Staff Working Paper/Document de travail du personnel — 2020-30

Last updated: July 6, 2020

Ten Isn't Large! Group Size and Coordination in a Large-Scale Experiment

by Jasmina Arifovic,[1] Cars Hommes,[2] Anita Kopányi-Peuker[3] and Isabelle Salle[4]

[1] Department of Economics, Simon Fraser University
Burnaby, BC

[2] Financial Markets Department
Bank of Canada, Ottawa, Ontario, Canada K1A 0G9
University of Amsterdam and Tinbergen Institute

chommes@bankofcanada.ca

[3] CPB Netherlands Bureau for Economic Policy Analysis and University of Amsterdam

[4] Financial Markets Department
Bank of Canada, Ottawa, Ontario, Canada K1A 0G9
University of Amsterdam

isalle@bankofcanada.ca

ISSN 1701-9397

Acknowledgements

This paper has benefited from fruitful discussions and helpful suggestions from Jason Allen, Gabriele Camera and Camille Cornand. We also thank the experimental team at the University of Valencia for their valuable support and Alba Ruiz for research assistance. We are also grateful for the discussions with the IBSEN-members and participants of the 2017 Experimental Finance conference in Nice, June 14–16; the 2018 CEF conference in Milan, June 17–19; the 2018 ESA World Meeting in Berlin, June 30–July 1; the 2019 workshop on Theoretical and Experimental Macroeconomics, June 25–26 at the Bank of Canada. Financial support from the H2020 IBSEN project ("Bridging the Gap: From Individual Behavior to the Socio-Technical Man", grant agreement 662725) is gratefully acknowledged. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Bank of Canada.

Abstract

This paper provides experimental evidence on coordination within genuinely large groups that could proxy the atomistic nature of real-world markets and organizations. We use a bank-run game where the two pure-strategy equilibria “run” and “wait” can be ranked by payoff and risk-dominance and a random sequence of public announcements introduces stochastic sunspot equilibria. We find systematic group size differences that theory fails to predict. In the presence of strategic uncertainty, the behavior of small groups is uninformative of behavior in large groups: in contrast to groups of 10, large groups only coordinate on the safest but Pareto-inferior “run” strategy and never coordinate on sunspots. Our results entail a series of theoretical and experimental implications.

Bank topic: Financial stability; Financial markets
JEL codes: C92, D83, D90, G20

1 Introduction

Undesirable economic outcomes, such as currency attacks, bank runs or liquidity runs in financial markets, may occur irrespective of the state of the fundamentals simply because agents coordinate their beliefs and related actions on such outcomes. This is true as soon as the context involves uncertainty and strategic complementary, as most economic and financial environments do: if everybody believes that others will attack a currency or withdraw their money from the bank or a fund, the optimal individual reaction is to do the same, even if there is no fundamental reason to do so, and the result is sub-optimal. Therefore, understanding what influences beliefs and drives agents' resulting actions is a critical step to resolve coordination issues and avoid efficiency losses.

Coordination games have been extensively studied in game theory, and the implied equilibrium predictions have been tested in many laboratory experiments.[1] These studies typically involve small groups of participants, even though coordination problems in the real world normally involve much larger numbers. Hence, there is a lack of experimental evidence on the possibility of coordination within genuinely large groups that could proxy the atomistic nature of real-world markets and organizations. This is the case irrespective of the type of coordination problems, i.e., whether in environments with multiple equilibria that are Pareto comparable or ranked by risk-dominance or multiple equilibria that include sunspots.

Clearly, large-scale laboratory experiments are prohibitively expensive. Fur-

[1] See Section 2 for a review of the relevant literature in the context of our experiment.

thermore, theory is indifferent about the number of players as soon as there are more than a few. In other words, theoretically, a group of 10 is already a large group. There is no reason, therefore, to fund large-scale experiments. We challenge this view by conducting experiments with 80+ people and report on the results in this paper. In our view, whether studying games with a few players is a good enough proxy for the dynamics of real markets or organizations that involve many actors is critical for policy design and at least deserves empirical investigation.

Hence, whether group size affects coordination is the first question that we study in this paper. To do so, we were able to couple two labs and run a group experiment by varying the group size between 10 and 80+, involving a total of more than 1,200 participants. Our experimental environment is a repeated bank-run game in which we study how group size influences whether participants coordinate on the risk-dominant but Pareto-inferior strategy 'run', or the Pareto optimal but risky 'wait' equilibrium. Importantly, no theoretical equilibrium selection criterion predicts a group size difference between 10 or 80+ participants in our setup.

Of course, in the real world, beliefs are not only influenced by information about the fundamentals but also by non-fundamental factors. In this context, a large and influential literature in macroeconomics has been concerned with *sunspots*:[2] sunspots are publicly observed variables that are unrelated to fundamentals but may affect agents' behavior if they believe these variables

[2]See Azariadis (1981) and Cass & Shell (1983) for early analyses of sunspot equilibria. On the importance of sunspots in macroeconomics and finance in general, we refer to Farmer (1999) for an earlier statement. The literature in game theory has also developed the concept of *correlated equilibria*, which is closely related (see Peck & Shell 1991).

are relevant, and therefore coordinate their actions on these extrinsic signals. This literature has remained largely theoretically-oriented. Empirical evidence of sunspot-induced coordination has remained sparse and has never been established in large-scale experiments, even though only large groups may adequately inform on the behavior of macroeconomic systems that involve many, if not a continuum, of individuals.

Our study is designed to investigate whether a large group of individuals is able to coordinate on a sunspot equilibrium. Specifically, before making their decision, subjects receive an announcement with a random forecast predicting whether few or many agents will withdraw their money from the bank in the coming period, so that each announcement clearly maps onto the strategy 'run' or 'wait'. All announcements are the same across subjects, and they are told that such forecasts have no fundamental value.

Our experiment is the first to show that group size *does* matter for both questions. As soon as the Pareto-optimal equilibrium is substantially riskier than the secure but Pareto-inferior 'run' equilibrium, small and large groups behave differently: while large groups systematically and rapidly coordinate on the Pareto-inferior equilibrium, groups of 10 may also coordinate on the Pareto-optimal 'wait' equilibrium, or even on sunspot-induced changes in strategies. Importantly, we never observe any instances of coordination on sunspot equilibrium in large groups. Therefore, our experiment shows that the variety of empirically relevant outcomes is drastically reduced when the group size is large enough.

Furthermore, bearing in mind the limitations of our study due to the limited number of experimental groups, we can say that coordination on sunspots appears empirically fragile even in small groups. We discuss the results in light of two plausible explanations – namely, the greater strategic uncertainty and the weaker individual market power in a large group with respect to a small one. The evidence from our post-experimental questionnaire support the former, while the latter is formalized using an individual evolutionary model.

The rest of the paper is organized as follows. Section 2 discusses the related literature, while Section 3 describes the experimental design. In Section 4, we present and discuss the experimental results, and Section 5 concludes with the implications from our study.

2 Related literature

Our study relates to several strands of the experimental literature. First, we contribute to the broad literature that investigates the possibility of coordination on bank runs in laboratory environments derived from the seminal theoretical contribution of Diamond & Dybvig (1983); see Duffy (2016) for a survey. In particular, we borrow the bank-run game from Arifovic et al. (2013) and Arifovic & Jiang (2019) who simplify the seminal game to facilitate the lab implementation while preserving its major feature of interest for our study, namely strategic complementarity. In this setup, bank runs are the result of pure coordination failures as the fundamentals (i.e., the rates of return) are non-stochastic and common knowledge. To the best of our knowledge, while real-life bank runs require coordination among a large number of depositors, none of the existing studies has considered groups larger than 10 players.

Our study also relates to the more specific experimental literature on sunspots. A few experiments have tried to reproduce sunspot-driven behavior in the lab. Marimon et al. (1993) introduce sunspots via blinking squares on the subjects' screen in an overlapping generation (OLG) experiment with multiple equilibria. They find evidence of price volatility induced by the sunspot variable, but not of long-lasting coordination. Duffy & Fisher (2005) find systematic group coordination on a sunspot in a stylized asset market with multiple equilibria, provided that the sunspot is phrased so as to clearly map onto the possible strategies available to the players. Yet, they only consider small groups – namely, five buyers and five sellers and, most importantly, the multiple equilibria in their setting cannot be Pareto-ranked. Similarly, Fehr et al. (2018) investigate coordination in a simple game where all equilibria yield the same payoff but they can be ranked by risk-dominance. They find evidence of sunspot-driven behavior but only consider coordination between pairs of subjects. Arifovic et al. (2020) report on an experiment in a macroeconomic model with two steady states and a sunspot variable. The authors find that coordination on sunspots may arise whether the two equilibria yield the same payoff or not, but coordination on sunspots is easier when they do. Once again, their experiment uses groups of only six subjects. The closest to our study is Arifovic & Jiang (2019), where the authors examined the same bank-run game as ours, but in groups of 10 players. They find experimental support for sunspots in risky environments. In the present study, we add two dimensions: we vary the transition probabilities of the sequence of announcements that serves as the sunspot variable, and most importantly, we scale the experiment up and consider groups of 80+ players.

