Bandits in the Lab*

Johannes Hoelzemann† Nicolas Klein‡

Current Version: March 3, 2019

Abstract

We experimentally implement a dynamic public-good problem, where the public good in question is the evolving information about agents' common state of the world. Specifically, we test Keller, Rady, and Cripps (2005)' game of strategic experimentation with exponential bandits in the laboratory. We find strong support for the prediction of free-riding because of strategic concerns. We also find strong evidence for behavior that is characteristic of Markov Perfect Equilibrium: non-cut-off behavior, lonely pioneers and frequent switches of action.

JEL Classification: C73, C92, D83, O32
Keywords: Dynamic Public-Good Problem, Strategic Experimentation, Exponential Bandits, Learning, Dynamic Games, Markov Perfect Equilibrium, Laboratory Experiments, Eye Tracking

---

*Financial support by the *UNSW Bizlab* and the *School of Economics at UNSW Sydney* is gratefully acknowledged. Nicolas Klein also gratefully acknowledges financial support from the *Fonds de Recherche du Québec - Société et Culture* and the *Social Sciences and Humanities Research Council of Canada*. This research has been approved by the Human Research Ethics Committee of UNSW Sydney under approval number HC 17069. We thank Francisco Alvarez-Cuadrado, Arthur Campbell, Juan Carlos Carbajal, Gary Charness, Michele De Nadai, Guidon Fenig, Denzil Fiebig, Gigi Foster, Ben Greiner, Raphael Godefroy, Gabriele Gratton, Yoram Halevy, Richard Holden, John Ledyard, Barton Lee, Hongyi Li, Erkut Ozbay, Sven Rady, Adam Sanjurjo, Andy Skrzypacz, Colin Stewart, and Tom Wilkening for helpful comments. A substantial part of this paper took shape while both authors enjoyed the hospitality of the Institute for Markets and Strategy at WU Vienna.

†University of Toronto. Mailing address: University of Toronto, Department of Economics, Max Gluskin House, 150 St. George Street; Toronto, ON M5S 2E9, Canada. e-mail: j.hoelzemann@utoronto.ca.

‡Université de Montréal and CIREQ. Mailing address: Université de Montréal, Département de Sciences Économiques, C.P. 6128 succursale Centre-ville; Montréal, H3C 3J7, Canada. e-mail: kleinnic@yahoo.com.

1 Introduction

Innovation and social learning are often the work of pioneers, who, by bearing the costs of experimenting with a new approach, create informational spill-overs for others. Whether we consider R&D, resource exploration, or the testing of a new drug, the information produced by a relatively small set of agents benefits a much larger group of agents. Indeed, R&D is universally recognized as an important factor of economic growth (Romer 1990; Grossman and Helpman 1993). An economy's productivity level depends on innovation, which is driven by knowledge emerging from cumulative R&D experience as well as an economy's overall knowledge stock (Griliches 1988; Coe and Helpman 1995). It is thus important for economists to analyze pioneers' incentives for information production in the presence of informational spill-overs.

There exists a vast experimental literature on games which examines the willingness to contribute to (linear continuous) public goods (for surveys, see Ledyard 1995 and Chaudhuri 2011; and for a meta-analysis, see Zelmer 2003). In these environments, payoff-maximizers' dominant strategy is to contribute none of their endowment to a group activity. The typical environment is such that it creates a social dilemma, leading to zero contribution to the group activity, while, in the efficient outcome, each player contributes his entire endowment. For many decades, economists have attempted to experimentally test this trade-off and to analyze factors that facilitate increased cooperation in such social-dilemma situations in the lab (e.g., Fehr and Gächter 2000 and Ambrus and Greiner 2012).[1]

In this paper, the public good agents contribute to is the dynamically evolving *information* about agents' common state of the world. Indeed, economic agents often endeavor to learn over time about some payoff-relevant aspect of their environment. Think, for instance, of a pharmaceutical company conducting costly clinical trials to find out the effectiveness of a drug. Learning often requires a costly investment in information acquisition, so that agents face a dynamically evolving trade-off on how much information to acquire. Indeed, in light of the signals it receives, the pharmaceutical company will revise its beliefs and decide whether to incur the costs necessary to acquire additional information by continuing its trials, or to give up.

Our setting differs from the previous experimental literature in that agents' incentives for free-riding depend on the information available to them, which evolves over time as a result of previous choices. In the theory literature, multi-armed bandit models have become canonical to study information producers' dynamic trade-offs. At each point in time, a decision maker either optimally exploits the information he already has, or he decides to invest in exploration in order to make better future decisions. Until fairly recently, the literature focussed on the trade-off of an individual decision maker acting in isolation. Bolton and Harris (1999) and Keller, Rady, and Cripps (2005) (subsequently: KRC) have extended the individual choice problem to a multi-player continuous-time framework. There now appears a strategic component to the information-acquisition problem, in that other players now also benefit from the information acquired at a cost by a given player. To make the

[1] For early experimental studies, see Kim and Walker (1984), Isaac, Walker, and Thomas (1984), Isaac, McCue, and Plott (1985), Isaac and Walker (1988a,b), and Andreoni (1988). For early studies embedded in the sociology literature, see Marwell and Ames (1979, 1980, 1981).

problem tractable, these papers are focussing on the choice between a safe arm, yielding a known payoff, and a risky arm, which yields payoffs following a stochastic process. The time-invariant quality of this risky arm can be good or bad. If it is good (bad), it dominates (is dominated by) the safe arm. Whether the risky arm is good or bad is initially unknown and can only be found out by trying it out over time. Trying it out is costly, however, as it means forgoing the safe payoff. As the quality of the risky arm is assumed to be the same across players, and players can observe each other's actions and payoffs, there is a positive informational externality associated with a player's use of the risky arm. This gives rise to a dynamic public-good problem in the form of dynamically evolving information about agents' common state of the world.

While the game-theoretical analysis of these problems will lead to multiple equilibria, it has nonetheless yielded many sharp qualitative behavioral predictions. Yet, empirical evidence for these predictions has thus far been scarce. Indeed, the dynamic nature of the problem and the continuous-time setting underlying its theoretical analysis raise some challenges both for the collection of field data and the experimental implementation in the laboratory. To the best of our knowledge, we are the first to implement an experimental test of continuous-time strategic-experimentation models in the laboratory. Our goal in doing so is twofold. Firstly, we want to test whether the bandit models correctly predict agents' behavior "in the model", by making our subjects face a setting closely resembling KRC's. This is of course a necessary, but by no means sufficient, condition for us to have "the right model" to approach these questions with. Secondly, we aim to shed some light on which of the multiple equilibria seem best-suited to capture actual behavior.

Our analysis relies on comparing the behavior of our experimental subjects in groups where the quality of the risky arm was known to be the same for all partners (which we call the *strategic treatment*) to that of groups where its quality was iid across members, the *control treatment*. When the quality of the risky arm is known to be the same across players, rational agents will take into account the result of their partners' experimentation when updating their beliefs. As they can learn from what others are doing, they have an incentive to induce others to behave in certain ways so they may learn from it. There is thus some strategic interaction across players, even though a player's payoffs depend only on his own action and the common state of the world, i.e., there are no payoff externalities.

Specifically, we use the simplest formalization of the continuous-time strategic-experimentation framework, KRC's exponential-bandit set-up, as our theoretical benchmark. In this setting, a bad risky arm never yields any payoff, while a good risky arm gives lump-sum payoffs at the jumping times of a Poisson process. Thus, whenever the risky arm is used without a success, players gradually grow pessimistic about its quality; as soon as they observe a success, they know for sure that the risky arm is good.

KRC analyze Markov perfect equilibria (MPE) with the players' common posterior belief as a state variable.[2] While there is a continuum of MPE, all equilibria make two fundamental qualitative predictions regarding players' behavior: As information is a public good, players will produce too little of it. Furthermore, it is predicted that all players will

[2]These are perfect Bayesian equilibria where a player's action choice depends on the history only via the common posterior belief.

not use a simple cut-off strategy in equilibrium. A cut-off strategy is defined by a unique threshold belief above which it prescribes risky play, while prescribing safe play below it.

Hörner, Klein, and Rady (2014) (subsequently: HKR) analyze non-Markovian equilibria, i.e., perfect Bayesian equilibria (PBE) in which a player's action choice can depend on the history in more complex ways. They show that free-riding prevails in all PBE as well. Moreover, the average-payoff maximizing PBE is strongly symmetric and has a particularly simple structure: Players play a cut-off strategy (on the path of play), applying the same cut-off as a single agent.

Our empirical tests are designed to contrast the qualitative predictions of MPE with those of the best PBE, which is a natural candidate for a focal equilibrium both because it maximizes players' average equilibrium payoffs and because it has a particularly simple structure. In a first step, we show that the informational externality indeed impacts subjects' behavior: the average experimentation intensity is lower, and subjects' payoffs are higher, in the strategic treatment. Secondly, we find strong evidence of the kind of qualitative behavior predicted by MPE as opposed to the simpler behavior predicted by the best PBE, with players' adopting more sophisticated behaviors than cut-off strategies in the strategic treatment. Indeed, players switch much more between safe and risky, and use cut-off strategies much less frequently, than they do in the control treatment. Moreover, there is a larger proportion of time during which exactly one player is playing risky in the strategic treatment. This is particularly striking, as the best PBE would predict this proportion to be nil, while it is always strictly positive in *any* of KRC's (infinitely many) MPEs. All of these effects are more pronounced for two-player groups than for groups of size three, and are fully consistent with the players' switching between the roles of pioneer and free-rider, which characterizes equilibrium play at intermediate beliefs in KRC and differentiates it, e.g., from the best PBE in HKR.

