Strategic Learning in Teams

Nicolas Klein*

This version: August 11, 2013

Abstract

This paper analyzes a two-player game of strategic experimentation with three-armed exponential bandits in continuous time. Players play bandits of identical types, with one arm that is safe in that it generates a known payoff, whereas the likelihood of the risky arms' yielding a positive payoff is initially unknown. When the types of the two risky arms are perfectly negatively correlated, the efficient policy is an equilibrium if and only if the stakes are high enough. If the negative correlation is imperfect and stakes are high, there exists an equilibrium that leads to efficiency for optimistic enough *prior* beliefs.

Keywords: Strategic Experimentation, Three-Armed Bandit, Exponential Distribution, Poisson Process, Bayesian Learning, Markov Perfect Equilibrium, R&D Teams.

JEL Classification Numbers: C73, D83, O32.

*Université de Montréal and CIREQ. Mailing address: Université de Montréal, Département de Sciences Économiques, C.P. 6128 succursale Centre-ville; Montréal, H3C 3J7, Canada. Telephone: +1-514-343-7908; email: kleinnic@yahoo.com.

1 Introduction

When faced with two competing hypotheses, economic agents often have to strike a balance between optimally using their current information on the one hand, and investing in the production of new information on the other hand. When doing so, they have to take into account the impact of their decisions not just on themselves, but on their partners and competitors also; indeed, the latter may benefit from the information a given agent produces. Think e.g. of two jurisdictions, or two hospitals, which are faced with a certain disease that could either be caused by a virus or by a bacterium.[1] Now, either hospital has to decide which of the two competing hypotheses to investigate, e.g. by administering either an antibiotic or an anti-viral drug. Once one of them has found out which hypothesis is true, both benefit from the discovery. Thus, while the costs of experimentation have to be borne privately, any information an agent produces is a public good. This makes for a situation in which a player's experimentation decisions are strategic, in that they affect the other player's payoffs.

I model this trade-off as a three-armed strategic bandit problem.[2] Specifically, I consider two players operating three-armed exponential bandits in continuous time. One arm is safe in that it yields a known flow payoff, whereas the other two arms are risky, i.e. they can be either good or bad. The risky arms are meant to symbolize two mutually incompatible hypotheses. In the baseline case, I assume that it is common knowledge that exactly one of them is good. In Section 6, I extend the analysis to the case in which there is some chance that both hypotheses may be bad, while maintaining the assumption that the two hypotheses cannot be true at the same time; i.e. it could be that the disease is neither caused by a virus nor by a bacterium (it could e.g. be genetic), but it is certainly not caused by both. Players are playing exact carbon copies of the same bandit machine; conditional on the state of the world, draws are iid between the players (i.e. players are playing so called *replica bandits*). The bad risky arm never yields a positive payoff, whereas a good risky arm yields positive payoffs after exponentially distributed times. As the expected payoff of a good risky arm exceeds that of the safe arm, players will want to know which risky arm is good. As either player's actions, as well as the outcomes of his experimentation, are perfectly publicly observable, there is an incentive for players to free-ride on the information the other player is providing; information is a public good.

Observability, together with a common prior, implies that the players' beliefs agree at all times. As only a good risky arm can ever yield a positive payoff, all the uncertainty is resolved as soon as either player has a breakthrough on a risky arm of his and beliefs become

[1] See Klein & Rady (2011).

[2] For an overview of the bandit literature, see Bergemann & Välimäki (2008).

degenerate at the true state of the world. As all the payoff-relevant strategic interaction is captured by the players' common belief process, I restrict players to use stationary Markov strategies with their common posterior belief as the state variable, thus making my results directly comparable to those in the previous strategic experimentation literature.

Players have to bear experimentation costs privately; the benefit, by contrast, is public. There are thus obvious incentives for players to free-ride on their partner's experimentation. Indeed, inefficiency because of free-riding has been a staple result of the literature on strategic experimentation with bandits. Using distributional assumptions similar to ours, Keller et al. (2005) e.g. have shown that in the game with positively correlated bandits, the efficient benchmark is never sustainable in equilibrium.[3] In Klein & Rady (2011), however, full efficiency, regarding both the amount and the speed of experimentation, is an equilibrium if, and only if, stakes are *below* a certain threshold. In the present setting, by contrast, I show that, if players know that exactly one risky arm is good, free-riding incentives can be completely overcome if and only if the stakes *exceed* a certain threshold; in this case, there exists an efficient equilibrium. The result extends to the imperfect negative correlation setting of Section 6 in the sense that, for high enough stakes, there exists an equilibrium that leads to players' behaving efficiently provided their *prior* beliefs are optimistic enough.

The rest of the paper is structured as follows: Section 2 reviews some related literature; Section 3 introduces the model; Section 4 analyzes the utilitarian planner's problem; Section 5 analyzes the non-cooperative game, exhibiting a necessary and sufficient condition for the existence of an efficient equilibrium; Section 6 investigates the robustness of our main result to a more general correlation structure; and Section 7 concludes. Some auxiliary results and proofs are provided in the Appendix.

2 Related Literature

Chatterjee & Evans (2004) analyze a treasure-hunting game in discrete time, where it is common knowledge that exactly one of several projects is good. As in my model, they allow players to switch projects at any point in time. The game ends as soon as one of the players finds the treasure. As in my model, they find that, for high enough stakes, there exists an efficient equilibrium. The forces leading to this result are very different from those at work in my model, though. The nub is that the Chatterjee & Evans (2004) game also involves payoff externalities, in the sense that if player 1 finds the treasure it is lost to player 2. In actual fact, their winner takes all, and hence internalizes all the social

[3] See also Bolton & Harris (1999,2000), and Keller & Rady (2010).

gains of his discovery. In my model, by contrast, externalities are purely informational in nature; when one agent makes a discovery, the other agent fully profits from the information generated by this breakthrough, as he will imitate his partner's successful approach in the future. Thus, Chatterjee & Evans' (2004) model may be better-suited e.g. to the analysis of experimentation by rival firms competing for market share; mine may be more appropriate to the case of different jurisdictions investigating the impact of various treatment options for a particular disease, or if e.g. one wants to analyze free-riding incentives by scientists working for the same firm or in the same lab, and the like.

My model lets agents choose themselves which hypothesis to explore; Klein & Rady (2011) by contrast assign one hypothesis each to either player. The comparison between my model and Klein & Rady (2011) would suggest that delegation was a good idea if the stakes at play were high; if the stakes are low, however, it can be better to assign one hypothesis each to either player, the comparison suggests. Indeed, it is notable that, depending on the circumstances, firms or institutions seem to pursue quite different approaches in this respect. For instance, subsequently to marked growth in the number of its research laboratories and facing increasing competitive pressures, 3M, which arguably makes more low-stakes products such as adhesives and abrasives, moved to restrict scientists' discretion over their work, which had traditionally been very vast (see Bartlett & Mohammed, 1995). By contrast, firms in the arguably more high-stakes pharmaceutical sector moved in the opposite direction. Swiss pharmaceutical giant Novartis, for instance, entered into a multi-million five-year agreement with the Department of Microbial and Plant Biology at Berkeley, CA, delegating project decisions to a committee being comprised of five experts, only two of whom were Novartis employees (see Lacetera, 2008)—a scheme that can reasonably be interpreted as a commitment device on the part of Novartis to delegate project choice to their scientific partners in academia. A somewhat similar deal had earlier been signed by Thousand Oaks, CA, based pharmaceutical company Amgen and MIT; Lawler (2003) quotes MIT biologist Nancy Hopkins: "There was no attempt by either side to change the direction of our basic research" in the aftermath of the agreement.[4]

The present paper belongs to the literature on strategic experimentation with bandits. While bandit models have been analyzed as early as the 1950s (see e.g. Bellman, 1956, Bradt et al, 1956, Robbins, 1952), their use in economics harks back to the discrete-time model of Rothschild (1974). Whereas the first papers analyzing strategic interaction featured a

[4]The optimal allocation of research projects between academia and the commercial sector is the subject of papers by Aghion et al. (2008), as well as by Lacetera (2008), who interpret academia as a commitment device for principals not to interfere with scientists' discretion. The frictions at the heart of both of these papers rely on the assumption that scientists' preferences diverge from those of economically oriented, profit-maximizing, firms.

Brownian motion model (Bolton & Harris, 1999, 2000), the exponential framework I use was first analyzed by Presman (1990) in a single-agent setting, and has proved itself to be more tractable (see Keller et al, 2005, Keller & Rady, 2010, Klein & Rady, 2011). In this literature, my paper is most related to Keller et al. (2005) and Klein & Rady (2011). Keller et al. (2005) show that with replica two-armed bandits there does not exist an equilibrium in cutoff strategies,[5] and that the amount, as well as the speed, of learning are inefficiently low in equilibrium. Klein & Rady (2011) show that with perfectly negatively correlated two-armed bandits there are equilibria in cutoff strategies; the long-run amount of experimentation is always at efficient levels, though the speed of experimentation may be too low. However, there exists an efficient equilibrium if, and only if, the stakes are below a certain threshold. These results extend to the case of imperfect negative correlation in the sense that there will always exist an equilibrium in cutoff strategies in which the long-run amount of experimentation will be at the efficient level. In my model, players play replica bandits, and both players will have access to both types of risky arm at any time.

While the afore-mentioned papers, as well as the present one, assume both actions and outcomes to be public information, Bonatti & Hörner (2011) analyze strategic interaction under the assumption that only outcomes are publicly observable, while actions are private information. Rosenberg et al. (2007), as well as Murto & Välimäki (2011), analyze the two-armed problem of public actions and private outcomes in discrete time, assuming action choices are irreversible. Recently, there has also been an effort at generalization of existing results in the decision-theoretic bandit literature. For example, Bank & Föllmer (2003), as well as Cohen & Solan (2009), analyze the single-agent problem when the underlying process is a general Lévy process.

Camargo (2007) also analyzes the problem of a single decision maker, who faces a two-armed bandit with correlated arms. He considers any number of possible states of the world as well as quite general distributional assumptions. He shows that if the set of outcomes can be ordered in such a way that higher outcomes are good news for one alternative and bad news for the other, a cutoff policy is optimal.[6] As, in general, beliefs are only partially ordered, cutoff beliefs need not be unique. In the baseline case of the present paper, strategic interaction between two players is analyzed in a setting with only two possible states of the world, so that beliefs can be represented by a single number in $[0, 1]$; Section 6 extends the analysis to three possible states of the world.

The present paper is also somewhat related to the Moral Hazard in teams literature, to

[5] A *cutoff strategy* is a strategy of the form "play risky if and only if my belief exceeds a given cutoff."

[6] In this more general setting, a *cutoff policy* is of the form "if arm A is used at a given belief, then arm A is used at higher beliefs as well."

which Holmström (1982) provided the seminal contribution. He found that the introduction of a principal acting as a budget breaker was apt to achieve first-best effort levels on the part of team members. Manso (2011) formalizes an agent's decision among an established production method of known yield, an innovative method with as yet unknown yields, and shirking, as a three-armed bandit in a two-period principal-agent model. His focus is on the wage schemes a principal would optimally offer the agent to induce him to choose the principal's preferred production method. In Klein (2012), I show that in a continuous-time model, the availability of the known production method either renders the implementation of the unknown method impossible, or does not distort its costs at all.

3 The Model

I consider a model of two players, each of whom operates a three-armed bandit in continuous time. One arm is safe in that it yields a known flow payoff of $s > 0$; both other arms, A and B, are risky, and in the baseline case, it is commonly known that exactly one of these risky arms is good and one is bad. The bad risky arm never yields any payoff. The good risky arm yields a positive payoff with a probability of $\lambda\, dt$ if played over a time interval of length dt; the appertaining expected payoff increment amounts to $g\, dt$. Players discount payoffs at the common discount rate $r > 0$.

The constants r, λ, s and g are common knowledge; the only uncertainty is which of the two risky arms is good. The common prior is that A is good with probability p_0. This belief evolves based on the history of experimentation and payoffs. These are commonly observable and so the players continue to have a common belief (probability that arm A is good), which we denote by p_t, at all times $t \geq 0$.

At each point in time, both players receive a flow endowment of one unit of a perfectly divisible resource. Either player's objective is to maximize his own expected discounted payoffs by choosing the fraction of his endowment flow that he wants to allocate to either risky arm. Specifically, either player $i \in \{1, 2\}$ chooses a stochastic process $\{(k_{i,A}, k_{i,B})(t)\}_{0 \leq t}$ which is measurable with respect to the information filtration that is generated by the observations available up to time t, with $(k_{i,A}, k_{i,B})(t) \in \{(a, b) \in [0, 1]^2 : a + b \leq 1\}$ for all t; $k_{i,A}(t)$ and $k_{i,B}(t)$ denote the fraction of the resource devoted by player i at time t to risky arms A and B, respectively.[7] Throughout the game, either player's actions and payoffs are

[7]Here, putting a fraction of the available resources on a risky project means that the probability of getting a lump-sum reward is reduced at any moment of time, but the size of the reward does not change. It can also be viewed as an approximation for a situation in which only one arm can be pulled at any moment but a policy may change arbitrarily quickly between the arms, spending fraction $k_{i,A}$ ($k_{i,B}$) of the time on arm

perfectly observable to the other player. Specifically, player i seeks to maximize his total expected discounted payoff

$$\mathrm{E}\left[\int_0^\infty r\, e^{-r\,t}\, \left[(1-k_{i,A}(t)-k_{i,B}(t))s + (k_{i,A}(t)p_t + k_{i,B}(t)(1-p_t))\, g\right]\, dt\right],$$

where the expectation is taken with respect to the processes $\{p_t\}_{t\in\mathbb{R}_+}$ and $\{(k_{i,A}, k_{i,B})(t)\}_{t\in\mathbb{R}_+}$. As can immediately be seen from this objective function, there are no payoff externalities between the players; the only channel through which the presence of the other player may impact a given player is via his belief p_t, i.e. via the information that the other player is generating. Thus, ours is a game of purely informational externalities.

As only a good risky arm can ever yield a lump sum, breakthroughs are fully revealing. Thus, if there is a lump sum on risky arm A (B) at time τ, then $p_t = 1$ ($p_t = 0$) at all $t > \tau$. If there has not been a breakthrough by time τ, Bayes' Rule yields

$$p_\tau = \frac{p_0 e^{-\lambda \int_0^\tau K_{A,t}\, dt}}{p_0 e^{-\lambda \int_0^\tau K_{A,t}\, dt} + (1-p_0) e^{-\lambda \int_0^\tau K_{B,t}\, dt}},$$

where $K_{A,t} := k_{1,A}(t) + k_{2,A}(t)$ and $K_{B,t} := k_{1,B}(t) + k_{2,B}(t)$. Thus, conditional on no breakthrough having occurred, the process $\{p_t\}_{t\in\mathbb{R}_+}$ will evolve according to the law of motion

$$\dot{p_t} = -(K_{A,t} - K_{B,t})\lambda p_t(1-p_t)$$

almost everywhere.

As all payoff-relevant strategic interaction is captured by the players' common posterior beliefs $\{p_t\}_{t\in\mathbb{R}_+}$, it seems quite natural to focus on Markov perfect equilibria with the players' common posterior belief p_t as the state variable. As is well known, this restriction is without loss of generality in the planner's problem, which is studied in Section 4. A Markov strategy for player i is any piecewise continuous function $(k_{i,A}, k_{i,B}) : [0,1] \to \{(a,b) \in [0,1]^2 : a + b \leq 1\}$, $p_t \mapsto (k_{i,A}, k_{i,B})(p_t)$, meaning that it is continuous at all but a finite number of points. Following Klein & Rady (2011), I say that a *pair* of Markov strategies is *admissible* if there exists at least one well-defined solution to the corresponding law of motion of beliefs; in case of multiple solutions, the unique solution that is consistent with a discrete-time approximation is selected.

