GOVERNANCE AND THE EFFICIENCY OF ECONOMIC SYSTEMS
GESY

Discussion Paper No. 387

Strategic Experimentation with Private Payoffs

Paul Heidhues*
Sven Rady**
Philipp Strack***

*European School of Management and Technology, Berlin
** University of Bonn
*** University of Bonn

September 2012

Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

Strategic Experimentation with Private Payoffs*

Paul Heidhues† Sven Rady‡ Philipp Strack‡

September 25, 2012

Abstract

We consider two players facing identical discrete-time bandit problems with a safe and a risky arm. In any period, the risky arm yields either a success or a failure, and the first success reveals the risky arm to dominate the safe one. When payoffs are public information, the ensuing free-rider problem is so severe that the equilibrium number of experiments is at most one plus the number of experiments that a single agent would perform. When payoffs are private information and players can communicate via cheap talk, the socially optimal symmetric experimentation profile can be supported as a perfect Bayesian equilibrium for sufficiently optimistic prior beliefs. These results generalize to more than two players whenever the success probability per period is not too high. In particular, this is the case when successes occur at the jump times of a Poisson process and the period length is sufficiently small.

JEL classification: C73, D83.

Keywords: Strategic Experimentation, Bayesian Learning, Cheap Talk, Two-Armed Bandit, Information Externality.

*Our thanks for helpful discussions and comments are owed to Simon Board, Patrick Bolton, Susanne Goldlücke, Nicolas Klein, Moritz Meyer-ter-Vehn, and Johannes Münster. We thank seminar participants at UCLA, UC Berkeley, Yale, the 2011 North American Winter Meeting of the Econometric Society in Denver and the April 2012 SFB/TR 15 Meeting in Mannheim. We thank the Economics Department at UC Berkeley and the Cowles Foundation for Research in Economics at Yale University for their hospitality. Financial support from the Deutsche Forschungsgemeinschaft through SFB/TR 15 is gratefully acknowledged.

†Lufthansa Chair in Competition and Regulation, ESMT European School of Management and Technology, Schlossplatz 1, D-10178 Berlin, Germany.

‡Department of Economics and Hausdorff Center for Mathematics, University of Bonn, Adenauerallee 24-42, D-53113 Bonn, Germany.

1 Introduction

In many real-life situations, economic agents face a trade-off between exploring new options and exploiting their knowledge about which option is likely to be best. A stylized model capturing this feature is a two-armed bandit problem in which a gambler repeatedly decides which of two different slot machines to play with the ultimate goal of maximizing his monetary reward. In the simplest version of this problem, one arm is safe, and the other risky. The expected payoff per round is known for the safe arm, whereas the consecutive payoffs of the risky arm are independent draws from some unknown distribution. When playing the risky arm, the agent learns more about this distribution—knowledge that is useful for future choices. Starting with Rothschild (1974), variants of this bandit problem have been studied in a wide variety of economic settings; see Bergemann and Välimäki (2008) for an overview, and the references cited below for specific applications.

Often, it is natural to presume that an agent can learn not only from her own exploration but also from the experiences of others. Experimental consumption is a case in point. As a stylized example, suppose there are two restaurant goers with common tastes. They choose between two items on a given menu and both know the quality of one regular item that has been on the menu for a long time. The other item is new on the menu; its quality critically depends on how well it is cooked. The restaurant's chef is believed to be either good or very good. A very good chef is able to sometimes prepare the item nearly perfectly, while a good chef is only able to prepare a dish of average quality. Each diner can now learn through three channels: he may experiment himself and try the new item, he may learn from observing what the other chooses, and finally he may ask the other customer about whether the item was well-prepared whenever she tried it.

To formalize these ideas, we consider a game in which two players face identical bandit problems with a safe and a risky arm. The payoff distribution of the risky arm is the same for both players, and can be either "good" or "bad". In each period, either type of arm produces a "success" or a "failure", with the latter being more likely when the arm is bad. While we are mainly interested in the problem of strategic experimentation with private payoffs, we begin by analyzing a benchmark case in which players' actions *and* payoffs are publicly observable, so that even absent any communication all players always share a common belief about the state of the world. This scenario has been studied extensively (Bolton and Harris 1999, 2000, Keller, Rady and Cripps 2005, Keller and Rady 2010). Focusing on continuous-time setups, the literature shows that, if the players condition their actions on their common belief only, i.e. if they use Markov perfect strategies, it is impossible to achieve the social optimum. Furthermore, if a single success on the risky arm fully reveals the good state of the world, then in any Markov perfect equilibrium players stop experimenting

once the common belief reaches the single-agent cut-off (Keller, Rady and Cripps 2005).

Maintaining the assumption of fully revealing successes, we introduce a discrete-time setup and considerably strengthen these existing results by showing that in *any* Nash equilibrium, players stop experimenting once the common belief falls beneath the single-agent cut-off. Furthermore, we prove that the total number of experiments performed in equilibrium differs from the single-agent optimum by at most one.

We then turn to the main focus of this paper: strategic experimentation under private information. Now, the players observe each other's behavior but not the realized payoffs. Instead, we allow players to communicate via cheap talk.[1] In such a context, players have three sources of information: their own signals, their observation of other players, and the cheap-talk messages. We begin by noting that cheap talk in this environment is very effective. In particular, we show that for every equilibrium with publicly observable payoffs, there exists an equilibrium with privately observable payoffs that yields the same distribution over experimentation paths, that is, sequences of experimentation choices and results.

The key intuition lies in the fact that truthful communication is easy to sustain. Following a success, a player is certain about the underlying state of the world and is willing to truthfully communicate her success. Furthermore, if a player believes that the other player is communicating truthfully, she believes that the state of the world is good with probability 1 upon hearing that her fellow player had a success. In this case, she is willing to play the risky arm forever without communicating her future payoffs. But then no player will want to wrongly announce a success because this will make it impossible to learn anything from one's fellow player in the future. Truthful revelation of payoffs, therefore, is always incentive compatible and any equilibrium outcome with observable payoffs can be replicated when payoffs are only privately observable.

More surprisingly, we also prove that if the initial belief is sufficiently optimistic, then the socially optimal symmetric strategy profile can be supported as a perfect Bayesian equilibrium.[2] To see the logic behind the equilibrium construction, suppose that the optimal symmetric solution of the social planner requires $\tau > 1$ periods of experimentation. Given an initial belief above the cut-off that

[1] It is very natural to let players communicate in a strategic environment in which there is only an information externality between them. A similar approach has been taken in collusion models with imperfect private monitoring—where communication however is illegal in contrast to the current setting; see e.g. Compte (1998) and Kandori and Matsushima (1998). Athey and Bagwell (2008) analyze collusive schemes in a framework with publicly observable pricing and output decisions and cheap-talk announcements of private cost shocks.

[2] Focussing on the socially optimal *symmetric* profile entails only a minor efficiency loss. In fact, we show that in the "unrestricted" social optimum there is at most one period in which a single player experiments.

a myopic agent would apply, both players are willing to experiment in the first period. At a later stage, player i may be tempted to pause while her fellow player j engages in costly experimentation. Following such a deviation to the safe arm by player i, however, player j believes that i had a prior success and, hence, j is willing to use the risky arm forever without communicating.[3] This implies that player i cannot learn anything from player j's subsequent actions, and thereby deters free-riding.

While these beliefs are consistent in the sense of sequential equilibrium, one may not find them fully convincing. This leads us to consider an ad-hoc refinement of *pessimistic beliefs*. Upon observing a deviation to the safe arm, pessimistic beliefs require a player—who is still uncertain as to whether the risky arm is good—to believe that the other player's experiments all failed. We prove that the socially optimal symmetric strategy profile can be supported as a perfect Bayesian equilibrium with pessimistic beliefs if the players' initial belief is sufficiently optimistic. To implement the socially optimal symmetric profile, we let both players experiment for as many rounds τ as the social planner would, and thereafter communicate whether they have had a success. Intuitively, we exploit the fact that a player observing only his own experiments learns at a slower rate than a social planner who observes both players' experiments. As players learn only from their own experiments prior to round τ, therefore, they remain above the myopic cut-off if initially they were sufficiently optimistic.

It is straightforward to embed our discrete-time model in a version of the continuous-time framework of Keller, Rady and Cripps (2005). The discrete-time model then corresponds to imposing an equally-spaced grid of times at which the players can choose actions and send messages. Taking the grid size to zero, we show that in our construction of perfect Bayesian equilibria with punishment via beliefs, the restriction to sufficiently optimistic prior beliefs can be dropped. Given any prior, therefore, a sufficiently small period length will ensure that the symmetric optimum can be achieved in an equilibrium with private payoffs and cheap-talk communication, in marked contrast to the case of public payoffs.

The proofs of our main results extend to an arbitrary number of players whenever the probability of observing a success in any given period is not too high. In particular, this is the case when we embed the model in continuous time as described and consider sufficiently short periods. Proceeding this way also ensures that the socially optimal *symmetric* strategy profile remains a good proxy for the asymmetric profile that an unrestricted social planner would like to implement.

Besides the papers on strategic experimentation already mentioned, our work is most closely related to Rosenberg, Solan and Vieille (2007). We follow these authors in studying a discrete-time experimentation game with bandits whose risky arm can be of two possible types, and with players who observe each other's

[3]This "punishment via beliefs" construction is similar in spirit to the one sustaining collusion in Blume and Heidhues (2006).

actions but not each other's payoffs. There are, however, two important differences: first, Rosenberg et al. assume the decision to stop experimenting to be irreversible, whereas we allow the players to freely switch from one arm to the other and back; second, we permit communication between the players. With irreversible stopping decisions, players cannot free-ride on an opponent's experimentation efforts, so if we assumed irreversible decisions in our framework, the socially optimal symmetric strategy profile could easily be supported as an equilibrium with truthful communication.[4] The ability to switch actions freely thus enriches the players' strategic possibilities considerably, and makes it more difficult to achieve efficiency.

