Bandits in the Lab *

Johannes Hoelzemann† Nicolas Klein‡

This version: September 12, 2017
Preliminary and Incomplete

Abstract

We conduct an experimental test of the main theoretical predictions of the model of strategic experimentation with exponential bandits by Keller, Rady, Cripps (2005). We find strong evidence for their prediction of free-riding because of strategic concerns. While experimental subjects are not able to update their beliefs precisely, we nonetheless find strong support for the equilibrium prediction of non-cutoff behavior as well.

Keywords: Strategic Experimentation, Exponential Bandits, Learning, Dynamic Games, Continuous Time, Laboratory Experiments, Eye Tracking.

JEL Classification Numbers: C73, C92, D83, O32

*Financial support by the UNSW Bizlab and the School of Economics at UNSW Sydney is gratefully acknowledged. Nicolas Klein also gratefully acknowledges financial support from the *Fonds de Recherche du Québec - Société et Culture* and the *Social Sciences and Humanities Research Council of Canada*. This research has been approved by the Human Research Ethics Committee of UNSW Sydney under approval number HC 17069. We thank Francisco Alvarez-Cuadrado, Denzil Fiebig, Ben Greiner, Richard Holden, Hongyi Li and Sven Rady 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.

†UNSW Sydney. Mailing address: UNSW Business School, School of Economics, Anzac Parade 2052 NSW, Australia; email: j.hoelzemann@unsw.edu.au.

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

1 Introduction

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

Multi-armed bandit models have become canonical in economics to study such dynamic trade-offs. At each point in time, a decision maker either optimally exploits the information he already has, or he decides to invest in exploration in order to make better future decisions. Until fairly recently, the literature focussed on an individual decision maker's trade-off acting in isolation. Bolton & Harris (1999) and Keller, Rady, Cripps (2005; subsequently: KRC) have extended the individual choice problem to a multi-player continuous-time framework. There now appears a strategic component to the information-acquisition problem, in that other players now also benefit from the information acquired at a cost by a given player. To make the problem tractable, these papers are focussing on the choice between a safe arm, yielding a known payoff, and a risky arm, which yields payoffs following a stochastic process. The time-invariant quality of this risky arm can be good or bad. If it is good (bad), it dominates (is dominated by) the safe arm. Whether the risky arm is good or bad is initially unknown and can only be found out by trying it out over time. Trying it out is costly, however, as it means forgoing the safe payoff. As the quality of the risky arm is assumed to be the same across players, and players can observe each other's actions and payoffs, there is a positive informational externality associated with a player's use of the risky arm. This gives rise to a dynamic public-good problem, where the public good in question is the dynamically evolving information about agents' common state of the world.

While the game-theoretical analysis of these problems has yielded many sharp insights, empirical evidence for its predictions has thus far been scarce. Indeed, the dynamic nature of the problem and the continuous-time setting underlying its theoretical analysis raise some challenges both for the collection of field data and the experimental implementation in the laboratory. To the best of our knowledge, we are the first to implement an experimental test of continuous-time strategic-experimentation models in the laboratory. Our analysis relies on comparing the behavior of our experimental subjects in groups where the quality of the risky arm was known to be the same for all partners (which we call the *strategic treatment*) to that of groups where its quality was iid across members, the *control treatment*. When the quality of the risky arm is known to be the same across players, rational agents will take into account the result of their partners' experimentation when up-

dating 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, i.e., there are no payoff externalities.

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

KRC make two fundamental qualitative predictions regarding players' equilibrium behavior. As information is a public good, players will produce inefficiently little of it. Furthermore, it is predicted that all players will not use a simple cutoff strategy in equilibrium. A cutoff strategy is defined by a unique threshold belief above which it prescribes risky play, while prescribing safe play below it.

We find strong evidence on both counts. Our experimental subjects exhibit a strong tendency for free-riding on the experimentation provided by their partner(s). Indeed, the risky arm is used at a significantly lower intensity in the strategic treatment. This is consistent with the game-theoretical analysis in KRC, who found that there exists a range of beliefs at which a single player will play risky but where safe and risky are mutually best responses in the strategic setting.

We furthermore find strong evidence for players' adopting more complex behaviors than cutoff strategies in the strategic treatment. Indeed, players switch much more between safe and risky, and use cutoff strategies much less frequently, than they do in the control treatment. Moreover, there is a larger proportion of time during which exactly one player is playing risky in the strategic treatment. This, combined with the high frequency of switches, is fully consistent with the players' switching between the roles of pioneer and free-rider, which characterizes equilibrium play at intermediate beliefs in KRC.

All these observations are in line with the complex coordination required by KRC equilibrium play. Yet, we of course cannot conclude that our experimental subjects fully adopt equilibrium behavior. In fact, we proceed to subdivide the players' beliefs into a region where risky is the dominant action choice, a region where safe and risky are mutually best responses and one where safe is a dominant action, and find no striking qualitative differences in players' behavior across these regions. Thus, while our subjects adopt some of the qualitative aspects of equilibrium behavior, they do not seem strictly to separate these different strategic regions. Furthermore, they will often extend experimentation even below the single-agent cutoff in both the strategic and the control treatments,

which leads us to conjecture that subjects are not able to compute beliefs and cutoffs precisely. Indeed, even in the region where safe is the dominant action, the average experimentation intensity is higher in the control treatment, which is inconsistent with the presence of an encouragement effect.[1]

The rest of the paper is organized as follows: Section 2 reviews some of the literature; Section 3 explains the KRC model in more detail; Section 4 sets out our experimental implementation; Section 5 discusses our findings, and Section 6 concludes. Appendix **??** exhibits and explains the interface our experimental subjects were using, while Appendix A 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, 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 & Jones (1974) showed that, if arms are stochastically independent of each other and the state of only one arm can evolve at any one time, an optimal policy in the multi-armed bandit problem is given by the so-called "Gittins Index" policy. For this policy, one can consider the problem of stopping on each arm in isolation from the other arms. The value of this stopping problem is the so-called *Gittins Index* for this arm. Now, an optimal policy consists of, at each point in time, using the arm with the highest Gittins Index. Presman (1990) calculated the Gittins Index for the case in which the underlying stochastic process is a Poisson process. Bergemann & Välimäki (2008) give a survey of this literature.

Bolton & Harris (1999, 2000) were the first to consider the multi-player version of the two-armed bandit problem. While they assumed that the underlying stochastic process was a Brownian motion, KRC analyzed the corresponding problem with exponential processes. This model proved to be more tractable and is underlying our theoretical hypotheses.[2]

[1] The *encouragement effect* has been identified by Bolton & 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 receiving 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.

[2] Many variants of the multi-player bandit problem have been analyzed since. In Keller & Rady (2010), a bad risky arm also sometimes yields a payoff. In Klein & Rady (2011), the quality of the risky arm is negatively correlated across players. Klein (2013) introduces a second risky arm, with a quality that is negatively correlated with that of the first.

The only papers we are aware of that conduct experimental tests of bandit problems are Meyer & Shi (1995), Banks, Olson, Porter (1997), Anderson (2001, 2012), and Gans, Knox, Crosson (2007). All these papers consider various single-agent problems. We are not aware of any previous experimental study of a *strategic*, multi-player, bandit problem.

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 at each instant. Payoffs are discounted at a rate $r > 0$. If they knew the quality of the risky arm, players would have a strictly dominant strategy always to pull a good risky arm and never to pull a bad one. They are initially uncertain whether their risky arm is good or bad. Yet, the only way to acquire information about the quality of the risky arm is to use it, which is costly as it implies forgoing the safe payoff flow s. The n players' risky arms are either all good or all bad. Players share a common prior belief $p_0 \in (0, 1)$ that their risky arms are good. Every player's actions as well as the outcomes of their actions are publicly observable; therefore, the information one player produces benefits the other players as well, creating incentives for players to free-ride on their partners' efforts. Players thus share a common posterior belief p_t at all times $t \in \mathbb{R}_+$. All the parameter values and the structure of the game are common knowledge.

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

In Keller & Rady (2015), the lump-sum payoffs are costs to be minimized. Rosenberg, Solan, Vieille (2007) and Murto & Välimäki (2011) analyze the case of privately observed payoffs, while Bonatti & Hörner (2011) investigate the case of privately observed actions. Bergemann & Välimäki (1996, 2000) consider strategic experimentation in buyer-seller settings. Hörner & Skrzypacz (2016) give a survey of this literature.

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

KRC go on to analyze the game of strategic information acquisition, where each player maximizes his own payoff, not taking into account that the information he produces is valuable to the other players as well. They analyze perfect Bayesian equilibria in Markov strategies (MPE), i.e., strategies where a player's action after any history can be written as a time-invariant function $k_i(p)$ of the common belief at that history. It is shown that, in any MPE with a finite number of switches, all players will set $k_i(p) = 0$ for all $p \leq p_1^*$ (see Proposition 6.1 in KRC). Moreover, it is shown that there exists no MPE in which all players play a cut-off strategy, i.e. a strategy that prescribes the use of the risky arm for beliefs above a single cutoff and that of the safe arm below. Thus, in stark contrast to the simple structure of the single-agent optimum, every MPE has the property that, for intermediate beliefs, players switch roles between experimenter and free-rider at least once. As a matter of fact, KRC show that, for any given number of switches greater than, or equal to, one, there exists an MPE with that number of switches. The behavioral prediction is thus that players switch roles for intermediate beliefs at least once.

KRC show that players' effort levels are strategic substitutes for intermediate beliefs. For beliefs close to 1 (0), playing risky (safe) is a dominant action.