Our game is of course very complicated, so that we cannot reasonably expect subjects to be able to compute equilibrium strategies. Yet, subjects' experimentation efforts are clearly decreasing with the incremental arrival of bad news in the form of unsuccessful previous experimentation. This would suggest that, even though subjects could of course not be expected to update using Bayes' rule precisely in continuous time, they were nonetheless reacting to the dynamically evolving incentives. Furthermore, we are documenting behavior that is very much in line with the sophisticated coordination required by MPE play, as opposed, e.g., to the simpler structure of the best PBE.

The rest of the paper is organized as follows: Section 2 reviews some additional related literature; Section 3 explains the KRC model in more detail; Section 4 sets out our experimental implementation; Section 5 discusses our main findings, and Section 6 concludes. Appendix A breaks down the analysis to the individual games subjects played and Appendix B exhibits and explains the interface our experimental subjects were using. Appendix C reproduces the instructions the subjects received.

2 Literature Review

The bandit problem as a stylized formalization of the trade-off between exploration and exploitation goes back to Thompson (1933) and Robbins (1952). It was subsequently analyzed, amongst others, by Bellman (1956) and Bradt, Johnson, and Karlin (1956). Its first application to economics was in Rothschild (1974), who analyzed the price-setting problem of a firm facing an unknown demand function. Gittins and Jones (1974) showed that, if arms are stochastically independent of each other and the state of only one arm can evolve at any one time, an optimal policy in the multi-armed bandit problem is given by the so-called "Gittins Index" policy. For this policy, one can consider the problem of stopping on each arm in isolation from the other arms. The value of this stopping problem is the so-called *Gittins Index* for this arm. Now, an optimal policy consists of, at each point in time, using the arm with the highest Gittins Index. Presman (1990) calculated the Gittins Index for the case in which the underlying stochastic process is a Poisson process. Bergemann and Välimäki (2008) give a survey of this literature.

Bolton and Harris (1999, 2000) were the first to consider the multi-player version of the two-armed bandit problem. While they assumed that the underlying stochastic process was a Brownian motion, KRC analyzed the corresponding problem with exponential processes. This model proved to be more tractable and is underlying our theoretical hypotheses. While the previous papers focussed on MPE, HKR extended the equilibrium concept beyond Markov perfect equilibrium.[3]

We are aware of only one other experimental investigation of a strategic-experimentation problem with bandits, by Boyce, Bruner, and McKee (2016). Their setting is specifically designed to test for strategic free-riding in a two-player, two-period context. Coordination issues are assumed away in that one player was known to have lower opportunity costs for playing risky than the other, so that it was clear which player ought to play the role of pioneer (and that of free-rider respectively) in the first period. Moreover, in Boyce, Bruner, and McKee (2016)'s experiment, subjects faced ambiguity concerning the type of the risky arm. Indeed, they were not told a prior probability of the risky arm's type, which allows for an explanation of subjects' behavior that relies on their priors and ambiguity attitudes. Our investigation, by contrast, is focussed on how players resolve the coordination problems arising from strategic interaction, and on comparing the predictive powers of different equilibrium concepts for this purpose. It was our goal to construct an experimental setting that was as close as possible to the continuous-time setting that has been extensively studied in the theoretical literature. Our subjects all face the same decision problem and are given a Bayesian prior at the outset. As they interact in (inertial) continuous time with a stochastic and unknown deadline, their action spaces are essentially continuous.

[3] Many variants of the multi-player bandit problem have been analyzed since. In Keller and Rady (2010), a bad risky arm also sometimes yields a payoff. In Klein and Rady (2011), the quality of the risky arm is negatively correlated across players. Klein (2013) introduces a second risky arm, with a quality that is negatively correlated with that of the first. In Keller and Rady (2015), the lump-sum payoffs are costs to be minimized. Rosenberg, Solan, and Vieille (2007) and Murto and Välimäki (2011) analyze the case of privately observed payoffs, while Bonatti and Hörner (2011) investigate the case of privately observed actions. Bergemann and Välimäki (1996, 2000) consider strategic experimentation in buyer-seller settings. Hörner and Skrzypacz (2016) give a survey of this literature.

The only other papers we are aware of that conduct experimental tests of bandit problems consider exclusively various single-agent problems without strategic interdependencies among experimental subjects. Banks, Olson, and Porter (1997) experimentally implement bandits with simple win-lose (Bernoulli) payout distributions, and test whether subjects value information gained through experimentation. In their experimental design the expected payoff of one arm is known, while the other is unknown. Experimentation is observed more in one treatment where initial selection of the unknown arm is optimal compared to the treatment where experimentation is suboptimal. These results suggest that subjects' behavior is consistent with the normative predictions and that subjects value the information gained through costly experimentation.

A couple of papers by Meyer and Shi (1995) and Gans, Knox, and Croson (2007) employ a different experimental approach, aiming at identifying choice patterns that are consistent with a list of simple decision rules. Meyer and Shi (1995) test decision under ambiguity and use experimental data to generate hypotheses about subject's possible heuristics. While observed choice behavior indicates Bayesian updating of priors, their experimental subjects also exhibit a strong bias toward myopic choices. Among all decision rules considered, the simple stick-with-a-winner strategy fits the data best. Gans, Knox, and Croson (2007) consider a list of simple discrete-choice models in a two-armed bandit set-up. The optimal choice model could not explain their experimental data well. To predict choice behavior, simpler heuristic models are proposed. Indeed, backward-looking strategies which predict switching arms after a fixed number of consecutive failures best explain the observed choices.

Anderson (2001, 2012) uses arms with payout distributions, e.g., simulated dice rolls and normally distributed rewards. He finds that subjects experiment less than would be optimal, and are willing to pay more for getting perfect information than theory would predict. In this set-up ambiguity aversion along with diffuse priors is identified as a driver of the observed behavior in the laboratory.

Rabanal, Rud, and Sharifova (2018) study an individual choice game of market entry and exit decisions under varying levels of uncertainty in an environment, where the trade-off is similar to a two-armed bandit problem with a safe arm and a risky arm. Hudja (2018) implements experimentally Strulovici (2010)'s collective experimentation model in continuous time. An individual experimentation problem is compared to a collective experimentation problem where groups of three players face a majority-vote. Fudenberg and Vespa (2018) analyze a signaling-game experiment and focus on the effect of how types are assigned. A bandit problem of their signaling game is employed as a robustness check in which subjects play against a computer.

3 The Theoretical Framework

We borrow our theoretical reference framework from KRC. There are $n \geq 1$ players, each of whom plays a bandit machine with two arms over an infinite horizon. One of the arms is safe, and yields a known flow payoff of $s > 0$ whenever it is pulled. The other arm is risky and can be either good or bad. If it is bad, it never yields any payoff. If it is good, it yields a lump sum of $h > 0$ at the jumping times of a Poisson process with parameter $\lambda > 0$. It is assumed that $g := \lambda h > s$. Players decide in continuous time which arm to pull. Payoffs are discounted at a rate $r > 0$. If they knew the quality of the risky arm, players would have a strictly dominant strategy always to pull a good risky arm and never to pull a bad one. They are initially uncertain whether their risky arm is good or bad. Yet, the only way to acquire information about the quality of the risky arm is to use it, which is costly as it implies forgoing the safe payoff flow s. The n players' risky arms are either all good or all bad. Players share a common prior belief $p_0 \in (0, 1)$ that their risky arms are good. Every player's actions as well as the outcomes of their actions are publicly observable; therefore, the information one player produces benefits the other players as well, creating incentives for players to free-ride on their partners' efforts. Players thus share a common posterior belief p_t at all times $t \in \mathbb{R}_+$. All the parameter values and the structure of the game are common knowledge.

The common posterior beliefs are derived from the public information via Bayes' rule. As a bad risky arm never yields any payoff, the first arrival of a lump sum fully reveals the quality of *all* players' risky arms. Thus, if a success on one of the players' risky arms is observed at instant $\tau \geq 0$, the common posterior belief satisfies $p_t = 1$ for all $t > \tau$. If no success has been observed until instant t, the common posterior belief satisfies

$$p_t = \frac{p_0 e^{-\lambda \int_0^t \sum_{i=1}^N k_{i,\tau} \, d\tau}}{p_0 e^{-\lambda \int_0^t \sum_{i=1}^N k_{i,\tau} \, d\tau} + 1 - p_0},$$

where $k_{i,\tau} = 1$ if player i uses the risky arm at instant τ and $k_{i,\tau} = 0$ otherwise.

KRC show in their Proposition 3.1 that, if players are maximizing the sum of their payoffs, all players $i \in \{1, \cdots, n\}$ choose $k_{i,t} = 1$ if $p_t > p_n^* := \frac{rs}{(r+n\lambda)(g-s)+rs}$, and $k_{i,t} = 0$ otherwise. Note that p_n^* is strictly decreasing in the number of players n. In particular, in the single-agent case ($n = 1$), the decision maker optimally sets $k_{1,t} = 1$ if $p_t > p_1^* := \frac{rs}{(r+\lambda)(g-s)+rs}$, and $k_{1,t} = 0$ otherwise.

