Quantitative Economics 12 (2021), 1021–1051 1759-7331/20211021

Bandits in the lab

Johannes Hoelzemann
Department of Economics, University of Toronto

Nicolas Klein
Département de Sciences Économiques, Université de Montréal and CIREQ

We experimentally implement a dynamic public-good problem, where the public good in question is the dynamically evolving information about agents' common state of the world. Subjects' behavior is consistent with free-riding because of strategic concerns. We also find that subjects adopt more complex behaviors than predicted by the welfare-optimal equilibrium, such as noncut-off behavior, lonely pioneers, and frequent switches of action.

Keywords. Dynamic public-good problem, strategic experimentation, exponential bandits, learning, dynamic games, laboratory experiments.

JEL classification. C73, C92, D83, O32.

1. Introduction

Economists have long been concerned with social dilemmas related to the production of public goods. There exists a vast experimental literature on games, which examines the willingness to contribute to 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

Johannes Hoelzemann: j.hoelzemann@utoronto.ca
Nicolas Klein: kleinnic@yahoo.com
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 three anonymous referees, Francisco Alvarez-Cuadrado, Heski Bar-Isaac, Arthur Campbell, Juan Carlos Carbajal, Gary Charness, Rahul Deb, 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, Yao Luo, Erkut Ozbay, Sven Rady, Bradley Ruffle, 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.

Available at http://qeconomics.org. https://doi.org/10.3982/QE1389

in such social-dilemma situations in the lab (e.g., Fehr and Gächter (2000), Ambrus and Greiner (2012), and Fenig, Gallipoli, and Halevy (2018)).[1]

Situations in which the public good is dynamically evolving, however, have received scant attention from experimental economists so far. This is the case, for instance, when the public good in question is information. By contrast, purely informational social dilemmas are the object of the so-called strategic multiarmed bandit problems, which have received a lot of attention in the recent theoretical literature. In multiarmed bandit problems, agents repeatedly choose among different options (or bandit *arms*) of initially unknown quality. A *strategic* bandit problem is one in which several agents solve a bandit problem each. In the simplest case, the underlying payoff parameters are the same for all the agents, so that the information produced by one agent is useful to the other agents as well. In particular, the other agents' actions do not impact a given agent's payoffs directly; the only strategic link across players is via the information they produce, which is useful to the other players' decision problems. By abstracting from payoff externalities, these settings reduce players' scope for intertemporal incentives, as punishments and rewards can only be informational in nature. In this paper, we offer one of the first experimental investigations of the provision of a dynamically evolving public good, and to the best of our knowledge, the first to implement an experimental investigation of a dynamic social dilemma where externalities are purely informational.

The trade-offs involved in the production of public information are important to understand. Indeed, innovation and social learning are often the work of pioneers, who, by bearing the costs of experimenting with a new approach, create informational spillovers 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. 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)). Situations in which the informational benefits of experimentation are shared abound: for example, the decision of where to fish when others can see one's boat and haul,[2] consumers searching for the right car or cell-phone to buy, farmers deciding whether to grow a traditional or a gene-modified crop, graduate students selecting their field of research, etc.

In the multiarmed bandit models, which have become canonical to study information producers' dynamic trade-offs, a decision maker, at each point in time, 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 focused 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 multiplayer continuous-time framework. The simpler exponential

[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).

[2]We are indebted to an anonymous referee for this example.

model of KRC especially has since been used to analyze a wide array of applications, such as, for instance, R&D races.[3]

Thus, the object of this paper is to experimentally examine behavior in a novel social dilemma over the exploration of information with common value. In order to do so, we base ourselves on KRC's exponential model. We chose this particular model for two reasons. First, its setup is simpler than that of the other strategic-experimentation papers, and, second, as Hörner, Klein, and Rady (2021) (subsequently: HKR) have shown, the welfare-optimal equilibrium[4] has a particularly simple structure in this model. Indeed, while it is non-Markovian, it is strongly symmetric and players play a cut-off strategy[5] (on the path of play), applying the same cut-off as a single agent. Given the simple structure of the best equilibrium, it can reasonably be expected to be focal among the continuum of equilibria that exist in the model. Moreover, it gives us a very clean empirical test: if they (want to) play the equilibrium, subjects should behave in the same way as when they solve the single-agent problem.

To make the problem tractable, the strategic-experimentation literature is by and large focusing on the choice between a safe option, yielding a known payoff, and a risky option, which yields payoffs following a stochastic process. The time-invariant quality of this risky option can be good or bad. If it is good (bad), it dominates (is dominated by) the safe option. Whether the risky option 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 option 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 option. This gives rise to a dynamic public-good problem in the form of dynamically evolving information about the agents' common state of the world.

As hinted at above, our analysis relies on comparing the behavior of our experimental subjects in groups where the quality of the risky option was known to be the same for all partners (which we call the *strategic treatment*) to that of groups where its quality was i.i.d. across members, the *control treatment*.[6] When the quality of the risky option

[3]See, for example, Besanko and Wu (2013), Akcigit and Liu (2016), or Das and Klein (2020). Besanko, Tong, and Wu (2018) used the exponential bandits framework to analyze optimal subsidies for R&D.

[4]We are committing a slight abuse of terminology here by referring to the "welfare-optimal," "average-payoff maximizing" or "best" equilibrium, which we shall maintain throughout this paper. HKR in fact analyze a setting in continuous time in which actions are frozen for small intervals of time of length $\Delta > 0$. They show that, for small enough Δ, there exists a perfect Bayesian equilibrium (PBE), which happens to be strongly symmetric, with payoffs that, as $\Delta \to 0$, converge to the payoff from all players playing risky above the single-agent threshold and safe below it, and that it is not possible to achieve higher limit average PBE payoffs. Our experimental implementation is, of course, strictly speaking, in discrete time, with information and action choices being updated every second.

[5]A cut-off strategy is defined by a unique threshold belief above which it prescribes risky play, while prescribing safe play below it.

[6]The fact that players' payoffs are realizations of compound stochastic processes in our setting considerably blows up their variances, so that the different solution concepts cannot be directly identified from subjects' realized payoffs. As we shall explain in detail below, we therefore conduct inference by comparing payoffs across treatments for *given realizations* of the stochastic processes and mostly by comparing *behavior* across treatments.

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, that is, there are no payoff externalities.

In a first step, we show that the informational externality impacts subjects' behavior. Average experimentation intensities are lower in the strategic treatment, and, in particular, in the belief region for which theory predicts free-riding to be an issue, subjects experiment significantly less in the strategic treatment, suggesting they are free-riding because of strategic concerns.[7] Moreover, subjects' payoffs are higher in the strategic treatment, suggesting that they are taking advantage of the information produced by their partners. Further, subjects are adopting more sophisticated behaviors in the strategic treatment than in the control treatment. Players switch much more between safe and risky, and use cut-off strategies much less frequently, in the strategic treatment. Additionally, there is a larger proportion of time during which exactly one player is playing risky in the strategic treatment.[8] If subjects were playing, or aiming to play, the best PBE, we should observe none of these differences. In this case, the only difference between the two treatments should consist in the increased speed of learning in the strategic treatment, with subjects' behavior otherwise the same across treatments.

We interpret our findings as showing that, while understanding the informational externality (since they achieved higher payoffs in the strategic treatment), subjects were *not* behaving, or aiming to behave, as in the best PBE. Rather, the behaviors we document are consistent with the qualitative predictions of *any* of KRC's (infinitely many) Markov perfect equilibria (MPE), which feature players' taking turns and alternating in the roles of free-riders and pioneers for some intermediate range of beliefs. Further, the differences between the two treatments tend to be more pronounced for two-player groups than for groups of size three.

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 continually to update their beliefs using Bayes' rule precisely at lightning speed, they were nonetheless reacting to the dynamically evolving incentives. Furthermore, we are documenting behavior that is very different from the simple structure of the best PBE, and arguably more in line with the sophisticated coordination required by MPE play.