There are a number of papers which introduce private information into a model of learning where, as in our setting, one type of risky "project" generates a perfectly informative signal at some random time while the other type never does.[5] Bergemann and Hege (1998, 2005) study the financing of a venture project in a dynamic agency model where the allocation of funds and the learning process are subject to moral hazard. Décamps and Mariotti (2004) analyze a duopoly model of irreversible investment with a learning externality and privately observed investment costs. Acemoglu, Bimpikis and Ozdaglar (2011) investigate the effect of patents on firms' irreversible decisions to experiment themselves or copy a successful rival; they allow for private information about a project's success probability, but assume observable actions and outcomes.[6] In a model with public actions, private payoffs and irreversible stopping decisions, Murto and Välimäki (2011) examine information aggregation through observational learning by a large number of players. Allowing for reversible experimentation choices, Bonatti and Hörner (2011) study a game in which the players' actions are private information and experimentation outcomes are public, which is precisely the opposite of what we assume here. Like Rosenberg et al., these authors all preclude communication between the players.

The remainder of the paper proceeds as follows. Section 2 sets up the model. Section 3 solves the social planner's problem. Section 4 studies strategic experimentation, first with public payoffs, then with private payoffs. Section 5 embeds

[4]In fact, a player who is meant to experiment under this strategy profile faces the same trade-off as the social planner: if she experiments, she bears the current cost of one experiment but learns the result of two experiments, while if she refrains from experimenting she obtains the outside payoff. Crucially, the players have no incentive to misrepresent their experimentation results when implementing this optimum: above the social planner's cut-off belief, both players experiment anyhow and benefit from each other's experimentation; below the cut-off, each player wants to cease experimentation even if the other experiments, and therefore has no incentive to misrepresent the fact that she is below the cut-off.

[5]With public information, this signal structure can be found in the models of R&D competition proposed by Malueg and Tsutsui (1997) and Besanko and Wu (2012).

[6]In their analysis of R&D races as preemption games with private information, Hopenhayn and Squintani (2011) instead let players' private information states increase stochastically over time according to a compound Poisson process.

the model in a continuous-time framework. Section 6 explains how our results generalize to more than two players. Section 7 concludes and discusses possible extensions.

2 The Model

There is an infinite number of periods $t = 0, 1, \ldots$ and there are two players who in each period choose between a safe and a risky action (or "arm"). Before making this choice, they can costlessly communicate with each other.

More precisely, at the beginning of period t, each player i chooses a cheap-talk message $m_i(t) \in [0, 1]$. Upon having observed the other player's cheap-talk message, each player i then chooses an action $k_i(t) \in \{R, S\}$. If $k_i(t) = S$, the player receives a safe payoff normalized to 0; if $k_i(t) = R$, the player receives a risky payoff $X_i(t)$ that is either low (X_L) or high (X_H), where $X_L < 0 < X_H$.

The distribution of the risky payoff depends on an unknown state of the world, which is either good ($\theta = 1$) or bad ($\theta = 0$). Conditional on the state of the world, payoffs are drawn independently across players and periods. In the good state of the world, the probability of the high payoff is $\mathbb{P}(X_H|\theta = 1) = \pi > 0$; in the bad state, it is $\mathbb{P}(X_H|\theta = 0) = 0$. Thus, a single draw of X_H proves that the state of the world is good. This makes our model a discrete-time analog of the model analyzed in Keller, Rady and Cripps (2005). We write E_θ for the conditional expectation $\mathbb{E}[X_i(t)|\theta]$ of the risky payoff in any given period, and assume that $E_0 < 0 < E_1$. We say that a player *experiments* if he chooses the risky action while still being uncertain about the true state of the world.

Our primary interest below is in analyzing the game in which actions are publicly observable but the realizations of the risky payoffs $X_i(t)$ are private information. When considering this game, we partition the set of private histories for each player into those after which he has to send a message and those after which he has to choose an action. Let

$$O = \{(R, X_L), (R, X_H), (S, 0)\}$$

be the set of possible combinations of a player's actions and payoffs that can occur in any period. Then, the set of *private message histories* of player i at time t is

$$H_{i,t}^m = \left(\underbrace{[0,1]^2}_{\text{messages}} \times \underbrace{O}_{\text{own action \& payoff}} \times \underbrace{\{R,S\}}_{\text{opponent's action}} \right)^t .$$

Similarly, the set of all *private action histories* of player i at time t is

$$H_{i,t}^a = \left(\underbrace{[0,1]^2}_{\text{messages}} \right)^{t+1} \times \left(\underbrace{O}_{\text{own action \& payoff}} \times \underbrace{\{R,S\}}_{\text{opponent's action}} \right)^t .$$

Finally, let $H_i^m = \bigcup_{t=0}^{\infty} H_{i,t}^m$ and $H_i^a = \bigcup_{t=0}^{\infty} H_{i,t}^a$.

A *pure strategy* is a mapping that assigns to each private message history $h_i^m \in H_i^m$ a message $m_i(h_i^m) \in [0,1]$ and to each private action history $h_i^a \in H_i^a$ an action $k_i(h_i^a) \in \{R, S\}$. Mixed strategies are defined in the usual way.

Given a probability $p_i(0) = p$ that player i initially assigns to the good state of the world, his expected payoff from a pure-strategy profile is

$$(1-\delta)\,\mathbb{E}_p\left[\sum_{t=0}^{\infty}\delta^t \mathbf{1}_{\{k_i(h_i^a(t))=R\}} X_i(t)\right],$$

where the factor $1-\delta$ serves to express the overall payoff in per-period units. Note that player j's strategy only enters through the expectation operator—there is just an informational externality at play here. We assume throughout that players start with a common non-degenerate prior: $p_1(0) = p_2(0) \in \,]0,1[\,$.

As a benchmark, we will also study the game with observable payoffs. The set of all *public message histories* at time t in this game is

$$H_t^m = \left(\underbrace{[0,1]^2}_{\text{messages}} \times \underbrace{O \times O}_{\text{actions \& payoffs}}\right)^t,$$

and the set of all *public action histories* at time t is

$$H_t^a = \left(\underbrace{[0,1]^2}_{\text{messages}}\right)^{t+1} \times \left(\underbrace{O \times O}_{\text{actions \& payoffs}}\right)^t.$$

With $H^m = \bigcup_{t=0}^{\infty} H_t^m$ and $H^a = \bigcup_{t=0}^{\infty} H_t^a$, a pure strategy of player i is now a mapping that assigns to each public message history $h^m \in H^m$ a message $m_i(h^m) \in [0,1]$ and to each public action history $h^a \in H^a$ an action $k_i(h^a) \in \{R, S\}$.[7]

Unless stated explicitly otherwise, our solution concept is *perfect Bayesian equilibrium*. Thus, a strategy profile is an equilibrium if there exists a belief system such that each player acts optimally after every history given her beliefs and the other player's strategy, and each player's beliefs are updated according to Bayes' rule whenever possible. Furthermore, we naturally require player i's beliefs about the state of the world to be independent of player j's actions and messages whenever every signal that player j observed, was observed by player i as well. In other words, we do not allow beliefs to change in response to observations which cannot contain new information.

[7] While communication obviously has no role in transmitting experimentation results when payoffs are observable, players could still use their messages to coordinate their continuation play.

Whether payoffs are public information or not, we call

$$H_t^e = (O \times O)^t$$

the set of all possible *experimentation paths* at time t, and we let $H^e = \bigcup_{t=0}^{\infty} H_t^e$. Using the canonical projections $H_{1,t}^a \times H_{2,t}^a \to H_t^e$ and $H_t^a \to H_t^e$, we associate an experimentation path with each pair of private action histories and with each public action history, respectively. We call two strategy profiles *path-equivalent* if they induce the same distribution over the set of all experimentation paths, H^e. Note that all path-equivalent strategy profiles give rise to the same payoff profile. What is more, this equivalence relation also applies in situations where one strategy profile is from the game with private payoffs, and the other from the game with public payoffs.

For future reference, we define

$$p^m = \frac{|E_0|}{|E_0| + E_1}\,.$$

This is the belief at which the expected current payoff from the risky option just equals zero, i.e. the safe payoff. A myopic player chooses the risky arm if and only if his posterior belief exceeds p^m. We therefore call p^m the *myopic cut-off belief.*

Given any probability $p \in [0,1]$ assigned to the good state of the world, the updated probability after n failed experiments is

$$B(n,p) = \frac{p(1-\pi)^n}{p(1-\pi)^n + 1 - p}\,.$$

In the planner's problem as well as in the game with public payoffs, we denote by $p(t)$ the public belief induced by the history up to time t; that is, $p(t) = 1$ if there was a success prior to period t, and $p(t) = B(n, p(0))$ if there were n failed experiments prior to period t.

3 The Planner's Problem

In this section, we discuss the problem of a social planner who chooses a strategy profile to maximize the average of the two players' objective functions. The optimum in this situation will serve as a benchmark for the case that individual players pursue their goals independently.

Let $k = (k_1, k_2)$ denote a pure strategy profile in the scenario with private payoffs. Then the players' expected average payoff, expressed in per-period units, is

$$u(p,k) = (1-\delta)\,\mathbb{E}_p\left[\tfrac{1}{2}\sum_{i=1}^{2}\sum_{t=0}^{\infty}\delta^t \mathbf{1}_{\{k_i(h_i^a(t))=R\}} X_i(t)\right],$$

where p denotes the probability that the planner initially assigns to the good state.