KRC go on to analyze the game of strategic information acquisition, where each player maximizes his own payoff, not taking into account that the information he produces is valuable to the other players as well. They analyze perfect Bayesian equilibria in Markov strategies (MPE), i.e., strategies where a player's action after any history can be written as a time-invariant function $k_i(p)$ of the common belief at that history. It is shown that, for beliefs close to 1 (0), playing risky (safe) is a dominant action; for intermediate beliefs, players' effort levels are strategic substitutes. In any MPE with a finite number of switches, all players will set $k_i(p) = 0$ for all $p \leq p_1^*$ (see Proposition 6.1 in KRC). Moreover, it is shown that there exists no MPE in which all players play a cut-off strategy, i.e., a strategy that prescribes the use of the risky arm for beliefs above a single cut-off and that of the safe arm below. The intuition for this result is best described in the context of a two-player game. Indeed, suppose to the contrary that there existed an equilibrium in cut-off strategies. As there is a region of beliefs in which safe and risky are mutually best responses, both players cannot use the same cut-off in equilibrium; i.e., one player plays the role of pioneer, while the other one free-rides, throughout the belief region where safe and risky are mutually best responses. As he gets all his information for free in the relevant belief region, the free-rider's payoff function will be higher than the pioneer's. As a player's propensity to play risky is increasing in his own payoff, however, this would imply that the free-rider entered the region in which risky is dominant at a more pessimistic belief than the pioneer. Thus, the roles of pioneer and free-rider must switch at least once in equilibrium.

HKR extend the analysis to non-Markovian PBE. They show that on the path of play in the average-payoff maximizing PBE, all players set $k_i(p) = 1$ for all $p > p_1^*$, and $k_i(p) = 0$ otherwise. Thus, in stark contrast to the simple structure of the single-agent optimum or HKR's average-payoff maximizing PBE, every MPE has the property that, for intermediate beliefs, players change roles between experimenter and free-rider at least once. As a matter of fact, KRC show that, for any given number of role changes greater than, or equal to, one, there exists an MPE with that number of role changes. A behavioral prediction of MPE is thus that players change roles for intermediate beliefs at least once.[4]

4 Parametrization and Experimental Design

4.1 Experimental Implementation

In our experimental treatments, the number of players will be $n = 2$ or $n = 3$. Indeed, as we have seen in the previous section, the overall amount of experimentation, as measured by the threshold belief p_1^*, at which all experimentation stops, is independent of the number of players n. By contrast, the efficient threshold p_n^* is decreasing in n. Moreover, the complexity of the coordination required to play MPE is higher the greater the number of players n. We choose the discount rate $r = \frac{1}{120}$. To implement the infinite-horizon game in the laboratory, we end the game at the first jump time of a Poisson process with parameter r.[5] With one unit of time corresponding to a second in our experimental implementation, games thus last 120 seconds in expectation. Ours being a rather complicated game that places high demands on subjects' concentration, our goal was to limit the duration of the game, while at the same time allowing for the collection of a wealth of data. We set the probability that the risky arm is good $p_0 = \frac{1}{2}$, the safe payoff $s = 10$, the lump-sum amount paid out by a good risky arm $h = 2500$, and the arrival rate of lump sums on the good risky arm $\lambda = \frac{1}{100}$. Thus, $25 = g > s = 10$. With this parametrization, the game starts in the belief region where risky is a dominant action; if no breakthrough arrives, play then moves into the belief region where safe and risky are mutually best responses, before entering the region where safe is dominant. The realizations of all random processes were simulated ahead of time.[6] We generated six different sets of realizations of the random parameters,

[4] KRC show that there is also a unique symmetric MPE, where players use the risky (safe) arm with an interior intensity $k(p) \in (0, 1)$ $(1 - k(p))$ throughout the belief region where risky and safe are mutually best responses. As we wanted to keep the decision problem as simple as possible, our subjects do not have the option of choosing interior experimentation levels. Please also see our discussion in the Conclusion.

[5] Subjects knew that the end time of the game corresponded to the first jumping time of a Poisson process with parameter r but did not know the realization of this process at any time before the game ended. In particular, the time axis they saw on their computer screens gradually grew longer as time progressed, so that they could not infer the end date. Please see Appendices B and C for details and for the instructions the subjects received.

[6] As all our stochastic processes are Lévy processes, simulating their realizations ahead of time is equivalent to simulating them as the game progresses. In order to increase the computational efficiency of the

corresponding to six different games each of our subjects played. To make our findings more easily comparable, we have kept the same realizations for both the strategic and the control treatments.[7] In keeping with the theoretical predictions, we have endeavored to implement our experimental investigation in continuous time, subject to the restrictions imposed by the available computing power.[8]

Subjects were randomly assigned to groups of $n = 2$ or $n = 3$ players. We used a between-subject design: Each group was randomly assigned either to a control treatment or to a strategic treatment, and played the six games in random order. To ensure a balanced data-collection process, we replicated any order of the six games that was used for k ($k \in \{1, \cdots, 10\}$) groups in the strategic treatment for k groups in the control treatment as well. Subjects could see their fellow group members' action choices and payoffs on their computer screens. They had to choose an action before the game started and could switch their action at any point in time by clicking on the corresponding button with their mouse.[9]

All experimental sessions took place in July and August 2017 at the BizLab Experimental Research Laboratory at UNSW Sydney. All subjects were recruited from the university's subject pool and administered by the online recruitment system ORSEE (Greiner 2015). All participants were native speakers of English. In total, 100 subjects, 46 of whom were female, participated in 60 sessions. The participants' age ranged from 18 to 35 years, with an average of 20.78 and a standard deviation of 2.43. Because the implementation was computationally very intensive and because we wanted to collect eye-tracking data, only between 2 and 3 subjects participated at a time in each session. Upon arrival, participants were seated in front of a computer at desks which were separated by dividers to minimize potential communication. Participants received written instructions and had the opportunity to ask questions.[10] After the subjects had successfully completed a simple comprehension test, the eye-tracking devices were calibrated, after which the subjects started the experiment. The experiment was programmed in zTree (Fischbacher 2007). At the end of the experiment, we collected some information on participants' demographic attributes and risk attitudes. They were then privately paid their cumulated experimental earnings from one randomly selected game in cash (with a conversion rate of E\$ 100 = AU\$ 1) plus a show-up fee of AU\$ 5. No subject was allowed to participate in more than one session. The average session lasted about 50 minutes, with average earnings of AU\$ 23.86 (with a standard deviation of AU\$ 9.95).

4.2 Behavioral Hypotheses

One of the main theoretical predictions of both MPE and PBE is that players use the risky arm less in a strategic setting than in a situation in which they are single players. This is because players *free-ride* on the information their partners are producing. Indeed, players are predicted to play safe at all beliefs $p \leq p_1^*$ in all these instances, while efficiency would

implementation, we chose to simulate them ahead of time.

[7] Details are available from the authors upon request.

[8] Thus, our implementation corresponds to the "Inertial Continuous-Time" setting in Calford and Oprea (2017).

[9] Please see the Appendix B for more details and screen shots.

[10] The instructions handed out to all participants can be found in Appendix C.

require that they play risky at all beliefs $p > p_n^*$, where $p_n^* < p_1^*$. Single players and players playing the best PBE should play risky at all beliefs $p > p_1^*$, i.e., in the average-payoff maximizing PBE, players on path adopt the same cut-off behavior as a single agent. In any MPE, by contrast, since at least one player is not playing a cut-off strategy, at least one player will play safe at some beliefs above p_1^*. Indeed, it is possible to derive a lower bound $p^\ddagger \in (p_1^*, p^m)$, where $p^m := \frac{s}{g}$ is a myopic player's cut-off belief, such that, for all beliefs in $(p_1^*, p^\ddagger)$, at least one player plays safe. Indeed, as KRC show (their Equation (6), p.49), it is a best response for player i to play safe if and only if his value function $u_i(p)$ satisfies $u_i(p) \leq s + K_{-i}(p)c(p)$, where $K_{-i}(p) := \sum_{j \neq i} k_j(p)$ is the number of players other than i who play risky at belief p, and $c(p) := s - pg$ is a player's myopic opportunity cost for playing risky, given the belief p. An upper bound on a player's equilibrium value function u_i is given by V_{n,p_1^*}, the value function of all players playing risky on $(p_1^*, 1]$, and safe on $[0, p_1^*]$. Thus, a lower bound $p^\ddagger$ is given by the unique root $V_{n,p_1^*}(p^\ddagger) - s - (n-1)c(p^\ddagger) = 0$. By the same token, we can derive an upper bound $\bar{p}$ on the lowest belief at which risky is a dominant action. For this, we use the fact that the single-agent value function V_1^* constitutes a lower bound on a player's equilibrium value function u_i, and find our upper bound $\bar{p}$ as the unique root $V_1^*(\bar{p}) - s - (n-1)c(\bar{p}) = 0$.