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 $r = 1/120$. 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.[3] We set $p_0 = \frac{1}{2}$, $s = 10$, $h = 2500$, $\lambda = 1/100$. Thus, $25 = g > s = 10$. The realizations of all random processes were simulated ahead of time.[4] We generated six different sets of realizations of the random parameters, corresponding to six different

[3] 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 Appendix **??** for details, and Appendix A for the instructions the subects received.

[4] 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.

games each of our subjects played. To make our findings more easily comparable, we have kept the same realizations for both the strategic and the control treatments.[5] One unit of time corresponds to a second in our experimental implementation.

In keeping with the theoretical predictions, we have endeavored to implement our experimental investigation in continuous time, subject to the restrictions imposed by the available computing power.[6]

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 the balance of the data-collection process, we replicated any order of the six games that was used for k ($k \in \{1, \cdots, 10\}$) groups in the strategic treatment for k groups in the control treatment as well. Subjects could see their fellow group members' action choices and payoffs on their computer screens. They had to choose an action before the game started and could switch their action at any point in time by clicking on the corresponding button with their mouse. (Please see the Appendix **??** for details and screen shots.)

All experimental sessions took place in July and August 2017 at the BizLab Experimental Research Laboratory at UNSW Sydney. All subjects were recruited from the university's subject pool and administered by the online recruitment system ORSEE (Greiner, 2015). All participants were native speakers of English. In total, 100 subjects, 46 of whom were female, participated in 60 sessions. The participants' age ranged from 18 to 35 years, with an average of 20.78 and a standard deviation of 2.43. Because the implementation was computationally very intensive and because we wanted to collect eye-tracking data, only between 2 and 3 subjects participated at a time in each session. Upon arrival, participants were seated in front of a computer at desks which were separated by dividers to minimize potential communication. Participants received written instructions and had the opportunity to ask questions.[7] After the subjects had successfully completed a simple comprehension test, the eye-tracking devices were calibrated, after which the subjects started the experiment. The experiment was programmed in zTree (Fischbacher, 2007). At the end of the experiment, we collected some information on participants' demographic attributes and risk attitudes. They were then privately paid their cumulated experimental earnings from one randomly selected game in cash (with a conversion rate of E$ 100 = AU$ 1) plus a show-up fee of AU$ 5. The average earning was AU$ 23.86, with a standard deviation of AU$ 9.95.

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

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

[7] The instructions handed out to all participants can be found in the Appendix A.

4.2 Behavioral Hypotheses

One of the main theoretical predictions is that players use the risky arm less in a strategic setting than in a situation in which they are single players. This is because players *free-ride* on the information their partners are producing. Indeed, while players are predicted to play safe at all beliefs $p \leq p_1^*$ in both instances, a single player should play risky at all beliefs $p > p_1^*$. By contrast, since at least one player is not playing a cut-off strategy in any MPE, at least one player will play safe at some beliefs above p_1^* in any MPE. 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 cutoff belief, such that, for all beliefs in $(p_1^*, p^\ddagger)$, at least one player plays safe. Indeed, as KRC show (their Equation (6), p.49), it is a best response for player i to play safe if and only if his value function $u_i(p)$ satisfies $u_i(p) \leq s + K_{-i}(p)c(p)$, where $K_{-i}(p) := \sum_{j \neq i} k_j(p)$ is the number of players other than i who play risky at belief p, and $c(p) := s - pg$ is a player's myopic opportunity cost for playing risky, given the belief p. An upper bound on a player's equilibrium value function u_i is given by V_{n,p_1^*}, the value function of all players playing risky on $(p_1^*, 1]$, and safe on $[0, p_1^*]$. Thus, a lower bound $p^\ddagger$ is given by the unique root $V_{n,p_1^*}(p^\ddagger) - s - (n-1)c(p^\ddagger) = 0$. By the same token, we can derive an upper bound $\bar{p}$ on the lowest belief at which risky is a dominant action. For this, we use the fact that the single-agent value function V_1^* constitutes a lower bound on a player's equilibrium value function u_i, and find our upper bound $\bar{p}$ as the unique root $V_1^*(\bar{p}) - s - (n-1)c(\bar{p}) = 0$.

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

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

Hypothesis 4.1 *The average experimentation intensity* $\frac{\int_0^{\hat{T}} \sum_{i=1}^n k_{i,t}\, dt}{n\hat{T}}$ *is significantly lower in the strate-*

gic treatment than in the control treatment.

As explained above, KRC predict that subjects will use cut-off strategies in the control treatment, whereas at least one player will not use a cut-off strategy in the strategic setting. *Cut-off behavior* consists in a player's playing risky at the outset, and continuing to play risky until his risky arm is revealed to be good, the game ends, or he switches to the safe action, and continues to play safe until the game ends or his risky arm is revealed to be good. We can now state the following

Hypothesis 4.2 *The frequency of cut-off behavior is significantly higher in the control treatment than in the strategic treatment.*

The same theoretical prediction moreover implies that players should switch arms more often in the strategic treatment. As noted above, learning also tends to be faster in the strategic setting, so that beliefs may more quickly reach the threshold at which the player will want to change his action. This effect would add to making switching more prevalent in the strategic treatment.

To control for the effect that, the longer the game goes on, the more time players have to switch actions, we define the *incidence of switches* as the number of a player's switches in a given game per unit of effective time, where *effective time* is understood as the time before the game ends or the player's risky arm is revealed to be good, whichever happens first. Thus, we have the following

Hypothesis 4.3 *The incidence of switches is significantly higher in the strategic treatment than in the control treatment.*

To find further evidence for the strategic rationale underlying the switching of roles between players, which characterizes the simple MPE in KRC, we measure the proportion of time (before a first breakthrough) during which exactly one of the players plays risky. We formulate the following

Hypothesis 4.4 *The proportion of time before a first breakthrough during which exactly one player plays risky is higher in the strategic treatment than in the control treatment.*

5 Experimental Results

5.1 Overview

Figures 1 and 2 display the evolution of players' action choices over all six games. Figure 1 shows Games 1 and 3 at the top left and top right, respectively, while exhibiting Game 2 at the bottom.

Figure 2 shows Game 4 at the top, and Games 5 and 6 at the bottom left and right, respectively. Players' actions are described by dots, the width of which corresponds to one second of time. For each of the six games, we conducted four treatments à ten groups each, the parameters of which (i.e. their duration, the quality of the risky arm and the timing of successes on the risky arm in case it was good) we had simulated ahead of time, as explained in Section 4. Groups 1-10 correspond to the strategic treatment for two-player groups; groups 11-20 are the corresponding control treatments. Groups 21-30 played the strategic treatment for three-player groups, while groups 31-40 were the corresponding control treatments. In each group, we refer to the lowermost player as 'player 1', while 'player 2' will denote the player right above, and 'player 3' is the uppermost player. The x-axis represents calendar time. A *red* dot indicates that a player is playing *risky* in a given second, while a *blue* dot indicates that the player is playing *safe*. A black square indicates a success.

As the figures show, the duration of the games ranged from 32 seconds for Game 5 to 230 seconds for Game 4. As is furthermore evident from the figures, players change their behaviors over time. While often playing risky at the beginning, players seem to grow less inclined to use the risky arm the longer it has unsuccessfully been used before. This is consistent with Bayesian updating of a prior belief. As we shall discuss in more detail below, this of course does not imply that players adjust their behaviors precisely at the equilibrium cutoff beliefs. Nevertheless, as our subsequent analysis will show, the main qualitative predictions of Markov Perfect Equilibrium are borne out by the experimental evidence.

5.2 Average Experimentation Intensities

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

Under Hypothesis 4.1, players will use the risky arm less in the strategic treatment. The data provides support for this hypothesis.

Result 5.1 *The average experimentation intensity* $\frac{\int_0^{\hat{T}} \sum_{i=1}^{n} k_{i,t}\, dt}{n\hat{T}}$ *is significantly lower in the strategic treatment, as compared to the control treatment. This result holds for both* $n = 2$ *and* $n = 3$.

As Table 1 reveals, the additional presence of one (two) perfectly positively correlated arms leads to lower experimentation intensities in all games. This is statistically significant for Games 1-5, but not for Game 6, in both settings with $n = 2$ and $n = 3$. The corresponding *p*-values in the case of