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

[7] *Free-riding* in our setting refers to a subject opportunistically using the safe option while efficiency would require the use of the risky option. Players have no incentives to do so at very optimistic beliefs, where risky is a dominant action. Subsequently, we shall therefore use the phrase only with respect to those belief regions where there is a strategic rationale for players to deviate from efficient behavior by playing safe.

[8] We refer to such players as *pioneers*.

experimental implementation; Section 5 presents our main findings; Section 6 provides additional results and robustness checks; and Section 7 concludes. Appendix A in the Online Supplementary Material (Hoelzemann and Klein (2021)) breaks down the analysis to the individual games subjects played. Appendix B exhibits and explains the interface our experimental subjects were using and 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 multiarmed 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) gave a survey of this literature.

Bolton and Harris (1999, 2000) were the first to consider the multiplayer 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.[9]

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's (2016) 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

[9]Many variants of the multiplayer 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) introduced 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) analyzed the case of privately observed payoffs, while Bonatti and Hörner (2011) investigated the case of privately observed actions. Bergemann and Välimäki (1996, 2000) considered strategic experimentation in buyer-seller settings. Hörner and Skrzypacz (2017) gave a survey of this literature.

ambiguity attitudes. Our investigation, by contrast, is focused on how players resolve the coordination problems arising from strategic interaction. Indeed, our subjects all face the same decision problem and are given a Bayesian prior at the outset. As they interact many times with a stochastic and unknown deadline, their action spaces are very rich.

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. Hudja and Woods (2020) experimentally investigates the single-agent version of KRC's exponential-bandit setting, finding that subjects tended to explore less than predicted. 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) tested decision-making under ambiguity and used experimental data to generate hypotheses about subjects' 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) considered 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) used arms with payout distributions, for example, 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.

Oprea, Charness, and Friedman (2014) studied experimentally a standard pubic-goods game with a rich communication protocol in both discrete and continuous time. They find that voluntary provision of the public good is higher in continuous than in discrete time. This, however, is only the case if subjects have the possibility to communicate freely to coordinate their contributions. Wilson and Vespa (2020) experimentally studied a cheap-talk game, while Reshidi, Lizzeri, Yariv, Chan, and Suen (2020) investigated agents' information-acquisition decisions before making an irreversible binary choice.

Battaglini, Nunnari, and Palfrey (2016) investigated a game of dynamic contributions to a *durable* public good in the laboratory; that is, the stock of the public good builds up over time. Theirs is thus a setting of conventional payoff externalities, while, in our setting, externalities are purely informational in nature; that is, other players' actions have no direct impact on a given player's payoffs—the presence of the other players impacts a given player only via the information they produce over time. Battaglini, Nunnari, and Palfrey (2016) find that subjects' qualitative behavior is by and large consistent with the predictions of the Markov Perfect Equilibria that were characterized in Battaglini, Nunnari, and Palfrey (2014), although they find some evidence of non-Markovian history dependencies.

Chernulich, Horowitz, Rabanal, Rud, and Sharifova (2020) studied an individual decision problem of market entry and exit decisions with and without counterfactual information in an environment where the trade-off is similar to a two-armed bandit problem with a safe arm and a risky arm. Hudja (2019) experimentally implemented Strulovici's (2010) collective experimentation model. An individual experimentation problem is compared to a collective experimentation problem where groups of three players face a majority-vote. Fudenberg and Vespa (2019) analyzed 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 test in which subjects play against a computer.

The prevalence of MPE-type behavior in dynamic games is investigated by Vespa and Wilson (2015, 2019) in the context of a prisoners' dilemma where payoffs depend on a binary state. Their game is set up in such a way that higher efficiency can be achieved by symmetric non-Markovian play, while full efficiency can be achieved by asymmetric SPE. They show that a substantial fraction of subjects behaved in a Markovian fashion. Those who did not tended to aim for higher efficiency. As the complexity of the coordination required to achieve more efficient outcomes than the best MPE increased, the prevalence of MPE play increased. Their findings thus suggest that one of the main draws of Markovian behavior is the simplicity of the coordination required. In our setting, by contrast, MPEs require more complex coordination than the best PBE.

On the other hand, MPEs feature turn-taking in our setting, while the best PBE does not. Cason, Lau, and Mui (2013) studied the emergence of turn-taking in the dynamic assignment game and highlight the importance of being able to teach dynamic strategies to others as well as the importance of using strategies that allow for teaching and learning. Leo (2017) studied both theoretically and experimentally flexible turn-taking. He shows that turn-taking leads to substantial efficiency gains and efficiency achieved by subjects is close to that expected in theory. Nevertheless, robust anomalies in subject behavior, which cannot be attributed to pro-social behavior or strategic concerns, are prevalent in his experimental implementation.

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 of 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 $i \in \{1, \dots, n\}$ 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), that is, strategies $k_i : [0, 1] \to \{0, 1\}$, $p \mapsto k_i(p)$, where a player's action after any history can be written as a time-invariant function of the common belief at that history.[10] Thus, the action of a player playing a Markov strategy depends on the previous history only via the current belief. 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 Section 6.1 in KRC). Moreover, it is shown that there exists no MPE in which all players play a cut-off strategy, that is, a strategy that prescribes the use of the risky arm for beliefs above a single cut-off and that of the

[10] In KRC, Markov strategies are actually defined as functions $k_i : [0, 1] \to [0, 1]$. In order to make the decision problem easier for our subjects, we have restricted the action space to $\{0, 1\}$ rather than $[0, 1]$. Of course, all equilibria in the game with the larger action space that only use actions in $\{0, 1\}$ (called "simple equilibria") remain equilibria in the game with the smaller action space. We, however, lose KRC's (unique) symmetric MPE, which involves interior action choices on some open subinterval of beliefs.

safe arm below. Thus, the roles of pioneer and free-rider must switch at least once in Markov equilibrium.[11] 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.

4. Parametrization and experimental design

4.1 *Experimental implementation*

In our experimental treatments, the number of players will be $n = 2$ or $n = 3$. 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.[12] 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 Markovian mutually best responses, before entering the region where safe is dominant.[13]

The realizations of all random processes were simulated ahead of time.[14] We generated six different sets of realizations of the random parameters controlling the length of the game, the quality of the risky arm, and the arrivals of the good risky arm. These corresponded to six different games that each of our subjects played. To make our findings more easily comparable, we have kept the same realizations for both the strategic and

[11]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; that is, 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.

[12]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.

[13]The fact that payoffs from a good risky arm are realizations of a compound Poisson process implies a large variance of payoff realizations. This makes it impossible to identify the various solution concepts from the realized payoffs (see Section 5.3 for a more detailed discussion).

[14]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 implementation, we chose to simulate them ahead of time.

the control treatments.[15] Participants' interfaces (see Appendix B) were updated every second.[16]

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 ($k \in \{1, \dots, 10\}$) 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.[17]

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 40 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 programmatically 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.[18] 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.[19] 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*

The best PBE (as well as all MPEs with a finite number of switches) predict players to play safe at all beliefs $p \leq p_1^*$, while efficiency would require that they play risky at all

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

[16]Thus, our setting approaches the "Inertial Continuous-Time" setting in Calford and Oprea (2017). We were able to optimize at the margin so as to decrease the lag times in zTree to be less than 250 milliseconds, that is, shorter than human reaction time, by constraining the maximum number of switches, game length, etc. allowed. These constraints were unproblematic in all of the sessions.

[17]See Appendix B for more details and screen shots.

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

[19]The theoretical treatment of strategic-experimentation problems has so far focussed on risk-neutral players only. Given the small stakes at play in the experiment, we did not expect subjects' risk attitudes to have an impact on their behavior. Consistently with this prediction, our data do not allow us to establish any effect of risk aversion on subjects' behavior. Details are available from the authors upon request.

Table 1. Belief thresholds.