With our numerical parameters, $p^m = 0.4$, $\bar{p} \approx 0.3578$ ($\bar{p} \approx 0.3742$) if $n = 2$ ($n = 3$), $p^\ddagger \approx 0.3428$ ($p^\ddagger \approx 0.3609$) if $n = 2$ ($n = 3$), $p_1^* \approx 0.2326$, $p_2^* \approx 0.1031$, and $p_3^* \approx 0.0535$. As $p_0 = 0.5 > 0.4 = p^m$, players start out with a belief that makes playing risky the dominant action. If, in the strategic treatment, n players were uninterruptedly playing risky and there was no breakthrough, the belief would drop to p^m after $40.6/n$ seconds, to our upper bound in the game with $n = 2$ players ($n = 3$ players) $\bar{p}$ after $58.5/n$ ($51.5/n$) seconds, to our lower bound in the game with $n = 2$ players ($n = 3$ players) $p^\ddagger$ after $65.0/n$ ($57.0/n$) seconds, to p_1^* after $119.4/n$ seconds, to p_2^* after $216.4/n$ seconds, and to p_3^* after $287.4/n$ seconds. For the control treatment, the same times apply with $n = 1$.

4.2.1 Free-Riding

Let $\hat{T}$ be the time of a first breakthrough or the end of the game, whichever arrives first. In order to measure the prevalence of free-riding, we investigate the behavior of the *average experimentation intensity*, where, following KRC, we define the *experimentation intensity at instant t* as $\sum_{i=1}^n k_{i,t}$. Note that, in the control treatment, a player conforming to the theoretical prediction will always play risky until his belief hits p_1^*. In the strategic treatment, at least one of them will switch to safe at a belief strictly above p_1^* if they play an MPE. In the best PBE, they both play risky until the belief p_1^* is reached. Furthermore, conditionally on no success arriving, beliefs will decrease faster in the strategic setting, as player i's belief also decreases in response to player j's hapless experimentation. As both effects go in the same direction, the average experimentation intensity should be lower in the strategic setting, whether players play MPE or the best PBE. We thus formulate the following

Hypothesis 1. *The average experimentation intensity* $\frac{\int_0^{\hat{T}} \sum_{i=1}^n k_{i,t}\, dt}{n\hat{T}}$ *is significantly lower in the strategic treatment than in the control treatment.*

Our game is one of purely (positive) informational externalities; i.e., players always have the option of ignoring the additional information they get for free from their partner(s). This observation motivates our following hypothesis

Hypothesis 2. *Players' average final payoffs are higher in the strategic treatment.*

4.2.2 MPE vs. Best PBE

As explained above, KRC predict that subjects will use cut-off strategies in the control treatment, whereas at least one player will not use a cut-off strategy in the strategic setting if MPE is played. By contrast, HKR show that cut-off behavior prevails on path in the strategic setting also if the best PBE is played. *Cut-off behavior* consists in a player's playing risky at the outset, and continuing to play risky until his risky arm is revealed to be good, the game ends, or he switches to the safe action, and continues to play safe until the game ends or his risky arm is revealed to be good. To investigate whether, qualitatively, the behavior predicted by MPE prevailed, we shall examine the following

Hypothesis 3. *The frequency of cut-off behavior is significantly higher in the control treatment than in the strategic treatment.*

In order further to discriminate between simple MPE and the best PBE, we measure the proportion of time (before a first breakthrough) during which exactly one of the players plays risky. Theory would predict this proportion to be nil both in the control treatment and in HKR's best PBE. In contrast, it is positive in any of KRC's MPEs, thus providing a sharp test of MPE-behavior. We formulate the following

Hypothesis 4. *The proportion of time before a first breakthrough during which exactly one player plays risky is higher in the strategic treatment than in the control treatment.*

The non-cut-off behavior predicted by MPE moreover implies that players should switch arms more often in the strategic treatment. Yet, as noted above, learning also tends to be faster in the strategic setting, so that beliefs may more quickly reach the threshold at which the player will want to change his action. While this effect would add to making switching more prevalent in the strategic treatment, a substantially higher number of switches in the strategic treatment would provide further evidence in favor of subjects' adopting MPE behavior. Indeed, recall that players are predicted to switch action at most once in both the control treatment and the best PBE, while, for any number of role changes, there exists an MPE with that number of role changes, as KRC show. For a two-player game, this e.g., implies that one of the players must switch actions at least twice, with the other one switching once, before p_1^* is reached.[11] To control for the effect that, the longer the game goes on, the more time players have to switch actions, we define the *incidence of switches* as the number of a player's switches in a given game per unit of effective time, where *effective time* is understood as the time before the game ends or the player's risky arm is revealed to be good, whichever happens first. Thus, we shall check the following

[11] Note that if players were to play the best PBE and the game happened to stop at a time such that p_1^* is only reached in the strategic treatment, we should observe exactly one switch per player in the strategic treatment and none in the control treatment. Therefore, a higher number of switches in the strategic treatment is not inconsistent with players' playing the best PBE. However, the magnitude of the effect, which we report in Section 5, cannot be accounted for by this explanation.

Hypothesis 5. *The incidence of switches is significantly higher in the strategic treatment than in the control treatment.*

The coordination required by MPE play is decidedly more complex than that which underlies the best PBE. Moreover, this complexity increases with the number of players for the former, while it remains unchanged for the latter. Indeed, recall that the latter implies cut-off behavior on the path of play, while the former is characterized by frequent role changes. This is inherently all the more complicated the more players there are. Therefore, one might expect that MPE-type behavior was more prevalent with groups of $n = 2$ players than with groups of size $n = 3$. Hence, our following

Hypothesis 6. *In the strategic treatment, there is more cut-off behavior and fewer switches, while single pioneers are less prevalent, if $n = 3$ than if $n = 2$.*

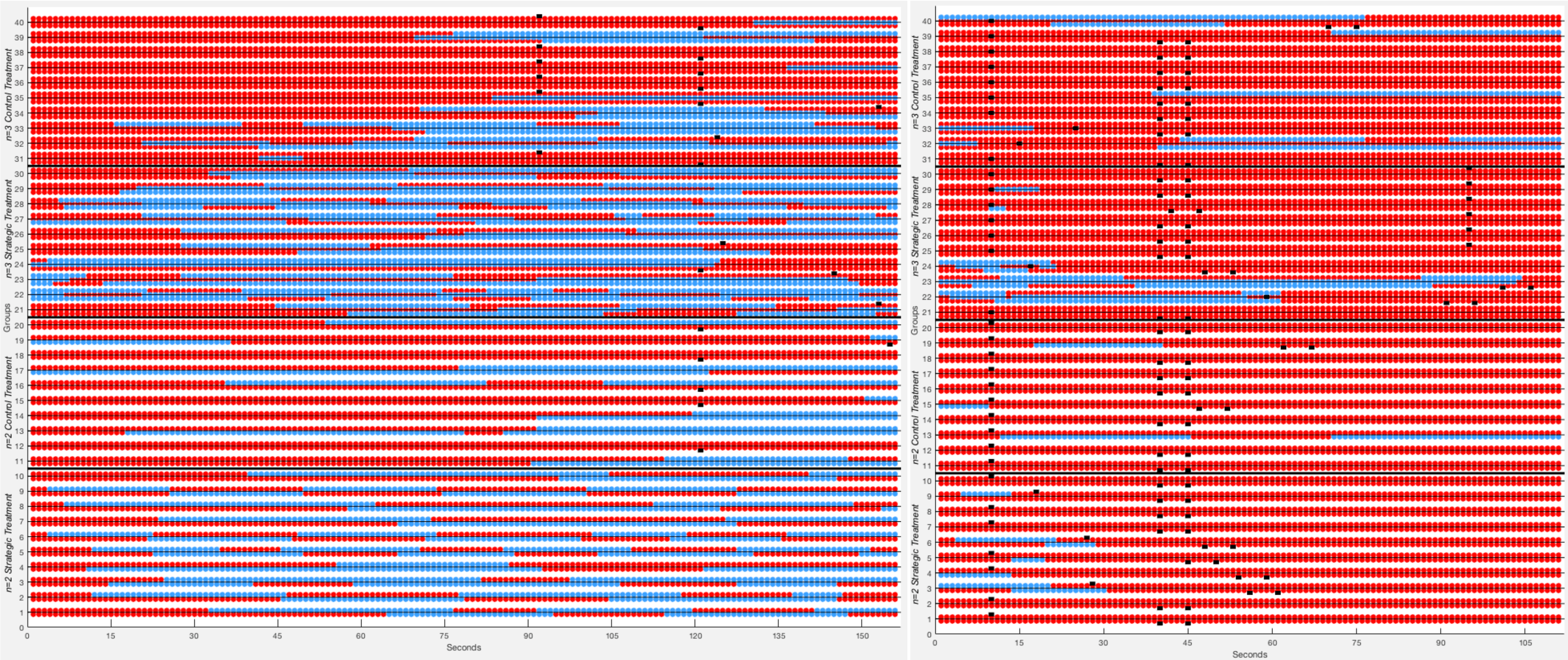
5 Experimental Results

5.1 Overview

Figures 1, 2 and 3 display the evolution of players' action choices over all six games. Players' actions are described by dots, the width of which corresponds to one second of time. For each of the six games, we conducted four treatments with ten groups each, the parameters of which (i.e., their duration, the quality of the risky arm and the timing of successes on the risky arm in case it was good) we had simulated ahead of time, as explained in Section 4. As the figures show, the duration of the games ranged from 32 seconds for Game 5 to 230 seconds for Game 4. As is furthermore evident from the figures, players change their behaviors over time. While often playing risky at the beginning, players seem to grow less inclined to use the risky arm the longer it has unsuccessfully been used before. This shows that our subjects adapted to the evolving information about their environment.