Figure 1: ACTION CHOICES BY PLAYERS OVER TIME: Games 1-3

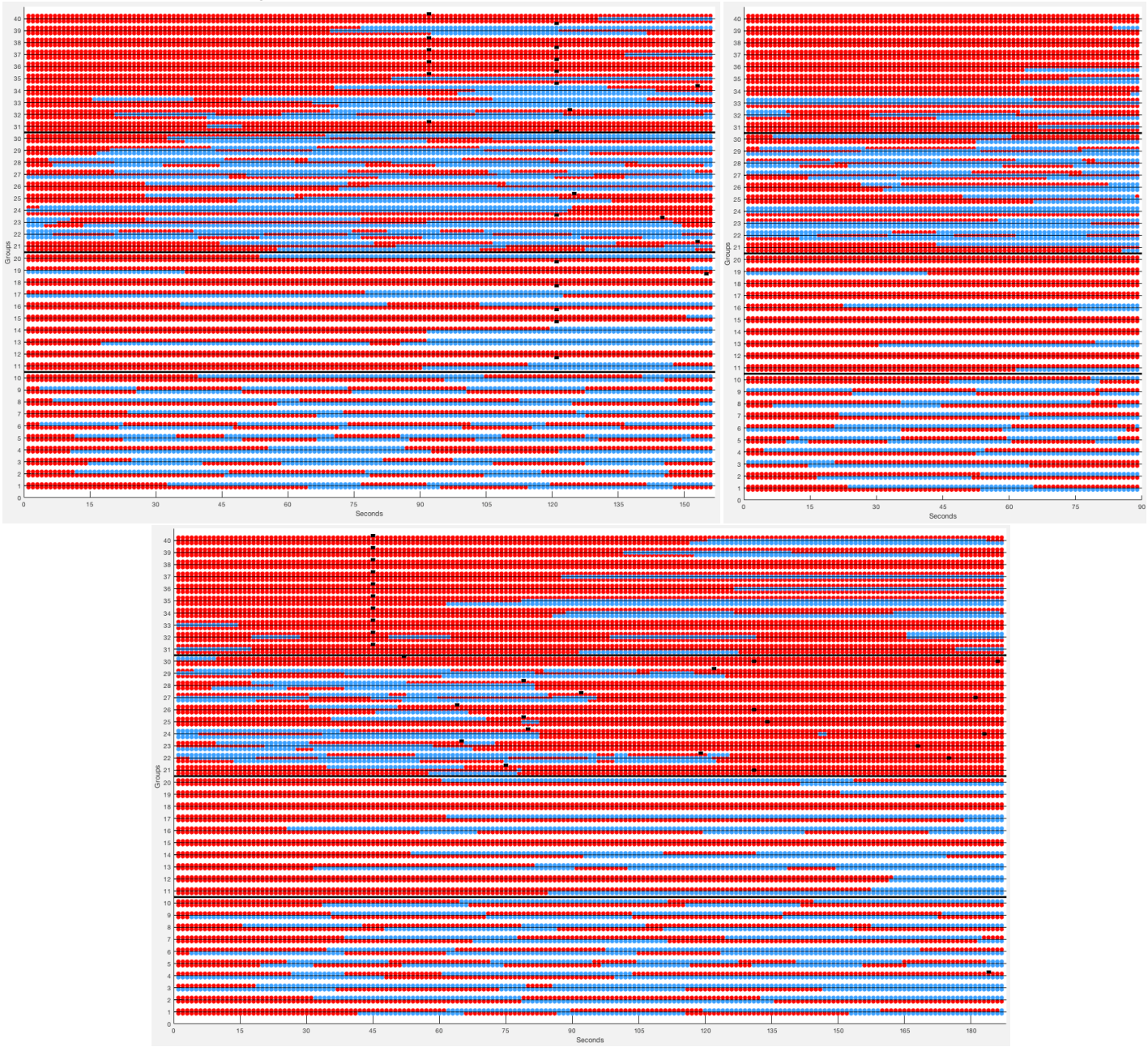
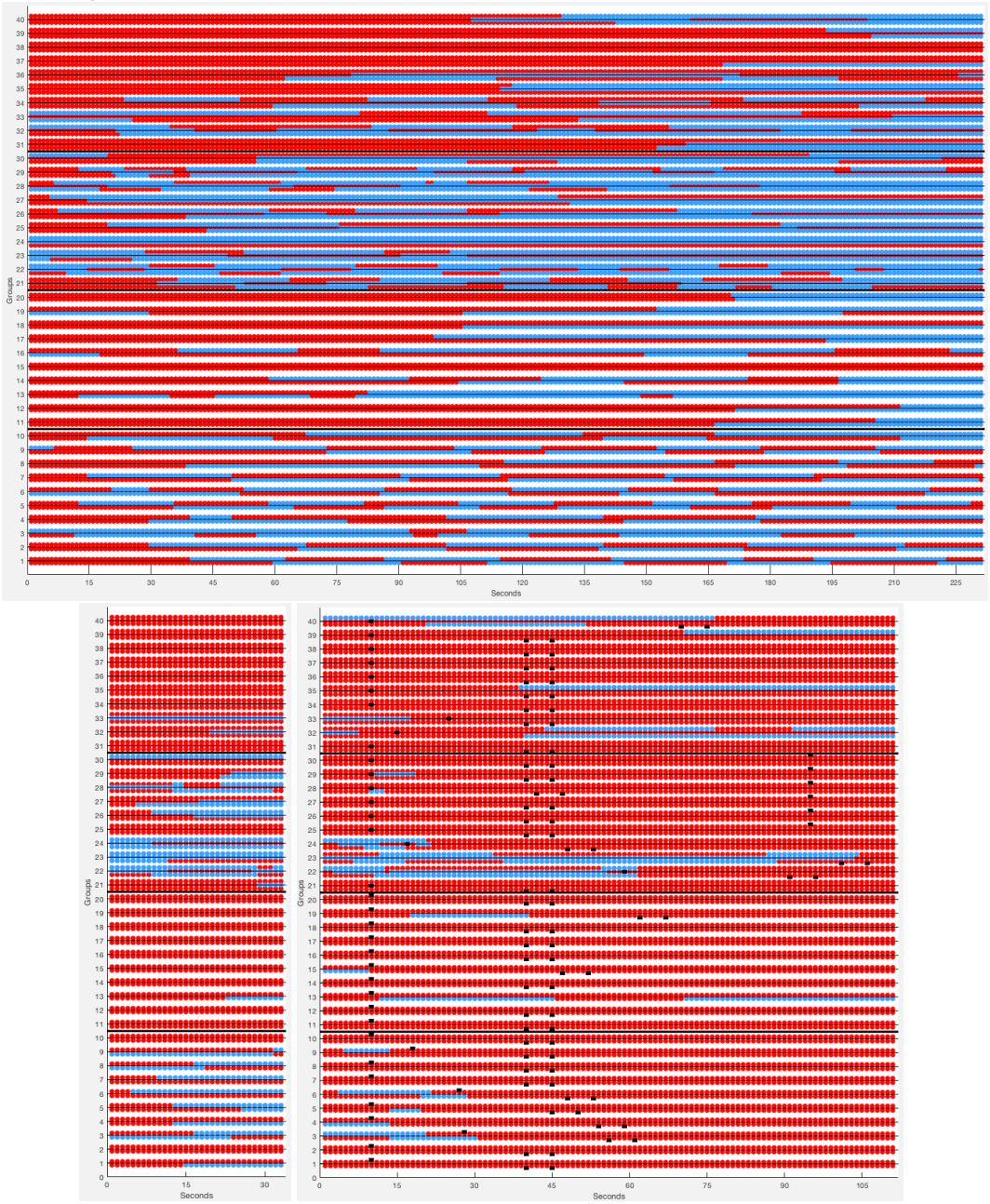


Figure 2: ACTION CHOICES BY PLAYERS OVER TIME: Games 4-6



$n = 2$ are 0.009, 0.035, 0.000, 0.000, 0.000, and 0.1218 for Games 1-6, respectively. In the setting with $n = 3$, the average experimentation intensity is also lower in the strategic treatment (p-values of 0.000, 0.000, 0.000, 0.000, 0.000, and 0.920 for Game 1, 2, 3, 4, 5, and 6, respectively). As Figure 2 highlights, Game 6 features an early success by Player 2 after 9 seconds of exploration, as well as successes by Player 1 after 39 and 44 seconds of exploration, respectively.

Table 1: AVERAGE EXPERIMENTATION INTENSITY

	$n = 2$				$n = 3$			
	Strategic Treatment		Control Treatment		Strategic Treatment		Control Treatment	
Game	Eff. Time	Av. Exp. Intensity	Eff. Time	Av. Exp. Intensity	Eff. Time	Av. Exp. Intensity	Eff. Time	Av. Exp. Intensity
1	155 [0]	.508 [.124]	146.5 [15.1]	.730 [.293]	147.7 [12.6]	.414 [.263]	134.6 [25.7]	.797 [.236]
2	185.9 [.2]	.512 [.157]	186 [0]	.696 [.283]	83.6 [21.9]	.543 [.269]	139 [67.6]	.833 [.218]
3	88 [0]	.565 [.126]	88 [0]	.878 [.235]	88 [0]	.457 [.305]	88 [0]	.866 [.248]
4	230 [0]	.519 [.134]	230 [0]	.678 [.239]	230 [0]	.383 [.245]	230 [0]	.728 [.243]
5	32 [0]	.653 [.349]	32 [0]	.984 [.072]	32 [0]	.596 [.381]	32 [0]	.953 [.196]
6	14.6 [7.6]	.810 [.314]	30 [256]	.941 [.167]	25 [30.4]	.800 [.336]	80.1 [34.6]	.857 [.250]

As we have mentioned above, information accumulation is potentially faster in the strategic treatment. Indeed, on account of the conditionally independent Poisson processes, the information acquired within a given unit of time is multiplied by the number of players currently playing risky. Therefore, conditionally on no success arriving, players' beliefs will tend to decrease more quickly in the strategic setting, implying that more time will be spent at more pessimistic beliefs. To ensure that Result 5.1 is not solely due to this effect, we conduct our parameter tests separately by belief region. Specifically, we consider the belief regions $[\bar{p}, \frac{1}{2}]$, where risky is a dominant action, and $(p_1^*, p^\ddagger)$, where risky and safe are mutually best responses in the strategic treatment.[8] In the control treatment, by contrast, all players should play risky in both regions. The following tables summarize our findings by belief region. As player 2 has a success after 9 seconds of using the risky arm, we omit Game 6 from these tables. We furthermore omit Game 5 from the tables for the "mutually BR" region, as this game lasts only 32 seconds, implying that the "mutually BR" region cannot be attained in the control treatment and only lasts for a few seconds in the strategic treatment, if it is attained at all. For Games 1-4, the missing observations for the "mutually BR" region correspond

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

Table 2: AVERAGE EXPERIMENTATION INTENSITY BY REGIONS FOR $n = 2$

	Strategic Treatment				Control Treatment			
Game	Obs.	Exp. Intensity	Min	Max	Obs.	Exp. Intensity	Min	Max
Panel A: R Dominant								
1	20	.648 [.315]	.094	1	20	.835 [.304]	.156	1
2	20	.723 [.291]	.211	1	20	.888 [.259]	.130	1
3	20	.617 [.281]	.125	1	20	.906 [.235]	.241	1
4	20	.732 [.323]	.117	1	20	.880 [.243]	.197	1
5	20	.653 [.349]	.065	1	20	.984 [.072]	.677	1
Panel B: Mutually BR								
1	20	.503 [.265]	.095	1	16	.758 [.343]	.114	1
2	20	.445 [.184]	0	.788	20	.783 [.362]	.118	1
3	20	.589 [.341]	0	1	16	.935 [.182]	.381	1
4	20	.484 [.254]	0	1	19	.765 [.343]	.092	1

to groups (in the strategic treatment) or individual players (in the control treatment) that have not reached the "mutually BR" region either on account of an early success or because they did not use the risky arm enough.