Symbol	Interpretation	Value
p_0	Prior belief	0.5
p^m	Myopic cutoff	0.4
p_1^*	Single-agent cutoff	0.2326
p_2^*	Efficient cutoff for $n=2$	0.1031
p_3^*	Efficient cutoff for $n=3$	0.0535
$\bar{p}$	$[p_1^*, \bar{p}]$ is a superset of the *free-riding region*	0.3578 (0.3742)
$p^\ddagger$	$(p_1^*, p^\ddagger)$ is a subset of the *free-riding region*	0.3428 (0.3609)

Note: The values for $\bar{p}$, $p^\ddagger$ are for $n=2$ $(n=3)$.

beliefs $p > p_n^*$, where $p_n^* < p_1^*$. Single players and players playing the best PBE should play risky at all beliefs $p > p_1^*$, that is, 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.[20] In the following, we refer to the belief region $(p_1^*, p^\ddagger)$ as the *free-riding region*. By the same token, we can derive an upper bound $\bar{p}$ on the lowest belief at which risky is a dominant action.[21] Table 1 provides an overview of belief thresholds, together with their numerical values given our parameters.

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 $n=1$. The bang-bang structure of the best PBE is highlighted in Figure 1 in Section 5.

4.2.1 *Free-riding* Let $\hat{T}_i$ be the time player i's risky arm is revealed to be good 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^*. The same holds true in the best PBE in the strategic setting, but conditionally on no success arriving, beliefs will decrease faster in the strategic setting, as

[20]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 $u_i(p) \le s + K_{-i}(p)c(p)$, 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 $V_{n,p_1^*}(p^\ddagger) - s - (n-1)c(p^\ddagger) = 0$.

[21]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 $V_1^*(\bar{p}) - s - (n-1)c(\bar{p}) = 0$.

player i's belief also decreases in response to player j's hapless experimentation. Since both effects go in the same direction, we hypothesize that average experimentation intensities are lower in the strategic treatment. To set the stage, we thus formulate the following.

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

Thus, a lower average experimentation intensity in the strategic treatment need not be due to subjects' strategic free-riding, since beliefs decrease faster in the strategic treatment. Strategic equilibrium free-riding can manifest itself in two ways: (i) some players play safe while the belief is above the single-agent cutoff p_1^*; (ii) players stop experimenting at the single-agent cutoff p_1^* (while efficiency would require them to experiment until the belief hits p_n^*). Effect (i) is *not* predicted to occur in the best PBE, whereas it is predicted to occur in any MPE. Effect (ii), by contrast, is predicted to arise in any equilibrium, Markovian or not. We can test for Effect (i) by comparing average experimentation intensities in the *free-riding belief region* where at least one player plays safe in *any* MPE. Theory would predict this intensity to be 1 in the control treatment; in the strategic treatment, the best PBE would predict it to be 1 as well, whereas it would be strictly less than 1 in any MPE. We therefore interpret a significantly lower average experimentation intensity for the strategic treatment in this belief region as evidence both for strategic free-riding and against the best PBE. These considerations lead us to formulate the following.

HYPOTHESIS 1. (a) *The average experimentation intensity* $\frac{\sum_{i=1}^{n}\int_0^{\hat{T}_i} k_{i,t}\,dt}{\sum_{i=1}^{n}\hat{T}_i}$ *in the* free-riding region $(p_1^*, p^\ddagger)$ *is strictly lower in the strategic treatment than in the control treatment.*

(b) *Moreover, it is no higher in the* safe-dominant *region* $[0, p_1^*]$.

Our game is one of purely (positive) informational externalities; that is, players always have the option of ignoring the additional information they get for free from their partner(s). Therefore, players should do better in the strategic treatment, which motivates our following

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

4.2.2 *Structural properties of the best PBE* *Cut-off behavior* consists in a player 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. As explained above, KRC predict that subjects will use cut-off strategies in the control treatment; by the same token, HKR show that cut-off behavior prevails on path in the strategic setting also if the best PBE is played.

This leads us to the following.

Hypothesis 3. *There is no difference in the frequency of cut-off behavior between the two treatments.*

Even if players were using cut-off strategies, they would not be conforming to the theoretical predictions if they were applying different cut-offs. Indeed, theory predicts that neither in the single-agent problem of the control treatment nor in the best PBE should exactly one of the players plays risky at any time. This is in contrast to any of KRC's "simple" MPEs, which all feature a pioneer who is experimenting alone on some interval of beliefs.

This motivates the following.

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

5. Experimental results

This section is, in the main, devoted to testing our behavioral hypotheses of Section 4.2. Throughout 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 (Hoelzemann and Klein (2021)). For each of the six games, we conducted four treatments (strategic and control treatment with $n = 2$ and $n = 3$), with ten groups each. We had simulated all the relevant parameters ahead of time, as explained in Section 4. These included separate processes for the games' duration, the quality of the risky arm and the timing of successes on the risky arm in case it was good. The duration of the games ranged from 32 seconds for Game 5 to 230 seconds for Game 4.

5.1 *Average experimentation intensities*

As we have argued in Section 4.2, we should expect average experimentation intensities to be lower in the strategic treatment. Recall that, in the strategic treatment, the experimentation intensity is calculated for each player until the time of a first breakthrough by *any* player in a group or the end of the game, whichever arrives first. In the control treatment, this measure is calculated until the time where the *individual* player observes a success or the game ends, whichever occurs first. Table 2 lists the observed mean experimentation intensities, using group averages across games for our four treatments.

In the strategic treatment, for groups of size $n = 2$ ($n = 3$), the ex ante expected experimentation intensity for the best PBE is 0.786 (0.712). For MPE, the ex ante expected experimentation intensity is between 0.739 and 0.750 (0.654 and 0.668).[22] Conditional on the random realizations of the stochastic processes in the experiment, the *average* over the game realizations for the best PBE is 0.609 (0.517), while, for MPE, it is 0.604 (0.481), for $n = 2$ ($n = 3$). By contrast, in the control treatment, the ex ante expected experimentation intensity would have been 0.903. Conditional on the random realizations

[22] Because of the multiplicity of MPE, it is not possible to give a point prediction of MPE experimentation intensities.

TABLE 2. Average experimentation intensities.

	Strategic Treatment		Control Treatment	
Group Size	Obs.	Experiment. Intensity	Obs.	Experiment. Intensity
$n=2$	60	0.594 [0.186]	60	0.818 [0.212]
$n=3$	60	0.539 [0.244]	60	0.839 [0.180]

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

of the stochastic processes, the *average* over the game realizations is 0.820 (0.853) for $n=2$ ($n=3$). Thus, the observed experimentation intensities are very much in line with the conditional predictions given the realizations of the random parameters, for either equilibrium concept.[23]

We now revisit Prediction 1 and Hypotheses 1–4. To test our behavioral hypotheses from Section 4.2 and treatment differences nonparametrically, we apply two-sided Wilcoxon rank-sum (Mann–Whitney) tests, using group averages as independent observations. We begin with Prediction 1. As Table 2 reveals, the additional presence of one (two) perfectly positively correlated risky arm(s) leads to lower experimentation intensities. This is highly statistically significant in both settings with $n=2$ and $n=3$. The corresponding p-values in both cases are 0.0001.[24] This is in line with Prediction 1, a finding we summarize in the following.

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

In Figure 1, we illustrate the evolution of experimenting subjects divided by all subjects across all games and all subjects. More precisely, the share of experimenting subjects, before their risky arm was revealed to be good or the game ended, whichever happened first, is indicated at each unit of calendar time. The denominator, accounting for all subjects still playing, decreases as calendar time progresses and individual action choices gain in relative weight. This is captured by the thickness of the line, which highlights the evolution of subjects still actively playing (the thicker the line the more subjects are still active). The theoretical predictions of both the optimal single-agent solution and the best PBE are included according to group size. These have a bang-bang

[23]Indeed, a one-sample t-test indicates that there is no evidence that the observed mean experimentation intensity is different from the predicted values (all p-values ≥ 0.4877), with the exception of the MPE prediction for $n=3$, with a p-value of 0.0704.

[24]The Wilcoxon rank sum 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 test, 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). In Section 6.4, we complement the nonparametric analysis by reporting results from ordinary least-square regressions with random effects and clustering of standard errors by group. We find no effect of the number of games previously played on subjects' behavior, and results reported throughout the paper remain robust.

