F) = 0.$$

We can use this constraint to capture the restriction that incentives (under a small amount of correlation) impose on the marginal distribution. In particular, we identify a distribution that we use to (locally) modify our equilibrium distribution while preserving incentives. We denote this distribution by $F_1(t; \rho)$.

Clearly, regardless of the degree of correlation $\rho$, no player can begin experimenting before $p^{**}$ or after $p^*$. The design variable is the degree of correlation but requires adjusting the support of the marginal distribution to match $p^*$ of the most pessimistic type. In particular, the mass point at time $\bar{\tau}$ is now a function of $\rho$. We then differentiate total costs under the distribution $F_1(t; \rho)$ in a neighborhood of $\rho = 0$.

For any value of the parameters, the derivative of the cost is negative, *i.e.*, positive correlation is beneficial. We conclude that some (possibly small) amount of positive correlation of switching times (subject to incentive compatibility) improves upon independence.[25]

However, the role of positive correlation (across switching times) must not be confused with a simple reduction in the dispersion of practices. Recall that only switching policies are optimal for any player. It is then important and immediately clear that the symmetric ("coordinated") pure-policy profile $\{\nu_t^i\}_{i=1}^I$ yields strictly higher costs than our mixed equilibrium.

[23] We are unable to solve for the optimal correlation scheme in the extensive form. In fact, even in the normal form, we are able to solve for it only in the special case of a particular parametrized family of correlation schemes, as described below. But this case suffices to show that independence is not optimal.

[24] For a marginal distribution $F(t)$, the Farlie-Gumbel-Morgenstern (FGM) copula is given by $F(t_1, t_2) = F(t_1)F(t_2)(1+\rho(1-F(t_1))(1-F(t_2)))$, with parameter $\rho \in [-1, 1]$. See Nelsen (2006). Throughout this case, we assume symmetry of this distribution, and we introduce an (arbitrarily small amount of) background learning, *i.e.*, $\dot{p}_t^i = -p_t^i(1 - p_t^i)(\bar{u} - u_t^i - \nu_t^j)$, with $\bar{u} > 2$.

[25] The details of the calculations leading to this comparative statics result are in the annotated Mathematica file `correlated.nb` on the authors' websites.

Again, from each player's perspective, it is irrelevant whether others randomize (holding $\nu^{-i}$ fixed). However, the best-reply problem admits only switching-policy solutions–it is costlier for a player to use the pure (non-extremal) policy $\nu_t^i$ than to use the distribution $F^i$ over stopping times that is equivalent to $\nu_t^i$, from the perspective of the other players.

### 6.2 The Role of Payoffs

The potentially beneficial role of government policy for technology adoption has been the subject of a large body of theoretical and empirical literature. We examine the role of policy interventions and their interaction with the monitoring structure.

We introduce payoff externalities in our model through a cross-subsidization or insurance scheme: when a breakdown occurs, we assume that the player who suffers this breakdown receives in turn a fraction $\alpha \leq 1$ as compensation, evenly shared by the other players (throughout this case, we assume that the identity of the player who suffers a breakdown is observable). Our baseline model corresponds to the special case $\alpha = 0$.

Formally, the total realized cost of player $i$ is now, given the realization of the process $\{N_t^i : t \geq 0\}_{i=1,\dots,I}$,

$$\int_0^\infty re^{-rt}\left(h\left(\frac{\alpha}{I-1}\textstyle\sum_{j\neq i}\mathrm{d}N_t^j + (1-\alpha)\mathrm{d}N_t^i\right) + su_t^i\mathrm{d}t\right).$$

To avoid equilibrium multiplicity at least under complete information,[26] we assume throughout that

$$\alpha < \hat{\alpha} := \gamma\frac{I+\mu-1}{I+\mu}.$$

Under cross-subsidies $\alpha$, the two bounds on experimentation we introduced earlier can be written as

$$\frac{p_\alpha^{**}}{1-p_\alpha^{**}} := \frac{1+\mu}{\mu}\frac{1}{\gamma-\alpha(1+\gamma)},$$

and

$$\frac{p_\alpha^{*}}{1-p_\alpha^{*}} := \frac{I+\mu}{(I+\mu)\left((1+\gamma)(1-\alpha)-1\right)-\gamma}.$$

[26]With bad news, when payoff externalities are strong, multiple equilibria exist when players are sufficiently patient, even when neither the actions nor the identity of the player suffering the breakdown is publicly observed. The absence of payoff externalities is key to the argument that behaviour after the first breakdown is trivial, as playing safe is then strictly dominant. For instance, if a breakdown entails the same cost to all players independent of who suffered it and if the state of the world is known to be bad, then one can follow Abreu, Milgrom and Pearce (1991) in constructing multiple Pareto-ranked equilibria.

It is easy to observe that $p^{**}_\alpha$ and $p^*_\alpha$ are strictly increasing and continuous in $\alpha$ (when $\alpha < \hat{\alpha}$ for $p^*_\alpha$ and when $\alpha < \tilde{\alpha}$, defined below, for $p^{**}_\alpha$). Furthermore, note that

$$p^{**}_\alpha > p^*_\alpha \iff \alpha < \alpha^* := \frac{\gamma}{1+\gamma}\frac{I-1}{I+\mu}. \tag{10}$$

We obtain the following theorem that generalizes Theorem 1.

**Theorem 4** *Consider the game with unobservable actions. There exists a unique symmetric equilibrium.*

1. *If* (10) *holds and $p^0 \in (p^*_\alpha, p^{**}_\alpha)$, then the equilibrium involves mixed policies. Specifically, player $i$ chooses a stopping policy $\pi_t$ among the set $t \in [0, \bar{\tau}_\alpha]$, with $\bar{\tau}_\alpha > 0$ and $p_{\bar{\tau}_\alpha} = p^*_\alpha$; this distribution is positive and continuous over $(0, \bar{\tau}_\alpha)$ and has an atom at times $t = 0, \bar{\tau}_\alpha$. If $p^0 \geq p^{**}_\alpha$, then equilibrium is pure and involves $F^i(t) = 0$ at all times.*

2. *If* (10) *does not hold, then the equilibrium involves pure policies. Specifically, player $i$ chooses $u^i_t = 1$ for all $t$ such that $p_t \geq p^*_\alpha$ and $u^i_t = 0$ otherwise.*

Thus, even for moderate levels of risk sharing, the equilibrium is in pure (cutoff) policies. To gain some insight into pure- vs. mixed-policy equilibria, consider the two effects of risk sharing: on the one hand, players are partially insured against their own breakdowns; on the other hand, they may need to compensate other players for theirs. When $\alpha$ is sufficiently large, it is no longer profitable to delay experimentation (free-ride) beyond $p^{**} = p^*$: if others experiment, the risk of needing to pay compensation is too high. Furthermore, it is not profitable to preempt others because risk sharing shifts the threshold beliefs, which makes experimentation less attractive.

A significant consequence of Theorem 4 is the following.

**Proposition 3** *The threshold $\alpha^*$ in* (10) *satisfies $p^*_{\alpha^*} = p^{FB}$. Furthermore, $\alpha^* < \hat{\alpha}$.*

In other words, *there exists a cross-subsidy that makes players use the first-best cutoff as their unique equilibrium policy.* In contrast to the analysis of moral hazard in teams by Holmström (1982), the players in our model can identify individual contributions and can condition payments on the identity of the agent suffering a breakdown.

Greater experimentation by others is always desirable in terms of the informational externality, as it results in more learning. However, more experimentation by others is *not* desirable in terms of payoff (cost) externality: a player who shares the cost of other players' breakdowns prefers less experimentation by others in terms of direct costs. The combination of positive informational externalities and negative payoff externalities is ambiguous *a* *priori*. The optimal level of the subsidy precisely offsets the payoff and informational externalities, yielding the socially efficient outcome. In particular, the optimal subsidy causes the two bounds on experimentation to coincide. Recall that the two bounds correspond to the single-agent thresholds under the two constant policies $\nu_t^{-i} \in \{0, I-1\}$. When the two thresholds coincide, player $i$ is willing to experiment from $p^*$ onward independent of the actions of others.