Given an admissible strategy pair $((k_{1,A}, k_{1,B})(p_t), (k_{2,A}, k_{2,B})(p_t))$, the players' belief is given by

$$p_\tau = \frac{p_0 e^{-\lambda \int_0^\tau K_A(p_t)\, dt}}{p_0 e^{-\lambda \int_0^\tau K_A(p_t)\, dt} + (1-p_0) e^{-\lambda \int_0^\tau K_B(p_t)\, dt}},$$

A (B), for instance.

if there has not been a breakthrough by time τ, with $K_A(p_t) := k_{1,A}(p_t) + k_{2,A}(p_t)$ and $K_B(p_t) := k_{1,B}(p_t)+k_{2,B}(p_t)$. Each admissible strategy pair $(k_1, k_2) = ((k_{1,A}, k_{1,B}), (k_{2,A}, k_{2,B}))$ induces a pair of payoff functions (u_1, u_2) with u_i given by

$$u_i(p|k_1, k_2) = \mathrm{E}\left[\int_0^\infty re^{-rt}\left\{(k_{i,A}(p_t)p_t + k_{i,B}(p_t)(1-p_t))g + [1-k_{i,A}(p_t)-k_{i,B}(p_t)]s\right\}dt \middle| p_0 = p\right]$$

for each $i \in \{1, 2\}$. For strategy pairs that are not admissible, I set $u_1 = u_2 = -\infty$.

In the subsequent analysis, it will prove useful to make case distinctions based on the stakes at play, as measured by the ratio of the expected payoff of a good risky arm over that of a safe arm ($\frac{g}{s}$), the players' impatience (as measured by the discount rate r), and the Poisson arrival rate of a good risky arm λ, which can be interpreted as the players' innate ability at finding out the truth: I say that the stakes are high if $\frac{g}{s} \geq \frac{4(r+\lambda)}{2r+3\lambda}$; stakes are intermediate if $\frac{2r+\lambda}{r+\lambda} < \frac{g}{s} < \frac{4(r+\lambda)}{2r+3\lambda}$; stakes are low if $\frac{g}{s} \leq \frac{2r+\lambda}{r+\lambda}$; they are very low if $\frac{g}{s} < \frac{2(r+\lambda)}{r+2\lambda}$.

4 The Planner's Problem

First, we investigate a benevolent utilitarian planner's solution to the two-player problem at hand. As the planner does not care about the distribution of surplus, and both players are equally apt at finding out the truth, all that matters to him is the sum of resources devoted to either type of risky arm, $K_A(p_t)$ and $K_B(p_t)$, respectively. Put differently, the planner's problem is equivalent to that of a single agent controlling twice the resources. In order to make his value function comparable to that of our players, we normalize the planner's value by the factor $\frac{1}{2}$. Now, straightforward computations show that the planner's Bellman equation is given by[8]

$$u(p) = s + \max_{\{(K_A,K_B)\in[0,2]^2 : K_A+K_B\leq 2\}} \left\{K_A\left[B_A(p,u) - \frac{c_A(p)}{2}\right] + K_B\left[B_B(p,u) - \frac{c_B(p)}{2}\right]\right\},$$

where $c_A(p) := s - pg$ and $c_B(p) := s - (1-p)g$ measure the myopic opportunity costs of playing risky arm A (risky arm B) rather than the safe arm. By contrast, $B_A(p,u) := \frac{\lambda}{r}p[g - u(p) - (1-p)u'(p)]$ and $B_B(p,u) := \frac{\lambda}{r}(1-p)[g - u(p) + pu'(p)]$ measure the value of information gleaned from playing risky arm A (or risky arm B, respectively).[9] As the Bellman equation is linear in the planner's choice variables, it is without loss of generality

[8] By standard arguments, if a continuously differentiable function solves the Bellman equation, it is the value function.

[9] By the standard principle of smooth pasting, the planner's payoff function from playing an optimal policy is once continuously differentiable.

for me to restrict attention to corner solutions, for which it is straightforward to derive closed-form solutions for the value function, which are exhibited in Appendix A.

The optimal policy depends on whether the stakes at play, as measured by the ratio $\frac{g}{s}$, exceed the threshold of $\frac{2(r+\lambda)}{r+2\lambda}$ or not. Note that $\frac{g}{s} \leq \frac{2(r+\lambda)}{r+2\lambda}$ if and only if $p_2^* \geq \frac{1}{2}$, where $p_2^* := \frac{rs}{(r+2\lambda)(g-s)+rs}$. It will be convenient furthermore to define the odds ratio $\Omega(p) := \frac{1-p}{p}$, and $\bar{u}_{11} := \frac{r+2\lambda}{2(r+\lambda)}g$, the planner's payoff when investing one unit each per risky arm. The planner's solution is summarized in the following proposition:

Proposition 4.1 (Planner's Solution) *If $\frac{g}{s} < \frac{2(r+\lambda)}{r+2\lambda}$, the planner will invest all his resources in arm A on $(p_2^*, 1]$; in arm B on $[0, 1-p_2^*)$; and in the safe arm on $[1-p_2^*, p_2^*]$. The corresponding payoff function is given by*

$$u(p) = \begin{cases} g\left[1 - p + \frac{2\lambda p_2^*}{2\lambda p_2^* + r} p \left(\Omega(p)\Omega(p_2^*)\right)^{-\frac{r}{2\lambda}}\right] & \text{if} \quad p \leq 1 - p_2^*, \\ s & \text{if} \quad 1 - p_2^* \leq p \leq p_2^*, \\ g\left[p + \frac{2\lambda p_2^*}{2\lambda p_2^* + r}(1-p)\left(\frac{\Omega(p)}{\Omega(p_2^*)}\right)^{\frac{r}{2\lambda}}\right] & \text{if} \quad p \geq p_2^*. \end{cases}$$

If $\frac{g}{s} > \frac{2(r+\lambda)}{r+2\lambda}$, the planner will invest all his resources in arm A on $(\frac{1}{2}, 1]$, and in arm B on $[0, \frac{1}{2})$. At $p = \frac{1}{2}$, he will split his resources equally between the risky arms. The corresponding payoff function is given by

$$u(p) = \begin{cases} g\left[1 - p + \frac{\lambda}{r+\lambda} p \Omega(p)^{-\frac{r}{2\lambda}}\right] & \text{if} \quad p \leq \frac{1}{2}, \\ g\left[p + \frac{\lambda}{\lambda+r}(1-p)\Omega(p)^{\frac{r}{2\lambda}}\right] & \text{if} \quad p \geq \frac{1}{2}. \end{cases}$$

Either solution is optimal if $\frac{g}{s} = \frac{2(r+\lambda)}{r+2\lambda}$.

PROOF: Optimality is established by a standard verification argument, which is in Klein (2011) and omitted here. ∎

Note that if the stakes are very low, there is no option value to the initially less promising risky arm, since the planner will never make use of it. As is easily verified, the optimal solution in this case implies incomplete learning. Indeed, let us suppose that it is risky arm A that is good. Then, if the initial prior p_0 is in $[0, 1-p_2^*)$, we have that $\lim_{t\to\infty} p_t = 1 - p_2^*$ with probability 1. If $p_0 \in [1-p_2^*, p_2^*]$, then $p_t = p_0$ for all t, since the planner will always play safe. If $p_0 \in (p_2^*, 1]$, it is straightforward to show that the belief will converge to p_2^* with probability $\frac{\Omega(p_0)}{\Omega(p_2^*)}$, while the truth will be found out (i.e. the belief will jump to 1) with the counter-probability. Hence, there is always a positive probability that the true state of the world will not be found out, i.e. learning is incomplete.

If $\frac{g}{s} > \frac{2(r+\lambda)}{r+2\lambda}$, however, which is the case if and only if $\bar{u}_{11} > s$, the planner will never avail himself of the option to play safe; his solution will ensure that learning be complete, i.e. that the truth will eventually be found out with probability 1. As the planner does not care which of the risky arms is good, the solution is symmetric around $p = \frac{1}{2}$. At the switch point $p = \frac{1}{2}$ itself, the planner's actions are pinned down by the need to ensure a well-defined law of motion of the state variable. Thus, there is now an option value to the initially less promising risky project, as the planner will make use of it with strictly positive probability, whatever his initial belief $p_0 \in (0, 1)$ may be.

At the knife-edge case of $\frac{g}{s} = \frac{2(r+\lambda)}{r+2\lambda}$, the planner is indifferent over all three arms at $p = \frac{1}{2}$. Yet, in order to ensure a well-defined time path of beliefs, he has to set $K_A(\frac{1}{2}) = K_B(\frac{1}{2}) \in [0, 1]$.

The single-agent optimum has the same structure as the planner's solution; all that changes in the relevant differential equations is that 2λ is replaced by λ. Of course, the relevant cutoffs will also change as a result: In the single-agent problem, complete learning will obtain for $\frac{g}{s} > \frac{2r+\lambda}{r+\lambda}$; for $\frac{g}{s} < \frac{2r+\lambda}{r+\lambda}$, the agent will switch from risky arm A to the safe arm at the cutoff belief $p_1^* := \frac{rs}{(r+\lambda)g-\lambda s} > p_2^*$, and from risky arm B to the safe arm at $1 - p_1^*$. Thus, whenever the stakes are below the relevant threshold, the second risky option does not play a role; hence, it is not surprising that the same cutoff will be applied as in the problem with two-armed exponential bandits, where p_1^* and p_2^* are the relevant cutoffs in the single-agent problem and the planner's problem with two replica bandits, respectively, as Proposition 3.1 in Keller et al. (2005) shows.

5 Equilibria of the Non-Cooperative Game

Proceeding as before, I find that the Bellman equation for player i ($i \neq j$) is given by[10]

$$u_i(p) = s + k_{j,A}B_A(p, u_i) + k_{j,B}B_B(p, u_i) + \max_{\{(k_{i,A},k_{i,B})\in[0,1]^2:k_{i,A}+k_{i,B}\leq 1\}} \{k_{i,A}\left[B_A(p, u_i) - c_A(p)\right] + k_{i,B}\left[B_B(p, u_i) - c_B(p)\right]\}.$$

As players are perfectly symmetric in that they are operating two replicas of the same bandit, the Bellman equation for player j looks exactly the same. On account of the linear

[10]By the smooth pasting principle, player i's payoff function from playing a best response is once continuously differentiable on any open interval on which $(k_{j,A}, k_{j,B})(p)$ is continuous. If $(k_{j,A}, k_{j,B})(p)$ exhibits a jump at p, $u_i'(p)$, which is contained in the definitions of B_A and B_B, is to be understood as the one-sided derivative in the direction implied by the motion of beliefs. In either instance, standard results imply that if for a certain fixed $(k_{j,A}, k_{j,B})$, the payoff function generated by the policy $(k_{i,A}, k_{i,B})$ solves the Bellman equation, then $(k_{i,A}, k_{i,B})$ is a best response to $(k_{j,A}, k_{j,B})$.

structure of the optimization problem, we can restrict our attention to the nine pure strategy profiles, along with three indifference cases per player. Each of these cases leads to a first-order ordinary differential equation (ODE). A leading case is exhibited in Appendix A; a full overview can be found in Klein (2011).

The linearity of the problem provides us with a powerful tool to derive necessary conditions for a certain strategy combination $((k_{1,A}, k_{1,B}), (k_{2,A}, k_{2,B}))$ to be consistent with mutually best responses on an open set of beliefs.[11] As an example, suppose player 2 is playing $(1, 0)$. If player 1's best response is given by $(1, 0)$, it follows immediately from the Bellman equation that it must be the case that $B_A(p, u_1) \geq c_A(p)$ and $B_A(p, u_1) - B_B(p, u_1) \geq c_A(p) - c_B(p)$ for all p in the open interval in question. Moreover, we know that in the open interval in question, the player's value function satisfies

$$2\lambda p(1-p)u_1'(p) + (2\lambda p + r)u_1(p) = (2\lambda + r)pg,$$

which can be plugged into the two inequalities above, yielding a necessary condition for $(k_{1,A}, k_{1,B}) = (1, 0)$ to be a best response to $(k_{2,A}, k_{2,B}) = (1, 0)$. Proceeding in this manner for the possible pure-strategy combinations gives us necessary conditions for a certain pure-strategy combination to be consistent with mutually best responses on an open interval of beliefs. I report these necessary conditions as an auxiliary result in Appendix A.

It is noteworthy that, as we can see immediately from the Bellman equation, a player only has to bear the opportunity costs of his own experimentation, while the benefits accrue to both, which indicates the presence of free-riding incentives. For future reference, I define the myopic cutoff belief $p^m := \frac{s}{g}$ by $c_A(p^m) = 0$. A player who was only interested in maximizing his current payoff would switch from risky arm A (B) to the safe arm at p^m $(1 - p^m)$.

Such free-riding incentives would also appear in a much simpler model without uncertainty.[12] Suppose the safe arm, which, when pulled, yielded a payoff flow of s, was paired with an arm that was known to be good, and to yield a lump sum of h, *to be shared equally by both players*, according to a known Poisson distribution with parameter $(k_{1,t} + k_{2,t})\lambda$, where $(k_{1,t}, k_{2,t}) \in [0, 1]^2$ denoted the proportion of his unit endowment flow that either player devoted to the Poisson arm at instant t. As is easy to verify, efficiency would require both players to pull the good arm if $\lambda h > s$; the players would be willing to do so in Markov equilibrium, however, only if $\frac{\lambda h}{2} \geq s$.[13] Hence, efficiency would prevail in equilibrium for

[11] As we keep player j's strategy $(k_{j,A}, k_{j,B})$ fixed on an open interval of beliefs, player i's value function u_i $(i \neq j)$ is of class C^1 on that open interval. Therefore, by standard arguments, u_i solves the Bellman equation on the open interval in question.

[12] I am indebted to an anonymous referee for pointing this out.

[13] As here the Poisson arm is known to be good, i.e. beliefs are degenerate, the Markov restriction essentially

very large and very small λh. For $s < \lambda h < 2s$, however, efficiency would require both players to use the good arm, but they would both prefer to play safe in equilibrium. Yet if there is some chance that the Poisson arm is bad, and never yields any payoffs, then in the case of many unsuccessful draws, players' beliefs may reach a level of pessimism such that the *expected* payoff from pulling the Poisson arm enters the free-riding region, however large λh may be. As even very optimistic players know that this is going to happen with positive probability at some point down the road, their equilibrium utility is lower than in the efficient benchmark. This is why in these sorts of problems, there is typically too little experimentation in equilibrium.[14]

By contrast, I find that the planner's solution is compatible with equilibrium if and only if stakes exceed a certain threshold. This may at first glance seem surprising given that, in contrast to Chatterjee & Evans (2004), a player does not fully internalize the benefits of his discovery. Indeed, players provide a positive informational externality through their experimentation; any information a player generates helps his partner make better decisions in turn. This is the reason why efficiency is not sustainable in equilibrium in Keller et al. (2005). In Klein & Rady (2011), this calculation changes, though, when the stakes are so low that the players' respective single-agent cutoffs do not overlap: In this case, the more pessimistic player will never play risky under any circumstances, which the more optimistic player will anticipate, and hence behave efficiently. However, the efficient equilibrium disappears as soon as the relevant single-agent cutoffs overlap and free-riding incentives kick in again.