Finally, while there are no large-scale experiments on coordination games such as bank-run environments or on sunspots, there exists an experimental literature on the effect of group size to which our study also contributes. Yet, only a handful of papers, that we discuss below, study the effect of group size in cash-incentivized experiments involving genuinely large groups.

So-called large groups in this literature do not typically encompass more than 14 or 16 subjects, while small ones usually involve pairs.[3] Sally (1995) provides a meta-analysis of experimental evidence in psychology on prisoners' dilemmas and finds that larger groups tend to cooperate less frequently. Using the minimum-effort game, Van Huyck et al. (1990) also show that groups of 14-16 players quickly converge to the worst equilibrium, while pairs of subjects may achieve coordination on the Pareto-optimal equilibrium. Using the same numbers of subjects, Weber (2006) shows that efficient outcomes may be maintained when a group size is gradually increased if newly added members are informed about the group history. Reporting on three groups with 24 subjects each, Riedl et al. (2016) show that efficiency may be achieved if subjects can choose with whom to play the game.

[3]There is a literature on classroom experiments where participants are students incentivized with extra credits, but there is no focus on the effect of group size (see, e.g., Williams & Walker (1993)). One of the largest experiments we are aware of has been conducted by Gracia-Lázaro et al. (2012), where participants are scattered over two networks of 600 players each and repeatedly play a 2-by-2 prisoner's dilemma with their neighbors. They show that the structure of the network does not affect cooperation that reaches similar levels to the ones reported in smaller lattices or unstructured populations of players. Yet, their game features decentralized, paired interactions among a large population of individuals, but not centralized group interactions among a hundred individuals, as is the case in the present study.

In a general class of coordination games, Heinemann et al. (2009) do not find any group size effect among groups of 4, 7 and 10 players. Duffy & Xie (2016) show that cooperation dramatically drops when the group size increases from 2 to 12 players. By contrast, in public goods games, the early survey by Ledyard (1995) concludes that group size has no systematic effect on contribution levels. Yet, using volunteer students incentivized with extra credits, Isaac et al. (1994) report that large groups, of up to 100 participants, contribute more than small groups, but the group level of contribution to the public good also depends on the impact of individual contributions on the group outcome. Diederich et al. (2016) report a similar result in an online experiment involving equally large groups. Weimann et al. (2019) revisit this environment with a much lower marginal per capita return in a lab setup that is similar to ours: the authors connect four labs and run public goods experiments with 8, 60 and 100 subjects. They find that there is no significant difference in contribution levels between large groups, where the impact of each individual contribution on the group outcome is negligible, and small groups, where each individual player has greater aggregate impact. Yet, in order for large groups to contribute as much as small ones, the aggregate benefit of individual contributions has to be salient enough in terms of payoff. One of our main results shares the same flavor, albeit obtained in a different environment: we show that large groups are sensitive to the riskiness of the mutually beneficial strategy.

In more complicated experimental environments, some studies have found it necessary to scale up the size of the group to obtain convergence towards equilibrium predictions; see, for instance, the trading experiments with multiple securities of Bossaerts & Plott (2004). As this review shows, there is a lack of experimental evidence on the possibility of coordination within genuinely large groups. We now detail our experimental design.

3 The experimental setup

3.1 The experimental game

We implement an experimental coordination game using a group design with N participants per group and within-session randomization. The game is a simplified version of the Diamond & Dybvig (1983) model. Participants repeatedly play the role of depositors at the experimental bank. In each round, they receive a deposit of one unit of experimental currency, which implies that the initial capital of the bank is equal to N. Their only task is then to simultaneously decide whether to withdraw their deposit, i.e., to 'run' on the bank, or to roll it over, which corresponds to the action 'wait'.

The bank promises to pay a short-term interest rate $r > 1$ for those who withdraw, and a long-term interest rate $R > r$ for those who wait. It does so in the following way. Once all participants have submitted their decision, the bank pays out the amount r to the depositors who decide to withdraw their deposit, within the limit of its initial capital N. If the bank has to use all its resources, they are equally distributed over all the withdrawals. Let us denote by $0 \leq e \leq N$ the number of early withdrawals. Hence, the payoff for withdrawing is given by:

$$\pi_{withdraw} = \min\left(r, \frac{N}{e}\right) \tag{1}$$

The remaining capital then generates a return R and is shared equally among depositors who decided to roll over. Hence, the amount of early withdrawals erodes the availability of funds and jeopardizes the payment of a higher rate of return for those who choose to wait. In particular, as soon as $e \geq \hat{e} \equiv \frac{N}{r}$, the bank's capital is fully exhausted by the early withdrawals, and participants who chose to roll over their deposit receive nothing. Therefore, the payoff for waiting is given by:

$$\pi_{wait} = \max\left(0, \frac{N - er}{N - e}R\right) \tag{2}$$

This game is characterized by strategic complementarity: it is in the interest of any participant to align her decisions with the others. This leads to multiple equilibria. In particular, the stage game has two symmetric pure-strategy Nash equilibria: the 'run'-equilibrium, where all participants withdraw their deposit ($e = N$) and receive one unit, and the 'wait'-equilibrium, where all participants roll over their deposit ($e = 0$) and receive the promised return R. There is one symmetric mixed-strategy equilibrium, where $\theta = (R - r)/[r(R - 1)]$ is the equilibrium probability of withdrawing. The repeated version of the game also features stochastic sunspot equilibria (SSE) along which participants switch between the two strategies conditionally on the realization of an extrinsic random variable.

To allow for the possibility of coordination on SSE in the experiment, we add such an extrinsic variable that is displayed to all participants and gives a forecast of the number of withdrawals at the beginning of each round. Following Arifovic & Jiang (2019), the forecast may take two values, namely either "The forecast is that e^* or more people will choose to withdraw" or "The forecast is that e^* or less people will choose to withdraw", where the value of e^* represents the number of people withdrawing that makes withdrawing and waiting equally beneficial. This number is adapted depending on the number of participants in a specific group (see Section 3.2). The transition between the two states is governed by a Markov process.

Importantly, forecasts are public knowledge but random. They do not relate to the fundamentals of the model, namely the returns on deposits r and R. The participants are aware of this absence of relation.[4] Hence, they may be understood as *sunspots* variables. Duffy & Fisher (2005) show that a sunspot variable is more likely to act as a coordination device if it clearly maps onto the subjects' decisions and if participants only receive aggregate information. In our experiment, this is the case and the forecasts unambiguously signal to 'run' or 'wait' participants. Hence, our experiment displays *a priori* the most favorable conditions for coordination on sunspots to occur. This type of context-led sunspots also correspond to illustrative examples given by Farmer (1999).

Before turning to the experimental treatments, let us remark that, from

[4]The exact wording in the instructions is as follows, see Appendix C: "*These announcements are forecasts, which can be right or wrong. The experimenter does not know better than you how many people will choose to withdraw (or wait) in each period. The actual number of withdrawals is determined by the decisions of all participants. Your actual payoff depends only on your own choice and the choices of other participants.*"

a game theoretic point of view, sunspot equilibria are nested in the notion of correlated equilibria. One such equilibrium could be one such that each player plays the following strategy: "choose 'wait' when the forecast is less than e^* withdrawals, choose 'withdraw' otherwise". Of course, it would be possible to construct other correlated equilibria in the stage game based on the forecast series (e.g., act exactly opposite to what the forecast predicts). However, we will not discuss these cases here, and the SSE that we consider in the repeated game is described by the repeated play of the above-mentioned correlated equilibrium.

3.2 Treatments and hypotheses

To study group coordination in the presence of multiple equilibria, we vary three parameters of the environment: the number of players N, the payoff structure via the short-term interest rate r, and the persistence of the forecast sequence via ρ.