For the social planner, it can never be strictly beneficial to have players hide information from each other because she can always choose a strategy profile that ignores unwanted information. Hence, when discussing the planner's problem, we will focus on strategy profiles in which all players truthfully communicate their past payoffs via their cheap-talk messages. The planner's problem then becomes a Markovian decision problem with the posterior belief as the single state variable.

A single success in the past fully reveals that the state of the world is good. It is then a dominant choice for the planner to have both players take the risky action in all following periods. Using this fact and restricting attention to symmetric strategy profiles, we can think of the planner as choosing a natural number $\tau \geq 0$ such that both players experiment in periods $t \leq \tau - 1$ and, in case all these experiments were unsuccessful, no player experiments in periods $t \geq \tau$. For any such τ, expected average payoffs are

$$\begin{aligned} u(p,\tau) &= (1-\delta)\left\{(1-p)\sum_{t=0}^{\tau-1}\delta^t E_0 + p\sum_{t=0}^{\tau-1}\delta^t E_1 + p[1-(1-\pi)^{2\tau}]\sum_{t=\tau}^{\infty}\delta^t E_1\right\} \\ &= (1-\delta^\tau)E_p + \delta^\tau p[1-(1-\pi)^{2\tau}]E_1 \end{aligned}$$

with $E_p = pE_1 + (1-p)E_0$. Since the difference

$$u(p,\tau+1) - u(p,\tau) = \delta^\tau\left\{(\delta-1)(1-p)|E_0| - (1-\pi)^{2\tau}pE_1(\delta(1-\pi)^2-1))\right\}$$

is positive if and only if

$$\tau < \frac{1}{2\ln(1-\pi)}\left(\ln\frac{\delta-1}{\delta(1-\pi)^2-1} + \ln\frac{1-p}{p} + \ln\frac{|E_0|}{E_1}\right),$$

the value function of the social planner is $v^{sc}(p) = u(p,\tau^{sc}(p))$ with[8]

$$\tau^{sc}(p) = \max\left\{\left\lceil\frac{1}{2\ln(1-\pi)}\left(\ln\frac{\delta-1}{\delta(1-\pi)^2-1} + \ln\frac{1-p}{p} + \ln\frac{|E_0|}{E_1}\right)\right\rceil, 0\right\}.$$

As $\tau^{sc}(p) = 0$ if and only if p lies below the cut-off

$$p^{sc} = \frac{(1-\delta)|E_0|}{(1-\delta)(E_1+|E_0|) + \delta(2-\pi)\pi E_1},$$

we have the following result.

Proposition 1 (**Optimal symmetric pure-strategy profile**). *Among all pure symmetric strategy profiles, the following maximizes the players' expected average payoff: both players always communicate their payoffs truthfully, both choose the risky arm when $p(t) \geq p^{sc}$, and both choose the safe arm otherwise.*

[8]The superscript "sc" indicates the "symmetric cooperative" solution. For any real number x, the *ceiling* $\lceil x \rceil$ is the smallest integer not less than x.

If the social planner can use asymmetric strategy profiles, it is optimal for her to experiment beyond the belief p^{sc}. In fact, the expected discounted payoff from letting one player experiment and stopping all experimentation thereafter, $(1-\delta)\frac{1}{2}E_p + \delta p\pi E_1$, is positive above the cut-off

$$p^c = \frac{(1-\delta)|E_0|}{(1-\delta)(E_1+|E_0|)+\delta 2\pi E_1} < p^{sc}.$$

For the full description of the social optimum, we will also need the following cut-off:

$$\tilde{p}^c = \frac{(1-\delta)|E_0|}{(1-\delta)(E_1+|E_0|)+\delta 2(1-\pi)\pi E_1}\,.$$

It is straightforward to see that $\tilde{p}^c > p^{sc}$ and that starting at $\tilde{p}^c$, one failed experiment takes the posterior belief below p^c.

Proposition 2 (**Optimal strategy profile**)**.** *There exists a socially optimal strategy profile in which both players always communicate their payoffs truthfully, both choose the risky arm when $p(t) \geq \tilde{p}^c$, one player chooses the risky arm when $\tilde{p}^c > p(t) \geq p^c$, and both choose the safe arm otherwise. Furthermore, there is at most one period in which just one player experiments.*

Proof. See Appendix. □

Restricting the planner to symmetric pure-strategy profiles thus entails only a small loss in expected average payoffs.

4 Strategic Experimentation

We now turn to the analysis of strategic experimentation, first with public payoffs, then with private payoffs. While the planner's solution identified in the previous section constitutes an upper bound on the average of the two players' equilibrium payoffs, the solution to the single-agent bandit problem constitutes a lower bound on each player's individual equilibrium payoff. In fact, each player always has the option to ignore the information contained in the opponent's actions and payoffs (if the latter are observable).

A simpler version of the arguments leading up to Proposition 1 establishes that the single-agent (or "autarky") solution is given by the cut-off belief

$$p^a = \frac{(1-\delta)|E_0|}{(1-\delta)(E_1+|E_0|)+\delta\pi E_1} > p^{sc}\,,$$

where—as in the planner's solution—we adopt the convention that the agent experiments when she is indifferent, that is, when her belief equals p^a. We shall show that with publicly observable payoffs, equilibrium experimentation cannot go beyond this cut-off. With privately observed payoffs and cheap-talk communication, by contrast, more efficient equilibria can be played.

4.1 Public Payoffs

As both players choose the risky arm after any history $h(t) \in H_t^a$ in which a success has been observed, we can restrict our attention to histories with no prior success. We begin our analysis of equilibrium behavior with the observation that in every Nash equilibrium of the game with public payoffs, both players choose the safe arm after any history that takes their common belief below the social planner's cut-off p^c. To see this, suppose to the contrary that there exists a Nash equilibrium in which a player experiments at some belief $p < p^c$. From the analysis of the planner's solution, we know that the average of the players' objective functions at p is negative. Consequently there needs to be at least one player who receives a negative expected payoff. By deviating and always choosing the safe arm this player can increase her payoffs.

To get a first intuition for why equilibrium experimentation cannot go beyond the single-agent cut-off, consider pure-strategy Nash equilibria first. Since players do not experiment below p^c, there are only finitely many periods in which a pure-strategy Nash equilibrium can require players to experiment in the absence of a prior success. In the last period in which a player is meant to experiment, the player knows that if she fails, no player will experiment in future. Hence, she will only be willing to experiment if this is individually optimal, that is, if the current belief is at least p^a. When both players are meant to experiment, the belief must be above p^a because the value of experimenting in this last period is lower if one's fellow player also experiments. Hence, in any pure-strategy Nash equilibrium there can be at most one more experiment than in the single-agent solution.

Conversely, it cannot be the case that both players permanently stop experimenting at a belief above the single-agent cut-off. The reason is simply that each player—believing that the other player stopped experimenting—would then face the single-agent trade-off. The following proposition exploits this logic and extends it to mixed-strategy Nash equilibria.

We call the number of times a player chooses the risky arm on the path of play when every experiment is unsuccessful the *amount of experimentation* performed by that player. The *total amount of experimentation* by both players is simply the sum of the individual amounts. The total amount will typically depend on the initial belief and, with mixed strategies, may be a random variable.

Proposition 3. *Given an initial belief, let the optimal amount of experimentation in the single-agent problem be K. In any Nash equilibrium of the experimentation game with public payoffs, the realized total amount of experimentation is $K-1$, K or $K+1$.*

Proof. First, consider any history of length t for which $p^a < p(t) < 1$. Since $p(t) < 1$, no prior experiment has been successful, and since $p(t) > p^a$, fewer than K experiments have been performed. We now argue that players experiment with

probability 1 at least one more time following any such history. Let $v^a(p(t)) > 0$ be the value of the single-agent problem at the belief $p(t)$. Let τ be the smallest integer such that $\delta^\tau E_1 < v^a(p(t))$, and ψ the probability that there will be at least one experiment in the periods $t, t-1, \ldots, t+\tau-1$. The period t continuation value of each player is then trivially bounded above by $(1-\psi)\delta^\tau E_1 + \psi E_1$. So ψ cannot be smaller than $\bar{\psi} = [v^a(p(t)) - \delta^\tau E_1]/[(1-\delta^\tau)E_1]$, because otherwise it would be profitable for each player to deviate to the single-agent solution. For $n = 1, 2, \ldots$, the probability that no player experiments in the next $n\tau$ periods is bounded above by $(1-\bar{\psi})^n$. Letting $n \to \infty$, we see that there will almost surely be another experiment on the path of play.

Next, consider a history of length t for which $p(t) = p < p^a$. Let ϕ be the probability with which player j experiments at time t. Suppose that the equilibrium requires player i to experiment with positive probability. Then he can do no better by switching to the strategy of playing safe now and, in case player j experiments and is unsuccessful, continuing to play safe forever. This implies

$$\delta\phi p\pi E_1 \;\leq\; (1-\delta)E_p + \delta\left\{p[\pi + \phi\pi - \phi\pi^2]E_1 + \left(1 - p[\pi + \phi\pi - \phi\pi^2]\right)v\right\},$$

where $\pi + \phi\pi - \phi\pi^2$ is the probability of at least one success, and v player i's continuation value after a double failure, that is, a payoff realization $X_1(t) = X_2(t) = X_L$. As $0 \leq v \leq E_1$, this in turn requires that

$$0 \leq (1-\delta)E_p + \delta\left\{p\pi E_1 + (1-p\pi)v\right\}.$$

As $p < p^a$, we have $(1-\delta)E_p + \delta p\pi E_1 < 0$, and hence $v > 0$. So some player must experiment with positive probability in round $t+1$ or later. Repeating this step until a time $t+\tau$ at which $p(t+\tau) < p^c$ in the absence of a success, we obtain a contradiction because no player can experiment below p^c in equilibrium.