While it is not surprising that the utilitarian planner, who now has more options, should always be doing better than the planner in Klein & Rady (2011), who could not transfer resources between the two types of risky arm, it may seem somewhat surprising that, for high stakes, the players should now be able to achieve even this *higher* efficient benchmark, while they could not achieve the *lower* benchmark in the perfectly negatively correlated two-armed model in Klein & Rady (2011). Indeed, with the stakes high enough, free-riding incentives can be overcome completely in non-cooperative equilibrium, as the following proposition shows.

Proposition 5.1 (Efficient Equilibrium) *There exists an efficient equilibrium if and only if* $\frac{g}{s} \geq \frac{4(r+\lambda)}{2r+3\lambda}$.

PROOF: See Appendix B. ∎

requires players to pick one arm and stick with it.

[14] See Bolton & Harris (1999, 2000) Keller et al. (2005), Keller & Rady (2010).

Since players are playing replica bandits, there will never arise a situation in which one player is optimistic while the other one is pessimistic; as soon as one player finds it optimal to experiment in isolation then so will the other player, and free-riding incentives enter the picture again. Therefore, the Klein & Rady (2011) channel effecting efficiency cannot be at work here, whatever the stakes might be. For high stakes, a different channel will kick in, though: On account of perfect negative correlation between the risky arms, players will never simultaneously be very pessimistic about both prospects. Hence, for stakes above a certain threshold, they would never consider the safe option. This is even though they are still exerting a positive externality, which they do not mind, however, as individual incentives to foreswear the safe option are strong enough. Moreover, since there are no switching costs in my model, players would use the risky arm that looks momentarily more promising if they were left to their own devices. Thus, in the absence of specific incentives to deviate from this policy, they would do what efficiency requires. In particular, if the other player behaves efficiently, a player's best response calls for behaving efficiently also; i.e. there exists an efficient equilibrium.[15]

In Chatterjee & Evans' (2004) efficient equilibrium, players behave myopically. Note that this is not necessarily the case here, as the relevant threshold above which free-riding incentives are totally eclipsed is lower than 2 (above which experimentation becomes costless, i.e. myopically optimal, at all beliefs). This is because players take the learning benefit of experimentation into account, at least to the extent it benefits the player himself. Thus, it is no surprise that the relevant threshold should be increasing in the players' impatience r, and decreasing in the informativeness of experimentation, as measured by λ.

For beliefs below this threshold, one can still construct a symmetric Markov perfect equilibrium.[16] If stakes are intermediate, i.e. $\frac{2r+\lambda}{r+\lambda} < \frac{g}{s} < \frac{4(r+\lambda)}{2r+3\lambda}$, such an equilibrium features complete learning, as efficiency requires. If the stakes are low, i.e. $\frac{g}{s} \leq \frac{2r+\lambda}{r+\lambda}$, there exists an equilibrium in which players exclusively use the safe arm on $[1-p_1^*, p_1^*]$. Hence, this equilibrium implies incomplete learning, while efficiency requires complete learning for $\frac{g}{s} > \frac{2(r+\lambda)}{r+2\lambda}$. Thus, while our analysis would unambiguously suggest that, if stakes were high, delegating project choice to the agents was a good idea since it increased experimentation intensities, this conclusion need not hold for $\frac{g}{s} < \frac{2r+\lambda}{r+\lambda}$. For this case, Klein & Rady (2011) have shown that if agents are assigned one of the projects each, the unique equilibrium

[15] Holmström (1982) shows that a team cannot produce efficiently in the absence of a budget-breaking principal, on account of payoff externalities between team members. By contrast, my analysis shows that, in a model with purely informational externalities in which players can choose whether to investigate a given hypothesis or its negation, the efficient solution is an equilibrium if the stakes at play exceed a certain threshold.

[16] See a previous version for details (Klein, 2011).

features an experimentation intensity of 1 for the more promising project throughout $[0, 1 - p_1^*) \cup (p_1^*, 1]$. By contrast, in the equilibrium for low stakes we have mentioned here and which is discussed in more detail in Klein (2011), the overall experimentation intensity increases continuously from 0 at p_1^* to 2 at some $\hat{p} > p_1^*$ (decreases continuously from 2 at $1 - \hat{p}$ to 0 at $1 - p_1^*$). Thus, for initial beliefs just above p_1^*, for instance, the rate of experimentation may be higher if scientists do not have the freedom to choose the hypothesis they are working on. Hence, if the stakes are low, as arguably they might be at a company like 3M, it might make sense to restrict scientists' discretion, whilst delegation would seem advisable in more high-stakes sectors, such as pharmaceutics, for instance.

6 Generalized Pessimism

Heretofore, we have assumed that players are certain that exactly one of their risky arms is good. The purpose of this section is to investigate whether our main result that free-riding incentives can be overcome for stakes exceeding a certain threshold extends to situations of generalized pessimism à la Klein & Rady (2011). In this setting, there are three possible states of the world: Either only arm A is good, or only arm B is good, or neither arm is good. Thus, the types of arms A and B are still negatively correlated, yet the negative correlation is no longer perfect. Following the constructive approach in Klein & Rady (2011), I show that the results for this case will be mixed: There does not exist an equilibrium in which players behave efficiently at all beliefs; however, if the stakes are high enough and *initial* general pessimism is low, it is possible to construct a symmetric Markov perfect equilibrium prescribing efficient behavior at all beliefs that are reached with positive probability along the path of play.

Our state variable (p_A, p_B) will now be two-dimensional, with $p_{I,t}$ denoting the players' (subjective) probability at time t that arm $I \in \{A, B\}$ is good. Thus, $\frac{p_{B,0}}{1-p_{A,0}}$ is an (inverse) measure of initial general pessimism. As before, players i will be restricted to stationary Markov strategies $(k_{i,A}, k_{i,B}) : [0,1]^2 \to \{(a,b) \in [0,1]^2 : a + b \leq 1\}$. Admissibility of strategies is defined in analogy to the case of perfect negative correlation. The laws of motion of beliefs are given by

$$\dot{p}_{A,t} = -\lambda p_{A,t} \left[K_{A,t}(1 - p_{A,t}) - K_{B,t} p_{B,t} \right],$$

and

$$\dot{p}_{B,t} = -\lambda p_{B,t} \left[K_{B,t}(1 - p_{B,t}) - K_{A,t} p_{A,t} \right].$$

As our state space is now two-dimensional, the derivation of explicit solutions for players' payoff functions now requires the solution of Partial Differential Equations (PDE). Details

are provided in Appendix A.

The following proposition summarizes the efficient benchmark for this case. The relevant cutoff is again given by p_2^*, as in the case of perfect correlation. When the probability of being good drops below this threshold for both arms, the planner will give up and henceforth play safe. At all other beliefs, the planner will devote all of his resources to the arm that is more likely to be good. When both arms are equally likely to be good, he uses both at equal intensity.

Proposition 6.1 (The Planner's Solution) *The planner sets* $(K_A, K_B) = (0,0)$ *if* $(p_A, p_B) \in [0, p_2^*]^2$*;* $(K_A, K_B) = (2,0)$ *if* $p_A > \max\{p_B, p_2^*\}$*;* $(K_A, K_B) = (0,2)$ *if* $p_B > \max\{p_A, p_2^*\}$*; and* $(K_A, K_B) = (1,1)$ *if* $p_A = p_B > p_2^*$.

PROOF: Proof is by a verification argument, see Appendix B. ∎

In the following proposition, I show that, if the stakes are high, i.e. if $\frac{g}{s} > \frac{4(r+\lambda)}{2r+3\lambda}$, there exists a region of prior beliefs which attach little enough weight to both arms' being bad, such that, starting out from these prior beliefs, players' behavior will be efficient with probability 1 in the equilibrium I construct. However, in marked contrast to the case of perfect correlation, there never exists an equilibrium in which players behave efficiently at all beliefs, no matter how high the stakes may be. Indeed, as in the case of perfect correlation, players are not willing to put in the required effort at beliefs close to p_2^*. Yet, with imperfect correlation, there are some points in the belief space at which the probability that *the more promising arm* will be good is discouragingly low, something that never happens for high enough stakes when the correlation is perfect. Hence we can have an efficient equilibrium in the latter case but not in the former.

The key to the construction of an equilibrium that achieves efficient behavior given optimistic prior beliefs is that once beliefs reach the 45-degree line in (p_A, p_B)-space, deviations are precluded by the admissibility requirement that strategies lead to a well-defined time path of beliefs which is consistent with Bayes' rule, a logic similar to that applying at the point $p = \frac{1}{2}$ in the case of perfect correlation. Thus, efficiency can be attained if and only if players are still optimistic enough *at the moment they reach the 45-degree line*, which is the case for high enough ratios $\frac{p_{B,0}}{1-p_{A,0}}$. In the limiting case when $\frac{p_{B,0}}{1-p_{A,0}} = 1$ (perfect correlation), the 45-degree line collapses to the point $p = \frac{1}{2}$, and players are still optimistic enough when they reach that point if and only if the stakes are high, i.e. $\frac{g}{s} \geq \frac{4(r+\lambda)}{2r+3\lambda}$. In case this inequality holds with strictness, there exists an $\eta > 0$ such that if $\frac{p_{B,0}}{1-p_{A,0}} \geq 1 - \eta$, players are still optimistic enough to do the work efficiency requires of them when beliefs reach the 45-degree line. The following proposition summarizes these findings.

Proposition 6.2 (Generalized Pessimism) *If $p_{B,0} < 1 - p_{A,0}$, there does not exist an efficient equilibrium. If $\frac{g}{s} > \frac{4(r+\lambda)}{2r+3\lambda}$, there exists an $\eta > 0$ such that, if $\frac{p_{B,0}}{1-p_{A,0}} \geq 1 - \eta$, there exists an equilibrium in which players' behavior coincides with the efficient benchmark with probability 1. If $\frac{g}{s} \leq \frac{4(r+\lambda)}{2r+3\lambda}$, players behave inefficiently with positive probability in all equilibria for all non-degenerate prior beliefs $p_{A,0} + p_{B,0} < 1$ not in $[0, p_2^*]^2 \cup \{(p_A, p_B) : p_A = p_B\}$.*

PROOF: Here, I give a very short sketch of the proof, the full proof being provided in Appendix B.

Player i's Bellman equation is given by $(i \neq j)$

$$u(p_A, p_B) = s + k_{j,A} B_A(p_A, p_B, u) + k_{j,B} B_B(p_A, p_B, u) + \max_{(k_{i,A}, k_{i,B})} \{k_{i,A} [B_A(p_A, p_B, u) - c_A(p_A)] + k_{i,B} [B_B(p_A, p_B, u) - c_B(p_B)]\}.$$

By symmetry, we can focus on the case $p_{A,0} \geq p_{B,0}$. We note that if only arm A is used on a time interval $[t, t + \Delta)$ (with $\Delta > 0$), $\frac{p_{B,\tau}}{1-p_{A,\tau}}$ remains constant for all $\tau \in [t, t + \Delta)$. Now, let $p_{A,0} > p_{B,0}$, and $x := \frac{p_{B,0}}{1-p_{A,0}} \leq \frac{p_2^*}{1-p_2^*}$. It is now straightforward to show that $B_A(p_A, p_B, u^P) < c_A(p_A)$ for p_A close to p_2^* (where u^P denotes the value function of the planner's problem). This already establishes that there is no efficient equilibrium if $p_{B,0} < 1 - p_{A,0}$.

Now let $\frac{g}{s} \leq \frac{4(r+\lambda)}{2r+3\lambda}$. For this case, one can show that, for each $x \in \left(\frac{p_2^*}{1-p_2^*}, 1\right)$, $B_A(p_A, p_B, u^P) < c_A(p_A)$ for $p_A > p_B$ and p_A close to p_B. This establishes that if $\frac{g}{s} \leq \frac{4(r+\lambda)}{2r+3\lambda}$, players behave inefficiently with positive probability in all equilibria for all non-degenerate prior beliefs $p_{A,0} + p_{B,0} < 1$ not in $[0, p_2^*]^2 \cup \{(p_A, p_B) : p_A = p_B\}$.

Now, let $\frac{g}{s} > \frac{4(r+\lambda)}{2r+3\lambda}$. One shows that there exists a neighborhood $\mathcal{N}$ of $x = 1$ such that, for all $x \in \mathcal{N}$, $B_A(p_A, p_B, u^P) \geq c_A(p_A)$ and $B_A(p_A, p_B, u^P) - c_A(p_A) \geq B_B(p_A, p_B, u^P) - c_B(p_B)$ for $p_A > p_B$ and $\frac{p_B}{1-p_A} \in \mathcal{N}$. The proof now constructs an equilibrium such that, starting from prior beliefs $(p_{A,0}, p_{B,0})$ satisfying $\frac{p_{B,0}}{1-p_{A,0}} \in \mathcal{N}$, players' behavior be efficient throughout with probability 1.

In this equilibrium, actions at beliefs $p_A \geq p_B$ are as follows: Both players play safe if $p_A \leq p_2^*$. If $p_A = p_B > p_2^*$, one player plays $(1, 0)$ while the other plays $(0, 1)$ until we reach (p_2^*, p_2^*). If $p_A > \max\{p_B, p_1^*\}$ and $\frac{p_B}{1-p_A} \in \mathcal{N}$, both players use arm A. If $p_A > \max\{p_B, p_1^*\}$ and $\frac{p_B}{1-p_A} \notin \mathcal{N}$, both players mix over arm A and the safe arm close to the 45-degree line in (p_A, p_B)-space (if $x \geq \frac{p_1^*}{1-p_1^*}$) or the line $p_A = p_1^*$ (if $x < \frac{p_1^*}{1-p_1^*}$), respectively. Away from the 45-degree line or the line $p_A = p_1^*$ respectively, both players use arm A if $p_A > \max\{p_B, p_1^*\}$. If $p_B < p_A \leq p_1^*$, both players play safe. At beliefs $p_A < p_B$, the symmetric actions prevail, with the roles of arm A and arm B reversed. That these strategies constitute mutually

best responses is established by a verification argument, the details of which are provided in Appendix B. ∎

7 Conclusion

I have analyzed a game of strategic experimentation with three-armed bandits. We have seen that when the two risky arms are perfectly negatively correlated, the efficient solution can be sustained as a Markov perfect equilibrium if and only if stakes are high. While making my results easily comparable with the existing literature, the restriction to Markovian equilibria of course rules out history-dependent play, which is familiar from discrete time, yet technically rather intricate to formalize in continuous time. Appropriate continuous-time concepts can e.g. be found in Bergin & MacLeod (1993). Hörner et al. (2013) extend the analysis of the experimentation game with replica two-armed Poisson bandits à la Keller & Rady (2010) to non-Markovian equilibria. While a full characterization and analysis of non-Markovian equilibria lies outside the scope of this paper, it seems clear that simple "grim trigger" equilibria could achieve the efficient outcome for a substantially larger set of parameters. Indeed, for stakes that are not very low, i.e. for $\frac{2(r+\lambda)}{r+2\lambda} < \frac{g}{s} < \frac{4(r+\lambda)}{2r+3\lambda}$, consider beliefs $\underline{p}, \bar{p} \in (0,1)$, with $\underline{p}$ and $\bar{p}$ very close to 0, and 1, respectively. On $(\underline{p}, \bar{p})$, the payoff in the planner's solution is bounded away from the payoff in the Markov perfect equilibrium for $\frac{g}{s} < \frac{4(r+\lambda)}{2r+3\lambda}$ that we have mentioned above. Thus, the threat of punishing any unilateral deviation with an immediate and indefinite reversion of play to the symmetric Markov perfect equilibrium gives players appropriate incentives to behave efficiently on $(\underline{p}, \bar{p})$. On $[0, \underline{p}] \cup [\bar{p}, 1]$, meanwhile, players want to behave efficiently even if deviations are ignored. If $\frac{g}{s} < \frac{2(r+\lambda)}{r+2\lambda}$, however, this simple construction fails, as it is now no longer the case that, away from $p = 0$ and $p = 1$, players' utility in the planner's solution is bounded away everywhere from that in the Markov perfect equilibrium.