The first four treatments vary the group size N and the payoff r. We consider two group sizes: $N = 10$ and $N > 80$. We set $R = 2$ in all treatments and vary the short-term interest rate between $r = 1.54$ and $r = 1.33$ to alter the relative attractiveness of the pure-strategy equilibria 'wait' versus 'run'. With $r = 1.54$, we have $e^* = 0.3N$, that is, if exactly 30% of the participants withdraw, then waiting and withdrawing give the same payoff. With $r = 1.33$, $e^* = 0.5N$, i.e., exactly half of the people have to withdraw for both strategies to yield the same payoff, which makes 'wait' more attractive than in the $r =$ 1.54 case. Figure 1 illustrates the payoff structure. In the four treatments,

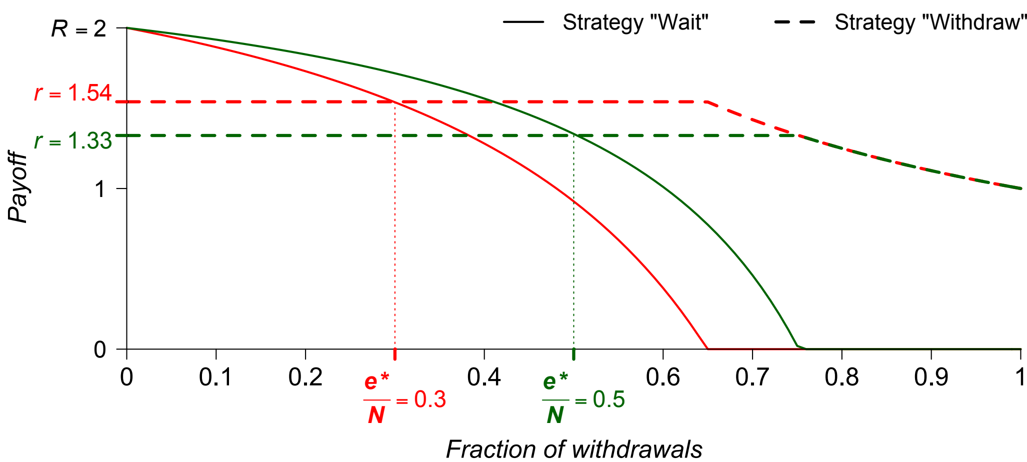


Figure 1: Payoff functions in the experiment

Notes: On the y-axis, payoff of the strategy 'Wait' (solid lines) and 'Withdraw' (dashed lines) as a function of the fraction of the players who choose to withdraw (x-axis) for $r = 1.54$ (in red, treatment *Run*) and $r = 1.33$ (in green, treatment *Wait*).

we set the transition probability $\rho = 0.8$, namely the announcement series is persistent.

We also consider small and large group sizes where the announcement series is not persistent and the occurrence of one or the other forecast does not depend on the previous announcement, both being purely equiprobable in any period, i.e., with $\rho = 0.5$. Given the cost of the experiment, we only use $r = 1.54$ in this set of treatments.

We first formulate hypotheses in the context of the group size N and the payoff structure r and the two pure-strategy equilibria 'wait' and 'run'. We discuss later hypotheses regarding coordination on SSE.

Table 1 presents equilibrium selection results using different game-theoretic refinements in the context of our experimental treatments. Note that most of the equilibrium selection criteria are usually defined for 2x2 games with 2

Payoff structure		$r = 1.54$		$r = 1.33$	
Group size		$N = 10$	$N = 100$	$N = 10$	$N = 100$
Equilibrium refinement criteria	Security/minmax	'Run'	'Run'	'Run'	'Run'
	Payoff-dominance	'Wait'	'Wait'	'Wait'	'Wait'
		($\forall N$)		($\forall N$)	
	Risk-dominance	'Run'	'Run'	'Run'	'Run'
		(for $N > 2$)		(for $N > 3$)	
	Maximization of potential	'Run'	'Run'	'Run'	'Run'
	Global payoff uncertainty	'Run'	'Run'	'Run'	'Run'
	Dynamic random matching	'Run'	'Run'	'Run'	'Run'
		(for $N > 2$)		(for $N > 6$)	
	Darwinian dynamics	'Run'	'Run'	'Wait'	'Wait'
		(for $N > 2$)		($\forall N$)	

Notes: Computations based on Kim (1996) and Monderer & Shapley (1996), see Appendix A. This table does not distinguish between the different persistence parameters, as they do not affect the selected equilibria. For simplicity we work with large groups of 100 in theory. However, the results do not differ if we work with $N \in [75, 90]$.

Table 1: Equilibrium refinements in the experimental treatments with respect to N and r

players. We relegate all the details of their generalization to an n-player game to Appendix A. The first five criteria correspond to a static selection process, while the last two are obtained within dynamic models of selection.

The first two refinements – the minmax criterion of von Neumann and payoff-dominance (Harsanyi & Selten 1988) – do not depend on the parameters r or N. It is easy to see that 'run' is always the *secure strategy*, i.e., the strategy that maximizes the minimum possible payoff. By 'running' players can ensure a minimum payoff of 1 independently from the behavior of the others, while the strategy 'wait' exposes players to the risk of a zero payoff. The 'wait'-equilibrium is clearly payoff-dominant, as it can deliver the maximum return $R = 2 > r$. However, the 'wait'-equilibrium is also more risky since it requires that a large enough fraction of the other players coordinate on ‘wait’.

This intuition is translated in the notion of risk-dominance that selects an equilibrium based on the relative profits from each action (Harsanyi & Selten 1988). This notion can be generalized to N-player games. Following Kim (1996), in all our treatments, the ‘run’-equilibrium risk-dominates the ‘wait’-equilibrium. The criterion of maximization of the potential of the game (Monderer & Shapley 1996) is another generalization of the notion of risk-dominance to an N-player game. In Appendix A, we show that the maximization of the potential, as well as the global payoff uncertainty approach of Carlsson & van Damme (1993) and the dynamic random matching framework of Matsui & Matsuyama (1995) all select the secure strategy in our N-player game, i.e., select the ‘run’-equilibrium.[5]

The literature on evolutionary game theory also provides predictions about the long-run outcome of the repeated version of the game. Interestingly, the Darwininan dynamics of Kandori et al. (1993) is the only criterion that predicts different outcomes across our treatments with respect to r values. According to this criterion, $r = 1.33$ is low enough for the selection of the ‘wait’-equilibrium, while $r = 1.54$ results in the ‘run’-equilibrium. The predictions hold for any $N > 2$. Yet, the value of $r = 1.33$ is exactly the watershed of coordination on ‘run’ versus ‘wait’ under this theory (see also Figure 9d in Appendix A, and Temzelides 1997). All six other selection criteria considered in Table 1

[5]Note that the equivalence between the selected equilibrium under those concepts and the risk-dominant equilibrium is established in a 2-player game but does not necessarily hold with $N > 2$ players. The evolutionary criterion of Foster & Young (1990) also always selects ‘run’ but applies to the limiting case of an infinite population and continuous time and is therefore not directly transportable in our experimental setting, especially in small groups. We report its outcomes in Appendix A for the sake of completeness.

fail to predict a difference in the outcome along the r treatment variable. In the absence of clear-cut theoretical predictions, it is useful to look at previous experimental evidence. In an environment without announcements and in small groups only, Arifovic et al. (2013) find support for the predictions of the Darwinian dynamics, i.e., subjects tend to coordinate on the 'wait'-equilibrium for $r = 1.33$ and on the 'run'-equilibrium for $r = 1.54$. We then formulate our first hypothesis:

Hypothesis 1 *Participants are more likely to coordinate on the 'run'-equilibrium in the risky environment ($r = 1.54$) and on the 'wait'-equilibrium with $r = 1.33$.*

As for the group size effect, all equilibrium refinement criteria considered give the same insight: $N = 10$ is already a 'large' number, and further increases in the group size should not make any difference in the selected equilibrium, which leads us to formulate our second hypothesis:

Hypothesis 2 *Group size does not affect the coordination of the participants in our experiment.*

We now formulate hypotheses on the effect of our three treatment variables on the likelihood of coordination on SSE. Concerning the effect of ρ, the literature on learning in dynamic models has discussed the condition for convergence on the type of SSE introduced in the experimental game, namely a two-state Markov SSE with constant transition matrix; see Evans & Honkapohja (2001, Chap. 4). In particular, coordination on an SSE near a pair of distinct equilibria may occur if those two equilibria are each *locally* stable under some simple adaptive learning rules. We can easily show that the 'wait' and the 'run' equilibria are each locally stable in the bank-run game, and this stability result is independent of the value of the transition matrix ρ (see also Arifovic et al. 2013).

Yet, convergence results are to be understood only locally, which means that agents' initial actions must be 'not too far away' from the SSE for coordination on the SSE to be possible. In other words, participants must initially believe in a correlation between the announcements and market outcomes. For this reason, and to give a higher chance of coordination on the SSE to occur, we introduce training periods during which participants are first individually conditioned to follow the announcements before entering the incentivized group game (Duffy & Fisher 2005, Arifovic et al. 2020). Of course, participants were explicitly told that the first 6 periods of the experiment did not count towards their earnings and only served the purpose of practicing the game while playing against $N-1$ 'robots'. The robot-players' behavior was pregenerated such that a majority of them follow the announcements. Participants were aware that their own decisions could not have any impact on them. They also knew that all their peers played against the same $N-1$ robots so that subjects shared a similar experience.

While there is no theoretical reason why parameter ρ may help or hinder coordination on the SSE, we can formulate some behavioral arguments. On the one hand, purely random announcements (i.e., when $\rho = 0.5$) vary more than persistent ones (i.e., when $\rho = 0.8$), which could catch the attention of the participants more than when the same announcement is repeatedly displayed. Additionally, it may be easier to coordinate on an equilibrium that corresponds to a long sequence of the same announcement and disregard the later changes in the announcements. On the other hand, more variable forecasts may generate more volatility in possible types of behavior and a greater strategic uncertainty from the players' point of view. Hence the treatment $\rho = 0.5$ may favor coordination on the safe strategy 'withdraw'. As arguments may go either way, we take a neutral stand when formulating our third hypothesis:

Hypothesis 3 *Coordination on SSE may arise independently from the values of the transition probability* ρ.