Finally, consider a history of length t for which $p(t) = p^a$, meaning that $p^a = B(K-1, p(0))$. Then, there can be no, one, or two further experiments on the path of play, which completes the proof. □

Keller, Rady and Cripps (2005) establish in a continuous-time setup that with fully revealing successes on the risky arm, the amount of experimentation in any Markov perfect equilibrium is limited by the single-agent amount. In our discrete-time setup, we can drop the assumption of Markov strategies and, with a minor qualification due to discrete rather than continuous time, establish that their finding extends to *all* Nash equilibria of the game with public payoffs.[9]

From now on, whenever we speak of an equilibrium, we mean a perfect Bayesian equilibrium that satisfies the restriction on beliefs set out in Section 2. Since such equilibria are also Nash equilibria, Proposition 3 immediately implies

[9]Note that a realized amount of experimentation equal to $K-1$ is non-generic because it requires the initial belief to lie in a countable subset of the unit interval.

Corollary 1. *Whenever the total amount of experimentation in the planner's optimal (or optimal symmetric) strategy profile exceeds the single-agent amount by more than 1, it cannot be implemented in an equilibrium of the experimentation game with observable payoffs.*

Above, we have fully characterized the equilibria with public payoffs in terms of the total amount of experimentation that is carried out on the path of play. These, equilibria, however, may differ in other dimensions such as *when* players experiment. For example, just above p^a players may engage in a war of attrition as to who has to carry out the final experiment. Thus, there may be periods in which no player experiments. Furthermore, players may use their communication to coordinate on whether a given player is meant to experiment in a given period. Below, we nevertheless show that for every equilibrium with observable payoffs we can find a path-equivalent equilibrium in the game with unobservable payoffs, that is, an equilibrium which induces the same distribution over experimentation paths. Moreover, we show that under certain conditions, higher amounts of experimentation can be supported when payoffs are unobservable.

For future reference, it is useful to establish the existence of a symmetric Markov perfect equilibrium for the game with observable payoffs. Since in all equilibria the total amount of experimentation is bounded by $K + 1$, we can restrict attention to finitely many beliefs when constructing equilibria. Using these finitely many beliefs as states (and thinking of the lowest belief as an absorbing state), it follows that there exist only finitely many pure Markov perfect strategies. We construct an auxiliary game in which players' finite action set corresponds to the set of pure Markov perfect strategies and payoffs are defined as in the original game. This is a symmetric finite game and the existence of a symmetric (possibly mixed-strategy) equilibrium in this auxiliary game follows from Nash (1951). The equilibrium in the auxiliary game corresponds to a symmetric Markov perfect equilibrium in the original game.

Below, we use the following bound on the payoffs of any symmetric Markov perfect equilibrium of the game with observable payoffs to construct punishment equilibria in the game with unobservable payoffs.

Lemma 1. *When payoffs are observable, the players' expected total payoff in any symmetric Markov perfect equilibrium is lower than the payoff from any strategy profile which has both players experiment with probability 1 if and only if* $p(t) \geq p^a$.

Proof. See Appendix. □

To see why this holds, fix a symmetric Markov perfect equilibrium. It follows from Proposition 3 that both players experiment with strictly positive probability at all beliefs above p^a. Since player i at least weakly prefers to experiment, his payoff remains unchanged if we modify the strategy profile so that he experiments with probability 1 at all beliefs above p^a. The crucial step is to show that if we,

in addition, change the strategy profile so that player j also experiments for sure at all these beliefs, player i's payoff weakly increases. This is obvious in the final period as player i now benefits from the greater information externality. The proof uses backward induction to establish this result in earlier periods as well.

4.2 Private Payoffs

We begin by noting that truthful communication can be sustained with privately observed, fully revealing payoffs. Suppose that after every period in which a player experimented, she announces a first success by sending the message $m_i(t) = 1$ and randomizes uniformly over all other messages otherwise. Furthermore, after the first success has been announced and both players know that the state of the world is good, suppose there are no meaningful messages any more, that is, both players always "babble" by randomizing uniformly over the interval $[0, 1]$. Similarly, a player who did not experiment randomizes uniformly over $[0, 1]$. Finally, on and off the equilibrium path, players believe that past communication by the other player was truthful. Intuitively, we are then back in the case of public payoffs.[10]

The key observation now is that if players anticipate this communication strategy, truthful communication is incentive compatible. Following a first success on player i's risky arm, player i knows the state of the world and hence is indifferent as to what player j believes. So truthfully announcing a success is optimal for player i following a success of his own. After such an announcement, player j believes with certainty that the state of world is good, and hence will play risky in all future periods irrespectively of what player i does after the announcement. If player i incorrectly announces a success, he cannot infer anything from player j's future behavior, so he is at last weakly better off telling the truth. We thus have the following result.

Proposition 4. *For every equilibrium of the game with public payoffs, there exists a path-equivalent equilibrium of the game with private payoffs.*

This shows that private information does not hurt players. Our next result establishes that players can often do better when payoffs are private. We construct equilibria in which players perform the optimal symmetric amount of experimentation whereafter, on the path of play, they once communicate and announce whether they had a prior success. If so, both players keep experimenting forever; otherwise, both stop experimenting. We punish early deviations (after the initial period) through beliefs: if a player refrains from experimenting at a time when the socially optimal symmetric strategy profile requires her to experiment, then

[10] One caveat here is that players may use the messages to create a controlled joint lottery to coordinate continuation play in the equilibrium with observable payoffs. But then we can encode the original message by using the odd digits only, while using the even digits to send the messages constructed above.

the other player reacts to this out-of-equilibrium event by assigning probability 1 to the good state of the world. Given the restriction on beliefs that we adopted in Section 2, our equilibrium concept does not allow us to assign the same optimistic beliefs to a player who observes a deviation at $t = 0$: as a player deviating in the first round cannot have seen a prior success, the other player cannot draw any inference on the state of the world from this deviation, and thus is not allowed to update her beliefs.[11]

Proposition 5. *There exists a threshold $p^\dagger < p^m$ such that for all initial beliefs $p \geq p^\dagger$, the experimentation game with private payoffs admits an equilibrium that is path-equivalent to the socially optimal symmetric pure-strategy profile.*

Proof. First, we specify the players' strategies, beginning with the behavior after all message histories and then turning to action histories. Play following any history in which at least one player did not experiment in period 0 is specified separately below; following such a history, we will prescribe players to play an equilibrium with truthful communication. For brevity, we do not explicitly specify behavior following observable simultaneous deviations as this is irrelevant for the incentives to deviate unilaterally.

If both players used the risky arm in all periods $t \leq \tau^{sc} - 1$, they both report truthfully in period τ^{sc}; following any other message history in which both players used the risky arm in period 0, they send babbling messages. Both players use the risky arm in any period $t \leq \tau^{sc} - 1$ as long as both did so in all prior periods. If both players used the risky arm in all periods $t \leq \tau^{sc} - 1$ and player j announced a success in period τ^{sc}, player i uses the risky arm in all periods $t \geq \tau^{sc}$. If both players used the risky arm in all periods $t \leq \tau^{sc} - 1$ and neither announced a success in period τ^{sc}, player i uses the risky arm in period $t' \geq \tau^{sc}$ if and only if he had a prior success himself or player j used the risky arm in a period $t'' \in \{\tau^{sc}, \ldots, t' - 1\}$ when neither player did so in periods $t = \tau^{sc}, \ldots, t'' - 1$. We are left to specify behavior for all action histories in which at least one player used the safe arm in any period $t \leq \tau^{sc} - 1$. If the first unilateral deviation occurs in a period $t' \in \{1, \ldots, \tau^{sc} - 1\}$, and player j is the one who deviates, then in all periods $t \geq t' + 1$ player i uses the risky arm and player j plays the autarky strategy (conditioning her behavior on her own signals only). Recall that the case of a first deviation in period $t = 0$ is handled separately below.

Second, we specify the players' beliefs if there was no deviation in period $t = 0$. It is convenient to specify beliefs about whether the state of the world is good (rather than the usual beliefs about nodes in an information set).[12] Both players use Bayesian updating on the path of play; in particular, if both players

[11] We could circumvent this constraint by letting the players observe one draw from the distribution of risky payoffs before the game starts. The following proposition would then hold with $p^\dagger = p^{sc}$.

[12] A probability distribution about possible nodes in a player's information set (where a node can be identified through whether and when the other player had a success when experiment-

used the risky arm in all periods $t \leq \tau^{sc} - 1$, both update under the assumption that the opponent reports truthfully in period τ^{sc}. If the first unilateral deviation occurs in a period $t' \in \{1, \ldots, \tau^{sc} - 1\}$, and player j is the one who deviates, then player i switches irrevocably to the belief that the state of the world is good with probability 1, while player j continues to apply Bayes' rule, conditioning on her own observations only. If the first unilateral deviation occurs in a period $t \geq \tau^{sc}$ by player j after a success was announced by player j, player i ignores this as well as all possible future deviations of player j and continues to believe that the state of the world is good with probability 1. If the first unilateral deviation occurs in a period $t \geq \tau^{sc}$ by player j after a success was announced by player i but not player j, player i ignores this deviation as well as all possible future deviations of player j and continues to apply Bayes' rule in every period based on his own experimentation results only. If the first unilateral deviation occurs in a period $t \geq \tau^{sc}$ by player j after no success was announced, player i switches irrevocably to the belief that the state of the world is good with probability 1 while player j updates using only her own experimentation results from thereon.