The comparison of the strategic treatment with the control treatment shows that the average experimentation intensity is substantially lower in the strategic treatment, for *both* belief regions. We first turn to the two-player setup and focus on the "R dominant" region, where the effect is statistically significant at least at the 10%-level in Games 1-5, the p-values of the two-sided Wilcoxon ranksum test amounting to 0.0367, 0.0371, 0.0010, 0.0646, and 0.0002, respectively.[9] Now, let us consider the "mutually BR" region. Here, the contrast between the strategic and the control treatment is even more pronounced and statistically significant at least at the 5%-level for all four games. The corresponding p-values are 0.0179, 0.0032, 0.0011, and 0.0071 for Games 1-4, respectively.

We now turn to the three-player setup, where we expect the same predictions to hold. When considering the "R dominant" region, we find the difference in average experimentation intensities between the two treatments to be highly statistically significant. The p-values are 0.0042, 0.0000,

[9]If we eliminate a single outlier in Game 4, namely player 1 in Group 13, our test yields a p-value of 0.0258, which would give us a 5% significance level for all games.

Table 3: AVERAGE EXPERIMENTATION INTENSITY BY REGIONS FOR $n = 3$

	Strategic Treatment				Control Treatment			
Game	Obs.	Exp. Intensity	Min	Max	Obs.	Exp. Intensity	Min	Max
Panel A: R Dominant								
1	30	.709 [.365]	0	1	30	.935 [.190]	.260	1
2	30	.649 [.340]	0	1	30	.976 [.076]	.689	1
3	30	.593 [.410]	0	1	30	.906 [.245]	0	1
4	30	.613 [.373]	0	1	30	.889 [.246]	.092	1
5	30	.596 [.381]	0	1	30	.953 [.196]	0	1
Panel B: Mutually BR								
1	30	.537 [.370]	0	1	29	.766 [.325]	.113	1
2	30	.549 [.369]	0	1	20	.673 [.340]	.023	1
3	30	.482 [.398]	0	1	25	.875 [.279]	.113	1
4	30	.471 [.353]	0	1	29	.815 [.275]	.299	1

0.0005, 0.0012, and 0.0000 for Games 1-5, respectively. The same is true for the "Mutually BR" region, with the exception of Game 2. This is most likely due to an early success by player 3 after only 44 seconds of exploration, which accounts for the sharp decrease in the number of observations for the control treatment, which we report in Table 3. The p-values are 0.0161, 0.2274, 0.0002, and 0.0002 for Games 1-4, respectively.

Since we are conditioning on the belief region, these results provide strong evidence that players are free-riding because of strategic considerations. Our analysis by belief region also shows that, while players tend to use the risky arm less in the "mutually BR" region than in the "R dominant" region, there does not appear to be any major qualitative difference between the two regions. This is true for both the strategic and control treatments. By contrast, theory would predict that, in the control treatment, players should play risky in both regions (i.e. we should observe average experimentation intensities of 1), while we should observe an average experimentation intensity of 1 only for the "R dominant" region in the strategic treatments. Our results would suggest that our experimental subjects did not distinguish between the two regions, possibly because they were not able to update their subjective beliefs with enough precision to tell them apart. We furthermore observe that, as far as free-riding is concerned, there do not seem to be any major differences between groups

Table 4: FREQUENCY OF CUT-OFF BEHAVIOR

	$n = 2$				$n = 3$			
	Strategic Treatment		Control Treatment		Strategic Treatment		Control Treatment	
Game	Eff. Time	Tot. (Rel.) Freq.	Eff. Time	Tot. (Rel.) Freq.	Eff. Time	Tot. (Rel.) Freq.	Eff. Time	Tot. (Rel.) Freq.
1	155 [0]	0 (0)	146.5 [15.1]	15 (.75)	147.7 [12.6]	3 (.1)	134.6 [25.7]	21 (.7)
2	185.9 [.2]	0 (0)	186 [0]	15 (.75)	83.6 [21.9]	3 (.1)	139 [67.6]	22 (.73)
3	88 [0]	5 (.25)	88 [0]	17 (.85)	88 [0]	11 (.37)	88 [0]	26 (.87)
4	230 [0]	0 (0)	230 [0]	14 (.7)	230 [0]	6 (.2)	230 [0]	19 (.63)
5	32 [0]	17 (.85)	32 [0]	20 (1)	32 [0]	17 (.57)	32 [0]	29 (.97)
6	14.6 [7.6]	13 (.65)	30 [256]	17 (.85)	25 [30.4]	19 (.63)	80.1 [34.6]	25 (.83)

of size two and groups of size three.

5.3 Cut-Off Behavior

As we have pointed out *supra*, optimality in the individual decision-making problem in our control treatment implies cut-off behavior, while KRC have shown that there does not exist a Markov Perfect Equilibrium in cut-off strategies in the strategic treatment. This prediction is confirmed by our experiment, where subjects often play cut-off strategies in the control treatment, while they hardly ever do so in the strategic setup.

Result 5.2 *The frequency of cut-off behavior is higher in the control treatment than in the strategic treatment. We find evidence for both* $n = 2$ *and* $n = 3$.

Indeed, Table 4 shows that the frequency of cut-off behavior is much higher in the control treatment than in the strategic treatment for both groups of size $n = 2$ and groups of size $n = 3$. While it increases sharply in Games 5 and 6 as compared to Games 1-4 in the strategic treatments, it is still higher in the corresponding control treatments (for either given group size). In Game 5, this sharp increase is most likely due to the short duration of that game. In Game 6, it is most likely driven by the resolution of uncertainty very early in the game, with Player 2 achieving a success after exploring for 9 seconds.

5.4 Switches of Actions

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

Result 5.3 *The incidence of switches is significantly higher in the strategic treatment than in the control treatment. This holds for both* $n = 2$ *and* $n = 3$*.*

Table 5 displays the average number of switches per player across games for our four treatments.[10] The incidence of switches in the strategic treatment is much higher than in the control treatment in all games except for Game 6 (all *p*-values=0.000 for Games 1-5 for either group size, with the exception of Game 4 in the setting with $n = 3$, where *p*-value=0.001). As noted above, the early success in Game 6 reveals the risky arm to be good and thus resolves all uncertainty at the very beginning of the game. While still marginally higher incidences of switches are observed in the strategic treatment for both $n = 2$ and $n = 3$, this is not statistically significant (*p*-values of 0.287 and 0.446, respectively).

To study the players' information acquisition processes further, we employ eye-tracking data obtained by using two (three) Tobii-TX300 eye trackers with a sampling rate of 300 Hz. Eye fixations are an unobtrusive measure and can provide information about the importance assigned by subjects to the different payoff streams (revealing both actions and payoffs).[11] The relative frequency of fixations corresponds to the relative importance of an information in the subject's decision-making process. We define a subject's fixation intensity as the total number of his 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.

Figure 3 displays (non-representative) heatmaps to illustrate the different information acquisition behavior in our four treatments. In the top-left corner, the strategic treatment with $n = 2$ is illustrated, with the corresponding control treatment represented just below. In the top-right corner, the strategic treatment with $n = 3$ is displayed, while the control treatment with $n = 3$ is shown at the bottom-right. As Figure 3 illustrates, players not only switch actions more frequently

[10] As there is rather little variation between the strategic and the control treatments for a given game, we have decided to report the average *number*, rather than the average *incidence* of switches in Table 5, as the former may be easier to interpret.

[11] Video recordings illustrating the use of the eye-tracking devices are available at www.jch.com.