The first-best cost-sharing level $\alpha^*$ is increasing in the cost of a breakdown $\gamma$ and in the number of players $I$. As we saw in Section 5.3, as either parameter increases, the cost-sharing rule must cancel the effect of a greater informational externality. A more striking result is that $\alpha^*$ is independent of the monitoring structure. In other words, for the right subsidy, whether actions are observed is irrelevant, and the unique symmetric (in the observable case, Markov) equilibrium achieves the first-best outcome.

When actions are observable, whether other players experiment too little or too much for a player's taste, he can nudge their action towards his preferred action by deviating accordingly. A policy of playing risky leads to accelerated learning, prompting others to experiment, whereas playing safe leads to slower learning and delayed experimentation by others. As a result, one would expect the equilibrium payoff and the amount of experimentation to be closer to their first-best levels when actions are observable, regardless of whether there is over-experimentation. In particular, when we obtain first-best outcomes in the unobservable case, we should also obtain the first-best outcomes with observable actions.

To substantiate this claim, we now turn to the observable case. We define

$$\frac{\bar{p}_\alpha}{1-\bar{p}_\alpha} := \frac{p_\alpha^*}{1-p_\alpha^*} \exp\left(-\frac{1+\mu}{\mu} - W_{-1}\left(-\gamma \frac{p_\alpha^*}{1-p_\alpha^*} e^{-\frac{1+\mu}{\mu}}\right)\right).$$

This threshold generalizes the threshold $\bar{p}$ introduced in Section 6.1 (that is, $\bar{p}_0 = \bar{p}$). It is easy to show that this threshold is strictly increasing in $\alpha$.

We have the following generalization of Theorem 3 (we omit the specification of the precise amount assigned to the risky arm in the interior region).

**Theorem 5** *Consider the game with observable actions.*

1. *For $\alpha \in [0, \alpha^*]$, a unique symmetric Markov equilibrium exists. The safe arm is pulled for $p \geq \bar{p}_\alpha$, the risky arm is pulled for $p \leq p_\alpha^*$, and the amount assigned to the safe arm is continuous and strictly increasing in the range of $p \in [p_\alpha^*, \bar{p}_\alpha]$.*

2. *For $\alpha \in (\alpha^*, \hat{\alpha})$, multiple symmetric Markov equilibria exist, indexed by an (arbitrary) value $p_* \in [p^{FB}, p_\alpha^*]$, such that the safe arm is pulled for $p \geq p_*$ and the risky arm for $p < p_*$.*

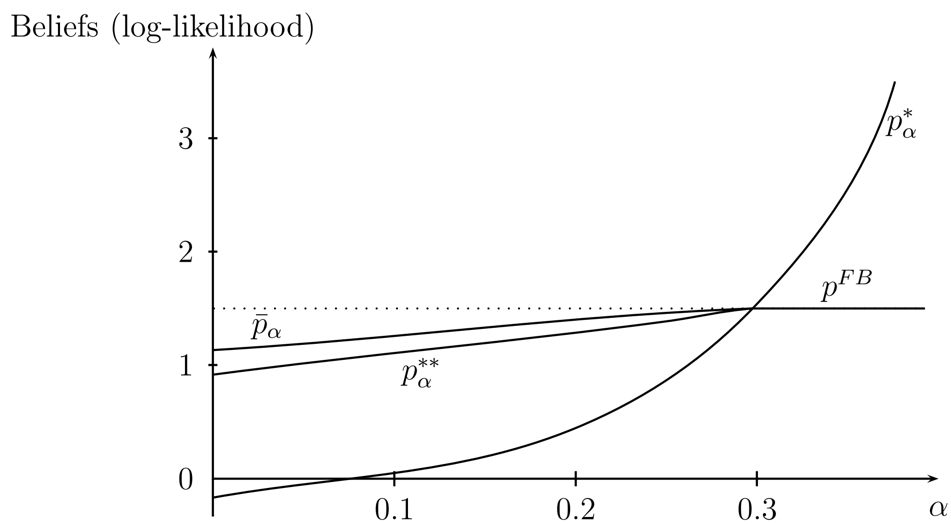



Figure 6: Pure- and mixed-policy equilibria with subsidies for $(I, \mu, \gamma) = (2, 1/4, 3)$.

The multiplicity for $\alpha > \alpha^*$ reflects complementarities among the players' actions: if other players switch to the risky action, then there are strict benefits of also doing so. The choice $p_* = p^B_\alpha$ is that for which smooth-pasting holds at the boundary. Figure 6 illustrates these boundaries.

The analysis substantiates our discussion: for $\alpha > \alpha^*$, over-experimentation is observed irrespective of whether actions are observed, but the extent of this over-experimentation is worsened by the lack of observability. Similarly, for $\alpha < \alpha^*$, under-experimentation is the result, and a lack of observability worsens this outcome. Thus, although better monitoring is beneficial independent of the magnitude of payoff externalities, our results suggest that with the correct intervention in the payoff environment, one can dispense with improvements in the monitoring technology.

Finally, our paper suggests a plausible explanation for the failure of third-party subsidies to foster the diffusion of agricultural innovations in developing countries (see Carter, Laajaj and Yang, 2013, for a discussion of low and heterogeneous adoption rates). In particular, we highlight an important difference between externally funded subsidies and budget-balanced (formal or relational) risk-sharing agreements. To understand this difference, contrast the role of risk sharing with a subsidy for experimentation that reduces $\gamma$ (for example, through an increase in the opportunity cost of playing safe or through an external insurance policy that reduces $g$).

This effect is best observed when we introduce *background learning*,[27] but also applies

[27]This means that beliefs never "freeze," as $\dot{p}^i_t = -p^i_t(1 - p^i_t)(\bar{u} - u^i_t - \nu^{-i}_t) < 0$, given $\bar{u} > I$, see ft. 12.

qualitatively in the baseline model. An important property of background learning is that the magnitude of the lump-sum losses affects only the timing of experimentation: as the relative cost of a breakdown $\gamma$ increases, the thresholds $(\underline{\tau}, \bar{\tau})$, with $\underline{\tau} > 0$, shift forward. In other words, a higher $\gamma$ delays the beginning of experimentation and the beginning of full experimentation by an equal amount without affecting the equilibrium distribution of stopping times. Consequently, provided that $p^0 > p^{FB}$, there is no subsidy that yields the first-best cutoff as an equilibrium policy.

Figure 7 shows the effect of doubling the risk-sharing rate $\alpha$ or halving the risk factor $\gamma$.

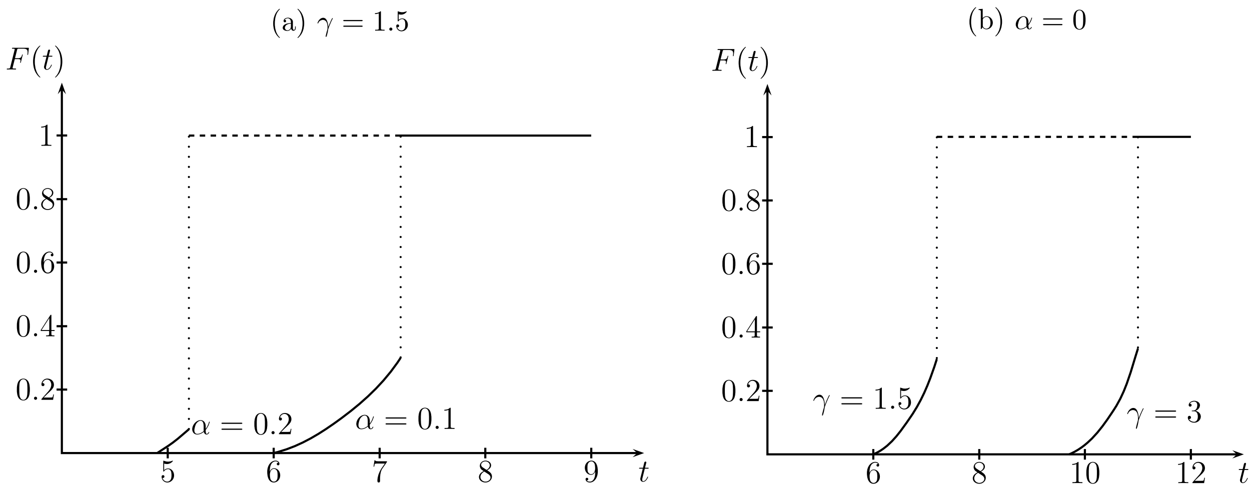



Figure 7: Risk sharing vs. subsidies for $(\bar{u}, I, \mu, p^0) = (2.2, 2, 0.25, 0.88)$.