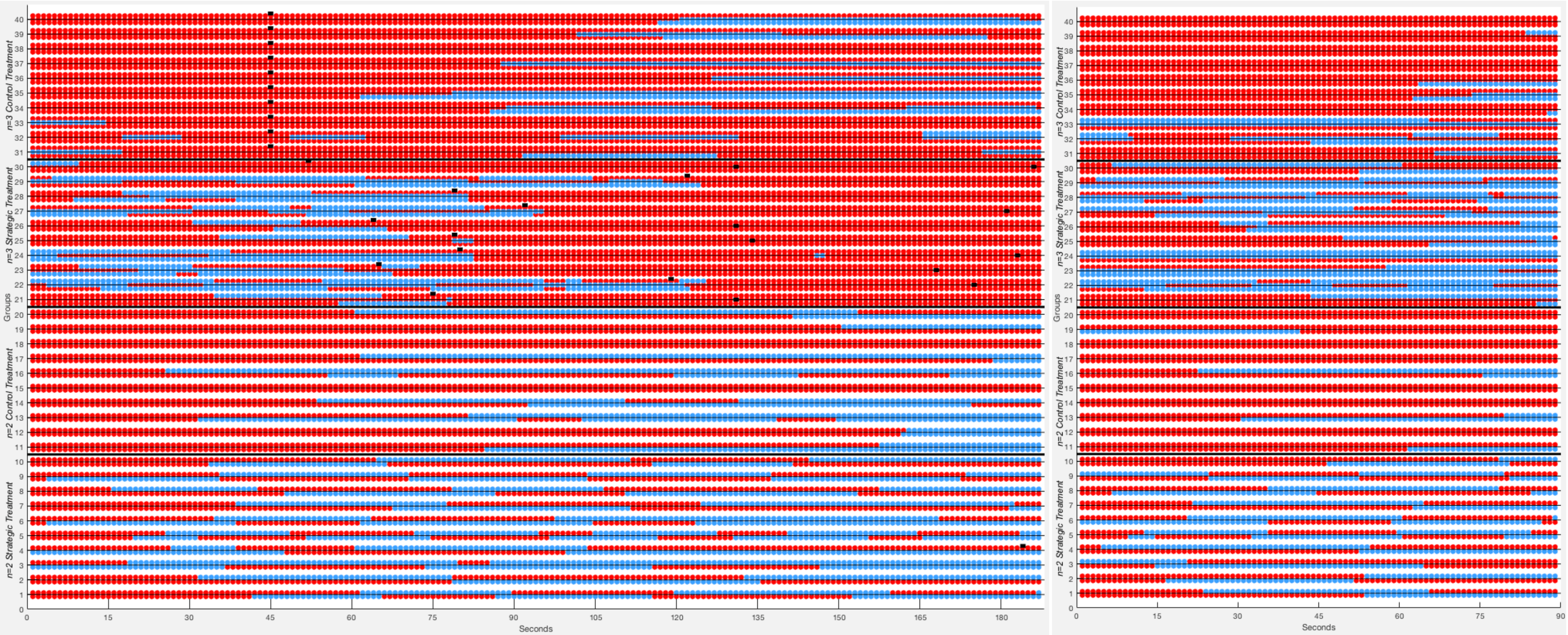
In this section, we conduct our analysis by averaging over the six games each subject played. Analysis of the individual games can be found in Appendix A.

Figure 1: Action Choices by Players over Time, Games 1 & 6



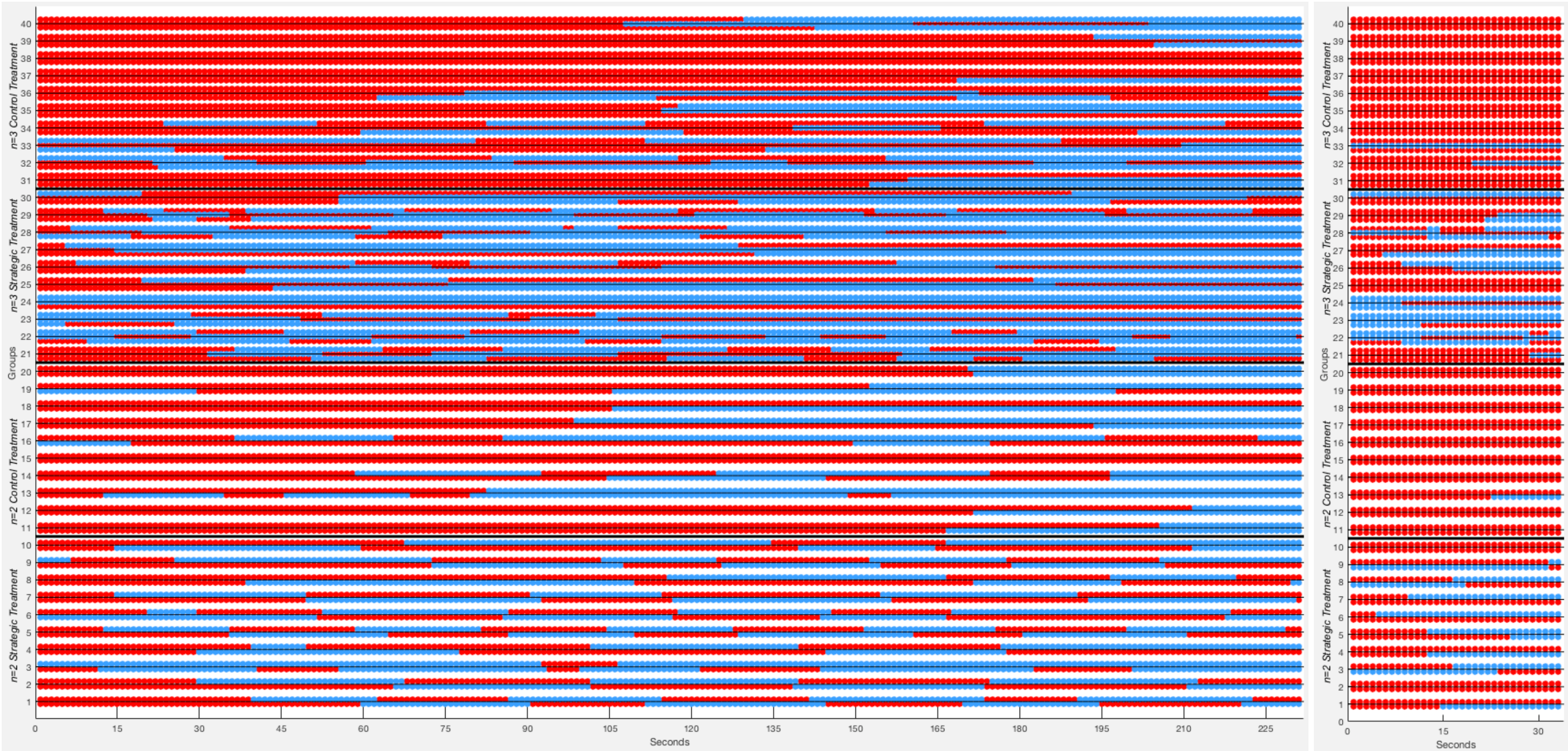
Games 1 and 6 are shown. Players' actions are described by dots, the width of which corresponds to one second of time. Groups 1-10 correspond to the strategic treatment for two-player groups; groups 11-20 are the corresponding control treatments. Groups 21-30 played the strategic treatment for three-player groups, while groups 31-40 were the corresponding control treatments. In each group, we refer to the lowermost player as 'player 1', while 'player 2' will denote the player right above, and 'player 3' is the uppermost player. The x-axis represents calendar time. A *red* dot indicates that a player is playing *risky* in a given second, while a *blue* dot indicates that the player is playing *safe*. A black square indicates a success.

Figure 2: Action Choices by Players over Time, Games 2 & 3



Games 2 and 3 are shown. Players' actions are described by dots, the width of which corresponds to one second of time. Groups 1-10 correspond to the strategic treatment for two-player groups; groups 11-20 are the corresponding control treatments. Groups 21-30 played the strategic treatment for three-player groups, while groups 31-40 were the corresponding control treatments. In each group, we refer to the lowermost player as 'player 1', while 'player 2' will denote the player right above, and 'player 3' is the uppermost player. The x-axis represents calendar time. A *red* dot indicates that a player is playing *risky* in a given second, while a *blue* dot indicates that the player is playing *safe*. A black square indicates a success.

Figure 3: Action Choices by Players over Time, Games 4 & 5



Games 4 and 5 are shown. Players' actions are described by dots, the width of which corresponds to one second of time. Groups 1-10 correspond to the strategic treatment for two-player groups; groups 11-20 are the corresponding control treatments. Groups 21-30 played the strategic treatment for three-player groups, while groups 31-40 were the corresponding control treatments. In each group, we refer to the lowermost player as 'player 1', while 'player 2' will denote the player right above, and 'player 3' is the uppermost player. The x-axis represents calendar time. A *red* dot indicates that a player is playing *risky* in a given second, while a *blue* dot indicates that the player is playing *safe*. A black square indicates a success.