As for the role of the group size N, we are not aware of any theoretical or empirical argument that would allow us to take a stand on the relative likelihood of coordination on SSE in large or small groups. Hence, we formulate the following hypothesis:

Hypothesis 4 *Coordination on SSE may arise independently from the group size* N.

The payoff structure may also affect coordination on the SSE. In a similar experiment but with small groups, Arifovic & Jiang (2019) do not find any instances of coordination on SSE when one or the other Nash equilibrium is 'salient enough' and provides a clear coordination device to the participants. This is the case either for values of r close to R that make 'run' more attractive or for values of r close to 1, which make 'wait' a valuable option. For an intermediate value of $r = 1.54$, the authors report instances of coordination on SSE ($R = 2$). Yet, the case of other intermediate values, such as $r = 1.33$, has not been investigated. We then again take a neutral stand and form our last hypothesis:

Hypothesis 5 *Coordination on SSE may arise independently from the values of r considered.*

3.3 Experimental procedures and implementation

The experiment was programmed in `PhP/MySql` and run between May and October 2017 by connecting the CREED lab of the University of Amsterdam and the LINEEX lab at the University of Valencia. In total, we recruited 1,246 participants (324 in Amsterdam and 922 in Valencia) over 16 experimental sessions.[6] Subjects were recruited from the IBSEN-pool, which recruits among the general population.[7] Most of our participants are students, in various fields, with an average age of about 22 years (with $\sigma = 5.13$, min=18, max=65). An experimental session lasted approximately two hours, including instructions, post-experimental questionnaire and payment. For large groups, we recruited as many participants as we could from both labs and pooled them together. Therefore, the number of subjects in large groups varies from 78 to 90. For small groups, we always recruited in total 60 participants per session

[6] Additionally we ran two pilot sessions in Amsterdam with four small groups in order to check the procedures before the large sessions. As these sessions substantially differ from the main experimental sessions in terms of procedures (no connection to the other lab was made), we do not consider these data in the analysis.

[7] The EU H2020 IBSEN project involved cooperation across 7 universities across Europe with the general aim to shed light on large-scale human behavior. For more information, see `https://ibsen-h2020.eu/`.

and randomly matched them, independently from their location, in groups of 10 players. This resulted in six independent groups for each *small* treatment and four independent groups for each *large* treatment. Subjects were not allowed to participate in more than one session.

At the beginning of the experiment, participants could choose between Spanish or English. They were then given the opportunity to read the instructions individually on their screen at their own pace, and had to answer a series of comprehension questions before starting the experiment. They were also given a print-out of the instructions that they could access at any time during the session. The complete instructions are given in Appendix C. Once all participants had correctly answered all questions, they entered the first period of the game.

After the six-period training phase, in period 7, participants started playing as a group. The group composition did not change throughout the experiment – which corresponds to fixed-group treatment – and participants were not allowed to communicate with each other. In each period, their payoff in units of experimental currency was calculated by Equations (1) or (2). Participants then played repeatedly for 50 periods, and their total payoff corresponded to their cumulative payoff over those 50 periods. The exchange rate between the experimental currency and euro was 4 to 1. The average earnings per participant were 22 euros (including a participation fee of 5 euros).

At the beginning of each period, the information available to participants only included three pieces: i) the total number of withdrawals up to the previous period, ii) their own earnings and iii) the announced forecasts up to

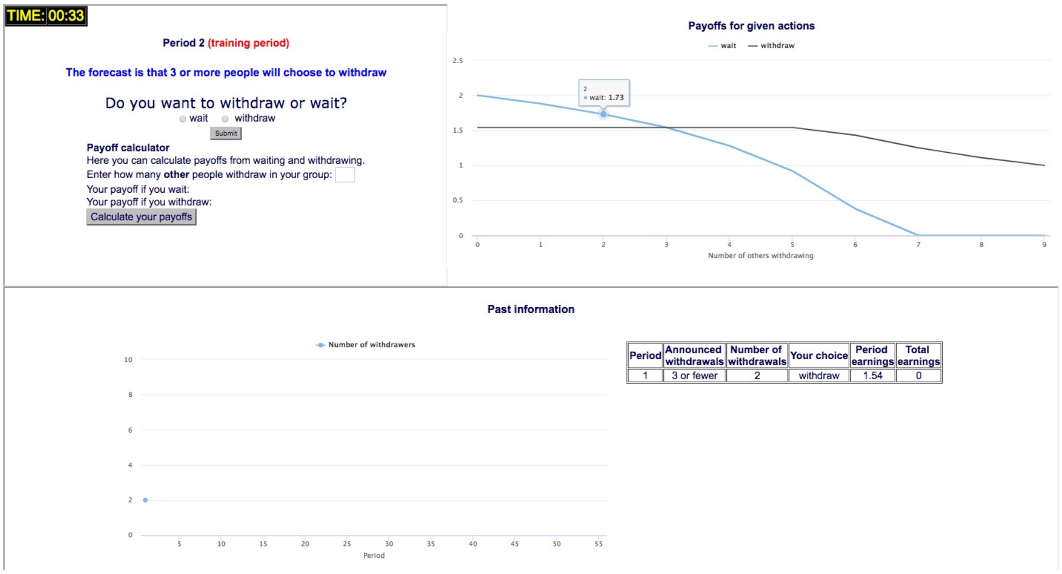


Figure 2: Decision screen in a large group

the current period; see their decision screen on Figure 2. We used the same sequence of announcements in all treatments and groups. Past information was shown in a table and in a graph, over which participants could hover their mouse to see the exact numbers. Moreover, subjects had access to the payoffs for waiting and withdrawing using a payoff graph and a payoff-calculator embedded in their screen. The calculator would return the payoff of each strategy based on any hypothetical forecast of *other* withdrawals that subjects would enter. Subjects had one minute to make a decision in each period (including the usage of the payoff-calculator). If they did not make a decision on time, they earned nothing for that period and their previous action (or a random action in period 1) counted towards the others' earnings.

At the end of the bank-run game, we conducted a post-experimental questionnaire. This questionnaire consisted of six questions and was unincentivized.[8] For the first five questions, subjects had to choose on a 5-point scale the extent to which they agree with the given statement. Question 6 was an open question where we asked subjects to specify their strategy if they used any. After the experiment, we asked a native Spanish research assistant who was also proficient in English to code all the answers from 1 to 6 using the following categories:

1. Try to wait, but then withdraw if others also do that.
2. Initially withdraw, as it is safer, but might switch to wait.
3. Follow the crowd: who do not specify a given strategy but emphasize that they want to do whatever the others do.
4. Follow the announcements.
5. Other rule.
6. Irrelevant item (e.g., description of what *should* have been done).

We now turn to our experimental results.

4 Experimental results

We first present an overview of our experiment results in Figures 3–5. Each figure displays the total number of withdrawals over time in selected groups. Those selected groups are representative of each type of observed behavior, and we defer the exhaustive list of figures to Appendix D. The solid black

[8] The exact questions are listed in Appendix C.

Payoff structure	$r = 1.54$	$r = 1.54$	$r = 1.33$
Persistence parameter	$\rho = 0.5$	$\rho = 0.8$	$\rho = 0.8$
Group size			
Small (10)	3/6 'run'; 2/6 SSE; 1/6 'wait'	4/6 'wait'; 2/6 'run'	6/6 'wait'
Large (75–90)	4/4 'run'	4/4 'run'	4/4 'wait'

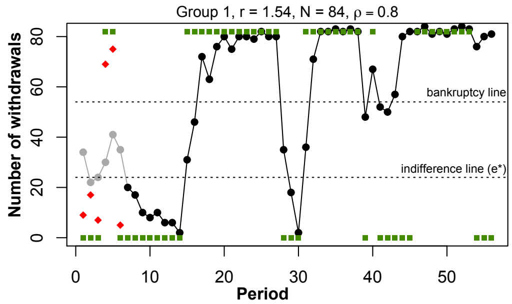
Table 2: Summary of the observed equilibria

line shows the total number of withdrawals in each group. The green squares represent the forecasts, where a value of 0 means that fewer than e^* participants were predicted to withdraw and a value equal to N indicates more than e^* predicted withdrawals. The red diamonds represent the pre-generated robot behavior (for large groups it is proportional to the group size). The lower horizontal line indicates e^*, the indifference line between the two strategies, and the higher horizontal line represents the 'bankruptcy line' ($\hat{e} = \lceil N/r \rceil$): above this threshold, those who wait earn nothing.

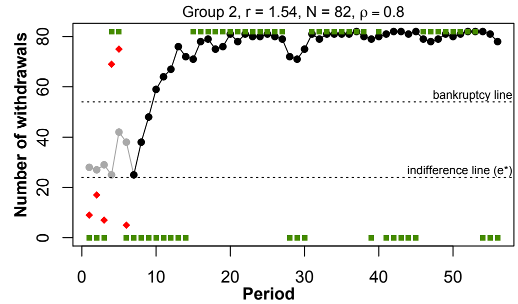
In order to quantify the results we obtained in the different sessions we provide the following criterion. We define coordination on a particular equilibrium, i.e., 'wait', 'run' or the SSE, as a configuration where at least 80% of the subjects play the corresponding strategy in the last 10 periods of the experiment. All experimental groups converge to one of the three equilibria, and we present the results in Table 2.[9]

Section 4.1 below discusses the group size differences across the two payoff treatments in the context of Hypotheses 1 and 2, and Section 4.2 focuses on the effect of the forecast announcements in the context of Hypotheses 3 and 5. Section 4.3 outlines several potential explanations behind our results.