We have seen that our results to some extent generalize when we additionally allow for the possibility of both risky arms' being bad; indeed, if players *initially* assess the likelihood of both arms being bad as low enough, there exists an equilibrium inducing efficient behavior with probability 1. In future research, it could be interesting to explore the additional trade-offs arising when players differed with respect to their innate learning abilities, as parameterized by the Poisson arrival rate of breakthroughs. Analyzing these additional trade-offs that would appear, if, say, player 1 was able to learn faster on risky arm A, while player 2 was faster with risky arm B might yield insights into conditions under which there is excessive, or insufficient, specialization in equilibrium.

Appendix

A Closed-Form Solutions And An Auxiliary Result

Closed-Form Solutions for the Case of Perfect Negative Correlation

If $((1,0),(1,0))$ is played, both players' value functions satisfy the following ODE:

$$2\lambda p(1-p)u'(p) + (2\lambda p + r)u(p) = (2\lambda + r)pg,$$

which is solved by

$$u(p) = pg + C(1-p)\Omega(p)^{\frac{r}{2\lambda}},$$

where C is some constant of integration. The same solution also applies to the planner's value function if $(K_A, K_B) = (2,0)$ prevails.

If $((0,1),(0,1))$ is played, the roles of p and $1-p$ are reversed, and both players' value functions satisfy the following ODE:

$$-2\lambda p(1-p)u'(p) + (2\lambda(1-p) + r)u(p) = (2\lambda + r)(1-p)g,$$

which is solved by

$$u(p) = (1-p)g + Cp\Omega(p)^{-\frac{r}{2\lambda}}.$$

The same solution also applies to the planner's value function if $(K_A, K_B) = (0,2)$ prevails.

If the players' belief freezes, as is e.g. the case when $((0,1),(1,0))$ obtains, the players' values are linear. Thus, in the case of $((0,1),(1,0))$ for instance, we have

$$u_1(p) = \frac{\lambda + r(1-p)}{\lambda + r}g;$$

$$u_2(p) = \frac{\lambda + rp}{\lambda + r}g.$$

In this case, the planner's value, being the average of the two, is constant, $\frac{r+2\lambda}{2(r+\lambda)}g = \bar{u}_{11}$.

Solutions for the other six pure strategy profiles, as well as the three indifference cases, are exhibited in Klein (2011).

An Auxiliary Result

The logic we discussed in Section 5 of the main text gives us the following auxiliary result, which will be useful in the proof of Proposition 5.1.

Lemma A.1 *Consider players $(i,j) \in \{1,2\}^2 \setminus \{(i,i) : i \in \{1,2\}\}$, and let $\mathcal{P} \subset (0,1)$ be an open interval of beliefs in which the action profile remains constant, and let $p \in \mathcal{P}$.*

Let $k_j(p) = (0,0)$. Then the following statements hold:

- *If player i's best response is given by $k_i(p) = (0,0)$, then $u_i(p) = s$.*

- *If player i's best response is given by $k_i(p) = (1,0)$ or $k_i(p) = (0,1)$, then $u_i(p) \geq \max\{s, \frac{r+\lambda}{2r+\lambda}g\}$.*

Let $k_j(p) = (1,0)$. Then the following statements hold:

- *If player i's best response is given by $k_i(p) = (0,0)$, then $\frac{\lambda+r(1-p)}{\lambda+r}g \leq u_i(p) \leq 2s - pg$.*
- *If player i's best response is given by $k_i(p) = (1,0)$, then $u_i(p) \geq \max\{\frac{\lambda+r(1-p)}{\lambda+r}g, 2s - pg\}$.*
- *If player i's best response is given by $k_i(p) = (0,1)$, then $u_i(p) = \frac{\lambda+r(1-p)}{\lambda+r}g$ and $p \leq \min\{1 - p^m, \frac{r+\lambda}{2r+3\lambda}\}$.*

Let $k_j(p) = (0,1)$. Then the following statements hold:

- *If player i's best response is given by $k_i(p) = (0,0)$, then $\frac{\lambda+rp}{\lambda+r}g \leq u_i(p) \leq 2s - (1-p)g$.*
- *If player i's best response is given by $k_i(p) = (1,0)$, then $u_i(p) = \frac{\lambda+rp}{\lambda+r}g$ and $p \geq \max\{p^m, \frac{r+2\lambda}{2r+3\lambda}\}$.*
- *If player i's best response is given by $k_i(p) = (0,1)$, then $u_i(p) \geq \max\{\frac{\lambda+rp}{\lambda+r}g, 2s - (1-p)g\}$.*

As $\frac{r+\lambda}{2r+3\lambda} < \frac{1}{2} < \frac{r+2\lambda}{2r+3\lambda}$, the lemma immediately implies that in no equilibrium $((1,0),(0,1))$ or $((0,1),(1,0))$ can arise on an open interval. If furthermore $\frac{g}{s} \geq 2$, and hence $2s - pg \leq \frac{\lambda+r(1-p)}{\lambda+r}g$ for all $p \in [0,1]$, then $((1,0),(0,0))$, $((0,0),(1,0))$, $((0,1),(0,0))$ and $((0,0),(0,1))$ cannot arise on an open interval either.

Explicit Solutions for the Case of Imperfect Correlation (Section 6)

If $(K_A, K_B) = (0,0)$, both players' payoff function u satisfies $u = s$.

For $(K_A, K_B) = (2,0)$, we verify that $\frac{p_B}{1-p_A}$ is constant. The PDE is given by

$$2\lambda p_A(1-p_A)\frac{\partial u}{\partial p_A} - 2\lambda p_A p_B \frac{\partial u}{\partial p_B} + (r + 2\lambda p_A)u = (r+2\lambda)p_A g,$$

which, as in Klein & Rady (2011), we find is solved by

$$u = p_A g + f_{20}\left(\frac{p_B}{1-p_A}\right)\check{u}_0(p_A),$$

with

$$\check{u}_0(p) := (1-p)\left(\frac{1-p}{p}\right)^{\frac{r}{2\lambda}}.$$

From Klein & Rady (2011), we know that

$$\check{u}_0'(p) = -\frac{\frac{r}{2\lambda} + p}{p(1-p)}\check{u}_0(p),$$

and $\check{u}_0'' > 0$.

For $(K_A, K_B) = (1,1)$, we verify that $\frac{p_B}{p_A}$ is constant and find that the players' average payoff is given by

$$u = \frac{r+2\lambda}{2(r+\lambda)}(p_A + p_B)g + f_{11}\left(\frac{p_B}{p_A}\right)u_0(p_A + p_B),$$

with $u_0(p) := (1-p)\left(\frac{1-p}{p}\right)^{\frac{r}{\lambda}}$. From Klein & Rady (2011), we know that

$$u_0'(p) = -\frac{\frac{r}{\lambda}+p}{p(1-p)}u_0(p),$$

and $u_0'' > 0$.

If a player is indifferent between arm A and the safe arm, his payoff function u satisfies

$$\lambda p_A(1-p_A)\frac{\partial u}{\partial p_A} - \lambda p_A p_B \frac{\partial u}{\partial p_B} + \lambda p_A u = (r+\lambda)p_A g - rs.$$

This is solved by

$$u = s + \frac{r+\lambda}{\lambda}(g-s) + \frac{r}{\lambda}s(1-p_A)\ln\left(\frac{1-p_A}{p_A}\right) + \tilde{f}\left(\frac{p_B}{1-p_A}\right)(1-p_A).$$

B Proofs

Proof of Proposition 5.1

Suppose $\frac{g}{s} \geq \frac{4(r+\lambda)}{2r+3\lambda}$. What is to be shown is that the action profiles $((1,0),(1,0))$ and $((0,1),(0,1))$ are mutually best responses on $(\frac{1}{2},1]$, and $[0,\frac{1}{2})$, respectively. At $p=\frac{1}{2}$, admissibility uniquely pins down a player's response to the other player's action. By the characterization of efficiency (see Proposition 4.1), both players' respective value function if efficiency prevails is given by:

$$u(p) = \begin{cases} g\left[1-p+\frac{\lambda}{r+\lambda}p\Omega(p)^{-\frac{r}{2\lambda}}\right] & \text{if} \quad p \leq \frac{1}{2} \\ g\left[p+\frac{\lambda}{r+\lambda}(1-p)\Omega(p)^{\frac{r}{2\lambda}}\right] & \text{if} \quad p \geq \frac{1}{2}. \end{cases}$$

Now, by Lemma A.1, it is sufficient to show that $u(p) > \max\{\frac{\lambda+r(1-p)}{\lambda+r}g, 2s-pg\}$ on $(\frac{1}{2},1]$, and $u(p) > \max\{\frac{\lambda+rp}{\lambda+r}g, 2s-(1-p)g\}$ on $[0,\frac{1}{2})$. I shall only consider the former interval, as the argument pertaining to the latter is perfectly symmetric.

Simple algebra shows that if $\frac{g}{s} \geq \frac{4(r+\lambda)}{2r+3\lambda}$, $w(p) := \frac{\lambda+r(1-p)}{\lambda+r}g \geq 2s-pg$ everywhere in $[\frac{1}{2},1]$. Since $u(\frac{1}{2}) = w(\frac{1}{2})$, and u is strictly increasing while w is strictly decreasing in $(\frac{1}{2},1)$, the claim follows.

Suppose $\frac{2(r+\lambda)}{r+2\lambda} \leq \frac{g}{s} < \frac{4(r+\lambda)}{2r+3\lambda}$, and define $\tilde{w}(p) := 2s-pg$. It is now straightforward to show that $\tilde{w}(\frac{1}{2}) > w(\frac{1}{2}) = u(\frac{1}{2})$, and, therefore, by Lemma A.1, there exists a neighborhood to the right of $p=\frac{1}{2}$ in which $(1,0)$ is not a best response to $(1,0)$.

Suppose that the stakes are very low, i.e. $\frac{g}{s} < \frac{2(r+\lambda)}{r+2\lambda}$. From our characterization of the efficient solution (see Proposition 4.1), we know that the planner's value function is given by

$$u(p) = \begin{cases} g\left[1-p+\frac{2\lambda p_2^*}{2\lambda p_2^*+r}p\left(\Omega(p)\Omega(p_2^*)\right)^{-\frac{r}{2\lambda}}\right] & \text{if} \quad p \leq 1-p_2^*, \\ s & \text{if} \quad 1-p_2^* \leq p \leq p_2^*, \\ g\left[p+\frac{2\lambda p_2^*}{2\lambda p_2^*+r}(1-p)\left(\frac{\Omega(p)}{\Omega(p_2^*)}\right)^{\frac{r}{2\lambda}}\right] & \text{if} \quad p \geq p_2^* \end{cases},$$

and that $B_A(p_2^*,u) = \frac{c_A(p_2^*)}{2}$. For the efficient actions to be a best response, it is necessary that $B_A \geq c_A$ on $(p_2^*,1]$. Yet, since u is of class C^1, we have that $\lim_{p\downarrow p_2^*} B_A(p,u) = \frac{c_A(p_2^*)}{2} < c_A(p_2^*)$, as $p_2^* < p^m$. ∎

Proof of Proposition 6.1

The planner's Bellman equation is given by

$$u = s + \max_{(K_A,K_B)} \left\{ K_A \left[B_A(p_A, p_B, u) - \frac{c_A(p_A)}{2} \right] + K_B \left[B_B(p_A, p_B, u) - \frac{c_B(p_B)}{2} \right] \right\}$$

with

$$B_A(p_A, p_B, u) := \frac{\lambda}{r} p_A \left[g - u - (1 - p_A) \frac{\partial u}{\partial p_A} + p_B \frac{\partial u}{\partial p_B} \right],$$

$$B_B(p_A, p_B, u) := \frac{\lambda}{r} p_B \left[g - u - (1 - p_B) \frac{\partial u}{\partial p_B} + p_A \frac{\partial u}{\partial p_A} \right],$$

$c_I(p_I) := s - p_I g$ for $I = A, B$.

By symmetry, we only need to look at beliefs $p_A \geq p_B$. We define $x := \frac{p_{B,0}}{1-p_{A,0}} < 1$, and note that whenever $K_{B,t} = 0$, $\frac{p_{B,t}}{1-p_{A,t}}$ remains constant, i.e. if $p_{A,0} > p_{B,0}$ and $(2,0)$ prevails, beliefs move toward the 45-degree line in (p_A, p_B)-space along the straight line $p_B = (1-p_A)x$. We make a case distinction depending on whether $x > \frac{p_2^*}{1-p_2^*}$ or $x < \frac{p_2^*}{1-p_2^*}$, dealing with the knife-edge case $x = \frac{p_2^*}{1-p_2^*}$ at the end.

Case $x > \frac{p_2^*}{1-p_2^*}$

We shall first analyze the case of $x > \frac{p_2^*}{1-p_2^*}$. As $x < 1$, this implies $p_2^* < \frac{1}{2}$, i.e. $\frac{g}{s} > \frac{2(r+\lambda)}{r+2\lambda}$. Along the 45-degree line, i.e. for $p_A = p_B =: p$, we have seen in Appendix A that the planner's payoff function for our conjectured solution is given by

$$u_{11}(p_A, p_B) = \frac{r + 2\lambda}{2(r + \lambda)} (p_A + p_B) g + f_{11}^P \left(\frac{p_B}{p_A} \right) u_0(p_A + p_B).$$

Value matching at (p_2^*, p_2^*) pins down $f_{11}^P(1) =: C^*$, and gives us

$$u_{11}^P(p_A, p_B) = \frac{r + 2\lambda}{2(r + \lambda)} (p_A + p_B) g + C^* u_0(p_A + p_B)$$

with $C^* = \frac{s - \frac{r+2\lambda}{r+\lambda} p_2^* g}{u_0(2p_2^*)}$. One verifies that $C^* > 0$ if $\frac{g}{s} > \frac{2(r+\lambda)}{r+2\lambda}$, i.e. if the stakes are not very low, and $p_2^* < \frac{1}{2}$.