5.2 Average Experimentation Intensities

One of the main qualitative predictions of the theoretical analysis is that players will tend to free-ride on the experimentation provided by their partners. To test for treatment differences non-parametrically, we apply two-sided Wilcoxon rank-sum (Mann-Whitney) tests, using group averages as independent observations. Table 1 lists the mean experimentation intensity observed in our four treatments.

Table 1: Average Experimentation Intensity

	Strategic Treatment		Control Treatment	
Group Size	Obs.	Experiment. Intensity	Obs.	Experiment. Intensity
$n = 2$	60	.594 [.186]	60	.818 [.212]
$n = 3$	60	.539 [.244]	60	.839 [.180]

Average [st. dev.] experimentation intensity using group averages.

As Table 1 reveals, the additional presence of one (two) perfectly positively correlated arms leads to lower experimentation intensities in all games. This is highly statistically significant in both settings with $n = 2$ and $n = 3$. The corresponding p-values in both cases are 0.0001.[12]

Under Hypothesis 1, players will use the risky arm less in the strategic treatment. The data provides support for this hypothesis.

Result 1. *The average experimentation intensity* $\frac{\int_0^{\hat{T}} \sum_{i=1}^{n} k_{i,t}\, dt}{n\hat{T}}$ *is significantly lower in the strategic treatment, as compared to the control treatment. This result holds for both* $n = 2$ *and* $n = 3$.

As we have mentioned above, information accumulation is potentially faster in the strategic treatment. Indeed, on account of the conditionally independent Poisson processes, the information acquired within a given unit of time is proportional to the number of players currently playing risky. Therefore, conditionally on no success arriving, players' beliefs will tend to decrease more quickly in the strategic setting, implying that more time will be spent at more pessimistic beliefs. To ensure that Result 1 is not solely due to this effect, we conduct our parameter tests separately by belief region. Specifically, we consider the belief regions $[\bar{p}, \frac{1}{2}]$, where risky is a dominant action, and $(p_1^*, p^\ddagger)$, where risky and safe are mutually best responses in MPE.[13] In the control treatment or if players were be-

[12]The Wilcoxon ranksum test treats group averages as independent observations. Yet, one might argue that players' action choices are not independent across subsequent games they play. As a robustness check, we additionally conduct a Wilcoxon test where we also average over all games for each group, thus yielding one independent data point across all games for each group of interacting subjects. The corresponding p-values for $n = 2$ ($n = 3$) are 0.0019 (0.0012).

[13]Besides the beliefs $(\frac{1}{2}, 1]$, which can never be reached in the absence of a success, the complementary set of these beliefs thus consists of the region $[0, p_1^*]$, where safe is a dominant action, and the (small) interval of beliefs $[p^\ddagger, \bar{p})$, which we have not assigned to either region. Indeed, as we explain in Section 4, we rely on conservative bounds in defining the "R dominant"and "Mutually BR" regions.

having according to the best PBE in the strategic setting, by contrast, all players should play risky in both regions. The following table (2) summarizes our findings by belief region.

As player 2 has a success after 9 seconds of using the risky arm, we omit Game 6 from these tables. We furthermore omit Game 5 from the tables for the "mutually BR" region, as this game lasts only 32 seconds, implying that the "mutually BR" region cannot be attained in the control treatment and only lasts for a few seconds in the strategic treatment, if it is attained at all. For Games 1-4, the missing observation for the "mutually BR" region corresponds to one three-player group in the control treatment that has not reached the "mutually BR" region.[14]

Table 2: Average Experimentation Intensity, by Belief Regions

		Strategic Treatment		Control Treatment	
Group Size	Belief Region	Obs.	Experiment. Intensity	Obs.	Experiment. Intensity
$n = 2$	R dominant	50	.675 [.222]	50	.899[.161]
$n = 2$	Mutually BR	40	.505 [.155]	40	.776 [.311]
$n = 3$	R dominant	50	.632 [.281]	50	.932 [.152]
$n = 3$	Mutually BR	40	.510 [.220]	39	.779 [.090]

Average [st. dev.] experimentation intensity using group averages.

The comparison of the strategic treatment with the control treatment shows that the average experimentation intensity is substantially lower in the strategic treatment, for *both* belief regions. The effect is statistically significant at the 1%-level for both belief regions, independently of group size, the p-values of the two-sided Wilcoxon ranksum test amounting to 0.0001. These results provide strong evidence that players are free-riding because of strategic considerations. Indeed, the information they provide in the strategic treatment is a public good; hence, they will provide too little of it. By contrast, the information they produce in the control treatment is a private good.

We now test whether players behave differently by belief region within a given regime. Recall that they should play risky in both in the control treatment. In the strategic treatment, however, while subjects should always play risky in the "R dominant" region, we should observe less risky play in the "Mutually BR" region, if they play MPE, and no difference if they play the best PBE. Our analysis shows that, in the strategic treatment for groups of $n = 2$, the average experimentation intensity is statistically significantly higher in the "R dominant"-region, with a p-value of 0.0001. In the corresponding control treatment, the difference is not statistically significant (p-value of 0.1088). In the setting with $n = 3$, while players tend to use the risky arm less in the "mutually BR" region than in the "R dominant" region, statistically no such effect can be established (p-value of 0.1296).

[14]In the control treatment, only some players of a given group may reach the "Mutually BR" region. We continue to include these groups in our data, without weighting the corresponding observations down. If we weighted groups in the "Mutually BR" region by the number of their members, average experimentation intensities would increase even further for both $n = 2$ and $n = 3$.

Thus, the contrast to groups of $n = 2$ would suggest that MPE-type behavior was more prevalent for the smaller group size.[15] The introduction of an additional player may have increased the complexity of coordination required, or subjects may not have been able to update their subjective beliefs with enough precision any longer to tell the regions apart. Overall, however, as far as free-riding is concerned, there do not seem to be any major differences between groups of size two and groups of size three.

5.3 Payoffs

Strategic interaction is predicted to arise among players as a result of (positive) informational externalities, i.e., the information produced by their partners allows players to make better decisions and hence to secure themselves higher payoffs. Thus, players' payoffs should be higher on average in the strategic treatment.

Table 3: Average Final Payoffs

	Strategic Treatment				Control Treatment			
Group Size	Obs.	Final Payoffs	Min	Max	Obs.	Final Payoffs	Min	Max
$n = 2$	60	1235.50 [1235.11]	0.00	3945.00	60	1030.75 [1272.16]	0.00	3870.00
$n = 3$	60	1420.28 [1045.41]	0.00	3363.33	60	981.22 [904.08]	0.00	2860.00

Average final payoffs using group averages.

Table 3 displays the average final payoffs using group averages across games for our four treatments. Average final payoffs are much higher in the strategic treatment than in the control treatment, for both group sizes. This is statistically significant. For $n = 2$ ($n = 3$), the corresponding p-values are 0.0674 (0.0001). Thus, our subjects indeed take advantage of the positive informational externalities in the strategic treatment, giving us

Result 2. *Players' average final payoffs are higher in the strategic treatment, for both group sizes.*

5.4 Eye-Tracking Data

As we have seen in the previous subsection, subjects' payoffs were markedly higher in the strategic treatment, which suggests that they were indeed able to take advantage of the positive informational externality. To study the players' information-acquisition processes further, we employ eye-tracking data obtained by using two (three) Tobii-TX300 eye trackers with a sampling rate of 300 Hz. The relative frequency of fixations corresponds to the relative importance of an information in the subject's decision-making process (Jacob and Karn 2003, Poole, Ball, and Phillips 2005). In our setting, eye fixations can thus provide information about the importance subjects assigned to the different payoff streams, which

[15] In the control treatment for $n = 3$, however, the difference between the belief regions is significant (p-value of 0.0005), for which there is no theoretical rationale.

revealed both a player's actions and payoffs.[16] We define a subject's fixation intensity as the total number of fixations on his own payoff stream, divided by the total number of all fixations (i.e., both on his own and on his partner's [partners'] payoff stream[s]) during a game before a breakthrough arrives or the game ends.

Table 4: Average Fixation Intensity

	Strategic Treatment		Control Treatment	
Group Size	Obs.	Fixation Intensity	Obs.	Fixation Intensity
$n = 2$	60	.614 [.087]	60	.865 [.090]
$n = 3$	60	.383 [.078]	60	.712 [.106]

Average [st. dev.] fixation intensity using group averages.

As Table 4 shows, the average fixation intensity is significantly lower in the strategic treatment. This is highly statistically significant for both group sizes (both p-values are 0.0001 for $n = 2$ and $n = 3$). The sophisticated coordination required by the switching of roles between pioneer and free-rider, which is characteristic of MPE and which we shall analyze in detail below, seems to force players to pay a lot of attention to their partner's (partners') behavior. This provides additional evidence that players behave strategically and try to learn from their partners' exploration efforts in the strategic treatments only. Furthermore, these results also indicate that subjects did indeed understand the simple, non-strategic, nature of the control treatment.

5.5 Cut-Off Behavior

As we have pointed out above, optimality in the individual decision-making problem in our control treatment implies cut-off behavior. The best PBE also features cut-off behavior on the path of play, while KRC have shown that there does not exist an MPE in cut-off strategies. This prediction of MPE is confirmed by our experiment, where subjects often play cut-off strategies in the control treatment, while they hardly ever do so in the strategic treatment. As it is not clear what it means for a group to engage in cut-off behaviour, we report each individual subject's decisions.