Third, we prove sequential rationality if there was no deviation in period $t = 0$. Any player who had a success in a period $t \leq \tau^{sc} - 1$ is willing to announce it truthfully in round τ^{sc}; given that the other player believes the announcement to be truthful (i.e. has the belief that the state of the world is good with probability 1), it is optimal for her to choose the risky action forever. If the first unilateral deviation occurs in a period $t' \in \{1, \ldots, \tau^{sc} - 1\}$, and player j is the one who deviates, then player i holds the belief that the state of the world is good and hence it is optimal for him to choose the risky action in all subsequent periods. As a consequence, player j ceases to learn anything from observing i's future behavior, hence finds herself in an autarky situation. It is optimal for her, therefore, to play the autarky strategy, and beliefs are consistent with this. If player j does not deviate, she obtains the value of the symmetric cooperative solution. Since the latter is always weakly larger than the autarky value, the deviation is unprofitable. We are left to consider behavior following histories in which the first deviation occurred in a period $t \geq \tau^{sc}$. If both players used the risky arm in all periods $t \leq \tau^{sc} - 1$, and if player i had a success or player j announced a success, it is obviously optimal for player i to always choose the risky action since he assigns probability 1 to the good state of the world. Exactly the same holds if both players used the risky arm in all periods $t \leq \tau^{sc} - 1$, neither announced a success and player j was the first to deviate by using the risky instead of the safe arm. If both players used the risky arm in all periods $t \leq \tau^{sc} - 1$ and player j neither announced a success nor was the first to deviate

ing in addition to one's own observations), can be constructed in the obvious way from the probability that the state of the world is good together with how often the other player used the risky arm; this probability distribution is unique but for the fact that we can arbitrarily prescribe when another player had a success following an out-of-equilibrium observation.

to the safe arm in a period $t \geq \tau^{sc}$, player i's belief is below p^{sc} if he had no prior success himself, and hence it is optimal for him to use the safe arm.

Fourth, we construct a continuation equilibrium that deters deviations at $t = 0$. Fix a symmetric Markov perfect equilibrium of the game with observable payoffs starting with the common prior $B(1, p)$. By Lemma 1, the corresponding equilibrium value is bounded above by $\tilde{v}(B(1, p))$, the value from both players always using the risky arm until their belief falls below the autarky cut-off p^a. Consider an equilibrium of the game with private payoffs that is path-equivalent to the Markov perfect equilibrium; such an equilibrium exists by Proposition 4. Following a unilateral deviation by player j in $t = 0$, we require player i to communicate truthfully whether he had a success; by the argument underlying Proposition 4, this is incentive compatible. If he announces a failure, both players' continuation play corresponds to the selected equilibrium of the game with private payoffs.

So, if player j experiments at $t = 0$, her expected overall payoff is

$$v^{sc}(p) = (1-\delta)E_p + \delta \underbrace{\left\{p[1-(1-\pi)^2]E_1 + \left(1 - p[1-(1-\pi)^2]\right) v^{sc}(B(2,p))\right\}}_{(I)};$$

if she deviates, this payoff is no more than

$$\delta \underbrace{\left\{p\pi E_1 + (1-p\pi)\,\tilde{v}(B(1,p))\right\}}_{(II)}.$$

By construction, $\tilde{v}(B(1,p)) \leq v^{sc}(B(1,p))$. Being the upper envelope of linear functions, v^{sc} is convex. This implies that $v^{sc}(B(1,p)) \leq B(1,p)\pi E_1 + (1 - B(1,p)\pi)\, v^{sc}(B(2,p))$. As the points $(B(1,p), v^{sc}(B(1,p)))$, $(B(2,p), v^{sc}(B(2,p)))$ and $(1, E_1)$ do not lie on a line, this inequality is in fact strict. Using this and the fact that $(1 - p\pi)B(1,p) = p(1-\pi)$, one has $(I) > (II)$. As $E_p \geq 0$ for $p \geq p^m$, moreover, we see by continuity that there exists a belief $p^\dagger < p^m$ such that deviating at $t = 0$ is suboptimal. □

In our equilibrium construction, deviations in the initial period are punished through a continuation equilibrium that is path-equivalent to a symmetric Markov perfect equilibrium. Of course, there may be harsher punishments for a player who deviates in the first period. For example, we could search for an equilibrium of the game with observable payoffs that minimizes the payoff of a given player. By playing a continuation equilibrium that is path-equivalent to this asymmetric equilibrium, we would punish initial deviations more severely and hence increase the range of initial beliefs for which the symmetric social optimal strategy profile can be sustained. In the above proposition as well as in Proposition 6 below, we refrain from doing so for ease of exposition.

The belief following a deviation in an early (but not the initial) round may seem somewhat unusual in the above equilibrium. Intuitively, upon observing

such a deviation, player i reasons as follows: "Clearly, player j was not careful and made a mistake. She must already know that the state of the world is good to be so careless." While this reasoning is compatible with the logic of sequential equilibrium, the equilibrium construction hinges crucially on this particular choice of out-of-equilibrium beliefs. Our next aim is therefore to show that private payoffs can lead to a more efficient outcome even under the stringent requirement that whenever a player—who is still uncertain as to whether the state of the world is good—observes a deviation to the safe action, her beliefs become as pessimistic as possible.

Definition 1 (Pessimistic Beliefs)**.** *We say that an equilibrium of the game with private payoffs has pessimistic beliefs if a player who observes a deviation to the safe action and does not yet assign probability 1 to the good state of the world, believes that the deviating player had only failures before the deviation.*

Recall from Proposition 4 that for any equilibrium with observable payoffs, we can construct an equilibrium with private payoffs that induces the same distribution over experimentation paths. In this equilibrium, players truthfully communicate after every period in which they experimented until a first success is announced.

Lemma 2. *The equilibrium constructed in Proposition 4 has pessimistic beliefs.*

Proof. If player i had a success, he believes that the state of the world is good with probability 1. Similarly, once player j has announced a success, the equilibrium is such that player i believes with probability 1 that there has been a success, and assigns probability 1 to the good state of world from then on. Consider a history, therefore, in which neither player j announced a success nor player i observed a success himself. As player i believes that player j communicated truthfully in the past, he believes that all of player j's experiments have been failures. So his beliefs are indeed pessimistic. □

Our next proposition shows that even if we restrict ourselves to pessimistic beliefs, we can find equilibria that implement the optimal symmetric strategy profile for sufficiently optimistic starting beliefs. To see the intuition, recall that in the symmetric optimum, the planner updates her beliefs on the basis of two experiments in every period until the belief falls below p^{sc}. Absent meaningful communication, however, each player updates her belief using only the result of her own experimentation, and hence beliefs decrease at a slower rate. The key observation is that for high enough starting beliefs, this slower learning implies that the social planner reaches p^{sc} before players who do not communicate reach p^m. Above p^m, players myopically prefer to experiment, and in the equilibrium that we will construct, they do so up to the period in which a social planner who has not observed a success would cease experimentation. At that point in

time, players truthfully communicate and continue experimenting only if at least one player had a prior success. A player who deviates by not experimenting prior to this point in time reduces her myopic payoff and induces a symmetric continuation equilibrium in which payoffs are weakly lower than in the optimal symmetric strategy profile; thus, such a deviation is unprofitable.

Let $n^\ddagger = \min\{n \in \mathbb{N} : B(n, p^m) < p^{sc}\}$. This is the minimal number of failed experiments that moves a player's belief from p^m to below p^{sc}. Next, define $p^\ddagger$ implicitly by $p^m = B(n^\ddagger, p^\ddagger)$, so that $B(2n^\ddagger, p^\ddagger) < p^{sc} \leq B(2n^\ddagger - 1, p^\ddagger)$. Intuitively, consider the case where both players always experiment and all experiments fail. Then, for initial beliefs $p \geq p^\ddagger$, a social planner who observes both players' failures reaches p^{sc} before a player who only observes her own failures reaches p^m.

Proposition 6. *For all initial beliefs $p \geq p^\ddagger$, the experimentation game with private payoffs has an equilibrium with pessimistic beliefs that is path-equivalent to the socially optimal symmetric pure-strategy profile.*

Proof. Recall that there exists a symmetric Markov perfect equilibrium in the case of observable payoffs. Choose such a symmetric Markov perfect equilibrium for any starting belief p. By Proposition 4 there exists an equilibrium with private payoffs that is path-equivalent. Denote this equilibrium by $\sigma(p)$.

We are now ready to specify strategies.[13] In all periods $t \leq n^\ddagger - 1$, both players babble and use the risky arm provided no player chose the safe arm in a previous period $t' \leq t - 1$. If a single player deviated and chose the safe arm in a period $\tau \leq n^\ddagger - 1$, the players truthfully communicate at the beginning of period $\tau + 1$. If player i communicated truthfully in period $\tau + 1$, he plays the strategy prescribed by $\sigma(q)$ in periods $t \geq \tau + 1$, where $q = B(2\tau - 1, p)$ if no player announced a success, and $q = 1$ otherwise. If player i did not communicate truthfully in period $\tau + 1$, he either had a success he did not announce or he incorrectly announced a success he did not have. In the former case, he uses the risky arm in periods $t \geq \tau + 1$; in the latter case, he plays the autarky strategy.

We are left to specify strategies in case both players used the risky arm in all periods $t \leq n^\ddagger - 1$. In this case, both players truthfully announce at the beginning of period $n^\ddagger$ whether they had a success in any of the previous rounds; and independent of how the play proceeds from period $n^\ddagger$ onwards, players babble in every period $t > n^\ddagger$. If player j announced a success, or if player i observed a success himself in any prior period, player i uses the risky arm in all periods $t \geq n^\ddagger$. If neither announced a success in period $n^\ddagger$, player i uses the risky arm in period $t' \geq \tau^{sc}$ if and only if he had a prior success himself or player j used the risky arm in a period $t'' \in \{n^\ddagger, \ldots, t' - 1\}$ when neither player did so in periods $t = n^\ddagger, \ldots, t'' - 1$.