Let us now turn to the $(2,0)$ region. As we have seen in Appendix A, the payoff function here is given by

$$u_{20}^P = p_A g + f_{20}^P \left(\frac{p_B}{1 - p_A} \right) \check{u}_0(p_A).$$

Value matching at $\left(\frac{x}{x+1}, \frac{x}{x+1} \right)$ gives us

$$f_{20}^P(x) \check{u}_0 \left(\frac{x}{x+1} \right) = \frac{\lambda}{r + \lambda} \frac{x}{x+1} g + C^* u_0 \left(\frac{2x}{x+1} \right),$$

i.e.

$$f_{20}^P(x) = \frac{\lambda}{r + \lambda} g x^{\frac{r+2\lambda}{2\lambda}} + C^* 2^{-\frac{r}{\lambda}} (1 - x)^{\frac{r+\lambda}{\lambda}} x^{-\frac{r}{2\lambda}} = x^{\frac{r}{2\lambda}} \left\{ \frac{\lambda}{r + \lambda} g x + C^* 2^{-\frac{r}{\lambda}} u_0(x) \right\}.$$

In order to show that the planner's behavior is indeed optimal, we fix an $x > \frac{p_2^*}{1-p_2^*}$ and check that $B_A \geq \frac{c_A}{2}$ and $B_A - \frac{c_A}{2} \geq B_B - \frac{c_B}{2}$ for all (p_A, p_B) in the $(2,0)$-region along the ray defined by x. From the closed-form solution, we have that

$$\frac{\partial u_{20}^P}{\partial p_A} = g - \frac{\frac{r}{2\lambda} + p_A}{p_A(1-p_A)} f_{20}^P(x)\check{u}_0(p_A) + \frac{x}{1-p_A} f_{20}^{P'}(x)\check{u}_0(p_A),$$

and

$$\frac{\partial u_{20}^P}{\partial p_B} = \frac{1}{1-p_A} f_{20}^{P'}(x)\check{u}_0(p_A).$$

Thus,

$$B_A = \tfrac{1}{2} f_{20}^P(x)\check{u}_0(p_A).$$

Hence,

$$B_A \geq^? \frac{c_A}{2} \iff p_A g + f_{20}^P(x)\check{u}_0(p_A) \geq^? s. \tag{B.1}$$

As the left-hand side is a strictly convex function in p_A, we are going (1.) to check optimality at $p_A = \frac{x}{x+1}$, and (2.) then check if the derivative of the left-hand side of (B.1) with respect to p_A at $p_A = \frac{x}{x+1}$ is non-negative (holding x fixed as we move along the ray given by our x). If both (1.) and (2.) hold, it follows by convexity that $B_A \geq \frac{c_A}{2}$ for all $p_A \geq \frac{x}{x+1}$ along our particular ray.

We note that by value matching (1.) is equivalent to $u_{20}^P\left(\frac{x}{x+1}, \frac{x}{x+1}\right) = u_{11}^P\left(\frac{x}{x+1}, \frac{x}{x+1}\right) > s$. We define $\tilde{u}(p) := u_{11}^P(p,p)$. Using the explicit expression for C^*, we have that

$$\tilde{u}(p) = \frac{r+2\lambda}{r+\lambda} pg + \left(s - \frac{r+2\lambda}{r+\lambda} p_2^* g\right) \frac{u_0(2p)}{u_0(2p_2^*)}.$$

As $\tilde{u}(p_2^*) = s$, and $\tilde{u}$ is a strictly convex function of p for $\frac{g}{s} > \frac{2(r+\lambda)}{r+2\lambda}$, with $\tilde{u}'$ approaching 0 as $p \downarrow p_2^*$, we can conclude that $\tilde{u}$ is strictly increasing on $(p_2^*, \frac{1}{2})$, and hence that $\tilde{u}(p) > s$ for all $p \in (p_2^*, \frac{1}{2}]$.

Regarding point (2.), we have that the derivative of the left-hand side of (B.1) with respect to p_A at $p_A = \frac{x}{x+1}$ is non-decreasing if and only if

$$\left(\frac{r}{2\lambda} + \frac{r+2\lambda}{2\lambda} x\right)\left[\frac{\lambda}{r+\lambda} g + C^* \frac{x+1}{x} u_0\left(\frac{2x}{x+1}\right)\right] \leq^? g. \tag{B.2}$$

Noting that $1 - 2p_2^* = \frac{r+2\lambda}{rs} p_2^* \left(g - \frac{2(r+\lambda)}{r+2\lambda} s\right)$, and that value matching at (p_2^*, p_2^*) implies that $C^* u_0(2p_2^*) = \frac{\lambda}{r} p_2^* \frac{r+2\lambda}{r+\lambda}\left[g - \frac{2(r+\lambda)}{r+2\lambda} s\right]$, we have that $\frac{C^* u_0(2p_2^*)}{1-2p_2^*} = \frac{\lambda}{r+\lambda} s$. Thus, we can re-arrange (B.2) to

$$\left(\frac{r}{2(r+\lambda)} + \frac{r+2\lambda}{2(r+\lambda)} x\right)\left[g - s + \frac{s}{x}\right] \leq^? g.$$

By direct calculation, one shows that this condition exactly binds at $x = \frac{p_2^*}{1-p_2^*}$ as well as at $x = 1$. The derivative of its left-hand side with respect to x works out as

$$\frac{r+2\lambda}{2(r+\lambda)}\left(g - s + \frac{s}{x}\right) - \frac{s}{x^2}\left(\frac{r}{2(r+\lambda)} + \frac{r+2\lambda}{2(r+\lambda)} x\right).$$

This derivative is strictly negative if and only if $x < \sqrt{\frac{p_2^*}{1-p_2^*}}$. Thus (B.2), and hence (B.1), hold for all $x \in \left[\frac{p_2^*}{1-p_2^*}, 1\right]$, so that we can conclude that $B_A \geq \frac{c_A}{2}$ for all $p_A \geq \frac{x}{x+1}$ along any given ray $x \in \left[\frac{p_2^*}{1-p_2^*}, 1\right]$.

It remains to verify that $2\left(B_A - B_B\right) \geq^? c_A - c_B = -(p_A - p_B)g$. After plugging in for $\frac{\partial u^P}{\partial p_A}$ and $\frac{\partial u^P}{\partial p_B}$ from the explicit solution to the payoff function, we find that

$$B_A - B_B = \frac{\lambda}{r}\left\{-p_B g + \left(\frac{r}{2\lambda} + \frac{r+2\lambda}{2\lambda}x\right) f_{20}^P(x)\check{u}_0(p_A) + x(1-x)f_{20}'^P(x)\check{u}_0(p_A)\right\} \geq^? -(p_A - p_B)\frac{g}{2}.$$

Again we want to check if this holds for every (p_A, p_B) along a given ray x. Thus, we can replace p_B by $p_A(1-x)$, and get

$$\begin{aligned} & B_A - B_B \geq^? \frac{c_A - c_B}{2} \\ \iff & g\left[\frac{p_A}{2} - (1-p_A)x\frac{r+2\lambda}{2r}\right] + \left(\tfrac{1}{2} + x\frac{r+2\lambda}{2r}\right) f_{20}^P(x)\check{u}_0(p_A) + \frac{\lambda}{r}x(1-x)f_{20}'^P(x)\check{u}_0(p_A) \geq^? 0. \end{aligned}$$

We have that

$$f_{20}'^P(x) = \frac{r}{2\lambda x}f_{20}^P(x) + x^{\frac{r}{2\lambda}}\left\{\frac{\lambda}{r+\lambda}g - \frac{\frac{r}{\lambda}+x}{x(1-x)}u_0(x)C^*2^{-\frac{r}{\lambda}}\right\}.$$

Using that, by the equation for $f_{20}^P(x)$, we have that $C^*2^{-\frac{r}{\lambda}}u_0(x) = x^{-\frac{r}{2\lambda}}f_{20}^P(x) - \frac{\lambda}{r+\lambda}xg$, we can rewrite this as

$$f_{20}'^P(x) = \frac{f_{20}^P(x)}{x}\left[\frac{r}{2\lambda} - \frac{\frac{r}{\lambda}+x}{1-x}\right] + \frac{\lambda}{r+\lambda}x^{\frac{r}{2\lambda}}g\left[1 + \frac{\frac{r}{\lambda}+x}{1-x}\right].$$

This gives us

$$\frac{\lambda}{r}x(1-x)f_{20}'^P(x) = -f_{20}^P(x)\left(\tfrac{1}{2} + \frac{r+2\lambda}{2r}x\right) + \frac{\lambda}{r}x^{1+\frac{r}{2\lambda}}g.$$

Thus, our condition simplifies to

$$B_A - B_B \geq^? \frac{c_A - c_B}{2} \iff \frac{p_A}{2} - (1-p_A)x\frac{r+2\lambda}{2r} + \frac{\lambda}{r}x^{1+\frac{r}{2\lambda}}\check{u}_0(p_A) \geq^? 0.$$

(1.) It is immediate to check that this condition binds at $p_A = \frac{x}{x+1}$, as expected. (2.) As the condition is strictly convex in p_A, we will again check whether the derivative of its left-hand side at $p_A = \frac{x}{x+1}$ is non-negative. At $p_A = \frac{x}{x+1}$, this derivative works out as

$$\tfrac{1}{2}\left[1 + \frac{r+2\lambda}{r}x\right] - x^{\frac{r}{2\lambda}}(x+1)^{\frac{1}{2}}\left[1 + \frac{r+2\lambda}{r}x\right]\check{u}_0\left(\frac{x}{x+1}\right),$$

which exactly equals 0 as $\check{u}_0\left(\frac{x}{x+1}\right) = \frac{x^{-\frac{r}{2\lambda}}}{x+1}$.

Thus, we have shown that arm A is strictly preferred to arm B for all $p_A > \frac{x}{x+1}$ along our ray x. Hence, we can conclude that arm A is indeed optimal in the region we conjectured.

Along the 45-degree line, where $(1,1)$ prevails in our conjectured solution, deviations to any action profile (a,b) with $a \neq b$ are ruled out by admissibility considerations. By the linearity of the Bellman equation, we thus only have to rule out a deviation to $(0,0)$, which is unprofitable as $\tilde{u}(p) > s$ for all $p > p_2^*$.

On $[0, p_2^*]^2$, $p_I \leq p_2^*$ implies that $B_I - \frac{c_I}{2} \leq 0$ for all $I \in \{A, B\}$. This completes the proof for the case $x > \frac{p_2^*}{1-p_2^*}$.

Case $x < \frac{p_2^*}{1-p_2^*}$

As before, in the $(2,0)$-region, the planner's value function is given by

$$u_{20}^P = p_A g + f_{20}^P\left(\frac{p_B}{1-p_A}\right)\check{u}_0(p_A).$$

Here, though, value matching at $(p_A, p_B) = (p_2^*, (1-p_2^*)x)$ gives us

$$f_{20}^P(x) = \frac{s - p_2^* g}{\check{u}_0(p_2^*)},$$

i.e. f_{20}^P is a constant that is independent of x. We have

$$\frac{\partial u_{20}^P}{\partial p_A} = g - \frac{\frac{r}{2\lambda} + p_A}{p_A(1-p_A)}(s - p_2^* g)\frac{\check{u}_0(p_A)}{\check{u}_0(p_2^*)},$$

and

$$\frac{\partial u_{20}^P}{\partial p_B} = 0.$$

Thus, we get

$$B_A = \tfrac{1}{2}(s - p_2^* g)\frac{\check{u}_0(p_A)}{\check{u}(p_2^*)} \geq^? \tfrac{1}{2}(s - p_A g) = \frac{c_A}{2} \iff p_A g + (s - p_2^* g)\frac{\check{u}_0(p_A)}{\check{u}_0(p_2^*)} \geq^? s.$$

(1.) This binds at $p_A = p_2^*$. (2.) As before, this condition is strictly convex in p_A, so that it is sufficient to verify that the derivative of its left-hand side with respect to p_A at $p_A = p_2^*$ is non-negative. Using the fact that $\frac{s - p_2^* g}{1-p_2^*} = \frac{2\lambda}{r+2\lambda}s$, we find that this derivative at $p_A = p_2^*$ works out as

$$g - \frac{s - p_2^* g}{1 - p_2^*}\left(1 + \frac{r}{2\lambda p_2^*}\right) = g - g = 0,$$

i.e. we indeed have $B_A > \frac{c_A}{2}$ for all $p_A > p_2^*$ along our ray x.

It remains to be shown that the planner prefers arm A over arm B as well. Here, we will show that the safe arm is preferred over Arm B, i.e. that $B_B \leq \frac{c_B}{2}$. Since we are checking optimality along a given ray x, we can again replace p_B by $(1-p_A)x$. We have that

$$B_B \leq^? \frac{c_B}{2} \iff (1-p_A)xg - (s - p_2^* g)\frac{\check{u}_0(p_A)}{\check{u}_0(p_2^*)}x \leq^? \frac{r}{r+2\lambda}s.$$

(1.) At $p_A = p_2^*$, this holds if and only if $x \leq \frac{p_2^*}{1-p_2^*}$. (2.) The left-hand side of the condition is strictly concave in p_A. Hence, it is enough to show that its derivative with respect to p_A at $p_A = p_2^*$ is non-positive. By plugging in the explicit expression for p_2^*, we find that this derivative works out as

$$-xg + (s - p_2^* g)\frac{\frac{r}{2\lambda} + p_2^*}{p_2^*(1-p_2^*)}x = 0,$$

i.e. we indeed have $B_B < \frac{c_B}{2}$ for all $p_A > p_2^*$ along our ray x.

As before, on $[0, p_2^*]^2$, $p_I \leq p_2^*$ implies that $B_I - \frac{c_I}{2} \leq 0$ for all $I \in \{A, B\}$. This completes the proof for the case $x < \frac{p_2^*}{1-p_2^*}$ as well.

As is easily verified, the planner's solution is continuous and smooth at $x = \frac{p_2^*}{1-p_2^*}$, and either argument for $x \downarrow \frac{p_2^*}{1-p_2^*}$ or $x \uparrow \frac{p_2^*}{1-p_2^*}$ can be extended to the case $x = \frac{p_2^*}{1-p_2^*}$. ∎

Proof of Proposition 6.2

Player i's Bellman equation is given by $(i \neq j)$

$$u(p_A, p_B) = s + k_{j,A} B_A(p_A, p_B, u) + k_{j,B} B_B(p_A, p_B, u) + \max_{(k_{i,A}, k_{i,B})} \{k_{i,A} [B_A(p_A, p_B, u) - c_A(p_A)] + k_{i,B} [B_B(p_A, p_B, u) - c_B(p_B)]\}.$$

Suppose there exists an efficient equilibrium. Then both players' payoff functions in this equilibrium are given by the planner's solution u^P, and neither player must have an incentive to deviate given this payoff function. Let $x \leq \frac{p_2^*}{1-p_2^*}$, and $p_{A,0} > p_{B,0}$. At $p_A = p_2^*$, we have that $B_A(p_A, p_B, u^P) = \frac{c_A(p_A)}{2} < c_A(p_A)$ where the last inequality follows from the fact that $p_2^* < p^m$. A unilateral deviation to $(k_{i,A}, k_{i,B}) = (0, 0)$ in the $(2, 0)$-Region of the presumed equilibrium is admissible since it does not alter the direction of the motion of beliefs; it only slows down its speed. Moreover continuity of $B_A(p_A, p_B, u^P)$ and c_A implies that such a unilateral deviation is profitable close to $p_A = p_2^*$, implying that there does not exist an equilibrium in which players' behavior is efficient at all beliefs.