# 7 Conclusions

Our results require a number of assumptions. Here, we briefly discuss how we expect them to extend in two important dimensions.

**Inconclusive bad news.** A complete analysis under a scenario of inconclusive bad news (that is, when a breakdown does not reveal the state) seems out of reach. However, we believe that the belief-disagreement result would become more pronounced. First, if all agents stop experimenting upon observing a breakdown, then learning stops and beliefs freeze at different levels depending on the agents' prior actions. Such endogenous belief heterogeneity has an effect on policy effectiveness, *e.g.*, if an external agent (the government) were to attempt to subsidize the risky arm to resume experimentation. Second, behaviour after the first breakdown can potentially differ across players. In particular, some players may revert to the risky arm for some time, whereas others may not. Conditional on the true state, this exacerbates performance differences, as the agents who were experimenting earlier are those who continue to do so.

**Monitoring and payoff externalities in games of strategic experimentation.** Despite the stylized nature of the learning processes involved, several economically relevant implications may be drawn by contrasting the good-news and bad-news models of strategic experimentation.

We have already remarked on the different welfare implications of observable actions in the two models. Observability improves welfare under bad news, whereas it is detrimental under good news. This contrast is easy to understand. Pulling the safe arm slows learning in both models. However, slowed learning results in more experimentation by others with good news, whereas it results in less experimentation with bad news.

However, unobservable actions have similar strategic effects under both signal structures. In the absence of payoff externalities, unobservable actions imply that every best reply involves frontloading risky (safe) actions in models with good (bad) news.[28] In other words, the pure interior action paths described by Keller, Rady and Cripps (2005), Keller and Rady (2015), and Bonatti and Hörner (2011) rely on either observable actions (the former two) or payoff externalities (the latter).

This result has differing implications for outcomes and provides a clear, if stylized, criterion to guide policy interventions depending on the nature of the technology. In particular, in the good-news case, unobservable actions eliminate inefficient delay but preserve the suboptimal *amount* of experimentation. Thus, subsidies can be used to augment the amount of experimentation. Under bad-news learning, we highlighted three sources of inefficiency: players experiment too little, with excessive dispersion in both practices and beliefs. Furthermore, in the presence of background learning (see Figure 7), subsidies are able to address the first source only, and group insurance may be more appropriate.

[28] For the good-news case, this result can be obtained by modifying Theorem 1 in Bonatti and Hörner (2011) to account for pure informational externalities. Instead, the welfare comparison above does not rely on the payoff externality.

# Appendix

## A Reformulation of the Objective

Here we reformulate each player's objective, and we keep track of additional cost terms that will be necessary for comparative statics. Each player minimizes

$$\int_{t\geq 0} e^{-rt}\left(rp_t g\left(1-u_t^i\right)+ru_t^i s+\lambda p_t(I-u_t^i-\nu_t^{-i})s\right)\frac{1-p^0}{1-p_t}\mathrm{d}t, \tag{11}$$

subject to

$$\dot{p}=-\lambda p_t(1-p_t)(I-u_t^i-\nu_t^{-i}), \quad p_0=p^0.$$

Let us do the transformations one by one, first rewriting the objective in terms of the log-likelihood ratio $\ell_t := \ln(p_t/(1-p_t))$. The objective becomes

$$\int_{t\geq 0} e^{-rt}\left(re^{\ell_t}g(1-u_t^i)+ru_t^i s(1+e^{\ell_t})+\lambda e^{\ell_t}(I-u_t^i-\nu_t^{-i})s\right)(1+e^{\ell^0})^{-1}\mathrm{d}t.$$

Next, we make the change of variable $t\mapsto t/\lambda$, and we define $\gamma := (g-s)/s$ and $\mu := r/\lambda$. Finally, we factor out $(1+e^{\ell^0})^{-1}$ to get

$$\int_{t\geq 0} e^{-\mu t}\left(\mu e^{\ell_t}g+\mu(s(1+e^{\ell_t})-ge^{\ell_t})(\dot{\ell}_t+I-\nu_t^{-i})-\dot{\ell}_t e^{\ell_t}s\right)\mathrm{d}t.$$

Integrating the last term yields

$$e^{\ell^0}s+\int_{t\geq 0} e^{-\mu t}\left(e^{\ell_t}(\mu g+\mu(s-g)(\dot{\ell}_t+I-\nu_t^{-i}))+\mu s(\dot{\ell}_t+I-\nu_t^{-i})-\mu s e^{\ell_t}\right)\mathrm{d}t.$$

Integrating the first two terms by parts, and factoring out $s$, we obtain the following expression for the expected cost:

$$W(\ell^0) := \frac{s(1+\mu\gamma)}{1+e^{-\ell^0}}+\frac{\mu s}{1+e^{\ell^0}}\int_{t\geq 0} e^{-\mu t}\left(\mu(\ell_t-\ell^0)-\gamma(I-\nu_t^{-i}-1+\mu)e^{\ell_t}+I-\nu_t^{-i}\right)\mathrm{d}t. \tag{12}$$

Therefore, ignoring constant terms, player $i$ minimizes

$$\int_{t\geq 0} e^{-\mu t}\left(\mu\ell_t-\gamma(I-\nu_t^{-i}-1+\mu)e^{\ell_t}\right)\mathrm{d}t,$$

subject to

$$\dot{\ell}_t = u_t^i+\nu_t^{-i}-I.$$

# B Proofs for Section 4

**Proof of Lemma 1.** The proof of this lemma relies on the proof of Lemma 2, proved next and independently (except for the last sentence of that next proof, which is not used here).

We apply the maximum principle to $\mathcal{P}$. It is easy to see that the program $\mathcal{P}$ is not abnormal (see Seierstad and Sydsæter 1987, Ch.2.4, Note 5).[29] The maximum principle implies that there exists an absolutely continuous $\psi : \mathbf{R}_+ \to \mathbf{R}$ such that (i) $\psi_t > 0 \Rightarrow u_t^i = 0$, (ii) $\psi_t < 0 \Rightarrow u_t^i = 1$, and (iii) almost everywhere

$$\dot{\psi}_t e^{\mu t} = \gamma(I - \nu_t^{-i} - 1 + \mu)e^{\ell_t} - \mu.$$

Because $\nu_t^{-i} \leq I - 1$, a sufficient condition for $\dot{\psi}_t > 0$ at any time $t$ such that $\ell_t \geq \ell^*$ is that

$$\gamma\mu e^{\ell^*} > \mu.$$

Using the definition of $\ell^*$, this is equivalent to

$$(\mu + I)\mu \geq \mu(\mu + I - 1),$$

which is true.

It follows that $\psi$ is strictly increasing at all times $t$ such that $\ell_t \in [\ell^*, \ell^0]$; hence, given (i), there exists $\bar{t} \geq 0$ such that any solution must specify $u_t^i = 1$ for all $t < \bar{t}$ and $u_t^i = 0$ for $t \geq \bar{t}$ (recall that $u_t^i = 0$ when $\ell_t < \ell^*$).