Result 3. *The frequency of cut-off behavior is higher in the control treatment than in the strategic treatment. We find evidence for both $n = 2$ and $n = 3$.*

Indeed, Table 5 shows that the frequency of cut-off behavior is much higher in the control treatment than in the strategic treatment for both groups of size $n = 2$ and groups of size $n = 3$. The difference is statistically significant, yielding p-values of 0.0001 in both settings. When, in the strategic set-up, one excludes Games 5 and 6, which are characterized by either a short duration (Game 5 lasted only 32 seconds) or a resolution of uncertainty that occurs very early in the game (with Player 2 achieving a success after exploring for 9

[16] Video recordings illustrating the use of the eye-tracking devices are available at the author's website: www.johanneshoelzemann.com.

Table 5: Average Frequency of Cut-Off Behavior

	Strategic Treatment		Control Treatment	
Group Size	Obs.	Total (Relative) Frequency	Obs.	Total (Relative) Frequency
$n = 2$	120	35 (.292)	120	100 (.833)
$n = 3$	180	59 (.328)	180	142 (.789)

Total number of cut-offs (number of cut-offs divided by total observations).

seconds in Game 6), the total number of cut-offs drops to 5 (23) out of 120 (180) for $n = 2$ ($n = 3$).

5.6 Pioneers

In the control treatment as well as in the best PBE, players are predicted to play risky on $(p_1^*, \frac{1}{2}]$; i.e., conditionally on no success arriving, players should switch from risky to safe only once, and do so at the same time, at which their beliefs reach p_1^*. By contrast, as KRC have shown, there is a range of beliefs containing $(p_1^*, p^\ddagger)$ such that safe and risky are mutually best responses in any Markov Perfect Equilibrium. In particular, there exists a range of beliefs in which just one pioneer should play risky while the other player(s) free-ride(s). The following result thus provides strong evidence that MPE seems to predict the qualitative features of subjects' behavior better, while confirming the prevalence of free-riding in our strategic treatment.

Result 4. *The proportion of time before a first breakthrough during which exactly one player plays risky is higher in the strategic treatment than in the control treatment.*

Table 6 shows the average proportion of time during which *exactly one* player is exploring before a first breakthrough by any player in his group. It is more than three times as large in the strategic treatment and highly statistically significant with p-values of 0.0001 for both $n = 2$ and $n = 3$.

Table 6: Proportion of Time with a Single Pioneer

	Strategic Treatment		Control Treatment	
Group Size	Obs.	Single Pioneer	Obs.	Single Pioneer
$n = 2$	60	.634 [.298]	60	.198 [.244]
$n = 3$	60	.497 [.338]	60	.080 [.168]

Average [st. dev.] proportion of time with a single pioneer in a group.

5.7 Switches of Action

Cut-off behavior implies at most a single switch of action from risky to safe per player in a given game. However, players should switch roles at least once in any Markov Perfect Equilibrium. Hence, if players' behavior is predicted by MPE, we should expect significantly more switches in the strategic treatment. Recall that we have defined the incidence of switches as the number of a player's changes in action choice in a given game per unit of effective time, defined as the time elapsed before the game ends or the player's risky arm is revealed to be good, whichever happens first.

Result 5. *The incidence of switches is significantly higher in the strategic treatment than in the control treatment. This holds for both $n = 2$ and $n = 3$.*

Table 7 displays the average number of switches per player across games for our four treatments.[17] The incidence of switches in the strategic treatment is much higher than in the control treatment for both $n = 2$ and $n = 3$ (both p-values of 0.0001).

Table 7: Average Number of Switches per Player

	Strategic Treatment		Control Treatment	
Group Size	Obs.	Switches per Player	Obs.	Switches per Player
$n = 2$	60	3.067 [2.450]	60	.792 [1.063]
$n = 3$	60	2.261 [2.040]	60	.778 [1.080]

Average [st. dev.] switches of players using group averages.

5.8 Groups of $n = 2$ vs. $n = 3$

The task of coordinating among larger groups is more challenging. Indeed, our eye-tracking data (see Table 4) shows that players pay significantly more attention to coordinating with their partners if $n = 3$ (p-value of 0.0001). In addition, our following result shows that evidence for the more complex forms of coordination required by MPE is significantly stronger for $n = 2$ than for $n = 3$.

Result 6. *The incidence of switches and the frequency of single pioneers are higher, while cut-off behavior is significantly less frequent, in the strategic treatment for $n = 2$ than for $n = 3$.*

The corresponding p-values are 0.2237, 0.0252, and 0.5096 for the average incidence of switches per player, for the proportion of time with a single pioneer, and the average frequency of cut-off behavior, respectively. Thus, the effect is statistically significant only with respect to the frequency of single pioneers. If we omit Games 5 and 6 (arguably outliers on account of their short length and the very early success, respectively), the difference in

[17] We report the average *number*, rather than the average *incidence*, of switches in Table 7, as the former may be easier to interpret.

cut-off behavior is highly significant as well (p-value of 0.0101).[18] Thus, while overall our subjects' behavior seems qualitatively to be best described by MPE play, the evidence for this conclusion is stronger for groups of size $n = 2$ than for $n = 3$. This may suggest that the complexity of the elaborate coordination required for MPE play increases quickly with group size.

6 Conclusion

We have analyzed a problem of dynamic public-good provision, where the public good in question is information about an uncertain state of the world. In particular, a group of several agents was facing the same decision problem, in which the optimal course of action depended on an unknown state of the world, which, in the strategic treatment, was common to everyone in the group. Therefore, the information produced by one agent benefited the other group member(s) as well. This informational externality constituted the only strategic link across players. Information, and hence agents' contribution incentives, evolve as the game progresses. We compare subjects' behavior in this strategic treatment to the behavior of subjects in the *control treatment*, where each agent's *individual* state of the world was iid, and there were therefore no strategic links across group members.

In a first step, we have exhibited strong evidence for strategic free-riding, as experimentation intensities are lower, and payoffs higher, in the strategic setting. Our eye-tracking data furthermore suggest that, in the strategic setting, subjects were paying keen attention to their partners' exploration efforts. Moreover, subjects seem to attempt to coordinate in rather complex ways, as evidenced, inter alia, by the much lower incidence of cut-off behavior and the higher incidence of switches in the strategic setting. This, together with the greater prevalence of single pioneers, suggests that the qualitative aspects of subjects' behavior is more accurately predicted by KRC's Markov Perfect Equilibrium than by HKR's average-payoff maximizing Perfect Bayesian Equilibrium, as Table 8 summarizes.

Table 8: Hypotheses, by Equilibrium Concept

	MPE	PBE
H3	$\surd$	$\times$
H4	$\surd$	$\times$
H5	$\surd$	$\times$

As we have seen in Result 6, this is particularly apparent in the two-player setting. Ours being a rather complex game, we of course cannot conclusively prove that subjects play, or aim to play, an MPE, as it is impossible to rule out other, more heuristic, forms of behavior. Yet, we believe that our results provide at least suggestive evidence for the main qualitative behavioral predictions of KRC's simple MPE.

KRC have also shown that there is a unique symmetric MPE in this game, which is characterized by players' using both arms at interior levels of intensity in the belief region

[18] If we analyzed the *number*, rather than the *incidence* of switches, the difference would be significant at the 10%-level (p-value of 0.0771).

where safe and risky are mutually best responses. In our experimental implementation, we do not allow subjects to pick experimentation intensities $k_i \in (0, 1)$. While one could in principle imagine an experimental set-up that does this (e.g., by letting subjects handle a gas pedal or a joystick), we have decided against doing so here in order to keep the already highly complex game as simple as possible for our subjects. Yet we think that allowing for interior experimentation intensities would be an interesting robustness check to perform in future research.

Recall that each of our subject groups played the six games in random order. As a robustness check, we also ran our analysis using only the last games played by each group. While this implies the loss of a large amount of data, and hence statistical power, our qualitative conclusions remain unaltered, although a few of our effects are no longer statistically significant.

As a further robustness check, one could in principle show subjects the current updated belief on their screens, in order to separate the task of belief updating from that of determining the cut-offs. We have decided against doing so here, as we were concerned about prodding subjects toward certain behaviors, which would have made the interpretation of our results more difficult.

We have confined our analysis to the exponential-bandit setting of KRC. While the tractability of the exponential-bandit setting will certainly have facilitated its experimental implementation, the model does have some special features. For instance, as successes are fully revealing, there is no encouragement effect in KRC, which our experimental investigation confirms.[19] Indeed, we can compute the average experimentation intensities in the region where safe is a dominant action, $[0, p_1^*]$, for Game 4 as well as for the two-player groups in Game 2.[20] Even in this region, the average experimentation intensity is lower in the strategic treatment: .511 [.042] in the strategic treatment for Game 4 with $n = 2$ vs. .655 [.237] in the control treatment; .325 [.091] vs. .756 [.220] in Game 4 for $n = 3$, and .511 [.063] vs. .696 [.251] in Game 2, where we report the standard deviation in square brackets. By contrast, if there were an encouragement effect, we should expect higher experimentation intensities in the strategic treatment for this belief region.

It might be interesting to test whether the encouragement effect can be shown in the laboratory for settings in which the theory would predict it to arise. This would be the case for instance in the Poisson setting with inconclusive breakthroughs à la Keller and Rady (2010), or in the Brownian-motion setting of Bolton and Harris (1999). It would also be intriguing to try and test the impact of privately observed actions or payoffs in the laboratory. We commend these questions for future research.