Table 5: AVERAGE NUMBER OF SWITCHES PER PLAYER

	$n=2$				$n=3$			
	Strategic Treatment		Control Treatment		Strategic Treatment		Control Treatment	
Game	Eff. Time	Switches Per Pl.	Eff. Time	Switches Per Pl.	Eff. Time	Switches Per Pl.	Eff. Time	Switches Per Pl.
1	155 [0]	4.45 [1.85]	146.5 [15.1]	.90 [.912]	147.7 [12.6]	3.4 [1.87]	134.6 [25.7]	1.13 [1.68]
2	185.9 [.2]	4.50 [1.91]	186 [0]	1.35 [1.5]	83.6 [21.9]	2.77 [1.85]	139 [67.6]	.97 [1.50]
3	88 [0]	2.20 [1.11]	88 [0]	.30 [.47]	88 [0]	1.73 [1.46]	88 [0]	.47 [.82]
4	230 [0]	6.05 [2.04]	230 [0]	1.85 [1.9]	230 [0]	4.00 [3.02]	230 [0]	1.7 [1.97]
5	32 [0]	.60 [.50]	32 [0]	.05 [.22]	32 [0]	.70 [.84]	32 [0]	.03 [.18]
6	14.6 [7.6]	.60 [.88]	30 [256]	.30 [.80]	25 [30.4]	.97 [1.47]	80.1 [34.6]	.37 [.72]

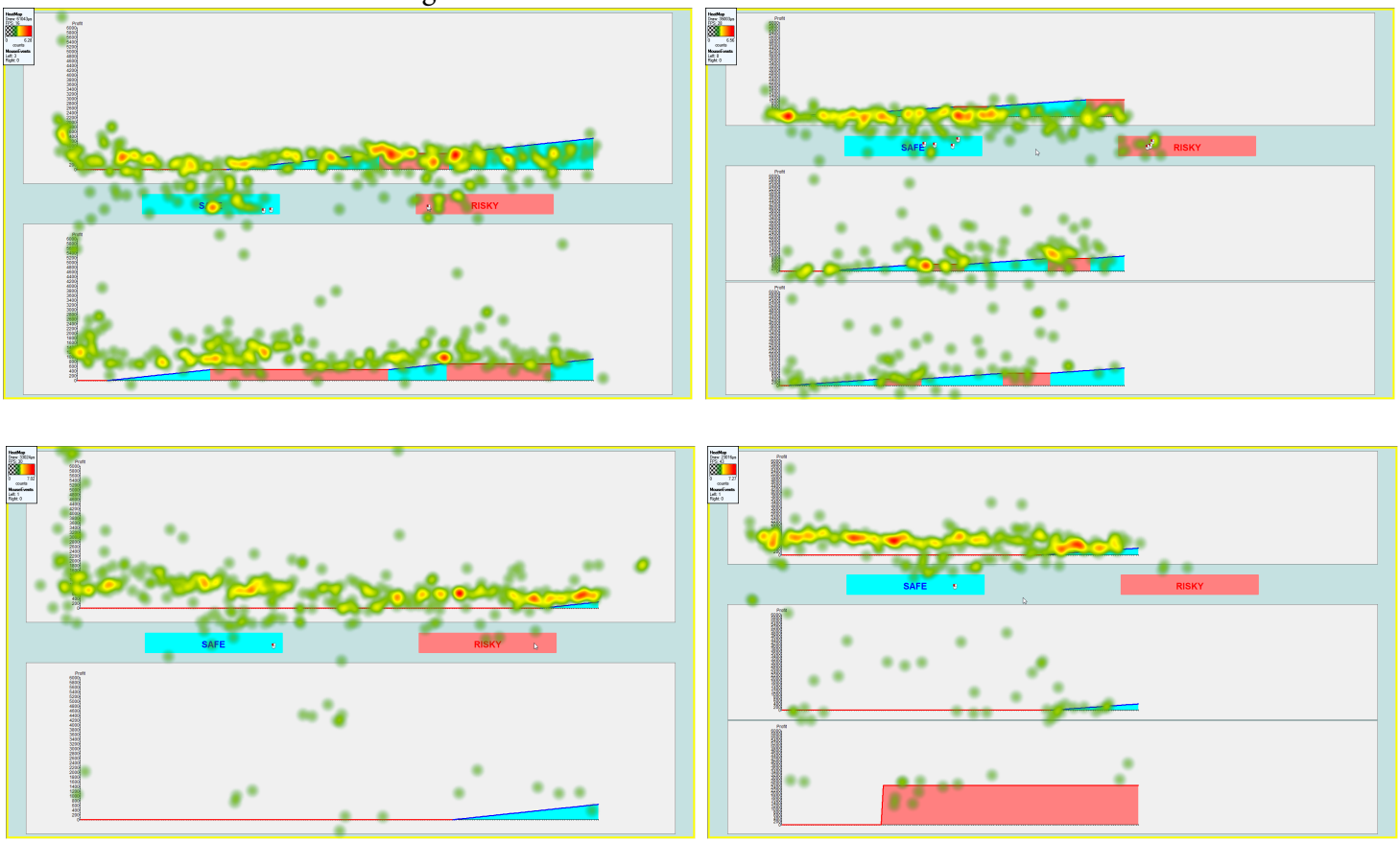
in the strategic treatment but also focus much more intensively on their partners' actions and payoffs. This is in sharp contrast to the corresponding control treatment, where players seem to focus almost exclusively on their own stream of payoffs. Indeed, a rational player should completely ignore a partner's actions and payoffs in the control treatments, as they are informative for his own problem only in the strategic setting. This observation is also consistent with the results presented in Tables 4 and 5. Indeed, in the control treatments, players' behavior requires less coordination, as significantly more cut-off behavior and less switching is observed.

As Table 6 shows, the average fixation intensity is significantly lower in the strategic treatment. This is statistically significant for all six games independently of the group size (all p-values=0.000 for $n = 2$ and $n = 3$). The complex coordination required by the switching of roles between pioneer and free-rider, which is characteristic of the strategic treatment, seems to force players to pay a lot of attention to their partner's (partners') behavior. This provides additional evidence that players behave strategically and try to learn from their partners' exploration efforts in the strategic treatments only.

5.5 Pioneers

In the control treatment, players are predicted to play risky on $(p_1^*, \frac{1}{2}]$; i.e., conditionally on no success arriving, players should switch from risky to safe only once, and do so at the same time, at which their beliefs reach p_1^*. By contrast, as KRC have shown, there is a range of beliefs containing $(p_1^*, p^\ddagger)$ such that safe and risky are mutually best responses in any Markov Perfect Equilibrium. In particular, there exists a range of beliefs in which just one pioneer should play risky while the

Figure 3: HEATMAPS OF FOUR TREATMENTS



other player(s) free-ride(s). The following result thus provides further evidence for the prevalence of free-riding in our strategic treatment.

Result 5.4 *The proportion of time before a first breakthrough during which exactly one player plays risky is higher in the strategic treatment than in the control treatment.*

Table 7 shows the average proportion of time during which *exactly one* player is exploring before a first breakthrough by any player in his group. In each game, it is more than twice as large in the strategic treatment.

Table 6: AVERAGE FIXATION INTENSITY

	$n=2$				$n=3$			
	Strategic Treatment		Control Treatment		Strategic Treatment		Control Treatment	
Game	Eff. Time	Fixation Intensity	Eff. Time	Fixation Intensity	Eff. Time	Fixation Intensity	Eff. Time	Fixation Intensity
1	155 [0]	.619 [.078]	146.5 [15.1]	.870 [.089]	147.7 [12.6]	.384 [.118]	134.6 [25.7]	.710 [.158]
2	185.9 [.2]	.620 [.121]	186 [0]	.882 [.131]	83.6 [21.9]	.365 [.113]	139 [67.6]	.709 [.176]
3	88 [0]	.600 [.086]	88 [0]	.874 [.111]	88 [0]	.392 [.164]	88 [0]	.762 [.112]
4	230 [0]	.615 [.095]	230 [0]	.875 [174]	230 [0]	.389 [.124]	230 [0]	.700 [.139]
5	32 [0]	.633 [.139]	32 [0]	.876 [.149]	32 [0]	.383 [.151]	32 [0]	.745 [.199]
6	14.6 [7.6]	.594 [.169]	30 [256]	.814 [.112]	25 [30.4]	.382 [.157]	80.1 [34.6]	.646 [.159]

Table 7: PROPORTION OF TIME WITH A SINGLE PIONEER

	$n=2$				$n=3$			
	Strategic Treatment		Control Treatment		Strategic Treatment		Control Treatment	
Game	Eff. Time	Single Pioneer	Eff. Time	Single Pioneer	Eff. Time	Single Pioneer	Eff. Time	Single Pioneer
1	155 [0]	.724 [.156]	146.5 [15.1]	.284 [.258]	147.7 [12.6]	.670 [.178]	134.6 [25.7]	.097 [.156]
2	185.9 [.2]	.708 [.176]	186 [0]	.315 [.254]	83.6 [21.9]	.425 [.352]	139 [67.6]	0 [0]
3	88 [0]	.745 [.156]	88 [0]	.187 [.253]	88 [0]	.563 [.348]	88 [0]	.136 [.256]
4	230 [0]	.757 [.175]	230 [0]	.294 [.214]	230 [0]	.741 [.171]	230 [0]	.249 [.198]
5	32 [0]	.581 [.360]	32 [0]	.029 [.092]	32 [0]	.361 [.304]	32 [0]	0 [0]
6	14.6 [7.6]	.288 [.399]	30 [256]	.078 [.246]	25 [30.4]	.219 [.369]	80.1 [34.6]	0 [0]

6 Conclusion

We have tested a game of strategic experimentation with bandits in the laboratory. We are able to provide evidence for the main qualitative behavioral predictions of KRC's equilibrium analysis. Indeed, we have exhibited strong evidence for strategic free-riding. Subjects seem to attempt to coordinate in rather complex ways, as evidenced, inter alia, by the much lower incidence of cutoff behavior and the higher incidence of switches.

We have confined our analysis to the exponential-bandit setting of KRC. While the tractability

of the exponential-bandit setting will certainly have facilitated its experimental implementation, the model does have some special features. For instance, as successes are fully revealing, there is no encouragement effect in KRC, which our experimental investigation confirms. Indeed, we can compute the average experimentation intensities in the region where safe is a dominant action, $[0, p_1^*]$, for Game 4 as well as for the two-player groups in Game 2.[12] Even in this region, the average experimentation intensity is lower in the strategic treatment: .511 [.138] in the strategic treatment for Game 4 with $n = 2$ vs. .660 [.249] in the control treatment; .325 [.310] vs. .736 [.302] in Game 4 for $n = 3$, and .510 [.270] vs. .743 [.262] in Game 2, where we report the standard deviation in square brackets. By contrast, if there were an encouragement effect, we should expect higher experimentation intensities in the strategic treatment for this belief region.