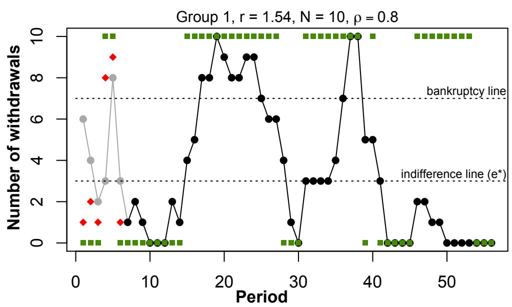
[9]This definition is only meant to help summarize the results. Alternative thresholds would not alter our main findings.



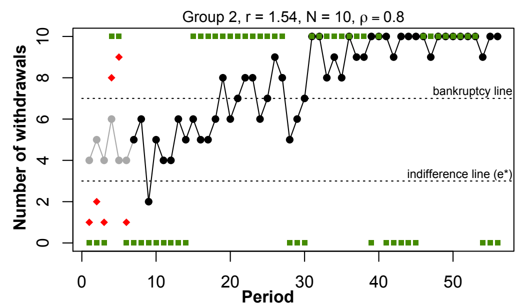
(a) A large group



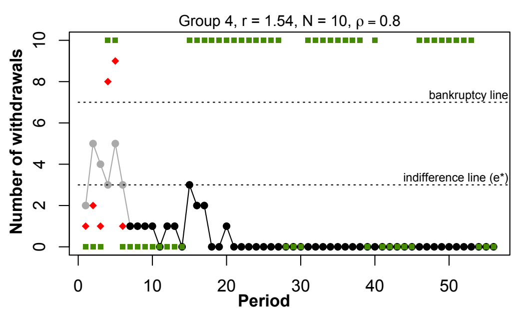
(b) A large group



(c) A small group



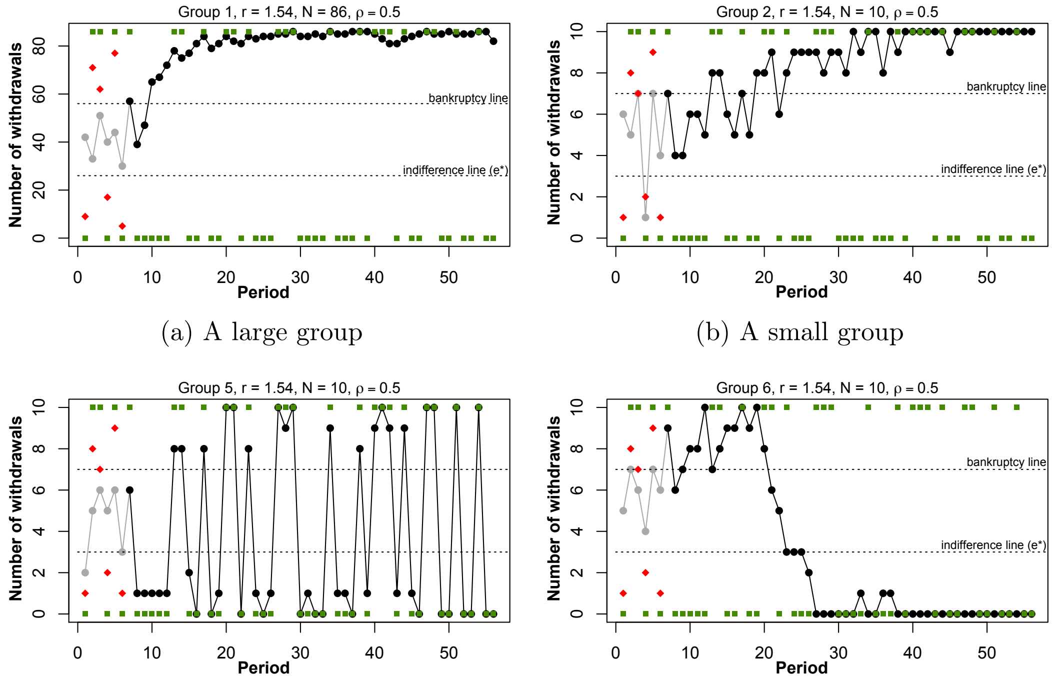
(d) A small group



(e) A small group

Notes: The dots represent the total number of withdrawals, the solid black line stands for periods counting towards participants' earnings, while the solid gray line stands for the training periods. The green squares represent the announcement (0 when less than e^* and N when more than e^* withdrawals are predicted). The red diamonds are pre-generated robot behavior (for large groups it is proportional to the group size). The lower horizontal line is at e^* (indifference line), the higher horizontal line at the 'bankruptcy' line, i.e., above this line those who wait earn nothing. For the large groups, the exact number of subjects in the given group is indicated in the title of the corresponding plot. The upper two figures are representative for the large groups, the further figures represent small groups. Additional figures can be found in Appendix D.

Figure 3: Withdrawals with $r = 1.54$ and $\rho = 0.8$



(a) A large group (b) A small group

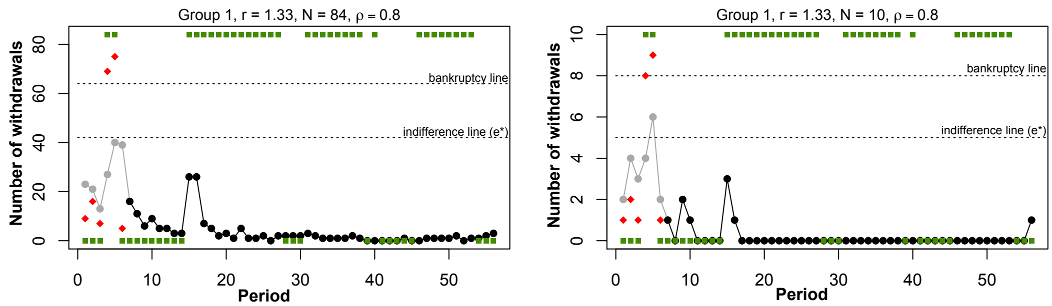
(c) A small group (d) A small group

Notes: See Figure 3. The upper-left figure is representative of the large groups; the other figures represent small groups. The rest of the groups can be found in Appendix D.

Figure 4: Withdrawals with $r = 1.54$ and $\rho = 0.5$

4.1 Group size and payoff structure

We first focus on the treatment with $r = 1.54$, while abstracting for the moment from the effect of ρ (cf. Figure 3 for $\rho = 0.8$ and Figure 4 for $\rho = 0.5$). A striking result is that behavior in *small* groups appears more heterogeneous than in *large* groups even though the payoff structure of the game is independent from the group size. All eight large groups invariably converge to the 'run'-equilibrium: seven groups do so within the first periods, while the remaining group (Group 1 with $\rho = 0.8$, Figure 3a) converges to 'run' after what is akin to a 'transient SSE'. This group mostly followed the first three announce-



(a) A large group (b) A small group

Notes: See Figure 3. The left-hand side figure is representative of a large group, and the right-hand side figure of a small group. The rest of the figures can be found in Appendix D.

Figure 5: Withdrawals with $r = 1.33$ (and $\rho = 0.8$)

ments but after the fourth change (in period 39), too few participants switched to 'wait' and the group eventually coordinated on the 'run'-equilibrium.

By contrast, among the 12 small groups with $r = 1.54$, only five converge to the 'run'-equilibrium (Groups 2 and 3 with $\rho = 0.8$ and Groups 2, 3 and 4 with $\rho = 0.5$), while another five converge to the 'wait'-equilibrium, two of them after a transient SSE (Groups 1, 4, 5 and 6 with $\rho = 0.8$ and Group 6 with $\rho = 0.5$) and the two remaining groups settle down on the SSE (Groups 1 and 5 with $\rho = 0.5$).

Results are dramatically different in the less risky environment, i.e., when $r = 1.33$: all groups, no matter their size, invariably converge to 'wait'; see Figure 5.

A formal analysis of the experimental data confirms that these group size differences are statistically significant for $r = 1.54$, and insignificant for $r = 1.33$. Table 3 compares the mean withdrawal rates between treatments using the non-parametric Mann-Whitney rank sum tests.[10] Panel A shows the mean

[10] In the sequel, all statistical tests are performed on matching group level unless otherwise

withdrawal rate for the whole data (from period 7 onwards), whereas Panel B depicts the same for period 7 only, i.e., the first paying period.