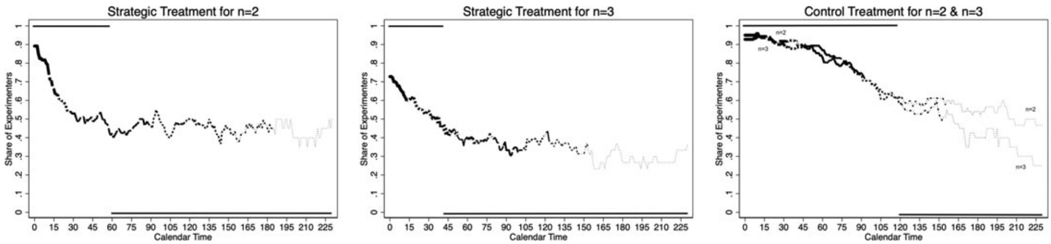

FIGURE 1. Share of experimenters by treatment. *Note*: The share of experimenters by treatment across all games and subjects is shown. The line thickness displays the evolution of subjects still in the game (the thicker the line the more subjects are in the game). Optimal, and best-PBE, predictions are also indicated for all group sizes and treatments. For the control treatment, share of experimenters is highlighted for both n in one panel. Labels indicate the corresponding lines for groups of size $n = 2$ and $n = 3$.

structure, with a cut-off at the point in time at which the belief threshold p_1^* would be reached if players adhered to the theoretical prediction, for any n.

As is evident from the figure, players change their behaviors over time. While often playing risky at the beginning, players' use of the risky arm decreases as time passes and no success is observed. This suggests that our subjects adapted to the evolving information about their environment. The stark bang-bang structure of the theoretical predictions, however, is not borne out by the average experimentation intensities. Note that the theoretically predicted cut-off only applies "on path," that is, it presupposes that everyone involved played risky with an intensity of 1 at all times before the cut-off. If this is not the case, the optimal "off-path" cut-off shifts to the right on the time axis. As we shall discuss in greater detail below, updated beliefs below p_1^* were reached in both the strategic and the control treatments only in two of the six games we simulated. As we shall also see below, subjects in the control treatment mostly followed cut-off strategies. As is apparent from the figure, the average proportion of risky play stays quite high for longer in the control treatment. This is consistent with theory, since subjects only have a single signal per unit of time to update their beliefs with in the control treatment. Meanwhile, they have two or three signals if their partner(s) also play(s) risky in the strategic treatment. Thus, information arrives faster, meaning that subjects become pessimistic faster (conditionally on no success being observed), in the strategic treatment. Subjects also alternate between risky and safe much more frequently and start to do so earlier. The figure also shows a remarkable similarity in subjects' behavior whether $n = 2$ or $n = 3$ in the control treatment, which is also consistent with theory; indeed, theory predicts subjects to behave like single agents, independently of group size n.

5.2 *Free-riding*

As we have discussed in Section 4.2.1, the reduced exploration intensity in the strategic treatment, which we have documented in the previous subsection, may well be partially, or even completely, owed to the faster information accumulation in the strategic treatment. Yet, from a theoretical standpoint, we are more interested in the phenomenon

Table 3. Average experimentation intensities in free-riding region.

	Strategic Treatment		Control Treatment	
Group Size	Obs.	Experiment. Intensity	Obs.	Experiment. Intensity
$n=2$	40	0.505 [0.155]	40	0.776 [0.311]
$n=3$	40	0.510 [0.220]	39	0.779 [0.243]

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

of strategic free-riding, that is, subjects' taking advantage of the information they receive for free from their partners in order to reduce their own exploration efforts. As we have discussed in Section 4.2.1, our setup allows us to identify Aspect (i) of strategic free-riding via the comparison of experimentation rates between the strategic and control treatments for beliefs in $(p_1^*, p^\ddagger)$, that is, a belief region in which at least one player is predicted to play safe and risky each in any MPE; by contrast, all players are predicted to play risky in the best PBE and in the efficient solution, as well as in the single-agent optimum. We therefore interpret a lower experimentation intensity in the strategic treatment for this belief region as evidence of both strategic free-riding and against the best PBE. Table 3 summarizes average experimentation intensities in the belief region $(p_1^*, p^\ddagger)$.[25]

The average experimentation intensity in the *free-riding region* is substantially and significantly lower in the strategic treatment. Independently of group size, the p-values of the two-sided Wilcoxon rank sum test amount to 0.0001.

Result 1a. *Subjects are* free-riding: *average experimentation intensities are lower in the strategic treatment over the* free-riding region $(p_1^*, p^\ddagger)$.

Result 1a, taken together with Remark 1, raises the question as to what extent subjects free-ride "correctly," that is, at the "right" beliefs. In order to investigate this question, we analyze subjects' experimentation intensities in the belief region $(\bar{p}, p_0]$, where risky is a dominant action (which we subsequently label the *risky dominant region*). For $n = 2$, the average experimentation intensity is lower in the strategic treatment with 0.675 [0.222] than in the control treatment with 0.899 [0.161], where we report the standard deviation in square brackets. The same applies to our three-player groups, where the average experimentation intensity amounts to 0.632 [0.281] in the strategic treatment, while it is substantially higher in the control treatment with 0.932 [0.152]. These differences are highly statistically significant, with the p-values of the two-sided

[25]We omit Games 5 and 6 from this table, since Player 2 has a success after 9 seconds on the risky arm in Game 6, and Game 5 lasted only 32 seconds, which implies that the *free-riding 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. The other missing observation corresponds to one three-player group in the control treatment that has not reached the *free-riding region*.

Wilcoxon rank sum test amounting to 0.0001 for both group sizes. While there is no theoretical rationale for the lower experimentation intensities in the *risky dominant region*, one may speculate that it may be due to subjects' aiming to reduce their experimentation intensities in the *free-riding region*, while not hitting this region precisely. Indeed, there is a theoretical rationale (namely, MPE) for different experimentation intensities across these two regions in the strategic treatment; there is no such rationale in the control treatment, where optimality would require an experimentation intensity of 1 in both belief regions.

In order to investigate this question further, we conduct a "difference-in-differences"-analysis, comparing the difference in intensities across belief regions and across treatments. As there is no a priori reason for imprecisions in belief updating or in the computation of the relevant thresholds to be more prevalent in the strategic treatment as compared to the control treatment, we interpret a bigger difference across belief regions in the strategic treatment as suggestive of MPE-style free-riding. For $n = 2$, this difference-in-differences is statistically significantly higher at the 5%-level in the strategic treatment, with a p-value of 0.0389. By contrast, no such evidence can be established for groups of size $n = 3$, where the effect is not statistically significant (p-value of 0.5118). Thus, there is stronger evidence that the free-riding we document is motivated by strategic rationales for the smaller group size $n = 2$. As we discuss in greater detail in Section 6.2, behavior was generally more reminiscent of MPE-play for the smaller group size $n = 2$.

To conclude our discussion of free-riding, we turn to the region $[0, p_1^*]$, where safe is a dominant action in both the strategic and control treatments. Indeed, the KRC model, which we have chosen for our experimental investigation on account of its tractability, exhibits no encouragement effect.[26] We can compute the average experimentation intensities in the region $[0, p_1^*]$, for Game 4 as well as for the two-player groups in Game 2.[27] Even in this region, the average experimentation intensity is lower in the strategic treatment: 0.511 [0.042] in the strategic treatment for Game 4 with $n = 2$ versus 0.655 [0.237] in the control treatment; 0.325 [0.091] versus 0.756 [0.220] in Game 4 for $n = 3$, and 0.511 [0.063] versus 0.696 [0.251] in Game 2. Thus, mean group averages are lower in the strategic treatment compared to the control. For Game 2, the effect is not significant (p-value of 0.1149). For Game 4, the p-value is 0.0849 (0.0015) for $n = 2$ ($n = 3$). This leads us to state the following.

Result 1b. *There is no encouragement effect in our data.*

[26]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.

[27]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.

Results 1a and 1b are fully in line with our Hypothesis 1. They also militate against our subjects' being motivated by social preferences.