**Proof of Lemma 2.** Consider the continuation cost corresponding to the objective (12), defined as

$$\mathcal{C}(\ell, t) := \int_{s \geq t} e^{-\mu s} \left(\mu(\ell + \chi_s) + \gamma(\nu_s^{-i} - I - \mu + 1)e^{\ell + \chi_s}\right) \mathrm{d}s,$$

where we define $\chi_s := \int_{\tau=t}^{s} (\nu_\tau^{-i} - I)\mathrm{d}\tau$ as the value from setting $u^i = 0$ (identically), given $\ell$ and $t$. Note that, integrating by parts,

$$\mathcal{C}(\ell, t) = e^{-\mu t}(\ell - \gamma e^{\ell}) + \int_{s \geq t} e^{-\mu s} \left(\mu\chi_s + \gamma e^{\ell + \chi_s}\right) \mathrm{d}s,$$

which is differentiable with respect to $\ell$, with

$$\frac{\partial \mathcal{C}(\ell, t)}{\partial \ell} = e^{-\mu t} \left(1 - \gamma e^{\ell} + \gamma e^{\ell} \int_{s \geq t} e^{\chi_s - \mu(s-t)} \mathrm{d}s\right),$$

[29] The argument given Seierstad and Sydsæter (1987) must be slightly modified, as it applies to a fixed horizon. The adjustment is straightforward.

which is minimized by setting $\nu_\tau^{-i} = 0$ for all $\tau \geq t$. In that case, the right-hand side is equal to

$$e^{-\mu t}\left(1 - \gamma e^{\ell} - \frac{\gamma e^{\ell}}{I+\mu}\right),$$

which is positive if *and only if* $\ell < \ell^*$. Hence, independently of $\nu^{-i}$, $\mathcal{C}(\ell, t)$ is strictly increasing in $\ell$ whenever $\ell < \ell^*$. It follows that, for $\ell < \ell^*$, $\mathcal{C}$ solves the Hamilton-Jacobi-Bellman ("HJB") equation

$$\frac{\partial \mathcal{C}(\ell,t)}{\partial t} + \min_{u^i}\left\{\frac{\partial \mathcal{C}(\ell,t)}{\partial \ell}(u_t^i + \nu_t^{-i} - I) + e^{-\mu t}\left(\mu \ell_t - \gamma(I - \nu_t^{-i} - 1 + \mu)e^{\ell_t}\right)\right\} = 0,$$

so that setting $u_t^i = 0$ is optimal. Because of the "if and only if" above, if $\nu_s^{-i} = 0$ for all $s \geq t$ (for which it suffices that $\nu_t^{-i} = 0$), yet $\ell_t = \ell > \ell^*$, it cannot be that $u_s^i = 0$ for all $s \geq t$ (and so it must be that $u_t^i > 0$).

**Proof of Lemma 3.** Ignoring some irrelevant constants, the cost can be rewritten as (abusing notation for $\mathcal{C}$)

$$\mathcal{C}_t^i := \frac{e^{-\mu t}}{\mu}\left(\mu\gamma e^{\ell_t^i}\int_t^\infty e^{\int_t^s (\nu_\tau^{-i} - (\mu+I))\mathrm{d}\tau}\mathrm{d}s - 1\right). \tag{13}$$

If $\nu_t^{-i} = I - 1$ and $t \in \operatorname{supp} F^i$, yet $\ell_t^i < \ell^{**}$, then because the time derivative $\mathcal{C}_t^{i'} = 0$,

$$\mathcal{C}_t^{i''} e^{\mu t} = \gamma\mu e^{\ell_t^i} - (\mu + 1) < 0,$$

by definition of $\ell^{**}$. It follows that for small enough $\varepsilon > 0$, $\mathcal{C}_{t-\varepsilon}^i < \mathcal{C}_t^i$, a contradiction.

**Proof of Lemma 4.** See Keller and Rady (2015, Proposition 1) for the cooperative solution $u^{FB}$. Note that if $\ell > \ell^{FB}$, $u^{FB}(\ell) \leq u'(\ell) \leq u''(\ell)$ implies that $u^{FB}(\ell) = u'(\ell) = u''(\ell) = I$ and costs are the same under all three policies. Hence, without loss, we assume $\ell^0 < \ell^{FB}$. Given some measurable $\underline{U}, \bar{U} : (-\infty, \ell^0] \to [0, I)$, with $0 \leq \underline{U}(\ell) < \bar{U}(\ell)$ and $\bar{U}$ bounded away from $I$, consider the program $\mathcal{P}^{FB}(\underline{u})$:

$$\min \int_t e^{-\mu t}\left(\mu \ell_t - \gamma(I - 1 + \mu)e^{\ell_t}\right)\mathrm{d}t$$

over all $\pi : \mathbf{R}_+ \to [0, I]$, measurable, subject to

$$\dot{\ell}_t = u_t - I, \quad \ell_0 = \ell^0,$$

with, for all $t \geq 0$ and $\ell_t \leq \ell^0$, $u_t \in [\underline{U}(\ell_t), \bar{U}(\ell_t)]$. By standard arguments, the optimal $u$ is measurable with respect to the belief $\ell$, and is the solution to the program

$$\min \int_\ell e^{-\mu t(\ell)}\left(\mu \ell - \gamma(I - 1 + \mu)e^{\ell}\right)\mathrm{d}\ell,$$

over all measurable $u : (-\infty, \ell^0] \to [0, I]$ such that $u(\ell) \in [\underline{U}(\ell), \bar{U}(\ell)]$, where $t(\ell)$ solves $t(\ell^0) = 0$ and

$$t'(\ell) = (u(\ell) - I)^{-1},$$

which is well defined because $u(\ell) < \bar{U}(\ell) < I$. A routine application of the maximum principle (Theorem 4.2, Cesari, 1983) yields that the optimal policy solves $u(\ell) = \underline{U}(\ell)$ a.e. Given $u', u''$ as stated in the lemma, the result follows if $u'' < I$ by setting $\underline{U} = u'$, $\bar{U} = u''$ and noting that $u''$ does not satisfy the necessary conditions. The same argument applies with trivial modifications if $\bar{U} = I$.

**Proof of Lemma 5.** Suppose that players $j \neq i$ stop at some fixed time $T \in \mathbf{R}_+$. For clarity, we use $I$ rather 2 for the number of players, as the arguments do not depend on it (though the statement of Lemma 5 is specialized to that case). Throughout, we assume that $\ell^0 \in [\ell^*, \ell^{**}]$, as the result is trivial otherwise. Then, inserting into the objective of player $i$, he chooses $\tau$ to minimize

$$\int_{t \leq T} e^{-rt} \left( \mu(\ell^0 + \lambda(t \wedge \tau - t)) - \gamma \mu e^{\ell^0 + l(t \wedge \tau - t)} \right) \mathrm{d}t$$
$$+ \int_{t \geq T} e^{-rt} \left( \mu(\ell^0 + \lambda(t \wedge \tau + (I-1)T - It)) - \gamma(I - 1 + \mu) e^{\ell^0 + \lambda(t \wedge \tau + (I-1)T - It)} \right) \mathrm{d}t.$$

This gives two expressions for the cost depending on $\tau \gtrless T$. Let us write $\mathcal{C}^1$ for the cost when $\tau \leq T$, and $\mathcal{C}^2$ for $\tau \geq T$ (the costs coincide when $\tau = T$). It is useful to use $x = \lambda\tau$ and $y = \lambda T$, instead of $(\tau, T)$. We explicitly compute both costs, which gives

$$\mathcal{C}^1(x) = -\frac{\gamma(I-1)e^{\ell^0 + x - (\mu+1)y}}{(\mu+1)(I+\mu)} - \frac{(I-1)e^{-\mu y}}{\mu} - \gamma e^{\ell^0} + \frac{\gamma e^{\ell^0 - \mu x}}{\mu + 1} + l - \frac{e^{-\mu x}}{\mu},$$

and

$$\mathcal{C}^2(x) = \frac{e^{-(I+\mu)x - (\mu+1)y} \left( \gamma \mu e^{\ell^0 + \mu y} \left( e^{Iy + x} - (I+\mu) e^{x(I+\mu) + y} \right) - (I+\mu) e^{Ix + y} \left( (I-1) e^{\mu x} - \mu \ell^0 e^{\mu(x+y)} + e^{\mu y} \right) \right)}{\mu(I+\mu)}.$$

It is readily checked that $\mathcal{C}^1$ is concave, and so minimized either at $x = 0$ or $x = y$, while $\mathcal{C}^2$ is convex, and minimized at

$$x^* := y + \frac{\ell^0 - \ell^*}{I - 1}.$$

Hence, we have only two candidates as global minimizer of the total cost, namely 0 and $x^*$. Note that (as shown in Figure 2) the candidate minimizer $x^*$ (resp., $\tau$) is affine in $y$ (resp.,

$T$). We compute the difference $\Delta := \mathcal{C}^2(x^*) - \mathcal{C}^1(0)$. Computing,

$$\Delta(y) := \gamma e^{\ell^0}\left(\frac{(I-1)e^{-(\mu+1)y}}{(\mu+1)(I+\mu)} - 1\right) + \frac{1 - \frac{(I-1)\left(\frac{\gamma(I+\mu-1)}{I+\mu}\right)^{\frac{\mu}{1-I}} e^{\mu\left(-\frac{\ell^0}{I-1}-y\right)}}{I+\mu-1}}{\mu} + \frac{\gamma\mu e^{\ell^0}}{\mu+1}.$$

We claim that $\Delta(y) < 0$ if and only if $y \leq \hat{y}$, for some $\hat{y} \geq 0$, and this will establish the result. First,

$$\lim_{y\to\infty} \Delta(y) = \frac{1}{\mu} - \gamma\frac{e^{\ell^0}}{1+\mu} > 0,$$

as $\ell^0 < \ell^{**}$. Second, $\Delta(0)$, viewed as a function of $e^{\ell^0}$, is concave, zero at $\ell^*$, with zero derivative at $\ell^*$. Hence, $\Delta(0) \leq 0$ for all $\ell^0 \in [\ell^*, \ell^{**}]$ (the inequality being strict for $\ell^0 > \ell^0$). Finally, with the change of variable $Y = e^{-(1+\mu)y}$, we get that $\Delta$ is convex in $Y$, and hence admits at most one root $Y$, hence $y$.