[19]The *encouragement effect* has been identified by Bolton and Harris (1999) and is not predicted to arise in the KRC setting. By virtue of this effect, players experiment more than if they were by themselves. They do so in the hope of producing public good news, which, in turn, makes their partners more optimistic. As their partners become more optimistic, they will be more inclined to experiment, thus providing some additional free-riding opportunities to the first player. This effect is absent in KRC, because here good news is conclusive: It resolves all uncertainty, so that, as soon as there is good news, players are not interested in free-riding any longer.

[20]These are the only settings in which this region is reached (and lasts for more than a few seconds) for both the strategic and the control treatments.

References

Ambrus, A., and B. Greiner (2012): “Imperfect Public Monitoring with Costly Punishment: An Experimental Study,” *American Economic Review*, 102(7), 3317–32.

Anderson, C. M. (2001): “Behavioral Models of Strategies in Multi-Armed Bandit Problems,” *PhD Thesis, California Institute of Technology.*

——— (2012): “Ambiguity Aversion in Multi-Armed Bandit Problems,” *Theory and Decision*, 72(1), 15–33.

Andreoni, J. (1988): “Why Free Ride? Strategies and Learning in Public Goods Experiments,” *Journal of public Economics*, 37(3), 291–304.

Banks, J., M. Olson, and D. Porter (1997): “An Experimental Analysis of the Bandit Problem,” *Economic Theory*, 10(1), 55–77.

Bellman, R. (1956): “A problem in the Sequential Design of Experiments,” *Sankhyā: The Indian Journal of Statistics (1933-1960)*, 16(3/4), 221–229.

Bergemann, D., and J. Välimäki (1996): “Learning and Strategic Pricing,” *Econometrica*, pp. 1125–1149.

——— (2000): “Experimentation in Markets,” *The Review of Economic Studies*, 67(2), 213–234.

Bergemann, D., and J. Välimäki (2008): “Bandit Problems,” The New Palgrave Dictionary of Economics, 2nd ed. Palgrave and Macmillan Press.

Bolton, P., and C. Harris (1999): “Strategic Experimentation,” *Econometrica*, 67(2), 349–374.

——— (2000): “Strategic Experimentation: The Undiscounted Case,” in *Incentives, Organizations and Public Economics–Papers in Honour of Sir James Mirrlees*, pp. 53–68. Oxford: Oxford University Press.

Bonatti, A., and J. Hörner (2011): “Collaborating,” *American Economic Review*, 101(2), 632–63.

Boyce, J. R., D. M. Bruner, and M. McKee (2016): “Strategic Experimentation in the Lab,” *Managerial and Decision Economics*, 37(6), 375–391.

Bradt, R. N., S. Johnson, and S. Karlin (1956): “On Sequential Designs for Maximizing the Sum of n Observations,” *The Annals of Mathematical Statistics*, 27(4), 1060–1074.

Calford, E., and R. Oprea (2017): “Continuity, Inertia, and Strategic Uncertainty: A Test of the Theory of Continuous Time Games,” *Econometrica*, 85(3), 915–935.

Chaudhuri, A. (2011): “Sustaining Cooperation in Laboratory Public Goods Experiments: A Selective Survey of the Literature,” *Experimental Economics*, 14(1), 47–83.

Coe, D. T., and E. Helpman (1995): "International R & D Spillovers," *European Economic Review*, 39(5), 859–887.

Fehr, E., and S. Gächter (2000): "Cooperation and Punishment in Public Goods Experiments," *American Economic Review*, 90(4), 980–994.

Fischbacher, U. (2007): "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments," *Experimental Economics*, 10(2), 171–178.

Fudenberg, D., and E. Vespa (2018): "Learning Theory and Heterogeneous Play in a Signaling-Game Experiment," *Working Paper, Massachusetts Institute of Technology*.

Gans, N., G. Knox, and R. Croson (2007): "Simple Models of Discrete Choice and Their Performance in Bandit Experiments," *Manufacturing & Service Operations Management*, 9(4), 383–408.

Gittins, J., and D. Jones (1974): "A Dynamic Allocation Index for the Sequential Design of Experiments," in *Progress in Statistics, European Meeting of Statisticians, 1972, 1*, pp. 241–266. Amsterdam: North Holland.

Greiner, B. (2015): "Subject Pool Recruitment Procedures: Organizing Experiments with ORSEE," *Journal of the Economic Science Association*, 1(1), 114–125.

Griliches, Z. (1988): "Productivity Puzzles and R & D: Another Nonexplanation," *Journal of Economic Perspectives*, 2(4), 9–21.

Grossman, G. M., and E. Helpman (1993): *Innovation and Growth in the Global Economy*. MIT Press.

Hörner, J., N. Klein, and S. Rady (2014): "Strongly Symmetric Equilibria in Bandit Games," *Cowles Foundation Discussion Paper No. 1956*.

Hörner, J., and A. Skrzypacz (2016): "Learning, Experimentation and Information Design," *mimeo*.

Hudja, S. (2018): "Voting for Experimentation: A Continuous Time Analysis," *Working Paper, Purdue University*.

Isaac, R. M., K. F. McCue, and C. R. Plott (1985): "Public Goods Provision in an Experimental Environment," *Journal of Public Economics*, 26, 51–74.

Isaac, R. M., and J. M. Walker (1988a): "Communication and Free-Riding Behavior: The Voluntary Contribution Mechanism," *Economic inquiry*, 26(4), 585–608.

——— (1988b): "Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism," *The Quarterly Journal of Economics*, 103(1), 179–199.

Isaac, R. M., J. M. Walker, and S. H. Thomas (1984): "Divergent Evidence on Free Riding: An Experimental Examination of Possible Explanations," *Public Choice*, 43(2), 113–149.

Jacob, R. J., and K. S. Karn (2003): "Eye Tracking in Human-Computer Interaction and Usability Research: Ready to Deliver the Promises," in *The Mind's Eye*, pp. 573–605. Elsevier.

Keller, G., and S. Rady (2010): "Strategic Experimentation with Poisson Bandits," *Theoretical Economics*, 5(2), 275–311.

——— (2015): "Breakdowns," *Theoretical Economics*, 10(1), 175–202.

Keller, G., S. Rady, and M. Cripps (2005): "Strategic Experimentation with Exponential Bandits," *Econometrica*, 73(1), 39–68.

Kim, O., and M. Walker (1984): "The Free Rider Problem: Experimental Evidence," *Public Choice*, 43(1), 3–24.

Klein, N. (2013): "Strategic Learning in Teams," *Games and Economic Behavior*, 82, 636–657.

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

Ledyard, J. O. (1995): "Public Goods: A Survey of Experimental Research," in *Handbook of Experimental Economics ed. by J. Kagel and A. Roth)*, pp. 111–194. Princeton: Princeton University Press.

Marwell, G., and R. E. Ames (1979): "Experiments on the Provision of Public Goods. I. Resources, Interest, Group Size, and the Free-Rider Problem," *American Journal of sociology*, 84(6), 1335–1360.

——— (1980): "Experiments on the Provision of Public Goods. II. Provision Points, Stakes, Experience, and the Free-Rider Problem," *American Journal of Sociology*, 85(4), 926–937.

——— (1981): "Economists Free Ride, Does Anyone Else? Experiments on the Provision of Public Goods," *Journal of Public Economics*, 15(3), 295–310.

Meyer, R. J., and Y. Shi (1995): "Sequential Choice under Ambiguity: Intuitive Solutions to the Armed-Bandit Problem," *Management Science*, 41(5), 817–834.

Murto, P., and J. Välimäki (2011): "Learning and Information Aggregation in an Exit Game," *The Review of Economic Studies*, 78(4), 1426–1461.

Poole, A., L. J. Ball, and P. Phillips (2005): "In Search of Salience: A Response-Time and Eye-Movement Analysis of Bookmark Recognition," in *People and Computers XVIII. Design for Life*, pp. 363–378. Springer.

Presman, E. L. (1990): "Poisson Version of the Two-Armed Bandit Problem with Discounting," *Theory of Probability & Its Applications*, 35(2), 307–317.

Rabanal, J. P., O. A. Rud, and M. Sharifova (2018): "Dynamic Entry and Exit Decisions with Noisy Information: An Experiment," *Working Paper, Monash University*.

Robbins, H. (1952): "Some Aspects of the Sequential Design of Experiments," .

Romer, P. M. (1990): "Endogenous Technological Change," *Journal of Political Economy*, 98(5, Part 2), S71–S102.

Rosenberg, D., E. Solan, and N. Vieille (2007): "Social Learning in One-Arm Bandit Problems," *Econometrica*, 75(6), 1591–1611.

Rothschild, M. (1974): "A Two-Armed Bandit Theory of Market Pricing," *Journal of Economic Theory*, 9(2), 185–202.

Strulovici, B. (2010): "Learning while Voting: Determinants of Collective eExperimentation," *Econometrica*, 78(3), 933–971.

Thompson, W. R. (1933): "On the Likelihood that One Unknown Probability Exceeds Another in View of the Evidence of Two Samples," *Biometrika*, 25(3/4), 285–294.

Zelmer, J. (2003): "Linear Public Goods Experiments: A Meta-Analysis," *Experimental Economics*, 6(3), 299–310.