Next, we specify beliefs about the state of the world. Any player who had a prior success believes the state of the world to be good with probability 1. In any

[13] We follow the usual convention again and ignore simultaneous deviations.

period $t \leq n^{\ddagger} - 1$ such that both players used the risky arm in all periods $t' \leq t$, the belief of player i is $p_i(t) = B(t-1, p)$ if all his experiments failed. If a player deviated to the safe arm in a period $\tau \leq n^{\ddagger} - 1$, she believes the state of the world is good with probability 1 if the other player announces a success in period $\tau + 1$ (or she had a prior success herself); otherwise her belief is $q = B(2\tau - 1, p)$. Thereafter, beliefs are as in the equilibrium $\sigma(q)$.

Now suppose that both players used the risky arm in all periods $t \leq n^{\ddagger} - 1$. Each player believes the other's message in period $n^{\ddagger}$ to be truthful, and all subsequent messages to be uninformative. If player i deviates in period $n^{\ddagger}$ by incorrectly announcing a success he did not have, and player j does not announce a success in period $n^{\ddagger}$, then player i's belief in period $\tau \geq n^{\ddagger}$ equals 1 if he experiences a success in one of the rounds $t = n^{\ddagger}, \ldots, \tau - 1$, and equals $B(2n^{\ddagger} + n, p)$ if he carries out n experiments in periods $t = n^{\ddagger}, \ldots, \tau - 1$ and they all fail. If player j is the first to deviate in period $t' \geq n^{\ddagger}$ after neither player announced a success, player i believes the state of the world to be good with probability 1. If player i is the first to deviate in period $t' \geq n^{\ddagger}$ after neither player announced a success, player i's belief in round $\tau \geq t' + 1$ equals 1 if he experiences a success in one of the rounds $t = t', \ldots, \tau - 1$, and equals $B(2n^{\ddagger} + n, p)$ if he carries out n experiments in periods $t = t', \ldots, \tau - 1$ and they all fail.

As the beliefs that we have specified follow Bayes' rule whenever possible, it remains to show sequential rationality. Each player uses the risky arm whenever he assigns probability 1 to the good state of the world, which is clearly optimal. If a single player deviates to the safe arm in a period $\tau \leq n^{\ddagger} - 1$, it is optimal for both players to communicate truthfully at the beginning of period $\tau + 1$ by the argument underlying Proposition 4. If player i does communicate truthfully in period $\tau + 1$, he believes with probability 1 that player j plays according to $\sigma(q)$ with q specified above. As $\sigma(q)$ constitutes an equilibrium, it is optimal for player i to play according to $\sigma(q)$ as well. If player i announces a success in period $\tau + 1$ that he did not have, player j uses the risky arm forever, so player i finds himself in an autarky situation and optimally plays the autarky strategy. If player i does not announce a success he had, it is clearly optimal for him to use the risky arm forever. If player i is the first to deviate in a period $\tau \leq n^{\ddagger} - 1$, his belief about the state of the world is weakly above p^m; by the same argument as the one used in the proof of Proposition 5 for deviations in period $t = 0$, such a deviation is unprofitable.

We are left to rule out deviations by player i in a period $t' \geq n^{\ddagger}$ following a history in which both players used the risky arm in all periods $t \leq n^{\ddagger} - 1$, player i had no success himself, neither player announced a success, and player j was not the first to deviate in a period $t'' \in \{n^{\ddagger}, \ldots, t' - 1\}$. In this case, player i's belief is below p^{sc}. If player i was the first to deviate in a period $t'' \in \{n^{\ddagger}, \ldots, t' - 1\}$, player j uses the risky arm in all future periods independent of her experimentation results. The same holds if there was no deviation in any

round $t \leq t' - 1$ and player i deviates to the risky arm in round t'. If there was no prior deviation and player i uses the safe arm in round t', finally, he expects player j to use the safe arm in all future periods. Whatever he does in round t', therefore, player i cannot learn anything from player j's future behavior and so finds himself in an autarky situation. As $p^{sc} < p^a$, it is thus optimal for player i to use the safe arm. □

Some straightforward computations show that

$$p^\ddagger \geq \frac{(1-\pi)[1-\delta+\delta(2-\pi)\pi]|E_0|}{(1-\pi)[1-\delta+\delta(2-\pi)\pi]|E_0| + (1-\delta)E_1} > p^m$$

and

$$p^\ddagger < \frac{[1-\delta+\delta(2-\pi)\pi]|E_0|}{[1-\delta+\delta(2-\pi)\pi]|E_0| + (1-\delta)E_1}\,.$$

Since we could again use harsher punishments after early deviations, the interval $[p^\ddagger, 1]$ constitutes just part of the set of initial beliefs for which the outcome of the socially optimal symmetric pure-strategy profile can be sustained in an equilibrium with pessimistic beliefs.

For a range of initial beliefs below $p^\ddagger$, the same logic as in Proposition 6 allows us to construct equilibria in which the players may not reach the symmetric social optimum but perform more experiments, and are better off, than in any equilibrium with observable payoffs.[14] Suppose that $B(2, p^a) > p^{sc}$, so that by Corollary 1 the efficient amount of experimentation cannot be reached with observable payoffs. Let $n' = \min\{n \in \mathbb{N} : B(n, p^m) < p^a\}$. This is the minimal number of failed experiments that moves a player's belief from p^m to below p^a. Next, define $p' < p^\ddagger$ implicitly by $p^m = B(n'+2, p')$, so that $B(2n'+4, p') = B(n'+2, p^m) < B(2, p^a)$. Then, starting from a prior $p \geq p'$ and defining strategies as in Proposition 6 but with $n^\ddagger$ replaced by n', we obtain an equilibrium of the game with private payoffs in which both players experiment some way below p^a. In view of the similarity with the above construction, we omit the details.

5 A Continuous-Time Limit

We now embed our discrete-time model in a continuous-time framework that coincides (up to a normalization of the safe payoff to zero) with the two-player version of the setup studied by Keller, Rady and Cripps (2005).

Let time be continuous and suppose that operating the risky arm comes at a flow cost of $s > 0$ per unit of time. In the good state ($\theta = 1$), the risky arm yields lump-sum payoffs which arrive at the jump times of a Poisson process with

[14] We established in the social planner's problem that the players' average payoff when both experiment for τ periods is single-peaked in τ. Here, we are below τ^{sc}, so the average payoff is still increasing in the number of experiments.

intensity $\lambda > 0$. These lump-sums are independent draws from a time-invariant distribution with known mean $h > 0$, and the Poisson processes in question are independent across the two players. In the bad state ($\theta = 0$), the risky arm never generates a lump-sum payoff. The safe arm does not produce any such payoffs either, but is free to use.

Given the common discount rate $r > 0$, a player's payoff increment from using a bad risky arm for a length of time $\Delta > 0$ is

$$\int_0^\Delta r\, e^{-r\,t}\,(-s)\,dt = \left(1 - e^{-r\,\Delta}\right)(-s).$$

The expected discounted payoff increment from a good risky arm is

$$\mathbb{E}\left[\int_0^\Delta r\, e^{-r\,t}\,(h\,dN_t - s\,dt)\right] = \int_0^\Delta r\, e^{-r\,t}\,(\lambda h - s)\,dt = \left(1 - e^{-r\,\Delta}\right)(\lambda h - s);$$

here N_t is a standard Poisson process with intensity λ, and the first equality follows from the fact that $N_t - \lambda t$ is a martingale. We assume $\lambda h > s$ so that a good risky arm dominates the safe arm. Finally, the probability of observing at least one lump-sum on a good risky arm during a time interval of length Δ is $1 - e^{-\lambda\Delta}$.

If we let the players adjust their actions only at the times $t = 0, \Delta, 2\Delta, \ldots$ for some fixed $\Delta > 0$, we are back in our discrete-time framework with $\pi = 1 - e^{-\lambda\Delta}$, $E_0 = -s$, $E_1 = \lambda h - s$, and $\delta = e^{-r\,\Delta}$. Letting $\Delta \to 0$, we can thus study the impact of vanishing time lags on the results we derived above.

First, we note that p^{sc} and p^c converge in a monotonically decreasing fashion to one and the same limit as Δ vanishes; this limit is the efficient two-player cut-off from Keller, Rady and Cripps (2005),

$$p_2^* = \frac{r|E_0|}{r(E_1 + |E_0|) + 2\lambda E_1}\,.$$

Thus, the difference between the socially optimal symmetric strategy profile and its unrestricted counterpart (which is small to start with by Proposition 2) completely disappears in the limit, and implementing the symmetric optimum as we do in Propositions 5 and 6 fully solves the free-rider problem.

Second, we observe that the restriction to sufficiently optimistic priors in Proposition 5 can be dropped when Δ becomes small.