**Proof of Lemma 6.** If any player $j$ uses a pure policy in equilibrium, it must be $t^j = \hat{t}$ so that, by the best-reply analysis in Lemma 5, player $i$ is indifferent between $t^i = 0$ and $t^i = \hat{t} + T$, where $T := (\ell^0 - \ell^*)$.

Lemma 5 further establishes that the best reply to $t^i = \hat{t} + T$ is $t^j = 0$ and that the best reply to $t^i = 0$ is $t^j = T$. We shall show that $T < \hat{t}$, so that $t^j = \hat{t}$ cannot be a best reply to any randomization over $t^i \in \{0, \hat{t} + T\}$.

It suffices to establish that the best reply to $t = T$ is, in fact, $\tau = 2T$. To do so, consider player $i$'s marginal cost $\partial\mathcal{C}^i/\partial t^i$ evaluated $t^i = 0$ when player $j$ uses $t^j = T$. This is proportional to

$$\left(\frac{\mu+2}{\gamma\mu+\gamma}\right)^{\mu}\left(-e^{-\mu\ell^0}\right) + \mu\left(\mu - \gamma(\mu+1)e^{\ell^0} + 2\right) + 1. \tag{14}$$

We want to show this expression is negative, so that switching to the risky arm later than $t = 0$ yields strict cost savings (hence that the best reply must be $2T$). Consider the derivative of the marginal cost with respect to $\ell^0$. This is given by

$$\mu\left(\frac{\mu+2}{\gamma\mu+\gamma}\right)^{\mu} - \gamma\mu(\mu+1)e^{\ell^0(1+\mu)}.$$

This expression is strictly decreasing in $\ell^0$ and negative (it is equal to $-\mu(1+\mu)$) when evaluated at $\ell^0 = \ell^*$. Therefore $\partial^2\mathcal{C}^i/\partial t^i\partial\ell^0 < 0$ for all $\ell^0$. To sign the marginal cost $\partial\mathcal{C}^i/\partial t^i$ evaluated at $t^i = 0$, it is sufficient to show that it is non-positive when $\ell^0 = \ell^*$. This is indeed the case, as the expression in (14) can be easily verified to be nil for $\ell^0 = \ell^*$.

# C Proofs for Section 5

**Proof of Theorem 1.** We first argue that in every symmetric equilibrium the support of the distribution is an interval: for all $i$, $\operatorname{supp} F^i = [\underline{\tau}, \bar{\tau}]$, for some $\underline{\tau} \leq \bar{\tau}$, with $\ell_{\bar{\tau}} = \ell^*$.

Using the same notation as in the proof of Lemma 2, let $\chi_t = \int_{\tau=0}^{t}(\nu_\tau^{-i} - I)\mathrm{d}\tau$. By stopping at time $t$, starting at time 0 with a "belief" $\ell$, player $i$'s cost is equal to (integrating (13) by parts)

$$\ell - \gamma e^{\ell} + \int_0^\infty e^{-\mu s}\mu\chi_s \mathrm{d}s + \frac{1 - e^{-\mu t}}{\mu} + \gamma \int_t^\infty e^{\ell + \chi_s - \mu s + t}\mathrm{d}s, \tag{15}$$

which is differentiable in $t$. If $t \in \operatorname{supp} F^i$, it must be that the derivative with respect to $t$ be zero, that is,

$$e^{-\mu t}\left(1 - \gamma e^{\ell + \chi_t + t}\right) + \gamma \int_t^\infty e^{\ell + \chi_s - \mu s + t}\mathrm{d}s = 0. \tag{16}$$

Furthermore, this expression being itself differentiable in $t$, the second derivative must be non-negative, which is equivalent to (differentiating and using the first-order condition)

$$\gamma(I - 1 - \nu_t^{-i} + \mu) - (1 + \mu)e^{-\ell_t} \geq 0. \tag{17}$$

Note that the left-hand side of (17) is decreasing in $t$ if $t \notin \cup_{j \neq i} \operatorname{supp} F^j$. Hence, if $t_1, t_2 \in \operatorname{supp} F^i$, with $t_1 < t_2$, it must be that $(t_1, t_2) \cap \operatorname{supp} F^j \neq \emptyset$ for at least one $j \neq i$. Otherwise, (15) must admit a local maximum at some $t \in (t_1, t_2)$, at which value the inequality of (17) is reversed. This is inconsistent with the monotonicity of the left-hand side of (17) over $(t_1, t_2)$, and the fact that it is positive as either $t \downarrow t_1$ or $t \uparrow t_2$. Because we focus on symmetric equilibria, this implies that, for any $t_1, t_2 \in \operatorname{supp} F^i$, $t_1 < t_2$, there exists $t \in (t_1, t_2)$ such that $t \in \operatorname{supp} F^i$. Hence, the support of $F^i$ (a closed set by definition) must be an interval, and by Lemma 2, we must have $\ell_{\bar{\tau}} = \ell^*$.

Because no pure-policy equilibrium exists, we know $\bar{\tau} > \underline{\tau}$. Assume for the time being that $\underline{\tau} = 0$ (we show later that $\underline{\tau} > 0$ cannot occur). Because the cost from stopping must be constant over $[0, \bar{\tau}]$, the second derivative given by (17) must be identically zero over $(0, \bar{\tau})$. Inequality (17) immediately gives $\nu_t^{-i}$ as a function of $\ell_t$. Because $\ell$ is differentiable, so must $\nu^{-i}$ be. Hence, defining $\xi_t^{-i} = (I - 1 - \nu_t^{-i})/\mu$ and differentiating (17) (eliminating $e^{\ell_t}$ by using (17)) gives that $\xi^{-i}$ obeys the differential equation

$$\dot{\xi}_t^{-i} = \mu\xi_t^{-i}(1 + \xi_t^{-i}),$$

and so $\xi_t^{-i} = (A_1 e^{-\mu t} - 1)^{-1}$ for some $A_1 > 0$ (because $\xi_t^{-i} > 0$), yielding

$$\nu_t^{-i} = I - 1 + \frac{\mu}{1 - A_1 e^{-\mu t}}, \tag{18}$$

for all $t \in (0, \bar{\tau})$. Hence,

$$\ln \mathbf{E}_{\tau^j}[e^{\int_0^t u_s^j \mathrm{d}s}] = \frac{1}{I-1} \int \left( I - 1 + \frac{\mu}{1 - A_1 e^{-\mu s}} \right) \mathrm{d}s = \frac{\ln(A_1 - e^{\mu t})}{I-1} + t + A_2,$$

for some $A_2 \in \mathbf{R}$. That is,

$$\begin{aligned}
\int_{s=0}^{t} e^s \mathrm{d}F(s) + (1 - F(t))e^t &= e^{A_2} e^{\frac{1}{I-1}\left(\ln(A_1 - e^{\mu t}) + (I-1)t\right)} \\
&= e^{A_2} (A_1 - e^{\mu t})^{\frac{1}{I-1}} e^t.
\end{aligned}$$

Differentiating both sides gives finally

$$1 - F(t) = \frac{e^{A_2}}{I-1} (A_1 - e^{\mu t})^{\frac{1}{I-1}} e^t \left( I - 1 - \frac{\mu}{A_1 e^{-\mu t} - 1} \right). \tag{19}$$

It remains to determine the constants $A_1, A_2$.