It might be interesting to test whether the encouragement effect can be shown in the laboratory for settings in which the theory would predict it to arise. This would be the case for instance in the Poisson setting with inconclusive breakthroughs à la Keller & Rady (2010), or in the Brownian-motion setting of Bolton & 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.

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

References

Anderson, C. (2012): "Ambiguity Aversion in Multi-Armed Bandit Problems," *Theory and Decision*, 72, 15–33.

Anderson, C. (2001): "Behavioral Models of Strategies in Multi-Armed Bandit Problems," PhD thesis, California Institute of Technology.

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

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

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

Bergemann, D. and J. Välimäki (2000): "Experimentation in Markets," *Review of Economic Studies*, 67, 213–234.

Bergemann, D. and J. Välimäki (1996): "Learning And Strategic Pricing," *Econometrica*, 64, 1125–1149.

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

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

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

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

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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Appendix

A Instructions

The order of the instructions is as follows:

1. $n = 2$: Strategic Treatment
2. $n = 2$: Control Treatment
3. $n = 3$: Strategic Treatment
4. $n = 3$: Control Treatment

Experiment Instructions

Ground Rules

Welcome to the experiment. Please read the instructions carefully. The earnings you make in this experiment will be paid to you, in cash, at the end of the session.

Your earnings will be determined by your choices and the choices of other participants.

Communication between participants is not allowed. Please use only the computer to input your decisions. Please do not start or end any programs, and do not change any settings.

How Groups are Organized

This experiment consists of six games in total. In the beginning of the first game, participants are randomly matched to pairs and the pairs stay the same in all six games. Therefore, in each game you will interact with the same participant.

How the Timing Works

Games will last on average ***120 seconds*** but may end at any time. The probability that the game ends is the same at each instant. Equivalently, the probability that the game ends during a given period of time depends only on the length of that period of time, and not on how long the game has already been going on. (Such processes are known as *exponential processes* in statistics.)

How the Game Works

In every game, you have to decide whether you want to play the "**safe**" or the "**risky**" option. You can switch between the two options at any time and as often as you like by clicking on the safe (Blue) or risky (Red) button on the screen.

Whenever you choose the **safe** option, your payoff will increase for sure at the rate ***E$ 10***. That means the **safe** option will give you a reward of ***E$ 10*** every second during which you use it.

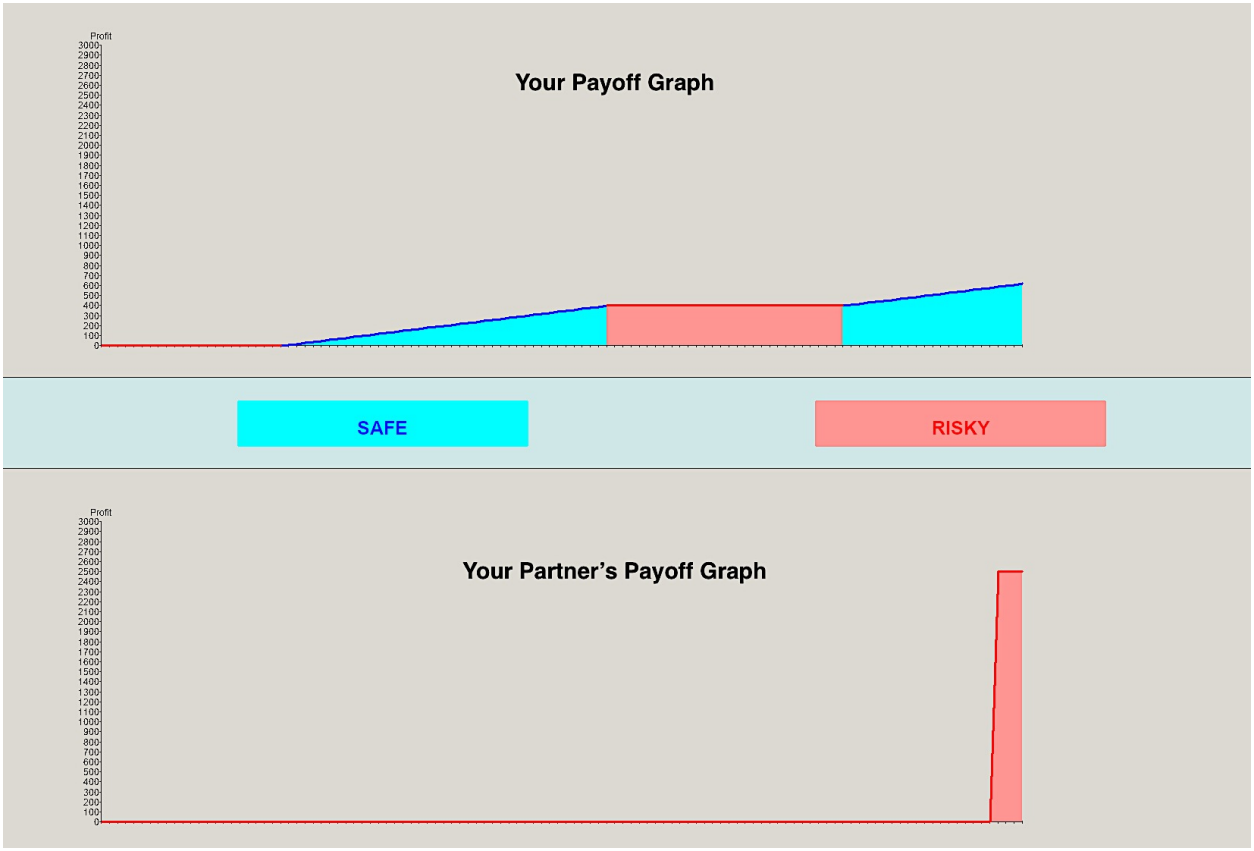
When you choose the **risky** option, however, what you will be getting depends on the quality of that risky option. The quality of the **risky** option is determined by the computer once and for all at the start of each game; it never changes during the course of the game. We have programmed the computer so that the risky option will be **good** or **bad** with equal probability in each of the six games. The quality of the risky option in later games is independent of its quality in previous games. That is, in each of your six games, with probability ½ your (and your partner's!) risky option will be **good**; with probability ½ they will be **bad**. Note that your risky option and that of your partner's will always be of the **same** quality.

If your risky option is **good**, it may give you a reward of ***E$ 2500***, but it will only ever do so if you use it. The probability that you get this reward from a good risky option during a given period of time during which you use it depends only on the length of that period of time; it does not depend on anything else, e.g. on how long the game has already been going on. Note that a good risky option may give you more than one reward of *E$ 2500* per game.

If your risky option is **bad**, it will never give you any reward.

You can switch back and forth between the risky option and the safe option at will and as many times as you like. All that matters for your chance of getting the reward is (1) the quality of the risky arm as determined by the computer before the game starts and (2) the overall amount of time you choose to spend on it.

The following graphic illustrates what you are going to see on your screen during the game. The graphs will be updated every second.



- The **upper** diagram always shows **your** actions and payoffs.
- In this example, you have started playing the risky option (highlighted in Red), then you have switched to the safe option (highlighted in Blue), then you have switched back again to the risky option, etc.
- The **lower** diagram always shows **your partner's** actions and payoffs.
- In this example, your partner has started playing the risky option and continues to do so.
- Note that, in this example, your partner's risky option was good and gave him once a reward of ***E$ 2500***. This means that your risky option was good too.

The parameters are chosen in such a way that, *if you knew* the risky option to be good, you would be best off by **always** choosing it. Yet, *if you knew* the risky option to be bad, you would be best off by **always** choosing the safe option. In short:

Good risky option > Safe option > **Bad** risky option

Your partner is solving the exact same problem as you and has read the exact same instructions. Note that by observing the behaviour of his risky option (provided he uses it) you can learn something about your risky option as well.

Payment

In the experiment you will be making decisions that will earn you E$ (Experimental Dollars). At the end of the experiment, the E$ you earned will be converted into Australian Dollars at an exchange rate of E$ 100 = AU$ 1, and paid out in cash. This amount will be added to your show-up fee of AU$ 5.

After completing the experiment, the computer will randomly select one out of the six games (this will be the same game for all participants), and this game will then be used to determine your payoffs.

Experiment Instructions

Ground Rules

Welcome to the experiment. Please read the instructions carefully. The earnings you make in this experiment will be paid to you, in cash, at the end of the session.

Your earnings will be determined by your choices and the choices of other participants.

Communication between participants is not allowed. Please use only the computer to input your decisions. Please do not start or end any programs, and do not change any settings.

How Groups are Organized

This experiment consists of six games in total. In the beginning of the first game, participants are randomly matched to pairs and the pairs stay the same in all six games. Therefore, in each game you will interact with the same participant.

How the Timing Works

Games will last on average ***120 seconds*** but may end at any time. The probability that the game ends is the same at each instant. Equivalently, the probability that the game ends during a given period of time depends only on the length of that period of time, and not on how long the game has already been going on. (Such processes are known as *exponential processes* in statistics.)

How the Game Works

In every game, you have to decide whether you want to play the "**safe**" or the "**risky**" option. You can switch between the two options at any time and as often as you like by clicking on the safe (Blue) or risky (Red) button on the screen.