Corollary 2. *For any initial belief $p > p_2^*$, there exists a $\bar{\Delta}(p) > 0$ such that for all $\Delta < \bar{\Delta}(p)$, the experimentation game with private payoffs admits an equilibrium that is path-equivalent to the socially optimal symmetric pure-strategy profile.*

Proof. The monotone convergence $p^{sc} \to p_2^*$ as $\Delta \to 0$ implies that $p_2^* < p^{sc} < p$ for sufficiently small Δ. Restricting ourselves to such Δ, we specify strategies and

beliefs exactly as in the proof of Proposition 5. According to that proof, these strategies and beliefs constitute an equilibrium if

$$v^{sc}(p) \geq \delta \left\{ p\pi E_1 + (1 - p\pi)\, \tilde{v}(B(1,p)) \right\}.$$

As $\tilde{v}$ is a non-decreasing function and $B(1,p) < p$, the inequality

$$v^{sc}(p) \geq \delta \left\{ p\pi E_1 + (1 - p\pi)\, \tilde{v}(p) \right\}$$

is a sufficient condition. And as $\delta \to 1$ and $\pi \to 0$ when $\Delta \to 0$, it is enough to show that

$$\lim_{\Delta \to 0} v^{sc}(p) \geq \lim_{\Delta \to 0} \tilde{v}(p)\,.$$

Starting from the explicit representation for the constrained planner's value function in Section 3, it is straightforward to compute

$$\lim_{\Delta \to 0} v^{sc}(p) = E_p - \frac{E_{p_2^*}}{w(p_2^*)}\, w(p)$$

with

$$w(p) = (1-p) \left(\frac{1-p}{p} \right)^{\frac{r}{2\lambda}};$$

this limit is the value of the planner's problem in continuous time. Similarly,

$$\lim_{\Delta \to 0} \tilde{v}(p) = \max \left\{ 0,\ E_p - \frac{E_{p_1^*}}{w(p_1^*)}\, w(p) \right\}$$

with

$$p_1^* = \frac{r|E_0|}{r(E_1 + |E_0|) + \lambda E_1}\,,$$

the autarky cut-off in continuous time. As $p_2^* < p_1^* < p^m$, we have $E_{p_2^*} < E_{p_1^*} < 0$. A simple computation reveals that

$$\frac{E_{p_2^*}}{w(p_2^*)} < \frac{E_{p_1^*}}{w(p_1^*)}\,,$$

so the two limits satisfy the desired inequality. □

There is no counterpart to this result for the equilibria constructed in Proposition 6. As Δ vanishes, $p^\ddagger$ converges to a belief strictly above p^m, so having ever shorter reaction lags does not help here. In fact, what these equilibria require is that it take a sufficiently long string of failures for the belief of a player learning only from his own experiments to pass p^m; more precisely, the minimal length of time this must take is $n^\ddagger \Delta$. When Δ decreases, each failure becomes less informative, and $n^\dagger$ increases so as to make $n^\ddagger \Delta$ converge to the length of time a player would have to experiment unsuccessfully in continuous time for his belief to fall from $\lim_{\Delta \to 0} p^\ddagger$ to p^m.

6 More Than Two Players

We focused on the two-player case but our main insights carry over to more players with only minor modifications.

With observable payoffs, our proof that in any equilibrium, no player is willing to experiment at a belief below the single-agent cut-off, and that some player must experiment above it, goes through almost unaltered. Hence, with observable payoffs and N players, the equilibrium amount of experimentation is generically at least K (the single-agent optimum) and never more than $K+N-1$. Whenever the social optimum requires a higher amount, therefore, it cannot be achieved in equilibrium.

With private payoffs, the observation that truthful communication is easy to sustain remains intact, and so does Proposition 4. Any equilibrium of the game with observable payoffs thus has a path-equivalent counterpart in the game with private payoffs.

The logic of the equilibrium constructions in Propositions 5 and 6 also continues to apply, though some adjustments are needed in the details of the argument. Regarding possible deviations in early rounds, our proofs of these propositions use the fact, established in Lemma 1, that for two players a symmetric Markov perfect equilibrium yields a lower payoff than if both players use the autarky cut-off, which in turn yields a lower payoff than the socially optimal symmetric pure-strategy profile. As our proof of Lemma 1 does not extend to more than two players, we cannot be sure that the threat to play a path-equivalent counterpart of a symmetric Markov perfect equilibrium is enough to deter first-round deviations.[15]

There are at least two ways to solve this problem. The first is to establish existence of *some* (possibly non-Markov and asymmetric) continuation equilibrium that gives a lower payoff to the deviating player than the symmetric optimum. The second is to consider the problem of a social planner who is restricted to symmetric (but possibly mixed) strategies prescribing use of the safe arm below the autarky cut-off, and to establish sufficient conditions under which this planner requires all players to use the risky arm with probability 1 at or above the autarky cut-off, so that all players applying this cut-off dominates any symmetric (and in particular, symmetric Markov perfect) equilibrium. One such condition is that the probability π of observing the high payoff on a good risky arm be

[15]Our proof of Lemma 1 does not generalize to $N > 2$ because it relies on the observation that two players using the autarky cut-off attribute a positive value to a free experiment, that is, to the observation of a draw from the distribution of risky payoffs before the game starts. This is no longer true for more than two players. The reason is that the observation of a bad signal X_L can "crowd out" one round of joint experimentation. When $N = 2$, this happens precisely when the one experiment thus lost overall would not have been worthwhile performing in the first place. When $N > 2$, however, a bad signal can lead to a loss of more than one signal overall, and this is detrimental to the players.

sufficiently low. In fact, when π is low enough, the set of beliefs at which a completely unrestricted social planner wants all players to use the risky arm is strictly larger than $[p^a, 1]$, and so a planner restricted in the above fashion can do no better than prescribe use of the risky arm with probability 1 on this interval.

Assuming a low π is natural given our emphasis on implementing the socially optimal *symmetric* strategy profile. This is because the more players there are, the larger is the potential performance loss of this strategy profile relative to the truly efficient profile that an unrestricted social planner would like to implement. In the continuous-time limit $\Delta \to 0$ of the previous section, however, the difference between the socially optimal symmetric strategy profile and its unrestricted counterpart vanishes. With more than two players, assuming a small period length Δ thus justifies our focus on the symmetric profile and at the same time ensures robustness to early deviations of the equilibria in Propositions 5 and 6. In particular, Corollary 2 holds with p_2^* replaced by

$$p_N^* = \frac{r|E_0|}{r(E_1 + |E_0|) + N\lambda E_1},$$

the efficient N-player cut-off from Keller, Rady and Cripps (2005).

Finally, the equilibrium construction used in Proposition 6 becomes more powerful as the number of players increases. Calculating the optimal stopping time for the symmetric social planner's solution as in Section 3, one finds that this stopping time is weakly decreasing in the number of players. Because the time it takes a single player to reach the myopic belief is independent of N by definition, the set of beliefs for which the optimal symmetric strategy profile can be implemented as in Proposition 6 increases in the number of players.

7 Conclusion

We analyzed a discrete-time experimentation game with two-armed bandits. For publicly observable payoffs, the free-rider problem is so severe that in any equilibrium, both players together perform at most one experiment more than a single agent would. Privately observed payoffs mitigate the free-rider problem to the point where for sufficiently optimistic prior beliefs, it becomes possible to sustain the socially optimal symmetric pure-strategy profile as an equilibrium with cheap-talk communication. We showed that these results are robust to letting players react very fast, and discussed how they extend to more than two players.

Throughout, we assumed that players cannot prove the results of their own experiments. If we suppose instead that a player can provide hard evidence of any prior success, our equilibrium constructions in Propositions 5 and 6 still carry through. Intuitively, whenever players communicate, they do so truthfully in these equilibria and, hence, it is unnecessary to show a proof. Moreover, a

player who has had a success is indifferent as to the other player's behavior and therefore willing not to reveal hard information.

In our model, only the good state of the world can be revealed. Suppose instead that the bad state of the world can also be fully revealed through an additional payoff X_B. In a research project, for example, this could be a surprising impossibility result. Whenever X_B is not drawn in the bad state, the payoff X_L is drawn, which can also materialize in the good state. As long as the high payoff X_H is drawn more often in the good state than X_B is drawn in the bad state, players become more pessimistic whenever they observe X_L. In this case, the social planner's solution can be characterized as in Section 3, and Propositions 4 and 6 can be adapted. Regarding the former, observe that a player who drew X_B has a continuation payoff of zero irrespectively of what the other player has done or said; such a player is clearly willing to communicate his information truthfully.

As a player who has observed a bad signal must cease to experiment, the equilibrium underlying Proposition 6 needs to be adjusted. On the path of play, rather than waiting to communicate until a player reaches the myopic cut-off based on his own experimentation results, we let players truthfully announce in the early periods whether they have received a bad signal or not. Whenever a bad signal is observed and announced, both players cease to experiment. When observing the opponent deviate to the safe arm in an early period (other than the first), a player concludes that the opponent has received a bad signal and, hence, stops experimenting. Otherwise, the logic of the equilibrium construction remains the same: initially, players are myopically motivated to experiment, and they truthfully communicate whether or not they had a success only at the point in time at which a social planner observing all experimentation results would reach the cut-off for the optimal symmetric strategy profile.

Although we believe that in most strategic experimentation problems the presumption that players can communicate is realistic, one may wonder exactly what role communication plays for our results. The answer is somewhat subtle. In the equilibria of Propositions 5 and 6, players only communicate at a single point in time on the path of play. That is, after a given number of rounds of experimentation—say 100—players announce truthfully whether or not they had a success. Intuitively, one could replace this communication by one round of play in period 101 in which each player uses the risky arm if and only if she had a prior success, ensuring that all necessary information is exchanged within one round. The role of communication may thus seem very limited. Truthful communication, however, plays another important role: it ensures the existence of a simple continuation equilibrium following a deviation—including one in the first period. What kind of equilibria exist absent communication remains, in our view, an interesting question for further research.