Now let $x > \frac{p_2^*}{1-p_2^*}$. Using the expression for B_A from the planner's solution (see proof of Proposition 6.1 above), we have that player i prefers arm A over the safe arm if and only if

$$\tfrac{1}{2} f_{20}^P(x) \check{u}_0(p_A) + p_A g \geq^? s,$$

with f_{20}^P the same function as in the planner's solution (see proof of Proposition 6.1 above). It is immediate to show that the derivative of the left-hand side of this condition with respect to p_A at $p_A = \frac{x}{x+1}$ (keeping x fixed) is $g + \frac{1}{2} f_{20}^P(x) \check{u}_0' \left(\frac{x}{x+1}\right) \geq g + f_{20}^P(x) \check{u}_0' \left(\frac{x}{x+1}\right) \geq 0$, where the second inequality follows from the proof of the planner's solution (Proposition 6.1). Thus, by strict convexity of $\check{u}_0$, a necessary and sufficient condition for player i to prefer arm A over safe is given by $B_A \left(\frac{x}{x+1}, \frac{x}{x+1}, u^P\right) \geq c_A \left(\frac{x}{x+1}\right)$, i.e.

$$\tfrac{1}{2} f_{20}^P(x) \check{u}_0 \left(\frac{x}{x+1}\right) + \frac{x}{x+1} g \geq^? s.$$

By value matching in the planner's solution, we know that $f_{20}^P(x) \check{u}_0 \left(\frac{x}{x+1}\right) + \frac{x}{x+1} g = \tilde{u} \left(\frac{x}{x+1}\right)$, so that the condition simplifies to

$$\tfrac{1}{2} \tilde{u} \left(\frac{x}{x+1}\right) + \tfrac{1}{2} \frac{x}{x+1} g \geq^? s \iff \tilde{u} \left(\frac{x}{x+1}\right) \geq 2s - \frac{x}{x+1} g. \tag{B.3}$$

The left-hand side is continuous and increasing in x. Thus for the condition to hold at any x, it is necessary that it hold at $x = 1$, which one shows to be the case if and only if $\frac{g}{s} \geq \frac{4(r+\lambda)}{2r+3\lambda}$. Thus, if $\frac{g}{s} > \frac{4(r+\lambda)}{2r+3\lambda}$, there exists a neighborhood of $x = 1$ such that for all x in that neighborhood and all p_A along the given ray x, arm A is preferred to the safe arm given the opponent plays $(1, 0)$, and given that efficiency will prevail once beliefs reach the 45-degree line. Thus, as symmetric arguments apply for beliefs $p_B > p_A$, we have shown that if $\frac{g}{s} \leq \frac{4(r+\lambda)}{2r+3\lambda}$, players behave inefficiently with positive probability in all equilibria for all non-degenerate prior beliefs $p_{A,0} + p_{B,0} < 1$ not in

$[0, p_2^*]^2 \cup \{(p_A, p_B) : p_A = p_B\}$. Thus, we henceforth assume that $\frac{g}{s} > \frac{4(r+\lambda)}{2r+3\lambda}$. Direct calculation shows that $\tilde{u}(p_1^*) < \frac{r+2\lambda}{r+\lambda}(p_1^* - p_2^*)g + s \leq 2s - p_1^* g$, where the second inequality is verified for all $\frac{g}{s} \geq \frac{2(r+\lambda)}{r+2\lambda}$. Thus, (B.3) is definitely violated at all $x \leq \frac{p_1^*}{1-p_1^*} + \phi$ for some $\phi > 0$.

It remains to show that, in the case that (B.3) holds, arm A is preferred over arm B as well. We have that

$$B_A - B_B \geq^? c_A - c_B \iff p_A - (1-p_A)x\frac{r+\lambda}{r} + \frac{\lambda}{r}x^{1+\frac{r}{2\lambda}}\check{u}_0(p_A) \geq^? 0.$$

Using the fact that $\check{u}_0\left(\frac{x}{x+1}\right) = \frac{x^{-\frac{r}{2\lambda}}}{x+1}$, one immediately verifies that this condition binds at $p_A = \frac{x}{x+1}$. Moreover, we have that the derivative of the left-hand side of this condition with respect to p_A at $p_A = \frac{x}{x+1}$ (keeping x fixed) is given by $\frac{1}{2}(1+x) > 0$. Thus, by convexity of $\check{u}_0$, arm A is indeed preferred over arm B for all $p_A \geq \frac{x}{x+1}$ and all $x \geq \frac{p_2^*}{1-p_2^*}$ (assuming that efficiency will prevail once beliefs reach the 45-degree line).

Suppose $\frac{g}{s} > \frac{4(r+\lambda)}{2r+3\lambda}$. Suppose that $(p_{A,0}, p_{B,0})$ is such that (B.3) holds . In the following we shall construct an equilibrium such that, starting from these prior beliefs, players' behavior be efficient throughout with probability 1.

In this equilibrium, actions at beliefs $p_A \geq p_B$ are as follows: Both players play safe if $p_A \leq p_2^*$. If $p_A = p_B > p_2^*$, one player plays $(1,0)$ while the other plays $(0,1)$ until we reach (p_2^*, p_2^*). If $p_A > \max\{p_B, p_1^*\}$ and (B.3) holds, both players use arm A. If $p_A > \max\{p_B, p_1^*\}$ and (B.3) does not hold, both players mix over arm A and the safe arm close to the 45-degree line (if $x \geq \frac{p_1^*}{1-p_1^*}$) or the line $p_A = p_1^*$ (if $x < \frac{p_1^*}{1-p_1^*}$), respectively. Away from the 45-degree line or the line $p_A = p_1^*$ respectively, both players use arm A if $p_A > \max\{p_B, p_1^*\}$. . If $p_B < p_A \leq p_1^*$, both players play safe. At beliefs $p_A < p_B$, the symmetric actions prevail, with the roles of arm A and arm B reversed.

We shall now focus on the situation where (B.3) is violated, and first turn to the case $x < \frac{p_1^*}{1-p_1^*}$. Near $p_A = p_1^*$, players are mixing over the safe arm and arm A. As we have seen in Appendix A, players' payoffs in this region satisfy

$$u = s + \frac{r+\lambda}{\lambda}(g-s) + \frac{r}{\lambda}s(1-p_A)\ln\left(\frac{1-p_A}{p_A}\right) + \tilde{f}\left(\frac{p_B}{1-p_A}\right)(1-p_A).$$

By value matching at $(p_1^*, (1-p_1^*)x)$, we find that

$$\tilde{f}(x) = -\frac{r+\lambda}{\lambda}\frac{g-s}{1-p_1^*} - \frac{r}{\lambda}s\ln\left(\frac{1-p_1^*}{p_1^*}\right) =: \tilde{C},$$

i.e. $\tilde{f}$ is independent of x. (This is not a surprise, since players are never using arm B, and hence do not care about p_B.) Thus, players' payoffs are independent of p_B and their payoff function can be written as $u(p_A, p_B) = \hat{u}(p_A)$, with $\hat{u}$ given by

$$\hat{u}(p_A) = s + \frac{r+\lambda}{\lambda}(g-s)\frac{p_A - p_1^*}{1-p_1^*} + \frac{r}{\lambda}s(1-p_A)\left[\ln\left(\frac{1-p_A}{p_A}\right) - \ln\left(\frac{1-p_1^*}{p_1^*}\right)\right].$$

This is strictly convex in p_A, and $\hat{u}'(p_A)$ approaches 0 as $p_A \downarrow p_1^*$. Hence, $\hat{u}$ is strictly increasing in p_A, for $p_A > p_1^*$.

In this mixing region, both players devote proportion $\tilde{k}_A$ of their unit endowment flow to arm A, with the rest going to the safe arm. The proportion is given by $\tilde{k}_A(p_A) = \frac{\hat{u}(p_A)-s}{c_A(p_A)}$. As $\hat{u}$ is strictly increasing with $\hat{u}(p_1^*) = s$, and $c_A \downarrow 0$ as $p \uparrow p^m$, the belief $\hat{p}_A < p^m$ which is implicitly defined by $\tilde{k}_A(\hat{p}_A) = 1$ is well-defined and unique. Players' payoffs are s on $[0, p_1^*]$, are given by $u(p_A, p_B) = \hat{u}(p_A)$ on $[p_1^*, \hat{p}_A]$, and on $[\hat{p}_A, 1]$ are given by $u = p_A g + f_{20}(x)\check{u}_0(p_A)$, with $f_{20}(x)$ defined by value matching at $(\hat{p}_A, (1-\hat{p}_A)x)$, i.e.

$$\hat{p}_A g + \check{u}_0(\hat{p}_A) f_{20}(x) = \hat{u}(\hat{p}_A) = 2s - \hat{p}_A g \Rightarrow f_{20}(x)\check{u}_0(\hat{p}_A) = 2c_A(\hat{p}_A).$$

As $\hat{u}$ is independent of x (i.e. independent of p_B), $\tilde{k}_A$ is independent of x (i.e. independent of p_B) as well. It thus follows that $\hat{p}_A$, and hence $f_{20}(x)$, are independent of x, i.e. $f_{20}(x) = const =: \hat{C}$. Put differently, in the $(2,0)$-region, $u(p_A, p_B) = \hat{u}_{20}(p_A)$ with

$$\hat{u}_{20}(p_A) := p_A g + \hat{C}\check{u}_0(p_A).$$

As $\hat{p}_A < p^m$, $\hat{C}\check{u}_0(\hat{p}_A) = 2c_A(\hat{p}_A) > 0$ implies that $\hat{C} > 0$ and hence that $\hat{u}_{20}$ is strictly convex.

By construction, $B_A = c_A$ in the mixing region, and players are indifferent between arm A and the safe arm. We shall now show that the safe arm is preferred to Arm B. As $\hat{u}$ is independent of p_B, we can use $\frac{\partial u}{\partial p_B} = 0$ to re-write the PDE for the indifference region as

$$\lambda p_A(1-p_A)\frac{\partial u}{\partial p_A} = (r+\lambda)p_A g - rs - \lambda p_A u.$$

Plugging this into the defining equation for B_B, we find that

$$B_B = p_A x g + \frac{\lambda}{r}x(g-u) - xs \leq^? s - (1-p_A)xg = c_B \iff \frac{\lambda}{r}[xg-u] + xg - (1+x)s \leq^? 0.$$

As $u \geq s$, for this it is sufficient that $x \leq \frac{(\lambda+r)s}{(\lambda+r)g-rs}$. It is straightforward to show that $\frac{(\lambda+r)s}{(\lambda+r)g-rs} \geq \frac{p_1^*}{1-p_1^*}$ if and only if $p_1^* \leq \frac{1}{2}$, i.e. if and only if $\frac{g}{s} \geq \frac{2r+\lambda}{r+\lambda}$, which holds by our assumption that $\frac{g}{s} > \frac{4(r+\lambda)}{2r+3\lambda}$. Thus, we can conclude that a deviation to arm B would be unprofitable in this mixing region.

Next, we turn to the $(2,0)$-region. As f_{20} is constant in x, and hence $\frac{\partial u}{\partial p_B} = 0$, we have that

$$B_A - B_B = \frac{\lambda}{r}(p_A - (1-p_A)x)(g-u) - \frac{\lambda}{r}p_A(1-p_A)(1+x)\frac{\partial u}{\partial p_A}.$$

By the PDE for the $(2,0)$-Region, we have that

$$2\lambda p_A(1-p_A)\frac{\partial u}{\partial p_A} = (r+2\lambda)p_A g - (r+2\lambda p_A)u$$

or

$$2\lambda p_A(1-p_A)\hat{u}_{20}'(p_A) = (r+2\lambda)p_A g - (r+2\lambda p_A)\hat{u}_{20}(p_A).$$

By convexity, for $\hat{u}_{20}$ to be increasing in the $(2,0)$-region, it is sufficient for it to be increasing at $p_A = \hat{p}_A$, which we shall verify by showing smooth pasting at $\hat{p}_A$, i.e. that $\hat{u}'(\hat{p}_A-) = \hat{u}_{20}'(\hat{p}_A+)$,

where $\hat{u}'(\hat{p}_A-)$ denotes the $\lim_{p_A\uparrow\hat{p}_A}\hat{u}'(p_A)$ and $\hat{u}'(\hat{p}_A+)=\lim_{p_A\downarrow\hat{p}_A}\hat{u}'(p_A)$ along the given ray x. Indeed, by the PDE for the indifference region,

$$\lambda\hat{p}_A(1-\hat{p}_A)\hat{u}'(\hat{p}_A-)=(r+\lambda)\hat{p}_Ag-rs-\lambda\hat{p}_A\hat{u}(\hat{p}_A),$$

whereas, by the PDE for the $(2,0)$-region,

$$\lambda\hat{p}_A(1-\hat{p}_A)\hat{u}'_{20}(\hat{p}_A+)=\left(\frac{r}{2}+\lambda\right)\hat{p}_Ag-\left(\frac{r}{2}+\lambda\hat{p}_A\right)\hat{u}_{20}(\hat{p}_A).$$

By plugging in $\hat{u}(\hat{p}_A)=\hat{u}_{20}(\hat{p}_A)=2s-\hat{p}_Ag$, one shows smooth pasting, i.e. $\hat{u}'(\hat{p}_A-)=\hat{u}_{20}(\hat{p}_A+)$. Yet, by the strict convexity of $\hat{u}$, we have that $\hat{u}'_{20}(\hat{p}_A)=\hat{u}'(\hat{p}_A)>\hat{u}'(p_1^*)=0$ since $\hat{p}_A>p_1^*$. We can thus conclude that, on $(p_1^*,1)$, players' payoffs are strictly increasing in p_A.

Plugging in the PDE gives us

$$\begin{aligned}B_A-B_B=-\frac{\lambda}{r}x(g-u)+\tfrac{1}{2}(1+x)(u-p_Ag)\geq^?xg-p_A(1+x)g=c_A-c_B\\ \iff\frac{\lambda}{r}\frac{x}{x+1}u-\frac{\lambda+r}{r}\frac{x}{x+1}g+\tfrac{1}{2}(u+p_Ag)\geq^?0.\end{aligned}$$

By monotonicity of u, we have that $u\geq 2s-\hat{p}_Ag\geq 2s-p_Ag$ (since $p_A\geq\hat{p}_A$). Thus a sufficient condition is given by $\frac{\lambda}{r}\frac{x}{x+1}u-\frac{\lambda+r}{r}\frac{x}{x+1}g+s\geq^?0$. As $u\geq s$, for this in turn it is sufficient that $\frac{x}{x+1}\leq p_1^*\iff x\leq\frac{p_1^*}{1-p_1^*}$. Thus, we can conclude that a deviation to arm B would be unprofitable in this region as well.

It remains to show that a deviation to the safe arm is not profitable either. Again using the PDE, we find that

$$B_A=\frac{u-p_Ag}{2}\geq^?s-p_Ag=c_A\iff\tfrac{1}{2}(u+p_Ag)\geq^?s.$$

As $u\geq 2s-p_Ag$, this holds, and we can conclude that a deviation to Safe would be unprofitable in this region as well.