If $\ell^0 < \ell^{**}$, combine (17) (with equality) at $t = 0$ with (18) to get

$$A_1 = \left( 1 - \frac{\mu}{1+\mu} \gamma e^{\ell^0} \right)^{-1}.$$

Moreover, note from (5) that $1 - F(0) = \nu_0^{-i}/(I-1)$. Plugging in (19) for $t = 0$ using (18) gives $A_2 = (A_1 - 1)^{-\frac{1}{I-1}}$. The resulting distribution is given by

$$\bar{F}(t) = \left( \frac{A_1 - e^{\mu t}}{A_1 - 1} \right)^{\frac{1}{I-1}} \left( 1 - \frac{\mu}{(I-1)(A_1 e^{-\mu t} - 1)} \right). \tag{20}$$

Let us make a few final remarks. First, note that this density is 0 at $\ell^0 = \ell^{**}$. That is, if the game starts with this belief, it never changes and the safe arm is pulled throughout. We must now rule out that $\underline{\tau} > 0$ for this special case. If $\ell^0 = \ell^{**}$, there is nothing to show (as the safe arm is pulled forever anyhow). If $\ell^0 > \ell^{**}$, the safe arm must be pulled throughout (the support of the distribution of stopping beliefs must be convex, yet the cost is strictly quasi-convex in $t$ for $\ell^0 > \ell^{**}$, yielding a contradiction if this region included a stopping time). Now suppose $\ell^0 < \ell^{**}$ and $\underline{\tau} > 0$. Given Lemma 1, the only potentially profitable deviations are stopping policies $\pi_\tau^i$ with $\tau < \underline{\tau}$. Note that, given that players $j \neq i$ use the stopping policy $F^j$, it holds that $\nu_t^{-i} = I - 1$ for all $t < \underline{\tau}$. Hence, a necessary condition for player $i$ to follow the equilibrium policy is that his cost be convex at $t = \underline{\tau}$. Note that the value of (17) at $t = \underline{\tau}$ is

$$\gamma(I - 1 + \mu - \nu_{\underline{\tau}}^{-i}) - (1+\mu)e^{-\ell_{\underline{\tau}}} = \gamma\mu - (1+\mu)e^{-\ell^0}, \tag{21}$$

which, using the definition of $\ell^{**}$, is negative. Because player $i$'s cost is constant over $(\underline{\tau}, \bar{\tau})$, we conclude that deviating to pulling the risky arm at time $\underline{\tau} - \varepsilon$ would be a profitable deviation for $\varepsilon > 0$ small enough.

**Proof of Theorem 2.** As mentioned, the proof of this theorem is rather tedious, and the interested reader might want to consult both the supplementary materials file and a Mathematica file with some of the omitted algebraic operations, available on the authors' websites (entitled `supplementary.pdf` and `theorem2proof.nb`).

The logic of the argument is as follows. Suppose another equilibrium exists. Because on any interval over which a player's opponent does not switch with positive probability, a player's cost is convex, there is at most one time during such an interval at which he is willing to switch. Because of Lemma 6, we know that each player's equilibrium policy must include in its support at least two switching times. If the support of a player's policy is a dense subset of some interval, then so must be his opponent's (because of convexity, as explained), and continuity of the cost function then implies that this support is precisely $[0, \bar{\tau}]$, as defined in Theorem 1, and the equilibrium is the one described there. Hence, we might assume that there exists at least two times $t_1, t_3$, with $0 < t_1 < t_3$, such that, say, player 1's policy assigns positive probability of switching at times $t_1$ and $t_3$, and at no time in between. This however implies (convexity again) that there is some time $t_2 \in (t_1, t_3)$ and some time $t_0 < t_1$ such that player 2 is willing to switch at time $t_0$ and $t_2$, but at no time in between (and 1 does not switch at any time in $(t_0, t_1)$ either).[30] We then derive a contradiction, showing that independently of how players behave at times not in $[t_0, t_4]$, the necessary (first- and second-order) conditions cannot hold simultaneously at those four dates. See `supplementary.pdf` for the details.

**Proof of Proposition 1.** (1.) The stopping-time distribution $F_t^I$ is given by

$$F_t^I = 1 - \frac{\nu_t}{I-1} e^{\int_0^t \frac{\nu_s}{I-1} \mathrm{d}s - t}.$$

For a given $t$, the first term is increasing in $I$. The second term is equal to

$$\left( \frac{e^{-\ell^0} \left(1 + \mu + e^{\mu t} \left(\mu \left(\gamma e^{\ell^0} - 1\right) - 1\right)\right)}{\gamma \mu} \right)^{\frac{1}{I-1}},$$

hence it is smaller than one and increasing in $I$. Therefore, the partial derivative of $F_t^I$ with respect to $I$ is positive for all $t < \bar{\tau}$. In addition, $1 - F_0^I = \nu_0/(I-1)$ which is increasing in $I$. Therefore the distributions $F_t^I$ are ranked by first-order dominance.

(2.) From the outside observer's perspective,

$$\nu_t^I = \frac{I}{I-1} \left( \mu - \frac{\mu(\mu+1)}{\mu + e^{\mu t} \left(\mu \left(\gamma e^{\ell^0} - 1\right) - 1\right) + 1} \right) + I.$$

[30] More precisely, either there is such a $t_0 < t_1$, or a $t_4 > t_3$ in the support of 2's policy, but relabeling the players if necessary, we may as well assume it is $t_0 < t_1$.

Notice that the first term is negative (as $\nu_t^I \leq 1$). This implies $\nu_t^I$ is increasing in $I$.
(3.) The speed of learning of the outside observer is

$$-\dot{\ell}_t^I = I - \nu_t^I,$$

which is decreasing by inspection of $\nu_t^I$. Therefore, during the mixing phase, beliefs decrease faster with a *lower* number of players. Furthermore, as $I$ increases, the length of the mixing phase increases. However, for $t > \bar{\tau}$, beliefs decrease at rate $I$, which implies faster learning for a higher number of players. Therefore, the outside observer's belief trajectories for $I' > I$ cross once at a time $t > \bar{\tau}_{I'}$.
(4.) Straightforward computations of the total cost yield expression (9) in the text. This cost is constant for any $I \geq 2$ and (because of positive informational externalities) strictly lower than the single-agent cost.

# D Proofs for Section 6

**Proof of Proposition 2.** The second inequality of the proposition ($\nu^o(p) \geq \nu^{fb}(p)$) being immediate given that $\bar{p} < p^{FB}$, it is the first inequality that must be established. Given $\ell^0$ and $\ell < \ell^0$, we let $t(\ell)$ denote the time at which the belief of the outside observer reaches belief $\ell$. Let $\nu_t^i$ denote the hazard rate of the outsider's belief at time $t$, *i.e.*, his belief satisfies

$$\dot{\ell}_t = -I(1 - \nu_t^i), \quad \ell_0 = \ell^0.$$

Now suppose towards a contradiction that there exists a "belief" $\hat{\ell}$ such that the outside observer's hazard rate in the unobservable case $\nu^n(\hat{\ell})$ is equal to the hazard rate in the observable case $\nu^o(\hat{\ell})$. We derive an ordinary differential equation for $\nu^{-i}(\ell) := (I-1)\nu_{t(\ell)}^i$ in both cases.

In the unobservable case, we know from the proof of Theorem 1 that

$$\nu_t^{-i} = -1 + I + \frac{\mu}{1 + \frac{e^{-\mu t}(1+\mu)}{e^{\ell^0}\gamma\mu - 1 - \mu}}.$$

Differentiating $\nu_t^{-i}$ with respect to $t$, we obtain

$$\frac{\mathrm{d}\nu_t^{-i}}{\mathrm{d}t} = \frac{\mu^2(\mu+1)e^{\mu t}\left(\mu\left(\gamma e^{\ell^0} - 1\right) - 1\right)}{\left(e^{\mu t}\left(\mu\left(\gamma e^{\ell^0} - 1\right) - 1\right) + \mu + 1\right)^2}.$$

Solving for $e^{\mu t}$ from the definition of $\nu_t^{-i}$ and plugging back into the derivative, we obtain

$$\frac{\mathrm{d}\nu^{-i}(\ell)}{\mathrm{d}\ell} = -(-1 + I - \nu^{-i}(\ell))(-1 + \mu + I - \nu^{-i}(\ell))t'(\ell),$$

where

$$t'(\ell) = \frac{1}{\frac{I}{I-1}\nu^{-i}(\ell) - I}.$$

Finally, we obtain the derivative

$$\frac{\mathrm{d}\nu^{-i}(\ell)}{\mathrm{d}\ell} = (\mu + I - \nu^{-i}(\ell) - 1)\frac{I - \nu^{-i}(\ell) - 1}{I - \nu^{-i}(\ell) - \frac{\nu^{-i}(\ell)}{I-1}}. \tag{22}$$

Note that $\nu^{-i}(\ell)$ is increasing in $\ell$, as expected. Also notice that the second term in (22) is smaller than one, because $\nu^{-i} \leq I - 1$.