5.3 *Payoffs*

Strategic interaction is predicted to arise among players as a result of (positive) informational externalities, that is, the information produced by their partners allows players to make better decisions, and hence to secure themselves higher payoffs. While certain courses of action are better than others from an ex ante expected point of view, the mapping from actions to payoffs is of course very stochastic in our setting, as it depends on the particular realizations of the random variables governing the length of the game, the quality of the risky arm and the timing of lump-sum arrivals from a good risky arm. Indeed, *conditionally* on a particular realization of the stochastic process, ex ante optimal behaviors may do very poorly, while ex ante very eccentric behaviors may well be optimal. Moreover, in our implementation, payoffs resulted from compound stochastic processes, implying a large ex ante variance in the mapping from behaviors to payoffs. We would thus caution against ascribing inferential value to payoff comparisons beyond what we do here, namely to compare payoffs between the strategic and control treatments, *for a given realization of the stochastic processes*.[28] For most of our inferences, though, we rely on our subjects' being ignorant of the realizations of the stochastic processes when they made their choices, which "thus filters out the noise" that prevails in the mapping from actions to payoffs.

Our test in this subsection is thus a simple one: Do subjects take advantage of the information they get for free from their partners in order to achieve higher payoffs in the strategic treatment? Theoretical ex ante expected per-capita payoffs are highest for the efficient solution, followed by the best PBE and then MPE, in the strategic treatment.[29] Any of these solution concepts implies, to varying degrees, players' taking advantage of the additional information they get from their partners, and thus leads to higher predicted payoffs than in the single-agent optimum. For groups of size $n = 2$ ($n = 3$), the ex ante expected per-capita payoff for the efficient solution is 1714.81 (1774.59), while, for the best PBE, it is 1699.00 (1734.90). For MPE, the *ex ante* expected per-capita payoff is between 1687.91 and 1690.68 (1714.49 and 1721.46).[30] For the single-agent optimum, the ex ante expected payoff is 1621.01. One can interpret the difference between the value of the efficient group solution and that of the single-agent optimum, 93.80 and 153.58 for groups of size $n = 2$ and $n = 3$, respectively, as measuring the size of the social dilemma we are analyzing. While entailing conspicuous differences in behavior, the impact on payoffs from the competing equilibrium concepts is small, roughly 10 (15) for groups of size $n = 2$ ($n = 3$). By contrast, even the lower bound for MPE payoffs entails an important payoff gain with respect to autarky, underlining the importance of the positive informational spillovers: 66.90 and 93.48 for groups of size $n = 2$ and $n = 3$, respectively.

[28] Recall from Section 4 that we have kept the same realizations of the random processes for the strategic and control treatments.

[29] This is not necessarily the case *given the particular realizations of the stochastic process*; please see Table A.3 in Appendix A.2 of the Online Supplementary Material (Hoelzemann and Klein (2021)).

[30] Because of the multiplicity of MPE, it is not possible to give a point prediction of MPE payoffs.

Table 4. Average final payoffs.

Group Size	Strategic Treatment Obs.	Final Payoffs	Min	Max	Control Treatment 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

Note: Average [st. dev.] final payoffs using group averages.

Table 4 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 (0.0103).[31] Thus, our subjects indeed take advantage of the positive informational externalities in the strategic treatment, lending support to Hypothesis 2.

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

5.4 *Cut-off behavior*

As we have pointed out above, optimality in the individual decision-making problem in our control treatment implies cut-off behavior, defined by a unique threshold belief above which it prescribes risky play, while prescribing safe play below it. 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. If subjects were trying to play the best PBE, therefore, we should observe roughly the same level of cut-off behavior in the strategic and the control treatments. Table 5 shows that the data emphatically reject this hypothesis, as cut-off behavior drops from roughly 80% in the control treatment to less than 33% in the strategic treatment. As it is not clear what it means for a group to engage in cut-off behavior, we report each individual subject's decisions.

The difference between the treatments 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 seconds in Game 6), the total number of cut-off observations drops to 5 (23) out of 120 (180) overall observations for $n=2$ ($n=3$).

To complement our binary measure of cut-off behavior, we are also analyzing a more "continuous" measure of cut-off behavior in order to capture the distance of a subject's

[31] To verify that our results are not driven by one particular game that may have unique features, we have computed our statistical tests each time excluding a different game. Differences in payoffs always remain statistically significant. This observation is confirmed by our ordinary least-square regressions with random effects controlling for learning effects (see Section 6.4), where results do not qualitatively change: we find a strong positive effect of the correlation structure—our strategic treatment—on payoffs. The same also holds true for groups of size $n=2$.

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 (0.292)	120	100 (0.833)
$n = 3$	180	59 (0.328)	180	142 (0.789)

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

behavior to a cut-off strategy.[32] In particular, we measure the proportion of time in which a subject plays safe before ever playing risky, or plays risky after they had previously switched from risky to safe, before his risky arm is revealed to be good or the end of the game, whichever arrives first. We define 1 minus this proportion of time as our continuous cut-off measure, so that a score of 1 indicates perfect cut-off behavior. While this measure has some shortcomings (e.g., a subject starting with risky, then switching repeatedly back to the risky arm for short amounts of time after having switched to safe, would be classified as being close to cut-off behavior), we think that, in conjunction with our cruder binary measure, it can serve as a useful robustness test. In the strategic treatment, this measure amounts to 0.7511 (0.7737) and in the control treatment to 0.9590 (0.9284) for groups of size $n = 2$ ($n = 3$). This difference is highly statistically significant for both n, with p-values of 0.0001.[33]

RESULT 3. *For both group sizes, the frequency of cut-off behavior is higher in the control treatment, contradicting Hypothesis 3.*

5.5 *Pioneers*

In the control treatment as well as in the best PBE, players are predicted to play risky on $(p_1^*, \frac{1}{2}]$; that is, 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^*. Thus, if subjects conformed to the best PBE, we should observe lonely pioneers for roughly the same proportion of time in both treatments.

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

[32]We are indebted to an anonymous referee for the suggestion.

[33]Our conclusion remains qualitatively unchanged if we instead use a more "lenient," less discriminating, measure where, after a switch from risky to safe, we focus only on a subject's second spell on the risky arm. Specifically, we measure the proportion of time in which a subject plays safe before playing risky plus the proportion of time taken up by the subject's second spell on the risky arm, before his risky arm is revealed to be good or the end of the game, whichever arrives first, and define 1 minus this proportion as our alternative measure of cut-off behavior. In the strategic treatment, this measure is 0.8341 (0.7975) while in the control treatment it is 0.9644 (0.9403) for groups of size $n = 2$ ($n = 3$). The corresponding p-values are both 0.0001. Thus, our conclusion that subject behavior was closer to cut-off behavior in the control treatment seems quite robust to how we measure the "distance to cut-off behavior."

Table 6. Proportion of time with a single pioneer.

Group Size	Strategic Treatment Obs.	Strategic Treatment Single Pioneer	Control Treatment Obs.	Control Treatment Single Pioneer
$n = 2$	60	0.634 [0.298]	60	0.198 [0.244]
$n = 3$	60	0.497 [0.338]	60	0.080 [0.168]

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

as large in the strategic treatment and the difference between treatments is highly statistically significant with p-values of 0.0001 for both $n = 2$ and $n = 3$. Thus, Hypothesis 4 is also emphatically rejected.

Result 4. *The proportion of time before a first breakthrough in the group during which exactly one player plays risky is higher in the strategic treatment, contradicting Hypothesis 4. This result holds for both $n = 2$ and $n = 3$.*

6. Discussion

In the previous section, we have seen that subjects' behavior differs starkly from the predictions of the best PBE. In this section, we provide additional results as well as robustness tests of the results presented in Section 5. We also discuss possible interpretations of subjects' behavior in light of some qualitative features of KRC's Markov perfect equilibria.