We now turn to the case $x>\frac{p_1^*}{1-p_1^*}$, and again first consider the mixing region. By value matching at $(\frac{x}{x+1},\frac{x}{x+1})$, we find that

$$\tilde{f}(x)=-(1+x)s-\frac{r+\lambda}{\lambda}(1+x)(g-s)+\frac{r}{\lambda}s\ln(x)+\frac{r+2\lambda}{r+\lambda}xg+C^*(x+1)u_0\left(\frac{2x}{x+1}\right)$$

in the players' payoff function

$$u=s+\frac{r+\lambda}{\lambda}(g-s)+\frac{r}{\lambda}s(1-p_A)\ln\left(\frac{1-p_A}{p_A}\right)+\tilde{f}\left(\frac{p_B}{1-p_A}\right)(1-p_A).$$

It thus follows that

$$\tilde{f}'(x)=\frac{\partial u}{\partial p_B}=-s-\frac{r+\lambda}{\lambda}(g-s)+\frac{r+2\lambda}{r+\lambda}g+\frac{rs}{\lambda x}-\frac{1+x}{1-x}\left(1+\frac{r}{\lambda x}\right)C^*u_0\left(\frac{2x}{x+1}\right).$$

In this mixing region, players devote $\tilde{k}_A$ of their unit endowment flow to arm A, with the rest going to the safe arm. The proportion is given by $\tilde{k}_A(p_A)=\frac{u_{Ind.}(p_A)-s}{c_A}$, where $u_{Ind.}(p_A)$ denotes the

players' payoff in the indifference region along the fixed given ray x under consideration. We define as $\hat{p}_A(x)$ the belief at which $\tilde{k}_A = 1$. As $u_{Ind.}\left(\frac{x}{x+1}\right) = \tilde{u}\left(\frac{x}{x+1}\right)$, $\hat{p}_A(x) > \frac{x}{x+1}$ by our assumption that Condition (B.3) is violated. I will now show that $\hat{p}_A(x)$ (defined by $u(\hat{p}_A(x)) + \hat{p}_A(x)g = 2s$) is well-defined and unique. In order to do so, we show that $p_A g + u_{Ind.}(p_A)$ is strictly increasing in $p_A \in (\frac{x}{x+1}, 1)$. First, we note that

$$\frac{\partial u_{Ind.}}{\partial p_A}|_{xfixed} = -\frac{r}{\lambda}s\left[\ln\left(\frac{1-p_A}{p_A}\right) + \frac{1}{p_A}\right] - \tilde{f}(x),$$

and hence that $u_{Ind.}$ is strictly convex in p_A for fixed given x. It is thus sufficient to show that $g + \frac{\partial u_{Ind.}}{\partial p_A}|_{xfixed} \geq 0$ at $p_A = \frac{x}{x+1}$, which, by using $C^* u_0\left(\frac{2x}{x+1}\right) = \tilde{u}\left(\frac{x}{x+1}\right) - \frac{r+2\lambda}{r+\lambda}\frac{x}{x+1}g$, one shows to be equivalent to

$$g\left[\frac{\lambda}{1+x} + r + \lambda\right] - rs\frac{x+1}{x} \geq^? \lambda\tilde{u}\left(\frac{x}{x+1}\right).$$

As (B.3) is violated, we have that $\tilde{u}\left(\frac{x}{x+1}\right) < 2s - \frac{x}{x+1}g$. Using this, one shows that for the above condition it is sufficient that $x \geq \frac{p_1^*}{1-p_1^*}$. Thus we can conclude that $p_A \mapsto p_A g + u_{Ind.}(p_A)$ ($\left[\frac{x}{x+1}, 1\right] \to \mathbb{R}$) is strictly increasing on $\left(\frac{x}{x+1}, 1\right)$. Next we show that this function reaches the level $2s$ at a belief $\hat{p}_A(x) < p^m$. Suppose to the contrary that there does not exist such a belief $\hat{p}_A(x) < p^m$. Then, by continuity of $u_{Ind.}$, the Mean Value Theorem implies that $p^m g + u_{Ind.}(p^m) \leq 2s \iff u_{Ind.}(p^m) \leq s$. As $\lim_{p\downarrow 0} u_{Ind.}(p) = +\infty$ and $\lim_{p\uparrow 1} u_{Ind.}(p) = g + \frac{r}{\lambda}(g-s) > u_{Ind.}\left(\frac{x}{x+1}\right) = \tilde{u}\left(\frac{x}{x+1}\right)$ (where the equality follows from value matching at $\left(\frac{x}{x+1}, \frac{x}{x+1}\right)$), $u_{Ind.}$ assumes a (unique) interior minimum at some $\tilde{p}_A \in (0,1)$. As $u_{Ind.}$ is strictly convex, this minimum is characterized by $\frac{\partial u_{Ind.}}{\partial p_A} = 0$, which is equivalent to $\tilde{f}(x) = -\frac{rs}{\lambda}\left[\ln\left(\frac{1-\tilde{p}_A}{\tilde{p}_A}\right) + \frac{1}{\tilde{p}_A}\right]$. Plugging this into the expression for $u_{Ind.}$ shows us that $u_{Ind.}(\tilde{p}_A) = s + \frac{r+\lambda}{\lambda}(g-s) - \frac{rs}{\lambda}\frac{1-\tilde{p}_A}{\tilde{p}_A} >^? s \iff \tilde{p}_A >^? p_1^*$. Now if $\tilde{p}_A > p_1^*$, we can conclude that $u_{Ind.}(p_A) > s$ for all $p_A \geq \frac{x}{x+1}$, and hence that $\hat{p}_A(x) < p^m$. Suppose, however, that $\tilde{p}_A \leq p_1^* \leq \frac{x}{x+1}$. We note that Condition (B.3) being violated implies that $\frac{x}{x+1} < p^m$, as, in the case that $p^m \leq \frac{1}{2}$, we have that $\tilde{u}(p^m) \geq \frac{r+2\lambda}{r+\lambda}s > s = 2s - p^m g$. Thus, we have that $u_{Ind.}(p^m) > u_{Ind.}\left(\frac{x}{x+1}\right) = \tilde{u}\left(\frac{x}{x+1}\right) > s$, where the last inequality follows from the fact that $\frac{x}{x+1} \geq p_1^* > p_2^*$. We have thus shown that there exists a unique $\hat{p}_A(x)$ such that $u_{Ind.}(\hat{p}_A(x)) + \hat{p}_A(x)g = 2s$, i.e. such that $\tilde{k}_A(\hat{p}_A(x)) = 1$. [We have also incidentally shown that $\tilde{k}_A > 0$, as $u_{Ind.} > s$ throughout the mixing region.]

Turning to the $(2,0)$-region, value matching at $(\hat{p}_A(x), (1-\hat{p}_A(x))x)$ gives us

$$\hat{p}_A(x)g + f_{20}(x)\check{u}_0(\hat{p}_A(x)) = 2s - \hat{p}_A g \Rightarrow f_{20}(x) = 2\frac{s - \hat{p}_A(x)g}{\check{u}_0(\hat{p}_A(x))}$$

in the players' payoff function

$$u = p_A g + f_{20}(x)\check{u}_0(p_A).$$

In the mixing region, $B_A = c_A$ by construction, and players are indifferent between arm A and Safe. It remains to show that they do not want to deviate to arm B either. We find that $B_B \leq c_B$ for all p_A along a given ray x if and only if

$$g(r+\lambda)x - u\lambda x - r(1+x)s - \lambda(1-p_A)x(1-x)\tilde{f}'(x) \leq^? 0.$$

Using $C^*u_0\left(\frac{2x}{x+1}\right) = \tilde{u}\left(\frac{x}{x+1}\right) - \frac{r+2\lambda}{r+\lambda}\frac{x}{x+1}g$ in the expression for $\frac{\partial u}{\partial p_B} = \tilde{f}'(x)$ derived above, we get

$$\tilde{f}'(x) = -s - \frac{r+\lambda}{\lambda}(g-s) + \frac{rs}{\lambda x} + \frac{r+2\lambda}{\lambda(1-x)}g - \frac{(r+\lambda x)(1+x)}{\lambda x(1-x)}\tilde{u}\left(\frac{x}{x+1}\right).$$

Plugging this into the above condition, one finds that $rB_B \leq rc_B$ for all p_A along a fixed ray x if and only if

$$gx\left[r+\lambda p_A - (r+\lambda)(1-p_A)x\right] - rs\left[2 - p_A + x\left(1 - x(1-p_A)\right)\right] - u\lambda x \\ + (r+\lambda x)(1-p_A)(1+x)\tilde{u}\left(\frac{x}{x+1}\right) \leq^? 0 \quad \text{(B.4)}$$

for all p_A along a fixed ray x. One shows that at $p_A = \frac{x}{x+1}$, Condition (B.4) is equivalent to $\tilde{u}\left(\frac{x}{x+1}\right) \leq 2s - \frac{x}{x+1}g$, which holds by our assumption that (B.3) is violated. As $\frac{\partial u_{Ind.}}{\partial p_A} = -\frac{r}{\lambda}s\left[\ln\left(\frac{1-p_A}{p_A}\right) + \frac{1}{p_A}\right] - \tilde{f}(x)$, we note that the left-hand side of Condition (B.4) is strictly concave in p_A. The derivative of the left-hand side of Condition (B.4) w.r.t. p_A (for fixed x) works out as

$$xg[\lambda + (r+\lambda)x] + rs(1-x^2) - \lambda x\frac{\partial u}{\partial p_A}|_{x fixed} - (r+\lambda x)(1+x)\tilde{u}\left(\frac{x}{x+1}\right).$$

One verifies that this is greater than 0 at $p_A = \frac{x}{x+1}$ if and only if $\tilde{u}\left(\frac{x}{x+1}\right) \leq 2s - \frac{x}{x+1}g$, which is the case as (B.3) is violated. By concavity, the maximum on $[0,1]$ of the left-hand side of (B.4) is thus either reached at the boundary point $p_A = 1$ or where the derivative of the left-hand side of (B.4) is 0. As the left-hand side of (B.4) tends to $-rs$, so that (B.4) holds as $p_A \uparrow 1$, it is enough to check (B.4) at the point where the derivative of the left-hand side is 0. Setting the derivative equal to 0, and plugging the expression into (B.4), one finds that the left-hand side of (B.4) at the point where its derivative equals 0 is given by $-1 - x + \frac{x}{p_A}$, which is less than 0 if and only if $p_A \geq \frac{x}{x+1}$, which is verified. We can thus conclude that a deviation to arm B would not be profitable in this region.

Next, we are turning our attention to the $(2,0)$-region. Using the PDE for the $(2,0)$-region,

$$\lambda p_A p_B \frac{\partial u}{\partial p_B} = \lambda p_A(1-p_A)\frac{\partial u}{\partial p_A} + \left(\frac{r}{2} + \lambda p_A\right)u - \left(\frac{r}{2} + \lambda\right)p_A g,$$

one finds that $rB_A \geq^? rc_A$ if and only if

$$p_A g + u \geq^? 2s \iff p_A g + (s - \hat{p}_A(x)g)\frac{\check{u}_0(p_A)}{\check{u}_0(\hat{p}_A(x))} \geq^? s$$

for all p_A along a given ray x under consideration. By strict convexity of $\check{u}_0$, the left-hand side is strictly convex in p_A, as $\hat{p}_A(x) < p^m$. The condition binds at $p_A = \hat{p}_A(x)$. By convexity of the left-hand side, a sufficient condition for $B_A \geq c_A$ for all p_A along the ray x under consideration is for the derivative of the left-hand side w.r.t. p_A at $p_A = \hat{p}_A(x)$ to be non-negative, i.e.

$$g - (s - \hat{p}_A(x)g)\frac{\frac{r}{2\lambda} + \hat{p}_A(x)}{\hat{p}_A(x)(1-\hat{p}_A(x))} \geq^? 0,$$

which one shows to be equivalent to $\hat{p}_A(x) \geq p_2^*$. We can thus conclude that a deviation to Safe is unprofitable in the $(2,0)$-region.

In order to argue that a deviation to arm B is also unprofitable in the $(2,0)$-region, we shall first show that there is smooth pasting at $(\hat{p}_A(x), (1-\hat{p}_A(x))x)$, and hence that B_B is continuous there. To do so, we first show that $\frac{\partial u_{20}}{\partial p_B} =^? \frac{\partial u_{Ind.}}{\partial p_B}$ at $(\hat{p}_A(x), (1-\hat{p}_A(x))x)$, which is equivalent to

$$\tilde{f}'(x) =^? f'_{20}(x)\frac{\check{u}_0(\hat{p}_A(x))}{1-\hat{p}_A(x)}. \tag{B.5}$$

Direct computation shows that

$$f'_{20}(x)\frac{\check{u}_0(\hat{p}_A(x))}{1-\hat{p}_A(x)} = 2\frac{\hat{p}'_A(x)}{1-\hat{p}_A(x)}\left[\frac{\frac{r}{2\lambda}+\hat{p}_A(x)}{\hat{p}_A(x)(1-\hat{p}_A(x))}(s-\hat{p}_A(x)g)-g\right].$$

We know that $\hat{p}_A(x)$ is implicitly defined via the function $F : \left(\frac{p_1^*}{1-p_1^*},1\right)\times\left(\frac{x}{x+1},1\right)\to\mathbb{R}$,

$$F(x,\hat{p}_A) := s+\frac{r+\lambda}{\lambda}(g-s)+\frac{rs}{\lambda}(1-\hat{p}_A)\ln\left(\frac{1-\hat{p}_A}{\hat{p}_A}\right)+(1-\hat{p}_A)\tilde{f}(x)-2s+\hat{p}_A g = 0,$$

by

$$F(x,\hat{p}_A) = 0.$$

Since, as we have already shown *supra*, $F_{\hat{p}_A} > 0$, we can apply the Implicit Function Theorem to find

$$\hat{p}'_A(x) = -\frac{F_x}{F_{\hat{p}_A}} = \frac{(1-\hat{p}_A(x))\tilde{f}'(x)}{\frac{rs}{\lambda}\left[\ln\left(\frac{1-\hat{p}_A(x)}{\hat{p}_A(x)}\right)+\frac{1}{\hat{p}_A(x)}\right]+\tilde{f}(x)-g}.$$

Plugging this into the right-hand side of (B.5), and using that $F(x,\hat{p}_A(x)) = 0 \iff \frac{rs}{\lambda}\ln\left(\frac{1-\hat{p}_A(x)}{\hat{p}_A(x)}\right) = -\tilde{f}(x)-\frac{s+\frac{r+\lambda}{\lambda}(g-s)}{1-\hat{p}_A(x)}+\frac{2s}{1-\hat{p}_A(x)}-\frac{\hat{p}_A(x)}{1-\hat{p}_A(x)}g$, one finds that (B.5) indeed holds. We can thus conclude that $\frac{\partial u_{20}}{\partial p_B} = \frac{\partial u_{Ind.}}{\partial p_B} = \tilde{f}'(x)$ at $(\hat{p}_A(x), (1-\hat{p}_A(x))x)$.
Moreover, by the PDE for the indifference region approaching $(\hat{p}_A(x), (1-\hat{p}_A(x))x)$, we have that (I am suppressing the dependence of $\hat{p}_A$ on x here)

$$\lambda\hat{p}_A(1-\hat{p}_A)\frac{\partial u_{Ind.}}{\partial p_A} = \lambda\hat{p}_A(1-\hat{p}_A)x\tilde{f}'(x)-\lambda\hat{p}_A(2s-\hat{p}_A g)+(r+\lambda)\hat{p}_A g-rs.$$

By the same token, the PDE for the $(2,0)$-region approaching $(\hat{p}_A(x), (1-\hat{p}_A(x))x)$ gives us

$$\lambda\hat{p}_A(1-\hat{p}_A)\frac{\partial u_{20}}{\partial p_A} = \lambda\hat{p}_A(1-\hat{p}_A)x\tilde{f}'(x)-\left(\frac{r}{2}+\lambda\hat{p}_A\right)(2s-\hat{p}_A g)+\left(\frac{r}{2}+\lambda\right)\hat{p}_A g.$$

It is immediate to verify that this implies $\frac{\partial u_{Ind.}}{\partial p_A} = \frac{\partial u_{20}}{\partial p_A}$ at $(\hat{p}_A(x), (1-\hat{p}_A(x))x)$, so that we indeed have smooth pasting at $(\hat{p}_A(x), (1-\hat{p}_A(x))x)$.