Appendix

Proof of Proposition 2

As $p^c \leq p^{sc}$ it follows that below p^c the social planner prefers not experimenting over one or two experiments. We calculate the payoff of performing one experiment and then stopping as

$$(1-\delta)\tfrac{1}{2}E_p + \delta p\pi E_1,$$

and the payoff of performing two experiments and then stopping as

$$(1-\delta)E_p + \delta p[1-(1-\pi)^2]E_1\,.$$

Subtracting the latter payoff from the former, we get

$$-(1-\delta)\tfrac{1}{2}E_p + \delta p\pi(1-\pi)E_1,$$

which is negative above $\tilde{p}^c$ and positive below. Because it is suboptimal to perform more than two experiments below $\tilde{p}^c$, it follows that between p^c and $\tilde{p}^c$, the planner prefers one experiment to two experiments.

Next, we establish that above $\tilde{p}^c$, the planner prefers the sequence 2–1 (two experiments this round followed by one experiment next round) to the sequence 1–2. Suppose it is optimal for the social planner to let one player experiment in one round, and two players in the subsequent round. We will prove that the planner can make herself strictly better off by first letting two players experiment, then one player, and afterwards using the same strategy as before. In the two rounds under consideration, the expected payoff of the social planner from 2–1 is

$$(1-\delta)\left\{E_p + \delta\left(\tfrac{1}{2}E_p + p[1-(1-\pi)^2]\tfrac{1}{2}E_1\right)\right\}\,.$$

The expected payoff from 1–2 is

$$(1-\delta)(\tfrac{1}{2}E_p + \delta E_p)\,.$$

Subtracting the latter payoff from the former, we get

$$(1-\delta)\tfrac{1}{2}\left\{\left(E_p + \delta p[1-(1-\pi)^2]E_1\right) - \delta E_p\right\}.$$

The part in parentheses is positive above p^{sc} by definition, and $-\delta E_p$ is positive above p^{sc} because $p^{sc} \leq p^m$, so the sequence 1–2 will never be used by the social planner above p^{sc}.

This means that if the planner ever switched to 1 (i.e. a single experiment) above $\tilde{p}^c$, she would have to continue with 1 until p^c is reached. In a last step, we rule this out by showing that the planner engages in at most one round of experimentation by a single agent before stopping.

If the planner finds it optimal at some belief p to engage in two rounds of experimentation by a single agent and then stop, her expected discounted payoff is

$$\begin{aligned}&(1-\delta)\tfrac{1}{2}E_p + \delta p\pi E_1 \\ &\quad +\delta(1-p\pi)\left\{(1-\delta)\tfrac{1}{2}\left[\frac{p(1-\pi)}{1-p\pi}E_1 + \frac{1-p}{1-p\pi}E_0\right] + \delta\frac{p(1-\pi)}{1-p\pi}\,\pi E_1\right\},\end{aligned}$$

which must be non-negative. The expression in braces must be non-negative as well or else it would not be optimal to perform one final experiment after a failure. The expected discounted payoff from performing two experiments at once and then stopping is

$$(1-\delta)E_p + \delta p[1-(1-\pi)^2]E_1.$$

Subtracting the former payoff from the latter, we obtain

$$\begin{aligned}&(1-\delta)\tfrac{1}{2}\left[p(1-\pi)E_1 + (1-p)E_0\right] + (1-\delta)\tfrac{1}{2}p\pi E_1 + \delta p(1-\pi)\pi E_1 \\ &\quad -\delta\left\{(1-\delta)\tfrac{1}{2}\left[p(1-\pi)E_1 + (1-p)E_0\right] + \delta p(1-\pi)\pi E_1\right\} \\ =\;& (1-\delta)\left\{(1-\delta)\tfrac{1}{2}\left[p(1-\pi)E_1 + (1-p)E_0\right] + \delta p(1-\pi)\pi E_1\right\} + (1-\delta)\tfrac{1}{2}p\pi E_1,\end{aligned}$$

which is positive since the term in braces is non-negative and $E_1 > 0$. □

Proof of Lemma 1.

Given $p(0) = p$, let $\tilde{v}(p)$ be the players' common value under the strategy profile where both experiment with probability 1 if and only if $p(t) \geq p^a$, and $v^M(p)$ the value in some symmetric Markov perfect equilibrium.

We first show that under the former strategy profile, the players assign positive value to a free signal, that is, for all beliefs $p \in]0,1[$,

$$\tilde{v}(p) < p\pi E_1 + (1-p\pi)\tilde{v}(B(1,p))\,.$$

As a first step, note that

$$\tilde{v}(p) = (1-\delta^T)E_p + \delta^T[1-(1-\pi)^{2T}]pE_1\,,$$

where $T = T(p)$ is the number of periods needed for the belief to fall below the autarky cut-off p^a when the players' experiments all fail; that is, $B(2T,p) < p^a \leq B(2T-2,p)$. Now, if $B(2T-1,p) \geq p^a$, the free signal does not alter the number of experiments that the players conduct under the strategy profile in question, and their expected payoff is

$$(1-\delta^T)E_p + \delta^T[1-(1-\pi)^{2T+1}]pE_1 > \tilde{v}(p).$$

If $B(2T-1,p) < p^a$, by contrast, the players engage in one less round of joint experimentation when all experiments fail, and their expected payoff is

$$\begin{aligned}&(1-\delta^{T-1})E_p + \delta^{T-1}[1-(1-\pi)^{2T-1}]pE_1\\ &\quad = \quad \tilde{v}(p) - \delta^{T-1}\left\{(1-\delta)E_p + \delta[1-(1-\pi)^{2T}]pE_1 - [1-(1-\pi)^{2T-1}]pE_1\right\}.\end{aligned}$$

Because the inequality $B(2T-1,p) < p^a$ means that an agent starting with the initial belief p in autarky strictly prefers not to experiment after $2T-1$ failed experiments, the expression in braces is negative. In either case, therefore, the players benefit from the free signal, which establishes the claimed inequality.

Let $p_n = B(p,n)$ be the belief that is reached after n failed experiments. We are now ready to show by backward induction that $v^M(p_n) \leq \tilde{v}(p_n)$ for all $n \in \mathbb{N}_0$. First observe that $v^M(p) = \tilde{v}(p) = 0$ for all $p < p^a$. For the induction step, let $p_n \geq p^a$ and assume that $v^M(p_{n'}) \leq \tilde{v}(p_{n'})$ for all $n' > n$. Let q be the probability with which the players experiment in the symmetric Markov perfect equilibrium at the belief p_n. As each player must weakly prefer to experiment at p_n, we have

$$\begin{aligned}v^M(p_n) \quad &= \quad (1-\delta)E_{p_n} + \delta[q(1-(1-\pi)^2) + (1-q)\pi]p_nE_1\\ &\qquad +\delta(1-q)(1-p_n\pi)v^M(p_{n+1}) + \delta q[(1-\pi)^2p_n + 1 - p_n]v^M(p_{n+2})\\ &\leq \quad (1-\delta)E_{p_n} + \delta[q(1-(1-\pi)^2) + (1-q)\pi]p_nE_1\\ &\qquad +\delta(1-q)(1-p_n\pi)\tilde{v}(p_{n+1}) + \delta q[(1-\pi)^2p_n + 1 - p_n]\tilde{v}(p_{n+2}).\end{aligned}$$

The right-hand side is linear in q and equals $\tilde{v}(p_n)$ for $q = 1$. It thus suffices to show that the right-hand side is increasing in q, which is equivalent to

$$[1-(1-\pi)^2-\pi]p_nE_1 - (1-p_n\pi)\tilde{v}(p_{n+1}) + [(1-\pi)^2p_n + 1 - p_n]\tilde{v}(p_{n+2}) > 0.$$

As $\tilde{v}(p_{n+1}) < p_{n+1}\pi E_1 + (1-p_{n+1}\pi)\tilde{v}(p_{n+2})$, moreover, it is enough to show that

$$\begin{aligned}&[1-(1-\pi)^2-\pi]p_nE_1 - (1-p_n\pi)p_{n+1}\pi E_1\\ &\quad + \ [(1-\pi)^2p_n + 1 - p_n - (1-p_n\pi)(1-p_{n+1}\pi)]\tilde{v}(p_{n+2}) \geq 0.\end{aligned}$$

As $(1-p_n\pi)p_{n+1} = p_n(1-\pi)$, this is is easily seen to hold as an equality. □

References

Acemoglu, D., K. Bimpikis and A. Ozdaglar (2011): "Experimentation, Patents, and Innovation," *American Economic Journal: Microeconomics*, 3, 37–77.

Athey, S. and K. Bagwell (2008): "Collusion with Persistent Cost Shocks," *Econometrica*, 76, 593–540.

Bergemann, D. and U. Hege (1998): "Venture Capital Financing, Moral Hazard and Learning," *Journal of Banking and Finance*, 22, 703–735.

Bergemann, D. and U. Hege (2005): "The Financing of Innovation: Learning and Stopping," *RAND Journal of Economics*, 36, 719–752.

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.

Besanko, D. and J. Wu (2012): "The Impact of Market Structure and Learning on the Tradeoff between R&D Competition and Cooperation," forthcoming at *Journal of Industrial Economics*.

Blume, A. and P. Heidhues (2006): "Private Monitoring in Auctions," *Journal of Economic Theory*, 131, 179–211.

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, 632–663.

Compte, O. (1998): "Communication in Repeated Games with Imperfect Private Monitoring," *Econometrica*, 66, 597–626.

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

Hopenhayn, H.A. and F. Squintani (2011): "Preemption Games with Private Information," *Review of Economic Studies*, 78, 667–692.

Kandori, M. and H. Matsushima (1998): "Private Observation, Communication and Collusion," *Econometrica*, 66, 627–652.

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.

Malueg, D.A. and S.O. Tsutsui (1997): "Dynamic R&D Competition with Learning," *RAND Journal of Economics*, 28, 751–772.

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

Nash, J. (1951): "Non-Cooperative Games," *Annals of Mathematics*, 54, 286-295.

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.