6.1 *Switches of action*

As pointed out above, in the best PBE, as well as in the single-agent optimum, players are predicted to switch from risky to safe at most once. Meanwhile, the turn-taking behavior predicted by MPE implies that players should switch arms more often in the strategic treatment. Yet, 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.[34] Recall that 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, for example, implies that one of the players must switch actions at least twice, with the other one switching once, before p_1^* is reached.[35]

[34]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 here, cannot be accounted for by this explanation. 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 suggestive evidence that subjects may indeed have endeavored to take turns, as predicted by KRC's MPEs.

[35]It is optimal for players to continue to play risky after observing a success until the game ends; in the strategic treatment, it does not matter whether the success has been achieved by the player himself or his

Table 7. Average number of switches per player.

Group Size	Strategic Treatment Obs.	Strategic Treatment Switches Per Player	Control Treatment Obs.	Control Treatment Switches Per Player
$n=2$	60	3.067 [2.450]	60	0.792 [1.063]
$n=3$	60	2.261 [2.040]	60	0.778 [1.080]

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

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.

Table 7 displays the average number of switches per player across games for our four treatments.[36] 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).

Result 5. *For both group sizes, the incidence of switches is higher in the strategic treatment.*

In addition, we examine and test for the difference in timing when the first switch from risky to safe in a given game occurred. For both group sizes, the first switch from risky to safe in calendar time is realized statistically significantly earlier in the strategic treatment (all p-values of 0.0001). As we have mentioned above, information accumulation is potentially faster in the strategic treatment. 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. To account for the fact that the decrease in beliefs may be up to n times faster in the strategic treatment, we compare n times the calendar time of the first switch in the strategic treatment to the calendar time of the first switch in the control treatment. For group size $n = 2$ ($n = 3$), the former was 36.31 (57.13) on average in the strategic treatment vs. 76.55 (76.14) in the control treatment. This difference is statistically significant for $n=2$ (p-value of 0.0007) and $n=3$ (p-value of 0.0330). Thus, for both n we can conclude that, on average, the first switch in the strategic treatment occurred at a statistically significantly more optimistic belief than in the control treatment. This is further evidence that participants attempted

partner, while in the control treatment, only a player's own successes are informative. Overall across both treatments, there were 65 (110) successes in the groups of size $n=2$ ($n=3$). Across all treatments and any n, only 6 subjects did not continuously explore until the end of the game after observing a success that resolves all uncertainty. Of these 6 players, 5 switched to safe for a few seconds and one subject reverted to playing safe after continuously playing risky for 120 seconds following his own success.

[36]While we run our hypothesis test with the average *incidence* of switches, we rather report the average *number* of switches in Table 7, as this may be easier to interpret. The number of switches is also statistically significantly higher in the strategic treatment for both n, with p-values of 0.0001.

to actively *free-ride* on the information generated by their partner(s). KRC's MPEs have the property that the first switch occurs at a belief strictly higher than the single-agent cut-off p_1^*. Taken together with the much higher frequency of cut-off play in the control treatment, this suggests that subjects' behavior may possibly be better predicted by MPE than by the best PBE. We explore this concept in greater depth in the following subsection.

6.2 *Groups of* $n = 2$ *versus* $n = 3$

As Figure 1 illustrates, behavior in the control treatment is remarkably similar across two- and three-player groups, as subjects do not have the opportunity to free-ride on the information generated by others. Meanwhile, in the strategic treatment, the coordination required by MPE play is decidedly more involved than that which underlies the best PBE. 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 role changes. Coordinating role changes is inherently more difficult in three-player groups. Therefore, one might expect that MPE-type behavior is more prevalent in groups of $n = 2$ players than in groups of size $n = 3$. Moreover, the size of the social dilemma is more than 50% larger in groups of size $n = 3$. Possible indicators of more MPE-like behavior in the strategic treatment would be less cut-off behavior, more switches and more single pioneers for the smaller group size $n = 2$.

Recall that our difference-in-differences analysis of experimentation intensities across belief regions and treatments (see Section 5.2) shows a statistically significantly larger difference between the *risky dominant* and *free-riding* regions in the strategic treatment only for groups of size $n = 2$. Moreover, the observed overall average experimentation intensity is furthest away from the MPE-prediction for groups of size $n = 3$, being significantly different from the observed value at the 10%-level (see Footnote 23). Our following result provides additional evidence that the more sophisticated forms of coordination required by MPE seem to be more prevalent for $n = 2$ than for $n = 3$.

Result 6. *The frequency of single pioneers is significantly higher in the strategic treatment for* $n = 2$ *than for* $n = 3$. *Furthermore, the incidence of switches is higher and cut-off behavior is less frequent in the strategic treatment for* $n = 2$ *than for* $n = 3$. *However, the latter two effects are not statistically significant.*

The p-value is 0.0252 for the proportion of time with a single pioneer. It is 0.2237 and 0.5096 for the average incidence of switches per player, and the average frequency of cut-off behavior, respectively. If we omit Games 5 and 6 (arguably outliers on account of their short length and the very early success, respectively), the difference in cut-off behavior is highly significant as well (p-value of 0.0101).[37] Thus, overall, our subjects are not behaving—or aiming to behave—as in the best PBE. Generally, our subjects' behavior seems qualitatively to be better described by MPE play, though the evidence for this conclusion is stronger for groups of size $n = 2$ than for $n = 3$.

[37] 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) for all six games, and even at the 1%-level for Games 1–4 only.

6.3 *Attention paid to partners' experimentation efforts*

As a robustness test, we would like to ensure that the differences in behavior and payoffs between the strategic and control treatments, which we are observing, are indeed due to the positive informational externality theory predicts. To do so, we study directly how much heed subjects paid to the information provided by their partner(s). We employ eye-tracking data obtained by 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 revealed both a player's actions and payoffs. While the use of this technology imposed subject constraints in the data-collection process, it allows us to gain additional insights into subjects' cognitive processes with the aim of better understanding subjects' behavior in the strategic treatment *relative* to the control treatment. If subjects were not making *any* use of the free information provided by their partners, then no statistically significant difference in observed attention should be detected. This in turn would invalidate any game-theoretical explanation of observed differences in behavior or payoffs between the strategic and control treatments, since theory predicts that the only source of strategic interaction in our game is the positive externality that arises because the information players produce is a public good.[38] 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.

As Table 8 shows, the average fixation intensity is much lower in the strategic treatment. This is highly statistically significant for both group sizes (both p-values are 0.0001 for $n = 2$ and $n = 3$). While a subject who is unsure about how to solve his decision problem might also be tempted to "copy" from his partner in the control treatment, the fact that players focus on each other much more in the strategic treatment very much suggests a strategic rationale for players' behavior, in that they are trying to learn about the quality of their own risky arms by observing their partners' exploration efforts. These results furthermore suggest that subjects do indeed understand the simple, nonstrategic, nature of the control treatment.

Table 8. Average fixation intensities.

	Strategic Treatment		Control Treatment	
Group Size	Obs.	Fixation Intensity	Obs.	Fixation Intensity
$n = 2$	60	0.614 [0.087]	60	0.865 [0.090]
$n = 3$	60	0.383 [0.078]	60	0.712 [0.106]

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

[38]Video recordings illustrating the use of the eye-tracking devices are available at the author's website: www.johanneshoelzemann.com. Heat maps spotlighting information search and attention behavior can be found in Appendix A (Hoelzemann and Klein (2021)).

6.4 *OLS estimations*

As a further robustness test and to complement the nonparametric analysis in Section 5 and key elements discussed so far in this section, we ran ordinary least-square regressions with random effects controlling for learning effects. In particular, we regressed experimentation intensity, payoffs, cut-off strategy, switches of action, and fixation intensity on the treatment dummy *Correlation*, which is 0 for the control treatment and 1 for the strategic treatment. Recall that subjects played the six games in random order and any order of these games that was used for k ($k \in \{1, \dots, 10\}$) groups in the strategic treatment was replicated for k groups in the control treatment. In order to verify that subjects treated the games they successively played as independent games rather than as parts of a larger super-game, we define a weighted learning function $\{g_o\} = \{\frac{1}{o}\}$ where o ($o \in \{1, \dots, 6\}$) corresponds to the random order in which each subject was exposed to each game. All regressions control for trends over time using this weighted learning function. The results do not qualitatively change when we replace the learning function with a linear version such that $\{g_o\} = \{o\}$. To account for the fact that behavior within

Table 9. OLS estimations with random effects of experimentation intensity, payoffs, cut-off strategy, switches of action, and fixation intensity.