If we compare the two payoff structures, we see that subjects wait significantly more often in case of $r = 1.33$ than in case of $r = 1.54$ (see Table 3, columns II and III; the p-value is 0.05 for the small groups and 0.02 for the large ones). Hence, with the following result, we cannot reject Hypothesis 1:

Result 1 *All experimental groups coordinated on the 'wait'-equilibrium in case of* $r = 1.33$*, whereas many groups coordinated on the 'run'-equilibrium for* $r = 1.54$*. Hence, the payoff structure affects coordination.*

Looking at the relatively riskier environment (namely $r = 1.54$), the mean withdrawal rate is indeed significantly higher in large groups than in small groups (the p-value is 0.03 for both ρ-values), while this difference becomes insignificant in the safer environment, namely $r = 1.33$ (the p-value is 0.52). This leads us to reject Hypothesis 2 with the following result:

Result 2 *For* $r = 1.54$*, coordination depends on the group size, but for* $r = 1.33$*, it does not. Thus, group size matters in the riskier environment.*

Interestingly, these treatment differences do not stem from different initial behaviors in period 7 across the different group sizes: subjects start with about the same withdrawal rate in both small and large groups (see Panel B of Table 3).[11] Moreover, those fractions are significantly different from 50% according

stated.

[11]Note that this is the only period where subjects cannot be influenced by each others' behavior. In the first six periods, they do not interact with each other, but they do have a common history (the robots) to share, and this history is the same in all groups. For this reason, all individual observations in period 7 are still independent and, hence, directly comparable.

	(I) $r = 1.54$ $\rho = 0.5$	(II) $r = 1.54$ $\rho = 0.8$	(III) $r = 1.33$ $\rho = 0.8$	(I) vs. (II) $r = 1.54$ only $p-$value	(II) vs. (III) $\rho = 0.8$ only $p-$value
Panel A. Whole data (from period 7)					
Small	0.61 (0.38)	0.39 (0.40)	0.03 (0.07)	0.26	0.05*
Large	0.92 (0.13)	0.87 (0.22)	0.04 (0.06)	1.00	0.02**
p-value	0.03**	0.03**	0.52		
Panel B. Period 7					
Small	0.68 (0.47)	0.23 (0.43)	0.15 (0.36)	0.00***	0.25
Large	0.64 (0.48)	0.31 (0.46)	0.23 (0.49)	0.00***	0.03**
p-value	0.52	0.25	0.17		

Notes: ∗∗∗: significant at the 1% level, **: significant at the 5% level, and *: significant at the 10% level. Standard deviations are reported in parentheses after the mean withdrawal rate, which is based on the paid periods (periods 7–56). All tests are two-sided Mann-Whitney rank sum tests. In Panel A the tests are performed on matching group level with 4 observations for large treatments and 6 for small treatments. In Panel B the tests are performed on the individual level with 60 observations for the small group treatments, 347 for the large group treatment with $\rho = 0.5$ and $r = 1.54$, 333 for large group treatment with $\rho = 0.8$ and $r = 1.54$, and 326 for the large group treatment with $r = 1.33$.

Table 3: Mean withdrawal rate over time, and for the first paid period

to a binomial test ($p = 0.00$ for all cases), which rules out the possibility that subjects decided purely randomly.

To explain the emergence of the group size difference from a similar initial state, we now proceed in two steps. First, we restrict our attention to the initial stage of the experiment, i.e., between periods 7 and 8, and show that group size differences already arise in period 8. Second, we show that behaviors are more sluggish in the subsequent periods of the experiment in the large groups than in the small groups, resulting in only small deviations from the selected 'run' equilibrium in large groups, but larger switches between the two equilibria and the SSE in the small groups.

At a first glance, a quick comparison of the fractions of subjects waiting in period 7 but changing for withdrawing in period 8 reveals significant group size differences. In large groups, 18% of the subjects switch from 'wait' to 'run', whereas only 10% do so in the small groups ($p = 0.043$ according to a 2-sample test for equality of proportions). Similarly, 34% of the subjects in large groups waited in period 7 and kept waiting in period 8, while 44% did so in small groups (the difference being significant with a p-value of 0.053). Looking at the other two strategy combinations between periods 7 and 8 (namely, 'run then wait' and 'run and keep running') does not lead to any significant group size difference. Hence, strategy switching from wait to withdraw between periods 7 and 8 is a critical difference between large and small groups.

Table 4 dives in greater detail into the decisions in period 8. The table presents logit models with the individual withdrawal decision in period 8 as the dependent variable. The first four columns look at the treatments with $r = 1.54$. It shows that the probability of choosing to withdraw in period 8 increases if more than e^* people withdrew in period 7 (even when controlling for participants' own past decisions), but also that subjects in large groups withdraw significantly more often than in small groups. This is also true for subjects in the $\rho = 0.8$ treatment.[12] The group size difference is robust to the inclusion of the treatments with $r = 1.33$ (column 5).

In columns 3 and 4, we separate the same data into two groups, namely subjects who waited and subjects who withdrew in period 7. This exercise confirms that the difference in group size originates from those subjects who waited in period 7 and switched to run in period 8, as no group size effect

[12] The effect of the announcement sequence is better understood by recalling that the announcement changes from 'run' to 'wait' between periods 7 and 8 in the $\rho = 0.5$-sequence, while it remains at 'wait' in the $\rho = 0.8$-sequence.

Dependent variable: withdrawal decision in period 8					
			waiters	withdrawers	all
	$r = 1.54$	$r = 1.54$	$r = 1.54$	$r = 1.54$	data
own withdrawal$_7$	-	0.63	-	-	0.77**
		(0.39)			(0.33)
$> e^*$ withdraw$_7$	0.94***	0.88***	1.00***	0.50	0.87***
	(0.30)	(0.30)	(0.30)	(0.63)	(0.30)
large	0.43**	0.43**	0.80***	0.03	0.46***
	(0.21)	(0.20)	(0.22)	(0.33)	(0.16)
$\rho = 0.8$	0.32**	0.51***	-0.31	1.30***	0.55***
	(0.13)	(0.16)	(0.40)	(0.35)	(0.15)
$r = 1.33$	-	-	-	-	-0.65**
					(0.29)
constant	-1.69***	-2.05***	-1.95***	-0.98	-2.16***
	(0.32)	(0.32)	(0.40)	(0.72)	(0.31)
# of observations	800	800	421	379	1186
-LogLikelihood	522.76	514.85	246.94	249.56	677.00

Notes: ***: significant at the 1% level and **: significant at the 5% level. Standard errors are reported below the coefficients in parentheses. The dependent variable is the withdrawal dummy (1 - withdraw) in period 8. Independent variables are the last decision (in period 7), a dummy to measure whether more than e^* actually withdrew in the last period (1 - yes), a group size dummy (1 - large), an announcement sequence dummy (1 - $\rho = 0.8$), and a payoff dummy (1 - $r = 1.33$). The first four columns include treatments with $r = 1.54$ only, the last column includes all six treatments. Column 3 (respectively 4) uses only data from those who waited (respectively withdrew) in period 7.

Table 4: Logit model on withdrawal decision in period 8

is found on the withdrawals separately. In particular, subjects who waited in period 7 are more likely to withdraw in period 8 in large groups than in small groups.

This regression shows that subjects react differently to the same initial proportion of withdrawals when the environment is riskier ($r = 1.54$): in large groups, subjects switch to 'run' more often than in small groups. Now comes the second stage of our argument to explain the subsequent different directions taken by large and small groups. After period 8, subjects in large groups change strategy significantly less than in small groups. From period 8 onward, more than 85% of subjects in large groups choose 'run' in two consecutive periods, while only 41% did so in small groups. The difference is significant (the p-value of the two-proportions z-test is 0.000). Such a difference was not observed in the less risky environment with $r = 1.33$: in both large and small groups, almost 85% of the subjects chose 'wait' in two consecutive periods.

This means that large groups exhibit a more sluggish behavior than small ones. At the aggregate level, this is confirmed by looking at the time series of the withdrawal rate. Their variance is 10 times larger in small groups than in large ones, and this difference is significant (the p-value of the Kruskal-Wallis test is 0.000). Hence, in the risky environment, the behavior of large and small groups is different, both at the individual and at the aggregate levels. Before discussing some possible explanations for the group size difference in Section 4.3, we take a look at the effect of the two announcement sequences on the individual behaviors in the context of Hypotheses 3–5.

4.2 The effects of the announcements

From Table 2 and Figures 3–5, we only observe two instances of coordination on the SSE, both in a small group playing in the risky environment ($r = 1.54$) with random announcement sequences ($\rho = 0.5$). This observation stands against Hypotheses 3–5. While the reported effect of a riskier payoff on the coordination on sunspot equilibria is in line with previous experimental results (Arifovic & Jiang 2019), the group size difference and the effect of the persistence parameter ρ are novel insights. Bearing in mind the intrinsic limitation of our results associated to the small number of groups, it seems that large groups are less likely to coordinate on the announcements, and it seems that a purely random sequence ($\rho = 0.5$) helps enforce coordination on such strategies rather than a persistent announcement sequence ($\rho = 0.8$).