Whenever you choose the **safe** option, your payoff will increase for sure at the rate ***E$ 10***. That means the **safe** option will give you a reward of ***E$ 10*** every second during which you use it.

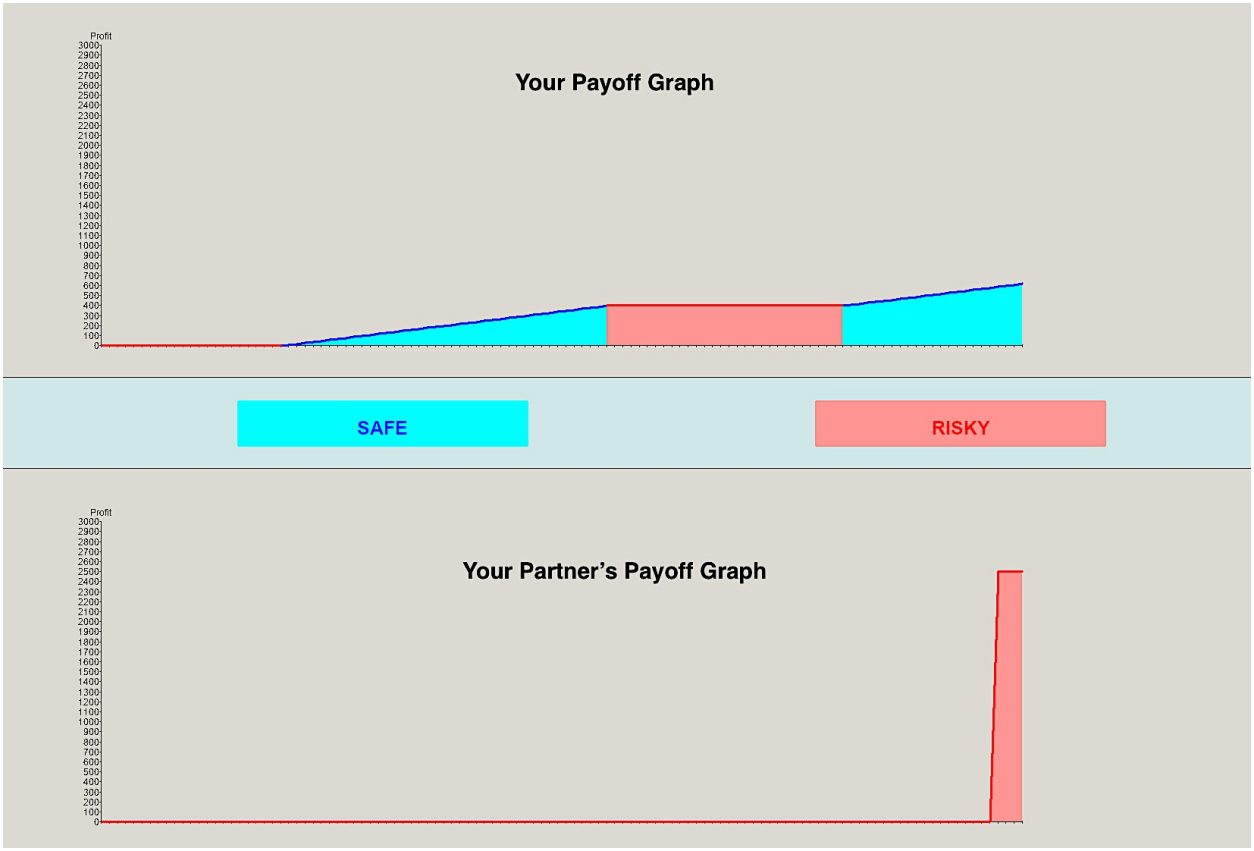
When you choose the **risky** option, however, what you will be getting depends on the quality of that risky option. The quality of the **risky** option is determined by the computer once and for all at the start of each game; it never changes during the course of the game. We have programmed the computer so that the risky option will be **good** or **bad** with equal probability in each of the six games. The quality of the risky option in later games is independent of its quality in previous games. That is, in each of your six games, with probability ½ your risky option will be **good**; with probability ½ it will be **bad**. The same is true for your partner. Note that your risky option and that of your partner's might or might not be of the same quality.

If your risky option is **good**, it may give you a reward of ***E$ 2500***, but it will only ever do so if you use it. The probability that you get this reward from a good risky option during a given period of time during which you use it depends only on the length of that period of time; it does not depend on anything else, e.g. on how long the game has already been going on. Note that a good risky option may give you more than one reward of *E$ 2500* per game.

If your risky option is **bad**, it will never give you any reward.

You can switch back and forth between the risky option and the safe option at will and as many times as you like. All that matters for your chance of getting the reward is (1) the quality of the risky arm as determined by the computer before the game starts and (2) the overall amount of time you choose to spend on it.

The following graphic illustrates what you are going to see on your screen during the game. The graphs will be updated every second.



- The **upper** diagram always shows **your** actions and payoffs.
- In this example, you have started playing the risky option (highlighted in Red), then you have switched to the safe option (highlighted in Blue), then you have switched back again to the risky option, etc.
- The **lower** diagram always shows **your partner's** actions and payoffs.
- In this example, your partner has started playing the risky option and continues to do so.
- Note that, in this example, your partner's risky option was good and gave him once a reward of ***E$ 2500***.

The parameters are chosen in such a way that, *if you knew* the risky option to be good, you would be best off by **always** choosing it. Yet, *if you knew* the risky option to be bad, you would be best off by **always** choosing the safe option. In short:

Good risky option > Safe option > **Bad** risky option

Your partner is solving the exact same problem as you and has read the exact same instructions.

Payment

In the experiment you will be making decisions that will earn you E$ (Experimental Dollars). At the end of the experiment, the E$ you earned will be converted into Australian Dollars at an exchange rate of E$ 100 = AU$ 1, and paid out in cash. This amount will be added to your show-up fee of AU$ 5.

After completing the experiment, the computer will randomly select one out of the six games (this will be the same game for all participants), and this game will then be used to determine your payoffs.

Experiment Instructions

Ground Rules

Welcome to the experiment. Please read the instructions carefully. The earnings you make in this experiment will be paid to you, in cash, at the end of the session.

Your earnings will be determined by your choices and the choices of other participants.

Communication between participants is not allowed. Please use only the computer to input your decisions. Please do not start or end any programs, and do not change any settings.

How Groups are Organized

This experiment consists of six games in total. In the beginning of the first game, participants are randomly matched to groups of three players and the groups stay the same in all six games. Therefore, in each game you will interact with the same participants.

How the Timing Works

Games will last on average ***120 seconds*** but may end at any time. The probability that the game ends is the same at each instant. Equivalently, the probability that the game ends during a given period of time depends only on the length of that period of time, and not on how long the game has already been going on. (Such processes are known as *exponential processes* in statistics.)

How the Game Works

In every game, you have to decide whether you want to play the "**safe**" or the "**risky**" option. You can switch between the two options at any time and as often as you like by clicking on the safe (Blue) or risky (Red) button on the screen.

Whenever you choose the **safe** option, your payoff will increase for sure at the rate ***E$ 10***. That means the **safe** option will give you a reward of ***E$ 10*** every second during which you use it.

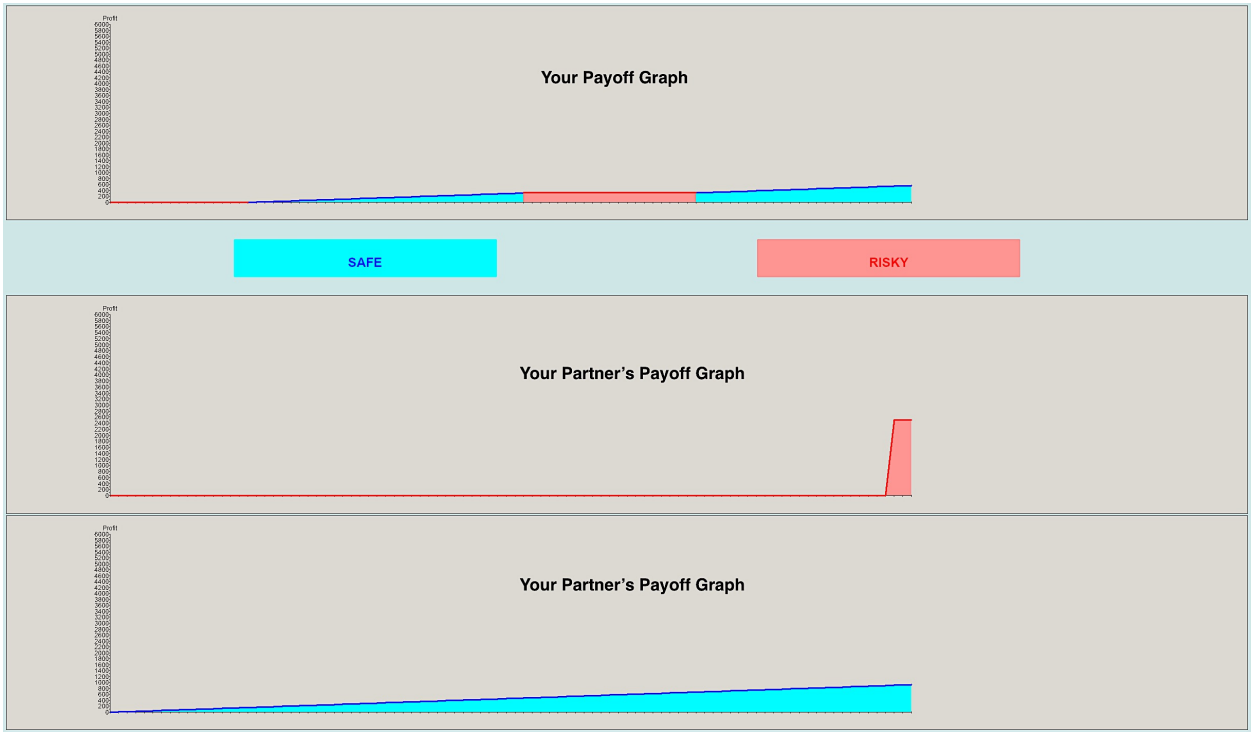
When you choose the **risky** option, however, what you will be getting depends on the quality of that risky option. The quality of the **risky** option is determined by the computer once and for all at the start of each game; it never changes during the course of the game. We have programmed the computer so that the risky option will be **good** or **bad** with equal probability in each of the six games. The quality of the risky option in later games is independent of its quality in previous games. That is, in each of your six games, with probability ½ your (and your partners'!) risky option will be **good**; with probability ½ they will be **bad**. Note that your and your partners' risky option will always be of the **same** quality.