	Experimentation Intensity						
	ALL	Risky Dom	Free-Riding	Payoffs	Cut-Off Strategy	Switches of Action	Fixation Intensity
			Panel A: $n = 2$				
Intercept	0.883	0.973	0.812	983.861	0.873	0.827	0.851
	(0.053)	(0.051)	(0.070)	(60.227)	(0.063)	(0.296)	(0.023)
Correlation	−0.223	−0.191	−0.299	204.750	−0.542	2.275	−0.251
	(0.050)	(0.050)	(0.066)	(55.838)	(0.058)	(0.378)	(0.024)
Learning	−0.148	−0.118	−0.016	106.566	−0.089	−0.081	0.033
	(0.119)	(0.115)	(0.141)	(103.356)	(0.103)	(0.659)	(0.035)
σ_ϵ	0.217	0.242	0.254	1474.627	0.411	1.946	0.103
σ_u	0.117	0.111	0.150	0	0.073	0.543	0.071
N	240	200	148	240	240	240	240
(Between) R-squared	0.420	0.319	0.375	0.116	0.705	0.602	0.716
			Panel B: $n = 3$				
Intercept	0.903	1.001	0.763	963.333	0.878	0.738	0.717
	(0.058)	(0.064)	(0.082)	(52.382)	(0.082)	(0.317)	(0.030)
Correlation	−0.300	−0.300	−0.282	439.056	−0.461	1.483	−0.330
	(0.062)	(0.090)	(0.073)	(44.900)	(0.085)	(0.438)	(0.030)
Learning	−0.145	−0.158	0.065	40.658	−0.203	0.089	−0.011
	(0.114)	(0.149)	(0.126)	63.533	(0.147)	(0.508)	(0.049)
σ_ϵ	0.253	0.216	0.292	1449.951	0.399	1.622	0.112
σ_u	0.126	0.205	0.177	0	0.182	0.918	0.103
N	360	300	223	360	360	360	360
(Between) R-squared	0.491	0.336	0.283	0.404	0.500	0.311	0.693

Note: For all estimations, robust standard errors are clustered at the group level and shown in brackets. The estimates for *Intercept* and *Correlation* are significant at the 1% level in all the reported instances; *Learning* is never significant at the 10% level.

groups of two (three) participants is not independent, we treat each group as our units of statistically independent observations and cluster standard errors by group.

Table 9 lists the results from this analysis where Panel A shows the results for two-player groups and Panel B displays the results for $n = 3$. We find a strong negative effect of the correlation structure, our *strategic treatment*, on experimentation intensity across both belief regions, cut-off strategy, and fixation intensity. By contrast, we find a strong positive effect of the additional presence of one (two) perfectly positively correlated arm(s) on stage game payoffs and switches of action.[39]

Thus, our OLS estimations with random effects confirm all of our previous, nonparametric results.[40] In particular, there is no evidence of super-game effects, as subjects' behavior does not change in response to the number of games they previously played.[41]

7. 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, evolved as the game progressed. 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 i.i.d., and there were therefore no strategic links across group members.

We have shown that experimentation intensities are lower in the strategic treatment. In particular, this is the case in the *free-riding belief region*, which points to strategic free-riding. Moreover, subjects seem to attempt to coordinate in rather elaborate ways, as evidenced, inter alia, by the much lower incidence of cut-off behavior and the higher incidence of lonely pioneers in the strategic setting. Overall, this leads us to reject the hypothesis that subjects played according to the best PBE. Indeed, the best PBE would predict no free-riding in the *free-riding region*, cut-off play and no lonely pioneers. While this does of course not constitute conclusive evidence in favor of MPE, it bears noting that these behaviors are fully in line with the qualitative predictions of MPE.

Why subjects should refrain from engaging in the simple cut-off behavior prescribed by the welfare-optimal equilibrium seems somewhat of a puzzle. The control treatment shows that subjects were not, in principle, averse to playing cut-off strategies. We conjecture that, in the strategic setting, the idea of taking turns (Cason, Lau, and Mui (2013),

[39] We do not report estimates of the proportion of time with a single pioneer as the interpretation of the lonely pioneers is only sensible at the individual group-level.

[40] Our results remain qualitatively unchanged if we use our "continuous" cutoff measure (see p. 19) instead.

[41] We also ran our nonparametric analysis using only the last (first) 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.

Leo (2017)), as evidenced by the greater prevalence of lonely pioneers and a greater number of switches, was attractive to subjects. Turn-taking is a feature of any of KRC's MPEs in this setting. Vespa and Wilson (2015, 2019) had already documented a tendency of a substantial fraction of subjects to adopt MPE behavior; in their setting, MPE play was simpler—in ours, it is not. Our investigation shows that, even in a setting like ours, where from a theoretical perspective one equilibrium is welfare-optimal and prescribes particularly simple behavior, making it an obvious candidate for a focal equilibrium, this equilibrium may well not describe subjects' behavior accurately. Our findings may thus counsel caution when supposing which, among a multitude of equilibria, may be considered focal by players. We commend the analysis of other dynamic games with an *a priori* "obvious" candidate for a focal equilibrium for future research. In particular, it would be interesting to analyze a game in which the welfare-optimal equilibrium was both particularly simple in structure *and* Markovian.

As a further robustness test, 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 nudging subjects toward certain behaviors, which would have made the interpretation of our results more difficult. It might also 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, such as, for instance, the Poisson setting with inconclusive breakthroughs à la Keller and Rady (2010), or 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.

References

Akcigit, U. and Q. Liu (2016), "The role of information in innovation and competition." *Journal of the European Economic Association*, 14 (4), 828–870. [1023]

Ambrus, A. and B. Greiner (2012), "Imperfect public monitoring with costly punishment: An experimental study." *American Economic Review*, 102 (7), 3317–3332. [1022]

Anderson, C. M. (2001), *Behavioral Models of Strategies in Multi-Armed Bandit Problems.* Ph.D. thesis, California Institute of Technology. [1026]

Anderson, C. M. (2012), "Ambiguity aversion in multi-armed bandit problems." *Theory and Decision*, 72 (1), 15–33. [1026]

Andreoni, J. (1988), "Why free ride? Strategies and learning in public goods experiments." *Journal of Public Economics*, 37 (3), 291–304. [1022]

Banks, J., M. Olson, and D. Porter (1997), "An experimental analysis of the bandit problem." *Economic Theory*, 10 (1), 55–77. [1026]

Battaglini, M., S. Nunnari, and T. R. Palfrey (2014), "Dynamic free riding with irreversible investments." *American Economic Review*, 104 (9), 2858–2871. [1027]

Battaglini, M., S. Nunnari, and T. R. Palfrey (2016), "The dynamic free rider problem: A laboratory study." *American Economic Journal: Microeconomics*, 8 (4), 268–308. [1026, 1027]

Bellman, R. (1956), "A problem in the sequential design of experiments." *Sankhyā: The Indian Journal of Statistics (1933–1960)*, 16 (3/4), 221–229. [1025]

Bergemann, D. and J. Välimäki (1996), "Learning and strategic pricing." *Econometrica*, 64 (5), 1125–1149. [1025]

Bergemann, D. and J. Välimäki (2000), "Experimentation in markets." *The Review of Economic Studies*, 67 (2), 213–234. [1025]

Bergemann, D. and J. Välimäki (2008), *"Bandit Problems," The New Palgrave Dictionary of Economics*, second edition. Palgrave and Macmillan Press. [1025]

Besanko, D., J. Tong, and J. J. Wu (2018), "Subsidizing research programs with "if" and "when" uncertainty in the face of severe informational constraints." *The RAND Journal of Economics*, 49 (2), 285–310. [1023]

Besanko, D. and J. Wu (2013), "The impact of market structure and learning on the trade-off between R&D competition and cooperation." *The Journal of Industrial Economics*, 61 (1), 166–201. [1023]