At a first glance, Table 3 does not report any significant differences between the mean withdrawal rates for $\rho = 0.8$ and $\rho = 0.5$ over the course of the whole experiment (Panel A). Yet, we observe a significantly higher withdrawal rate in the first paid period (Panel B, Table 3) for $\rho = 0.5$, where the announcement in period 7 predicted more than e^* people withdrawing than for $\rho = 0.8$, where the announcement predicted fewer than e^* people withdrawing. Of course, we cannot disentangle from this data whether those who behaved in accordance with the announcement did follow it, or simply took this specific action disregarding the announcement. We need a detailed analysis of the effect of the announcements on the individual behavior of the participants.

To do so, we look at periods where the announcements changed and compute the fraction of people who behaved according to the announcement both *before* and *after* the announcement change (see Table 5).[13] Looking at the whole data (Panel A), the only significant group size difference is observed for $\rho = 0.5$ (the p-value is 0.08), which is the only treatment where we observe long-lasting sunspot-driven behavior in small groups. Otherwise, the announcements affect small and large groups in the same way.

Comparing the two sequences of announcements reveals that subjects fol-

[13]Note that our measurement of following the announcements is rather conservative, as we restrict ourselves to only those observations where an individual changed his/her behavior according to the announcement *exactly* when the announcement changed. We only consider the first two announcement changes, as the results of Section 4.1 show that the effect of the announcement fades away over time. By looking at the first two changes, we also look at both types of changes (i.e., from a 'wait'-forecast to a 'withdraw'-forecast and the other way around).

	(I) $r = 1.54$ $\rho = 0.5$	(II) $r = 1.54$ $\rho = 0.8$	(III) $r = 1.33$ $\rho = 0.8$	(I) vs. (II) $r = 1.54$ only $p-$value	(II) vs. (III) $\rho = 0.8$ only $p-$value
Panel A. Whole data					
Small	0.33 (0.47)	0.11 (0.32)	0.04 (0.19)	0.20	0.04**
Large	0.06 (0.25)	0.11 (0.31)	0.04 (0.20)	0.25	0.77
p-value	0.08*	0.39	0.52		
Panel B. First change					
	Period 8	***Period 15***			
Small	0.40 (0.49)	0.25 (0.44)	0.22 (0.42)	0.05*	0.25
Large	0.39 (0.50)	0.15 (0.36)	0.27 (0.44)	0.08*	0.55
p-value	0.67	0.39	0.52		
Panel C. Second change					
	Period 13	***Period 28***			
Small	0.25 (44)	0.17 (0.38)	0 (0)	0.74	0.04**
Large	0.16 (0.37)	0.20 (0.40)	0.02 (0.13)	0.25	0.06*
$p-$value	0.39	0.54	0.02**		

Notes: **: significant at the 5% level and *: significant at the 10% level. Std deviation is in parentheses after the mean following rate. All tests are two-sided Mann-Whitney rank sum tests performed on matching group level with $n = 4$ for large treatments and $n = 6$ for small treatments.

Table 5: Proportion of subjects following the announcements

low significantly more often the first change for $\rho = 0.5$ than for $\rho = 0.8$, both in large and small groups (Panel B; the p-values are, respectively, 0.05 and 0.08 for small and large groups). Note that in the $\rho = 0.5$-sequence, the announcement already changed in period 8, when subjects may still have considerable uncertainty about others' behavior. By contrast, in the $\rho = 0.8$-sequence, the first announcement change occurred later, in period 15, when uncertainty about others' behavior is mostly resolved. The difference in compliance with the announcements between the two ρ sequences vanishes over time: for the second change, there are no significant differences across ρ-values.

These observations lead us to the following companion results:

Result 3 *Subjects tend to follow the announcements for* $\rho = 0.5$ *more often than for* $\rho = 0.8$*. We do observe two instances of coordination on SSE with* $\rho = 0.5$*, but none with* $\rho = 0.8$*.*

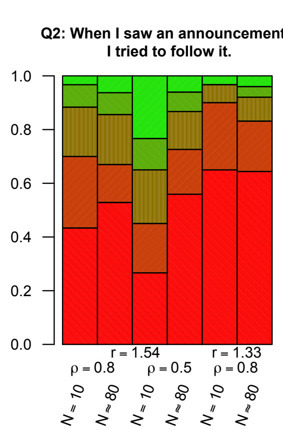
Result 4 *Subjects tend to follow the announcements in small groups more often than in large groups. Two small groups coordinate on SSE while none of the large groups do.*

Comparing across payoff structures, we observe that subjects follow the announcements significantly less often in later periods with $r = 1.33$ than with $r = 1.54$ (Panel C; the p-values are 0.04 for small groups and 0.06 for large groups). The reason is that subjects are settled on the 'wait'-equilibrium at an earlier stage of the experiment in $r = 1.33$ than on any of the pure-strategy equilibria in $r = 1.54$, see again Figures 3–5. We then obtain our last result:

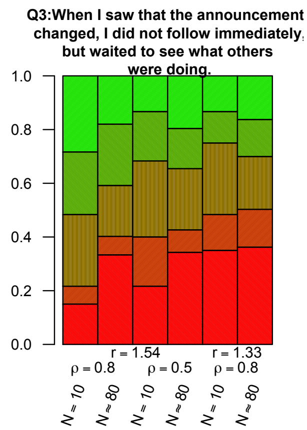
Result 5 *Subjects tend to follow the announcements for* $r = 1.54$ *more often than for* $r = 1.33$*, at least in the early stage of the experiment. We do observe two instances of SSE when* $r = 1.54$*, but none when* $r = 1.33$*.*

The results in this section are also in line with what the participants reported themselves in the post-experimental questionnaire. Note that these questions were not incentivized,[14] and participants answered the questions after the completion of the group experiment, so that the observations from the

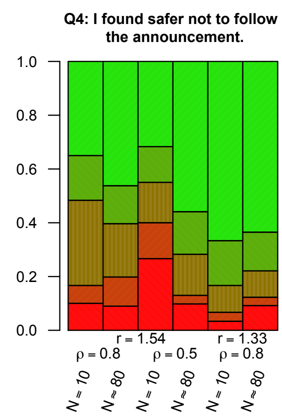
[14] Recall that the experimental sessions were already long, demanding in terms of logistics to connect the two labs, and costly due to the high number of participants and average payment. Therefore, collecting further incentivized data would have been difficult.



(a) Question 2



(b) Question 3



(c) Question 4

Notes: The answers are scaled from 1, corresponding to 'strongly disagree', to 5, i.e., 'strongly agree', with 3 being a 'neutral' statement.

Figure 6: Responses to questions regarding the announcements

questionnaire are not independent. Yet, bearing in mind those limitations to a formal analysis of the answers, they still constitute suggestive evidence to shed further light on our experimental findings.

Questions 2, 3 and 4 concern the announcements. Figure 6 shows the exact questions and the distribution of the answers for these questions. Overall, there is no substantial difference across the different treatments, and most subjects reported not following the announcements, which is in line with the observed behavior. The only exception here is the risky treatment ($r = 1.54$) with $\rho = 0.5$ and small groups where we see that subjects reported following the announcements to a greater extent (see Questions 2 and 4). This is indeed the treatment where two out of the six groups did coordinate on an SSE.

Similarly, the answers to the open question were also treatment-dependent (see Figure 7c on page 39): strategy 4 ('Follow the announcements') was not popular (about 5 to 10% of the participants), except in the above-mentioned treatment where a more substantial fraction (about 25%) did report to have used this strategy. These answers are consistent with the observed behavior of discarding the announcements. Throughout the experiment, subjects seem to mostly condition their decisions on the overall group history.

Therefore, we reject Hypotheses 3, 4 and 5. A small group or a riskier payoff structure or a purely random sequence of announcements lead to a stronger influence of the announcements on participants' behavior. We could not find any announcement-driven behavior in large groups, no matter the sequence of announcements or the payoff structure. Hence, group size and incentives matter for the emergence of coordination on SSE, although the small number of successful coordination on SSE suggests that this outcome may be fragile. We now flush out several potential explanations behind the reported group size differences.

4.3 Why does group size matter?

None of the theoretical criteria of equilibrium selection reviewed in Section 3.2 could predict the observed group size differences in the experiment. As reported above, a similar initial proportion of withdrawals triggers more frequent 'run' behavior in a large group than in a small one. We discuss two possible lines of explanation. We do not claim that these are the only ones, and we do not take a stand on which one may be more relevant, as our experiment was not designed to discriminate between the two.

Strategic uncertainty Subjects need to form beliefs on which strategy the others will choose and act accordingly. Data from the end questionnaire suggest that participants perceive a greater strategic uncertainty in a large than in a small group.[15] The two questions directly related to strategic uncertainty (i.e., Question 1, see Figure 7a) reveal an equally high level of awareness of this strategic uncertainty between treatments, no matter the group size. Yet, while all participants tried to form beliefs about others' behavior, there is a striking difference in the strategies adopted in face of this uncertainty in large and in small groups.

In Question 6, strategy 2 – "Initially withdraw, as it is safer, but might switch to wait" – was mentioned by a substantial share of the subjects in the large groups (16% with $\rho = 0.8$ and even 22% when $\rho = 0.5$), but not or barely mentioned in the small ones (see the blue areas in Figure 7c). A similar group size difference, albeit of milder magnitude, is observed when $r = 1.33$. In this treatment, the most popular strategy was the 'wait and see' strategy; about 43% in small groups and 45% in large groups reported to use a similar strategy, which confirms that the perception of risk also depends on the payoff structure.