In the observable case, we already have the expression for the hazard rate

$$\nu^{-i}(\ell) = \mu + I - 1 - \frac{1 + (\ell - \ell^*)\mu}{e^{\ell}\gamma - 1}.$$

Differentiating with respect to $\ell$ and replacing $e^{\ell}$ with the solution to the previous equation, we obtain the following differential equation

$$\frac{\mathrm{d}\nu^{-i}(\ell)}{\mathrm{d}\ell} = (\mu + I - \nu^{-i}(\ell) - 1)\frac{I - \nu^{-i}(\ell) + \mu(\ell - \ell^*)}{1 + \mu(\ell - \ell^*)}. \tag{23}$$

Notice that $I - \nu^{-i} > 1$, and therefore the ratio in (23) is larger than one. Furthermore, the first term $(\mu + I - \nu^{-i} - 1)$ is identical in the two expressions (22) and (23). Thus, if the two paths $\nu^o(\ell)$ and $\nu^n(\ell)$ cross, the observable path must be steeper. This yields a contradiction, because

$$\nu^o(\ell^{**}) < \nu^n(\ell^{**}) = I - 1,$$

and therefore if the paths $\nu(\ell)$ cross, the unobservable path must be steeper at the crossing point closest to $\ell^{**}$.

**Proof of Theorem 4.** (1.) Under a risk-sharing rule $\alpha$, player $i$'s expected cost can be written as

$$\int_{t\geq 0} e^{-rt}\left(rp_t g\alpha\left(1 - \frac{\nu_t^{-i}}{I-1}\right) + rp_t g(1-\alpha)\left(1 - u_t^i\right) + ru_t^i s + \lambda p_t(I - u_t^i - \nu_t^{-i})s\right)\frac{1-p^0}{1-p_t}\mathrm{d}t.$$

With transformations analogous to those performed in Appendix A, the objective becomes

$$\int_{t\geq 0} e^{-\mu t}\left(e^{\ell_t} g\alpha\left(1 - \frac{\nu_t^{-i}}{I-1}\right) + \mu s\ell_t - e^{\ell_t}(g(1-\alpha) - s)(I - \nu_t^{-i} - 1 + \mu)\right)\mathrm{d}t. \tag{24}$$

Substituting a stopping policy $\pi_t$ for player $i$, differentiating twice with respect to $t$, and setting the first derivative equal to zero yields

$$\gamma\left(I + \mu - \nu_t^{-i} - 1\right) - (1+\mu)e^{-\ell_t} - \alpha(1+\gamma)\left(I\left(1 - \frac{\nu_t^{-i}}{I-1}\right) + \mu\right), \tag{25}$$

which is the analogue of expression (17).

The construction of the equilibrium distribution then mirrors the proof of Theorem 1. In particular, let $\xi := (I - \nu_t^{-i} - 1)/\mu$, and differentiate with respect to time to obtain the differential equation

$$\dot{\xi} = \mu\xi(1+\xi) + \xi B,$$

where

$$B := \frac{\alpha(1+\gamma)(I+\mu-I)}{\gamma(I-1) - \alpha(1+\gamma)I}.$$

This yields as solution

$$\xi_t = \frac{1}{A_1 e^{-(B+\mu)t} - \frac{B}{B+\mu}},$$

hence

$$\nu_t^{-i} = I - 1 - \frac{\mu}{A_1 e^{-(B+\mu)t} - \frac{B}{B+\mu}}. \tag{26}$$

The same steps as in the proof of the main theorem yield

$$1 - F_t = \frac{e^{A_2+t}}{I-1}\left(A_1\left(1+\frac{B}{\mu}\right) - e^{(B+\mu)t}\right)^{\frac{1}{I-1}}\left(I - 1 - \frac{B+\mu}{A_1(1+B/\mu)e^{-(B+\mu)t} - 1}\right),$$

which reduces to our baseline $F_t$ if $\alpha$ and hence $B$ are equal to zero. To solve for $A_1, A_2$ we then combine (25) (equal to zero) and (26), with $\ell_t = \ell^0$, to solve for $\nu_0^{-i}$. Finally, we impose $F_0 = 1 - \nu_0^{-i}/(I-1)$.
(2.) Let players $-i$ follow the threshold policy $p_\alpha^*$. It follows from Lemma 2 that player $i$ wishes to begin experimentation immediately whenever $p^0 < p_\alpha^*$ (and hence players $-i$ pull the risky arm). When $p^0 \geq p_\alpha^*$, players $-i$ never experiment. We know from Lemma 3 that player $i$ is willing to experiment for all $p \leq p_\alpha^{**}$. But when (10) does not hold, $p_\alpha^{**} < p_\alpha^*$, hence player $i$ is not willing to experiment.

**Proof of Proposition 3.** Note that the social planner's problem is unaffected by the transfers associated with risk-sharing. Thus, we still have

$$\frac{p^{FB}}{1-p^{FB}} = \ln\frac{\mu+I}{\mu\gamma}.$$

We can then solve for the value of $\alpha$ that equates the two beliefs $p^{FB}$ and $p_\alpha^*$, obtaining the expression for $\alpha^*$ in equation (10). It is then immediate to verify that

$$\alpha^* = \frac{\gamma(I-1)}{(1+\gamma)(I+\mu)} < \frac{\gamma(I-1+\mu)}{I+\mu} = \hat{\alpha}.$$

**Proof of Theorem 5.** Let $\tilde{\gamma} := (g(1-\alpha)-s)/s$, and write the expression for expected costs (24) recursively. The HJB equation for player $i$ is given by

$$\mu\mathcal{C}^o(\ell) = \mu\ell - \nu^{-i}(\ell) - \tilde{\gamma}(I-\nu^{-i}(\ell)-1+\mu)e^{\ell} + e^{\ell}\alpha\frac{g}{s}\left(1-\frac{\nu^{-i}(\ell)}{I-1}\right) + \min_{u^i}\left\{(u^i+\nu^{-i}(\ell)-I)\frac{\partial\mathcal{C}^o(\ell)}{\partial\ell}\right\}. \tag{27}$$

This corresponds to the objective in Keller and Rady (2015), up to an additional term (the validity of this approach can be validated by a verification theorem, the HJB equation admitting a closed-form solution). The value at the threshold $\ell^*_\alpha$ is given by

$$\mathcal{C}^o_1 = \ell^*_\alpha - \frac{\tilde{\gamma}}{\mu}(I-1+\mu)e^{\ell^*_\alpha} + \frac{\alpha g}{s\mu}e^{\ell^*_\alpha} = \ell^*_\alpha - 1 - \frac{I}{\mu}.$$

Therefore, during the interior-action phase, we have

$$\mu\ell^*_\alpha - \mu - I = \mu\ell - (I-1)u^i(\ell) - \tilde{\gamma}((I-1)(1-u^i(\ell)) + \mu)e^{\ell} + e^{\ell}\alpha\frac{g}{s}\left(1-u^i(\ell)\right).$$

Solving for $u^i(\ell)$, and setting equal to one, we obtain the following equation for the threshold $\bar{\ell}_\alpha$:

$$\mu\ell^*_\alpha - \mu - I = \mu\ell - (I-1) - \tilde{\gamma}\mu e^{\ell},$$

yielding the expression in the text. Further, it is easy to verify that $\alpha^*$ and $\ell^*$ solve the above equation, *i.e.*, the two thresholds $\ell^*$ and $\bar{\ell}$ coincide.