We want to show that in the $(2,0)$-region $r(B_A-B_B) \geq^? r(c_A-c_B)$ for all $p_A \geq \hat{p}_A(x)$ along any given ray $x \in \left(\frac{p_1^*}{1-p_1^*},1\right)$. Fix such a ray x. We have that

$$rB_B = \lambda(1-p_A)xg-\frac{\check{u}_0(p_A)}{\check{u}_0(\hat{p}_A(x))}\left[\left(\lambda+\frac{r}{2}\right)f_{20}(x)+\lambda(1-x)f'_{20}(x)\right].$$

Let $\xi(x)$ be defined as the limit of B_B as $p_A \downarrow \hat{p}_A(x)$ along the ray x. Then we can write B_B as

$$rB_B = \lambda(1-p_A)xg - \frac{\check{u}_0(p_A)}{\check{u}_0(\hat{p}_A(x))}\left[\lambda(1-\hat{p}_A(x))xg - r\xi(x)\right].$$

By smooth pasting at $(\hat{p}_A(x), (1-\hat{p}_A(x))x)$, B_B is continuous there, which implies that $\xi(x) \leq c_B((1-\hat{p}_A(x))x)$. Using furthermore that $B_A = (s-\hat{p}_A(x)g)\frac{\check{u}_0(p_A)}{\check{u}_0(\hat{p}_A(x))}$, we find that $r(B_A - B_B) \geq^? r(c_A - c_B)$ if and only if

$$\frac{\check{u}_0(p_A)}{\check{u}_0(\hat{p}_A(x))}\left[r(s-\hat{p}_A(x)g) + \lambda(1-\hat{p}_A(x))xg - r\xi(x)\right] - \lambda(1-p_A)xg + rp_Ag - r(1-p_A)xg \geq^? 0.$$

As $r\xi(x) \leq rc_B((1-\hat{p}_A(x))x) = r(s-(1-\hat{p}_A(x))xg)$, a sufficient condition for $r(B_A - B_B) \geq^? r(c_A - c_B)$ is given by

$$\frac{\check{u}_0(p_A)}{\check{u}_0(\hat{p}_A(x))}\left[(r+\lambda)(1-\hat{p}_A(x))x - r\hat{p}_A(x)\right] - (r+\lambda)(1-p_A)x + rp_A \geq^? 0 \tag{B.6}$$

for all $p_A \geq \hat{p}_A(x)$ along the given ray x. It is immediate to verify that this condition binds at $p_A = \hat{p}_A(x)$. Depending on the sign of $(r+\lambda)(1-\hat{p}_A(x))x - r\hat{p}_A(x)$, (B.6) is either concave or convex in p_A (for a fixed x). Thus, if we can show that the derivative of the left-hand side of (B.6) with respect to p_A,

$$(r+\lambda)x + r - \frac{\frac{r}{2\lambda}+p_A}{p_A(1-p_A)}\frac{\check{u}_0(p_A)}{\check{u}_0(\hat{p}_A(x))}\left[(r+\lambda)(1-\hat{p}_A(x))x - r\hat{p}_A(x)\right],$$

is greater than, or equal to, 0, at both $p_A = \hat{p}_A(x)$ and $p_A = 1$, we are done. As $\check{u}_0(1) = 0$, this is immediate at $p_A = 1$. At $p_A = \hat{p}_A(x)$, this is equivalent to

$$\hat{p}_A(x) \geq^? \frac{(r+\lambda)x}{(r+\lambda)x + r + 2\lambda}.$$

As $\hat{p}_A(x) > \frac{x}{x+1}$, for this it is sufficient that $\frac{x}{x+1} \geq^? \frac{(r+\lambda)x}{(r+\lambda)x+r+2\lambda}$, which is equivalent to $\lambda \geq^! 0$. We have thus shown that $B_A - B_B \geq^! c_A - c_B$ for all $p_A \geq \hat{p}_A(x)$ along a given ray x, and hence that a deviation to arm B is unprofitable in the $(2,0)$-region as well.

We now turn to the case $x = \frac{p_1^*}{1-p_1^*}$, $\frac{g}{s} > \frac{4(r+\lambda)}{2r+3\lambda}$ and $p_A > p_B$. For this case, we shall first solve the constrained problem in which players do not have access to arm B. Since $\left(\dot{\frac{p_B}{1-p_A}}\right) = -K_B\frac{\lambda p_B(1-p_A-p_B)}{(1-p_A)^2}$, this implies that x cannot be altered in this constrained problem. Then, in a second step, I will argue that any use of arm B would lead to an instantaneous downward jump in utility , by virtue of the decrease in $\frac{p_{B,t}}{1-p_{A,t}}$, so that the solution to our constrained problem coincides with the solution to the original problem for $\frac{p_{B,0}}{1-p_{A,0}} = \frac{p_1^*}{1-p_1^*}$. In fact, the solution will coincide with the limit to our solution for the case $x > \frac{p_1^*}{1-p_1^*}$ as $x \downarrow \frac{p_1^*}{1-p_1^*}$.

As we have seen that (B.3) is always violated at $x = \frac{p_1^*}{1-p_1^*}$, players mix over the safe arm and arm A close to (p_1^*, p_1^*). Thus, their payoff function is given by

$$\tilde{\hat{u}}(p_A) = s + \frac{r+\lambda}{\lambda}(g-s) + \frac{rs}{\lambda}(1-p_A)\ln\left(\frac{1-p_A}{p_A}\right) + \check{C}(1-p_A).$$

Value matching at $p_A = p_1^*$ gives us

$$\check{C} = -\frac{rs}{\lambda}\ln\left(\frac{1-p_1^*}{p_1^*}\right) + \frac{1}{1-p_1^*}\left[\tilde{u}(p_1^*) - s - \frac{r+\lambda}{\lambda}(g-s)\right],$$

so that

$$\tilde{\hat{u}}(p_A) = \hat{u}(p_A) + \frac{1-p_A}{1-p_1^*}\left(\tilde{u}(p_1^*) - s\right).$$

As $\tilde{u}(p_1^*) > s$, it follows that $\tilde{\hat{u}}(p_A) > \hat{u}(p_A) \geq s$ for all $p_A \in [p_1^*, 1)$, so that $\tilde{k}_A(p_A) := \frac{\tilde{\hat{u}}(p_A)-s}{s-p_A g} > 0$ for all $p_A < p^m$. We will now show that the belief $\tilde{\hat{p}}_A$, implicitly defined by $\tilde{k}_A(\tilde{\hat{p}}_A) = 1$, is well-defined and unique. In order to do so, as before, we show the monotonicity of $p_A \mapsto \zeta(p_A) := p_A g + \tilde{\hat{u}}(p_A)$. We can now equivalently define $\tilde{\hat{p}}_A$ via $\zeta(\tilde{\hat{p}}_A) = 2s$. We have that for $p_A \geq p_1^*$

$$\zeta'(p_A) = g + \tilde{\hat{u}}'(p_A) \geq g - \frac{\tilde{u}(p_1^*) - s}{1-p_1^*} >^? 0 \iff (1-p_1^*)g + s >^? \tilde{u}(p_1^*),$$

where the first inequality follows by the convexity of $\hat{u}$ and the fact that $\hat{u}'(p_1^*) = 0$. As $\tilde{u}(p_1^*) < s + \frac{r+2\lambda}{r+\lambda}(p_1^* - p_2^*)g$, for this it is sufficient that

$$1 - p_1^* \geq^? \frac{r+2\lambda}{r+\lambda}(p_1^* - p_2^*),$$

which one shows to be verified for $\frac{g}{s} \geq \frac{4(r+\lambda)}{2r+3\lambda}$. Thus, ζ is strictly increasing, and $\tilde{\hat{p}}_A$ is well-defined and unique. Suppose that $\tilde{\hat{p}}_A \geq p^m$. Then, $\zeta(p^m) = s + \tilde{\hat{u}}(p^m) \leq 2s$, a contradiction to $\tilde{\hat{u}}(p_A) > s$ for all $p_A \in [p_1^*, 1)$. Thus, we can conclude that $\tilde{\hat{p}}_A < p^m$.

As $B_A = c_A$ by construction in the mixing region, all that remains for us to show is that $B_A \geq c_A$ for $p_A \geq \tilde{\hat{p}}_A$. The players' payoff function in this region is given by

$$\tilde{\hat{u}}_{20}(p_A) = p_A g + \check{C}_{20}\check{u}_0(p_A),$$

where value matching at $\tilde{\hat{p}}_A$ gives us

$$\check{C}_{20} = 2\frac{s - \tilde{\hat{p}}_A g}{\check{u}_0(\tilde{\hat{p}}_A)}.$$

As $\tilde{\hat{p}}_A < p^m$, $\check{C}_{20} > 0$, and $\tilde{\hat{u}}_{20}$ is strictly convex. We find that $B_A = \frac{1}{2}\check{C}_{20}\check{u}_0(p_A)$, and hence that

$$B_A \geq^? c_A \iff p_A g + \tilde{\hat{u}}_{20}(p_A) \geq^? 2s.$$

As this condition is convex and binds at $p_A = \tilde{\hat{p}}_A$, for this it is sufficient that $g + \tilde{\hat{u}}'_{20}(\tilde{\hat{p}}_A) \geq^? 0$, i.e. that

$$2g - 2\left(s - \tilde{\hat{p}}_A g\right)\frac{\frac{r}{2\lambda} + \tilde{\hat{p}}_A}{\tilde{\hat{p}}_A(1-\tilde{\hat{p}}_A)} \geq^? 0,$$

which one shows to be verified for $\tilde{\hat{p}}_A \geq p_2^*$.

We have thus verified the solution to our constrained problem. We will now show that players will never unilaterally want to deviate to arm B if $x = \frac{p_1^*}{1-p_1^*}$ lest they experience an immediate downward jump in utility. We shall first consider the region close to $p_A = p_1^*$, where the constrained

solution for $x = \frac{p_1^*}{1-p_1^*}$ and the solution for $x \uparrow \frac{p_1^*}{1-p_1^*}$ both prescribe mixing between arm A and the safe arm. (We write u_+ and u_- for the payoff functions pertaining to the former and latter cases, respectively.) For every p_A in this region, we have that $\hat{u}(p_A) < \tilde{\hat{u}}(p_A)$, since $\tilde{C} < \check{C} \iff \tilde{u}(p_1^*) >^! s$. From this, it immediately follows that $\tilde{\hat{p}}_A < \hat{p}_A$. Next, we consider the region $\tilde{\hat{p}}_A \leq p_A < \hat{p}_A$. Here, $u_- < 2s - p_A g \leq u_+$. Turning to the region $\hat{p}_A \leq p_A < 1$, we find that $u_-(p_A) = p_A g + \hat{C}\check{u}_0(p_A) <^? p_A g + \check{C}_{20}\check{u}_0(p_A) = u_+(p_A) \iff \hat{C} <^? \check{C}_{20} \iff h(\hat{p}_A) <^? h(\tilde{\hat{p}}_A)$, where the function h is defined by $h(y) := \frac{s-yg}{\check{u}_0(y)}$ for $y \in (p_1^*, 1)$. Direct calculation shows that $h'(y) < 0$ if and only if $y > p_2^*$, which is verified. We can thus conclude that $u_- < u_+$ in this region as well, so that our constrained solution indeed coincides with the solution to our original problem. As $\check{C} = \lim_{x \downarrow \frac{p_1^*}{1-p_1^*}} \tilde{f}(x)$, we can furthermore conclude that our equilibrium payoff function is right-continuous at $x = \frac{p_1^*}{1-p_1^*}$.

The arguments for the beliefs $p_B > p_A$ are symmetric, with the roles of arm A and arm B reversed. Along the 45-degree line, deviations are ruled out by admissibility considerations. ∎

Acknowledgements

Earlier versions of this paper were circulated under the titles "Free-Riding and Delegation in Research Teams" and "Free-Riding And Delegation in Research Teams—A Three-Armed Bandit Model." I am grateful to two anonymous referees for very helpful comments and suggestions, to the Department of Economics at Yale University and the Cowles Foundation for Research in Economics for their hospitality, and to Sven Rady for advice and encouragement, as well as Dirk Bergemann, Ludwig Ensthaler, Johannes Hörner, Daniel Krähmer, Jo Thori Lind, Benny Moldovanu, Tymofiy Mylovanov, Frank Riedel, Larry Samuelson, Klaus Schmidt, Lones Smith, Eilon Solan, and seminar participants at Bielefeld, Penn State, Southern Methodist University, Yale University, and various conferences and workshops for helpful comments and suggestions.

Role of Funding Source

Financial support from the Deutsche Forschungsgemeinschaft through GRK 801 and SFB/TR 15 is gratefully acknowledged for 2008-09; for 2010, the author thanks the National Research Fund, Luxembourg, for its support of this project. These funding sources had no involvement in the actual writing or submission of this article.

References

Aghion, P., M. Dewatripont and J. Stein (2008): "Academic Freedom, Private-Sector Focus, and the Process of Innovation," *RAND Journal of Economics*, 39 (3), 617–635.

Bank, P. and H. Föllmer (2003): "American Options, Multi-armed Bandits, and Optimal Consumption Plans: A Unifying View," in: *Paris-Princeton Lectures on Mathematical Finance 2002*, ed. by R. A. Carmona et al. Springer-Verlag, Berlin and Heidelberg.

Bartlett, C. and A. Mohammed (1995): "3M: Profile of an Innovating Company," Harvard Business School Case Study 9-395-016.

Bellman, R. (1956): "A Problem in the Sequential Design of Experiments," *Sankhya: The Indian Journal of Statistics (1933–1960)*, Vol. 16, No. 3/4, 221–229.

Bergemann, D. and J. Välimäki (2008): "Bandit Problems," in: *The New Palgrave Dictionary of Economics*, 2nd edition. ed. by S. Durlauf and L. Blume, Basingstoke and New York: Palgrave Macmillan Ltd.

Bergin, J. and W.B. MacLeod (1993): "Continuous Time Repeated Games," *International Economic Review*, 34, 21–37.

Bolton, P. and C. Harris (1999): "Strategic Experimentation," *Econometrica*, 67, 349–374.

Bolton, P. and C. Harris (2000): "Strategic Experimentation: the Undiscounted Case," in: *Incentives, Organizations and Public Economics – Papers in Honour of Sir James Mirrlees*, ed. by P.J. Hammond and G.D. Myles. Oxford: Oxford University Press, 53–68.

Bonatti, A. and J. Hörner (2011): "Collaborating," *American Economic Review*, 101(2), 632–663.

Bradt, R., S. Johnson and S. Karlin (1956): "On Sequential Designs for Maximizing the Sum of n Observations," *The Annals of Mathematical Statistics*, 27, 1060–1074.

Camargo, B. (2007): "Good News and Bad News in Two-Armed Bandits," *Journal of Economic Theory*, 135, 558–566.

Chatterjee, K. and R. Evans (2004): "Rivals' Search for Buried Treasure: Competition and Duplication in R&D," *RAND Journal of Economics*, 35, 160–183.

Cohen, A. and E. Solan (2009): "Bandit Problems with Lévy Payoff Processes," working paper, University of Tel Aviv, archived at http://arxiv.org/abs/0906.0835v1.

Holmström, B. (1982): "Moral Hazard in Teams," *Bell Journal of Economics*, 13, 324–40.

Hörner, J., N. Klein and S. Rady (2013): "Strongly Symmetric Equilibria in Bandit Games," mimeo.

Keller, G. and S. Rady (2010): "Strategic Experimentation with Poisson Bandits," Theoretical Economics, 5, 275–311.

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

Klein, N. (2011): "Strategic Learning in Teams," mimeo, available at http://nicolasklein.com/data/documents/3ab052511.pdf.

Klein, N. (2012): "The Importance of Being Honest," mimeo, available at www.nicolasklein.com.

Klein, N. and S. Rady (2011): "Negatively Correlated Bandits," *Review of Economic Studies*, 78(2), 693–732.

Lacetera, N. (2008): "Different Missions and Commitment Power in R & D Organization: Theory and Evidence on Industry-University Alliances," *Organization Science*, published online before print, September, 17, 2008.

Lawler, A. (2003): "Last of the big-time spenders?," *Science*, 299, 330-333.

Manso, G. (2011): "Motivating Innovation," *Journal of Finance*, 66, 1823–1860.

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

Presman, E.L. (1990): "Poisson Version of the Two-Armed Bandit Problem with Discounting," *Theory of Probability and its Applications*, 35, 307–317.

Robbins, H. (1952): "Some Aspects of the Sequential Design of Experiments," *Bulletin of the American Mathematical Society*, 58, 527–535.

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

Rothschild, M. (1974): "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory*, 9, 185–202.