If your risky option is **good**, it may give you a reward of ***E$ 2500***, but it will only ever do so if you use it. The probability that you get this reward from a good risky option during a given period of time during which you use it depends only on the length of that period of time. It does not depend on anything else, e.g. on how long the game has already been going on. Note that a good risky option may give you more than one reward of *E$ 2500* per game.

If your risky option is **bad**, it will never give you any reward.

You can switch back and forth between the risky option and the safe option at will and as many times as you like. All that matters for your chance of getting the reward is (1) the quality of the risky arm as determined by the computer before the game starts and (2) the overall amount of time you choose to spend on it.

The following graphic illustrates what you are going to see on your screen during the game. The graphs will be updated every second.



- The **upper** diagram always shows **your** actions and payoffs.
- In this example, you have started playing the risky option (highlighted in Red), then you have switched to the safe option (highlighted in Blue), then you have switched back again to the risky option, etc.
- The **lower** diagram always shows **your partners'** actions and payoffs.
- In this example, one of your partners has started playing the risky option and continues to do so. The other partner has started and continues playing the safe option.
- Note that, in this example, your partner's risky option was good and gave him once a reward of ***E$ 2500***. This means that your risky option was good too.

The parameters are chosen in such a way that, *if you knew* the risky option to be good, you would be best off by **always** choosing it. Yet, *if you knew* the risky option to be bad, you would be best off by **always** choosing the safe option. In short:

Good risky option > Safe option > **Bad** risky option

Your partners are solving the exact same problem as you and have read the exact same instructions. Note that by observing the behaviour of their risky option (provided they use it) you can learn something about your risky option as well.

Payment

In the experiment you will be making decisions that will earn you E$ (Experimental Dollars). At the end of the experiment, the E$ you earned will be converted into Australian Dollars at an exchange rate of E$ 100 = AU$ 1, and paid out in cash. This amount will be added to your show-up fee of AU$ 5.

After completing the experiment, the computer will randomly select one out of the six games (this will be the same game for all participants), and this game will then be used to determine your payoffs.

Experiment Instructions

Ground Rules

Welcome to the experiment. Please read the instructions carefully. The earnings you make in this experiment will be paid to you, in cash, at the end of the session.

Your earnings will be determined by your choices and the choices of other participants.

Communication between participants is not allowed. Please use only the computer to input your decisions. Please do not start or end any programs, and do not change any settings.

How Groups are Organized

This experiment consists of six games in total. In the beginning of the first game, participants are randomly matched to groups of three players and the groups stay the same in all six games. Therefore, in each game you will interact with the same participants.

How the Timing Works

Games will last on average ***120 seconds*** but may end at any time. The probability that the game ends is the same at each instant. Equivalently, the probability that the game ends during a given period of time depends only on the length of that period of time, and not on how long the game has already been going on. (Such processes are known as *exponential processes* in statistics.)

How the Game Works

In every game, you have to decide whether you want to play the "**safe**" or the "**risky**" option. You can switch between the two options at any time and as often as you like by clicking on the safe (Blue) or risky (Red) button on the screen.

Whenever you choose the **safe** option, your payoff will increase for sure at the rate ***E$ 10***. That means the **safe** option will give you a reward of ***E$ 10*** every second during which you use it.

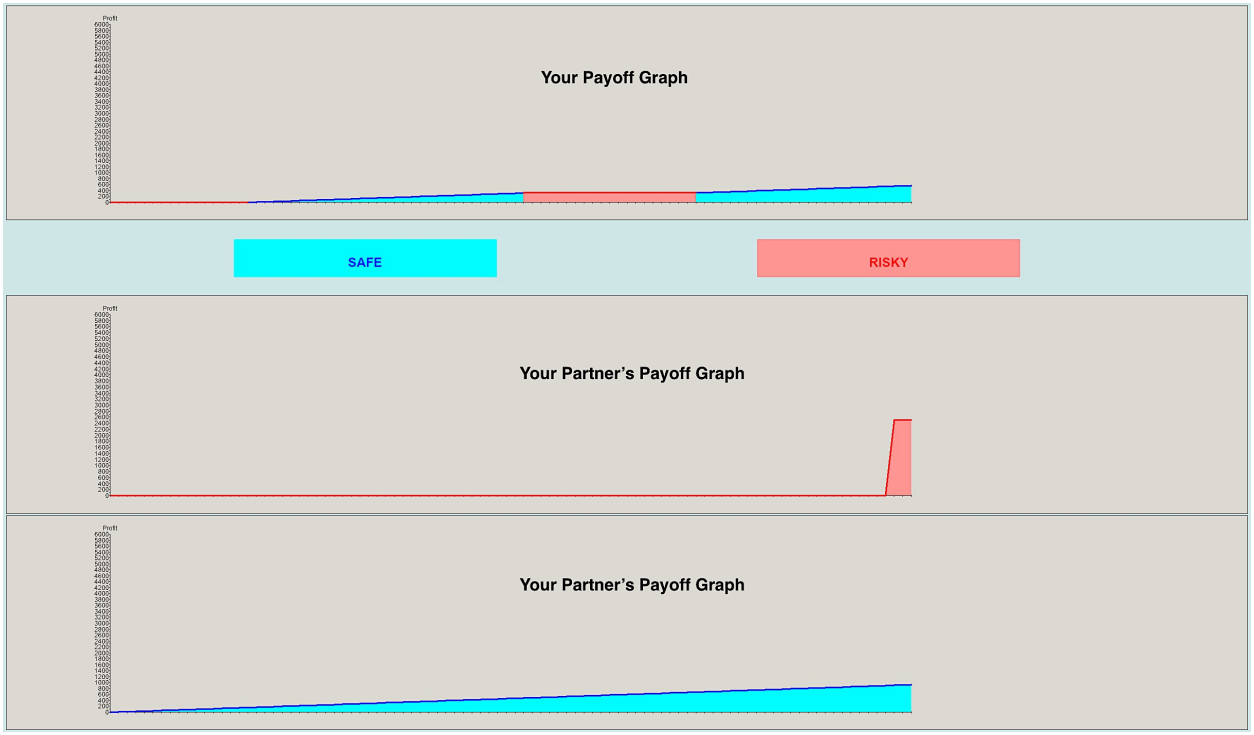
When you choose the **risky** option, however, what you will be getting depends on the quality of that risky option. The quality of the **risky** option is determined by the computer once and for all at the start of each game; it never changes during the course of the game. We have programmed the computer so that the risky option will be **good** or **bad** with equal probability in each of the six games. The quality of the risky option in later games is independent of its quality in previous games. That is, in each of your six games, with probability ½ your risky option will be **good**; with probability ½ it will be **bad**. The same is true for your partners. Note that your risky option and that of your partners' might or might not be of the same quality.

If your risky option is **good**, it may give you a reward of ***E$ 2500***, but it will only ever do so if you use it. The probability that you get this reward from a good risky option during a given period of time during which you use it depends only on the length of that period of time. It does not depend on anything else, e.g. on how long the game has already been going on. Note that a good risky option may give you more than one reward of *E$ 2500* per game.

If your risky option is **bad**, it will never give you any reward.

You can switch back and forth between the risky option and the safe option at will and as many times as you like. All that matters for your chance of getting the reward is (1) the quality of the risky arm as determined by the computer before the game starts and (2) the overall amount of time you choose to spend on it.

The following graphic illustrates what you are going to see on your screen during the game. The graphs will be updated every second.



- The **upper** diagram always shows **your** actions and payoffs.
- In this example, you have started playing the risky option (highlighted in Red), then you have switched to the safe option (highlighted in Blue), then you have switched back again to the risky option, etc.
- The **lower** diagram always shows **your partner's** actions and payoffs.
- In this example, one of your partners has started playing the risky option and continues to do so. The other partner has started and continues playing the safe option.
- Note that, in this example, at least one of your partner's risky option was good and gave him once a reward of ***E$ 2500***.

The parameters are chosen in such a way that, *if you knew* the risky option to be good, you would be best off by **always** choosing it. Yet, *if you knew* the risky option to be bad, you would be best off by **always** choosing the safe option. In short:

Good risky option > Safe option > **Bad** risky option

Your partners are solving the exact same problem as you and have read the exact same instructions.

Payment

In the experiment you will be making decisions that will earn you E$ (Experimental Dollars). At the end of the experiment, the E$ you earned will be converted into Australian Dollars at an exchange rate of E$ 100 = AU$ 1, and paid out in cash. This amount will be added to your show-up fee of AU$ 5.

After completing the experiment, the computer will randomly select one out of the six games (this will be the same game for all participants), and this game will then be used to determine your payoffs.