What happens for $\alpha > \alpha^*$? Consider a player's expected cost when all other players experiment starting from a belief $\ell$. This is given by

$$\mathcal{C}(\ell) = \int_0^\infty e^{-\mu t}\left(\mu(\ell - It) - \tilde{\gamma}(I-1+\mu)e^{\ell-It} + \alpha\frac{g}{s}e^{\ell-It}\right)\mathrm{d}t. \tag{28}$$

Now compute the belief such that $\partial\mathcal{C}^o/\partial\ell = 0$ from the HJB equation (27) when all players are playing safe and the continuation payoff is given by $\mathcal{C}(\ell)$ in (28). This belief $\ell^B_\alpha$ satisfies

$$\mu\mathcal{C}(\ell) = \mu\ell - (I-1) - \tilde{\gamma}\mu e^{\ell},$$

which can be solved in closed form to yield

$$\frac{p^B_\alpha}{1-p^B_\alpha} = \frac{\mu+I}{\gamma\mu} =: p^{FB}.$$

Therefore, for $\alpha > \alpha^*$ there is a single switching belief $p_*$. At the first-best belief, players are indifferent given that all others play safe. At the threshold $p^*_\alpha$, they are indifferent given that all others play risky. Thus, any symmetric equilibrium with observable actions involves strictly *less* experimentation than the unique equilibrium under non-observable actions (which involves pure policies with threshold $p^*_\alpha$) but nevertheless a (weakly) excessive amount of experimentation from the planner's perspective.

# References

[1] Abreu, D., P. Milgrom and D. Pearce (1991). “Information and Timing in Repeated Partnerships,” *Econometrica*, **59**, 1713–1733.

[2] Akcigit, U. and Q. Liu (2014). “The Role of Information in Innovation and Competition,” working paper, Columbia University.

[3] Aumann, R.J. (1964). “Mixed and Behavior Policies in Infinite Extensive Games,” in *Advances in Game Theory*, Annals of Mathematics Studies Vol. 52, edited by M. Dresher, L.S. Shapley and A.W. Tucker, 627–650, Princeton University Press, Princeton.

[4] Bandiera, O. and I. Rasul (2006). “Social Networks and Technology Adoption in Northern Mozambique,” *Economic Journal*, **116**, 869–902.

[5] BenYishay, A. and A.M. Mobarak (2014). “Social Learning and Communication,” working paper, NBER.

[6] Bloom, N. and J. Van Reenen (2007). “Measuring and Explaining Management Practices across Firms and Countries,” *Quarterly Journal of Economics*, **122**, 1351–1408.

[7] Board, S. and M. Meyer-ter-Vehn (2014). “A Reputational Theory of Firm Dynamics,” working paper, UCLA.

[8] Bolton, P. and C. Harris (1999). “Strategic Experimentation,” *Econometrica*, **67**, 349–374.

[9] Bonatti, A. and J. Hörner (2011). “Collaborating,” *American Economic Review*, **101**, 632–663.

[10] Carter, M., R. Laajaj and D. Yang (2013). “The Impact of Voucher Coupons on the Uptake of Fertilizer and Improved Seeds: Evidence from a Randomized Trial in Mozambique,” *American Journal of Agricultural Economics*, **95**, 1345–1351.

[11] Cesari, L. (1983). *Optimization: Theory and Applications. Problems with ordinary differential equations*, Applications of Mathematics, 17, Heidelberg: Springer-Verlag.

[12] Coleman, J.S., E. Katz and H. Menzel (1966). *Medical Innovation: A Diffusion Study*, Indianapolis: Bobbs-Merrill Company.

[13] Conley, T. and C.R. Udry (2010). “Learning about a New Technology: Pineapple in Ghana,” *American Economic Review*, **100**, 35–69.

[14] Covert, T. (2014). "Experiential and social learning in firms: the case of hydraulic fracturing in the Bakken Shale," working paper, University of Chicago.

[15] Décamps, J.-P., and T. Mariotti (2004). "Investment Timing and Learning Externalities," *Journal of Economic Theory*, **118**, 80–102.

[16] Duflo, E., M. Kremer and J. Robinson (2006). "Understanding Technology Adoption: Fertilizer in Western Kenya. Evidence from Field Experiments," unpublished manuscript.

[17] Fleming, W.H. and H.N. Soner (2005). *Controlled Markov Processes and Viscosity Solutions*, Applications of Mathematics, 25, Heidelberg: Springer-Verlag.

[18] Foster, A.D. and M.R. Rosenzweig (2010). "Microeconomics of Technology Adoption," *Annual Review of Economics*, **2**, 395–424.

[19] Holmström, B. (1982). "Moral Hazard in Teams," *The Bell Journal of Economics*, **13**, 324–340.

[20] Holmström, B. (1999). "Managerial Incentive Problems: A Dynamic Perspective," *The Review of Economic Studies*, **66**, 169–182. (Originally published in 1982 in Essays in Honor of Professor Lars Wahlbeck.)

[21] Jack, B.K. (2013). "Constraints on the adoption of agricultural technologies in developing countries," literature review, Agricultural Technology Adoption Initiative, J-PAL (MIT) and CEGA (UC Berkeley).

[22] Karlin, S. (1959). *Mathematical Methods and Theory in Games, Programming and Economics*, Dover Publications.

[23] Keller, G. and S. Rady (2003). "Price Dispersion and Learning in a Dynamic Differentiated-Goods Duopoly," *The RAND Journal of Economics*, **34**, 138–165.

[24] Keller, G. and S. Rady (2015). "Breakdowns," *Theoretical Economics*, **10**, 175–202.

[25] Keller, G., S. Rady and M. Cripps (2005). "Strategic Experimentation with Exponential Bandits," *Econometrica*, **73**, 39–68.

[26] Mansfield, E (1961). "Technical change and the rate of imitation," *Econometrica*, **29**, 741–766.

[27] Milgrom, P.R. and R.J. Weber (1985). "Distributional policies for games with incomplete information," *Mathematics of Operations Research*, **10**, 619–632.

[28] Murto, P. and J. Välimäki (2011). "Learning and Information Aggregation in an Exit Game," *The Review of Economic Studies*, **78**, 1426–1461.

[29] Nelsen, R.B. (2006). *An Introduction to Copulas*, 2nd Ed., Springer Series in Statistics, Heidelberg: Springer-Verlag.

[30] Presman, E.L. and I.N. Sonin (1990). *Sequential Control with Incomplete Information*, in: Economic Theory, Econometrics, and Mathematical Economics, San Diego, CA: Academic Press.

[31] Rosenberg, D., A. Salomon and N. Vieille (2013). "On Games of Strategic Experimentation," *Games and Economic Behavior*, **82**, 31–51.

[32] Rosenberg, D., E. Solan and N. Vieille (2007). "Social Learning in One-Arm Bandit Problems," *Econometrica*, **75**, 1591–1611.

[33] Seierstad, A. and K. Sydsæter (1987) *Optimal Control Theory with Economic Applications*, North-Holland.

[34] Shmaya, E. and E. Solan (2014). "Equivalence between Random Stopping Times in Continuous Time," working paper, Kellogg School of Management.

[35] Skinner, J. and D. Staiger (2007). "Technology Adoption from Hybrid Corn to Beta-Blockers," in *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches*, University of Chicago Press.

[36] Syverson, C. (2011). "What Determines Productivity?," *Journal of Economic Literature*, **49**, 326–365.

[37] Touzi, N. and N. Vieille (2002). "Continuous-time Dynkin Games with Mixed Policies," *SIAM Journal of Control and Optimization*, **41**, 1073–1088.

[38] Townsend, R.M. (1994). "Risk and Insurance in Village India," *Econometrica*, **62**, 539–591.

[39] Townsend, R.M. (1995). "Consumption Insurance: An Evaluation of Risk-Bearing Systems in Low-Income Economies," *Journal of Economic Perspectives*, **9**, 83–102.

[40] Vives, X. (1999). *Oligopoly Pricing, Old Ideas and New Tools*, Cambridge, MA: MIT Press.

[41] Weizsäcker, H. (1974). “Zur Gleichwertigkeit zweier Arten der Randomisierung,” *Manuscripta Mathematica*, **11**, 91–94.

[42] Yushkevich, A.A. (1988). “On the two-armed bandit problem with continuous time parameter and discounted rewards,” *Stochastics*, **23**, 299–310.