Bolton, P. and C. Harris (1999), "Strategic experimentation." *Econometrica*, 67 (2), 349–374. [1022, 1025, 1037, 1047]

Bolton, P. and C. Harris (2000), "Strategic experimentation: The undiscounted case." In *Incentives, Organizations and Public Economics–Papers in Honour of Sir James Mirrlees*, 53–68, Oxford University Press, Oxford. [1025]

Bonatti, A. and J. Hörner (2011), "Collaborating." *American Economic Review*, 101 (2), 632–663. [1025]

Boyce, J. R., D. M. Bruner, and M. McKee (2016), "Strategic experimentation in the lab." *Managerial and Decision Economics*, 37 (6), 375–391. [1025]

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. [1025]

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. [1030]

Cason, T. N., S.-H. P. Lau, and V.-L. Mui (2013), "Learning, teaching, and turn taking in the repeated assignment game." *Economic Theory*, 54 (2), 335–357. [1027, 1046]

Chaudhuri, A. (2011), "Sustaining cooperation in laboratory public goods experiments: A selective survey of the literature." *Experimental Economics*, 14 (1), 47–83. [1021]

Chernulich, A., J. Horowitz, J. P. Rabanal, O. A. Rud, and M. Sharifova (2020), "Entry and exit decisions under public and private information: An experiment." Working Paper. [1027]

Coe, D. T. and E. Helpman (1995), "International R&D spillovers." *European Economic Review*, 39 (5), 859–887. [1022]

Das, K. and N. Klein (2020), "Do stronger patents lead to faster innovation? The effect of duplicative search." Working Paper. [1023]

Fehr, E. and S. Gächter (2000), "Cooperation and punishment in public goods experiments." *American Economic Review*, 90 (4), 980–994. [1022]

Fenig, G., G. Gallipoli, and Y. Halevy (2018), "Piercing the "payoff function" veil: Tracing beliefs and motives." University of Toronto, Department of Economics Working Paper No. 625. [1022]

Fischbacher, U. (2007), "z-tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics*, 10 (2), 171–178. [1030]

Fudenberg, D. and E. Vespa (2019), "Learning theory and heterogeneous play in a signaling-game experiment." *American Economic Journal: Micro*, 11 (4), 186–215. [1027]

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. [1026]

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*, Vol. 1, 241–266, North Holland, Amsterdam. [1025]

Greiner, B. (2015), "Subject pool recruitment procedures: Organizing experiments with ORSEE." *Journal of the Economic Science Association*, 1 (1), 114–125. [1030]

Griliches, Z. (1988), "Productivity puzzles and R&D: Another nonexplanation." *Journal of Economic Perspectives*, 2 (4), 9–21. [1022]

Grossman, G. M. and E. Helpman (1993), *Innovation and Growth in the Global Economy*. MIT Press. [1022]

Hoelzemann, J. and N. Klein (2021), "Supplement to 'Bandits in the lab'." *Quantitative Economics Supplemental Material*, 12, https://doi.org/10.3982/QE1389. [1025, 1033, 1038, 1044]

Hörner, J., N. Klein, and S. Rady (forthcoming), "Overcoming free-riding in bandit games." *The Review of Economic Studies*. [1023]

Hörner, J. and A. Skrzypacz (2017), "Learning, experimentation, and information design." In *Advances in Economics and Econometrics: Eleventh World Congress*, 63–98, Cambridge University Press. [1025]

Hudja, S. (2019), "Voting for experimentation: A continuous time analysis." Working Paper. [1027]

Hudja, S. and D. Woods (2020), "Behavioral bandits: Analyzing the exploration versus exploitation trade-off in the lab." Working Paper. [1026]

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. [1022]

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. [1022]

Isaac, R. M. and J. M. Walker (1988a), "Communication and free-riding behavior: The voluntary contribution mechanism." *Economic Inquiry*, 26 (4), 585–608. [1022]

Isaac, R. M. and J. M. Walker (1988b), "Group size effects in public goods provision: The voluntary contributions mechanism." *The Quarterly Journal of Economics*, 103 (1), 179–199. [1022]

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*, 573–605, Elsevier. [1044]

Keller, G., S. Rady, and M. Cripps (2005), "Strategic experimentation with exponential bandits." *Econometrica*, 73 (1), 39–68. [1022]

Keller, G. and S. Rady (2010), "Strategic experimentation with Poisson bandits." *Theoretical Economics*, 5 (2), 275–311. [1025, 1047]

Keller, G. and S. Rady (2015), "Breakdowns." *Theoretical Economics*, 10 (1), 175–202. [1025]

Kim, O. and M. Walker (1984), "The free rider problem: Experimental evidence." *Public Choice*, 43 (1), 3–24. [1022]

Klein, N. (2013), "Strategic learning in teams." *Games and Economic Behavior*, 82, 636–657. [1025]

Klein, N. and S. Rady (2011), "Negatively correlated bandits." *The Review of Economic Studies*, 78 (2), 693–732. [1025]

Ledyard, J. O. (1995), "Public goods: A survey of experimental research." In *Handbook of Experimental Economics*, 111–194, Princeton University Press, Princeton. [1021]

Leo, G. (2017), "Taking turns." *Games and Economic Behavior*, 102, 525–547. [1027, 1047]

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. [1022]

Marwell, G. and R. E. Ames (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. [1022]

Marwell, G. and R. E. Ames (1981), "Economists free ride, does anyone else? Experiments on the provision of public goods." *Journal of Public Economics*, 15 (3), 295–310. [1022]

Meyer, R. J. and Y. Shi (1995), "Sequential choice under ambiguity: Intuitive solutions to the armed-bandit problem." *Management Science*, 41 (5), 817–834. [1026]

Murto, P. and J. Välimäki (2011), "Learning and information aggregation in an exit game." *The Review of Economic Studies*, 78 (4), 1426–1461. [1025]

Oprea, R., G. Charness, and D. Friedman (2014), "Continuous time and communication in a public-goods experiment." *Journal of Economic Behavior & Organization*, 108, 212–223. [1026]

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*, 363–378, Springer. [1044]

Presman, E. L. (1990), "Poisson version of the two-armed bandit problem with discounting." *Theory of Probability & Its Applications*, 35 (2), 307–317. [1025]

Reshidi, P., A. Lizzeri, L. Yariv, J. Chan, and W. Suen (2020), "Individual and collective information acquisition: An experimental study." Working Paper. [1026]

Robbins, H. (1952), "Some aspects of the sequential design of experiments." *Bulletin of the American Mathematical Society*, 58, 527–535. [1025]

Romer, P. M. (1990), "Endogenous technological change." *Journal of Political Economy*, 98 (5, Part 2), S71–S102. [1022]

Rosenberg, D., E. Solan, and N. Vieille (2007), "Social learning in one-arm bandit problems." *Econometrica*, 75 (6), 1591–1611. [1025]

Rothschild, M. (1974), "A two-armed bandit theory of market pricing." *Journal of Economic Theory*, 9 (2), 185–202. [1025]

Strulovici, B. (2010), "Learning while voting: Determinants of collective eExperimentation." *Econometrica*, 78 (3), 933–971. [1027]

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. [1025]

Vespa, E. and A. J. Wilson (2015), "Dynamic incentives and Markov perfection: Putting the 'conditional' in conditional cooperation." CESifo Working Paper Series. [1027, 1047]

Vespa, E. and A. J. Wilson (2019), "Experimenting with the transition rule in dynamic games." *Quantitative Economics*, 10 (4), 1825–1849. [1027, 1047]

Wilson, A. J. and E. Vespa (2020), "Information transmission under the shadow of the future: An experiment." *American Economic Journal: Microeconomics*, 12 (4), 75–98. [1026]

Zelmer, J. (2003), "Linear public goods experiments: A meta-analysis." *Experimental Economics*, 6 (3), 299–310. [1021]

Co-editor Christopher Taber handled this manuscript.

Manuscript received 12 July, 2019; final version accepted 27 February, 2021; available online 16 March, 2021.