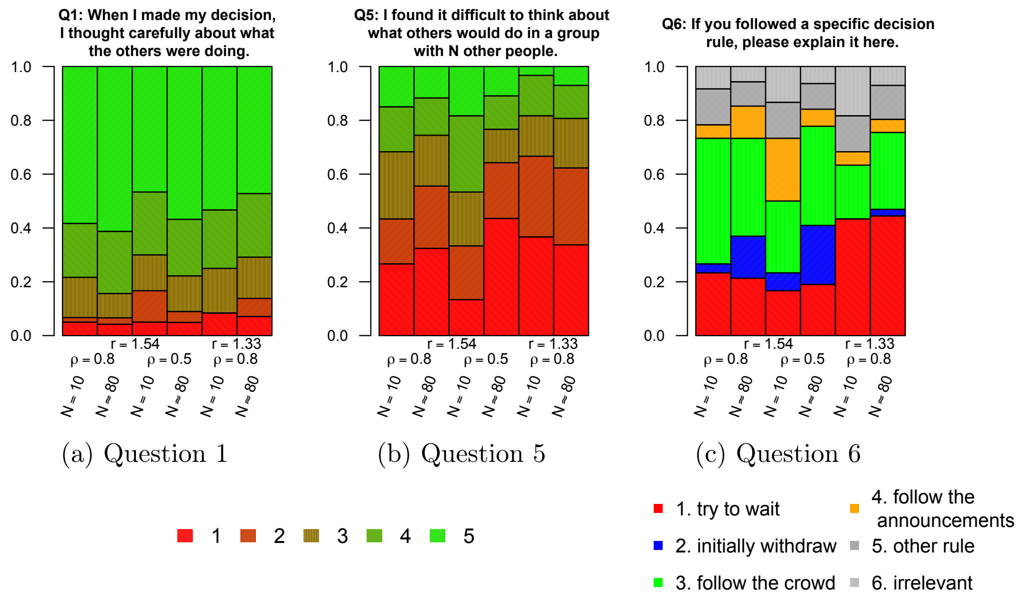
It is possible to formalize the idea of a greater perceived strategic uncer-

[15] Interestingly, Duffy & Xie (2016) identify greater strategic uncertainty in large than in small groups as the driver of the drop in cooperation in a prisoners' dilemma. However, the group sizes considered in their experiment range between 2 and 12 players. We show here that this increase in strategic uncertainty has a dramatic effect between 10 and 80+ subjects.

tainty in large groups by using an evolutionary argument. Kandori et al. (1993, Theorem 3) state that the equilibrium with the larger basin of attraction is selected in the repeated version of a coordination game. In our case, the basin of attraction of the 'wait'-equilibrium corresponds to a share of early withdrawals $[0, \hat{\theta})$, while the basin of attraction of the 'run'-equilibrium is $(\hat{\theta}, 1]$ where $\hat{\theta}$ is the mixed-strategy Nash-equilibrium withdrawal probability. As illustrated in Figure 9d in Appendix A, as r increases, the 'run' equilibrium gains a relatively larger basin of attraction at the expense of the 'wait' equilibrium. However, 10 or more than 80 participants do not influence the relative likelihood of each equilibrium. Stated differently, the theoretical threshold value $\hat{\theta}$ is independent of N, the group size.

Our experimental results suggest that this is in fact not the case: subjects perceive the basin of attraction of the 'run'-equilibrium to be larger in a large than in a small group. A given initial share of withdrawals, even if it lies below the threshold $\hat{\theta}$, is then perceived as riskier in a large than in a small crowd, and participants tend to choose the safe but sub-optimal 'run'-strategy in a large group more often than they do in a small one.

As a corollary, if subjects perceive a greater strategic uncertainty in a group of more than 80 than in a group of 10, this can explain the absence of coordination on SSE in large groups as well. In presence of strategic uncertainty, subjects tend to prefer a payoff stream that does not fluctuate and, thus, appears safer to one that does and depends on the ability of the entire group to coordinate on the corresponding strategy changes, as is the case along



(a) Question 1 (b) Question 5 (c) Question 6

Notes: On Fig. 7a–7b, the answers are scaled from 1, corresponding to 'strongly disagree', to 5, i.e., 'strongly agree', with 3 being a 'neutral' statement. On Fig. 7c, the detail of the classification is given in the main text.

Figure 7: Questions regarding strategic uncertainty (Q1–5) and strategies (Q6).

an SSE.[16] Looking at Question 5 on Fig. 7b, it indeed seems that subjects perceive strategic uncertainty to be higher along an SSE than in the case of coordination around a pure-strategy equilibrium: more than half of the subjects in the treatment where SSE were observed (i.e., $N = 10$, $r = 1.54$ and $\rho = 0.5$) found it difficult to think about what the others were doing, and this is the highest proportion among all treatments. In the other treatments, this proportion does not exceed 25%.

Signal and power of the pivotal player The second, related, class of explanations for the observed group size differences relates to the relative im-

[16]See Arifovic et al. (2019) for a related discussion in an OLG experiment with six players coordinating on a two-cycle or a steady state allocation.

portance of each individual player in a small versus a large group. In fact, from the point of view of an individual, we could say that there is an 'uncertainty of order $1/N$' around the size θ of the basin of attraction of the 'wait' equilibria.

Indeed, in treatments with $r = 1.54$, all groups, large and small ones, start close to the indifference line $\hat{\theta} \simeq 0.3$ between the two strategies; see the time series of the withdrawal rates in Fig. 3 and 4. However, in a small group, as a single participant weights 10% of the group, *ceteris paribus*, an individual who decides to switch from 'run' to 'wait' will be pivotal and will move the group from the basin of attraction of 'run' to the one of 'wait'. In that case, in a small group, participants' actions have a signaling value towards the payoff-dominant 'wait'-equilibrium.[17]

By contrast, in a group of about 100, a participant weights a marginal 1% of the market and cannot send a clear coordination signal when switching to 'wait'. Hence, in the next period, looking again at the payoff function in Figure 1, the difference in expected payoff from switching from one strategy to the other when the group is close to the indifference threshold is potentially large in a small group if one individual can be pivotal, but infinitesimal in a large group. Hence the uncertainty around the threshold $\hat{\theta}$ is of order $1/N$. Taking into account the entire future course of the experiment results in an even greater difference in payoff for the pivotal player. Bearing the cost of waiting in earlier periods does not potentially reward this player by inducing repeated coordination on the payoff-dominant strategy in a large group, while

[17] In a pivotal voter model, Duffy & Tavits (2008) find that subjects tend to overstate the probability of being pivotal in a group of 10, especially at early stages of the experiment, which may reinforce the strategic effort towards 'wait' in our small groups.

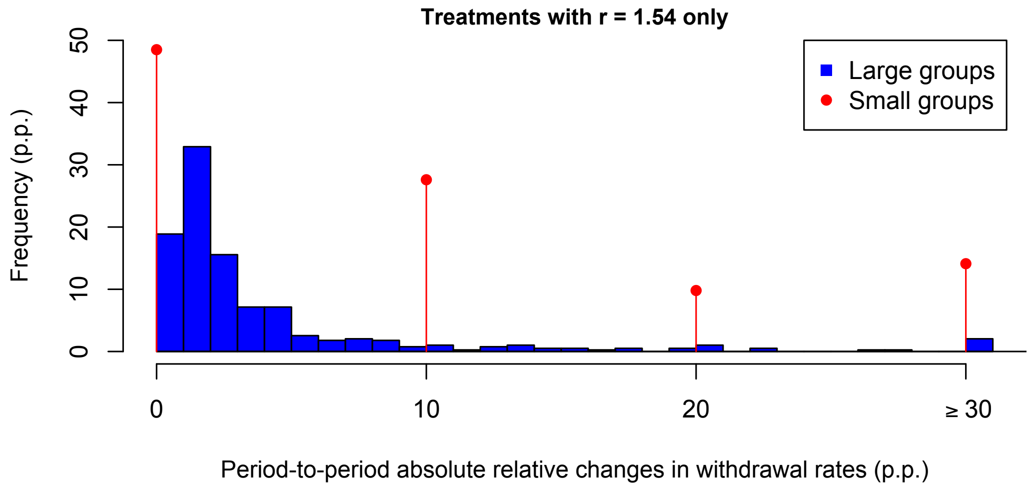
it may be the case in a small group. The same reasoning holds for signaling an intention to coordinate on following the announcements, which still provides a higher payoff than 'run' and may appear worthy in a small group but costly in a large one.

This point is illustrated in Figure 8, which shows the histogram of the absolute relative changes in withdrawal rates for both small and large groups. In large groups, especially in the risky environment ($r = 1.54$), changes in withdrawal rates are of much smaller amplitude than in small groups. In large groups, instances where at least 10% of the participants change strategies from one period to the next rarely occur (in less than 10% of the periods), while it is fairly common in small groups (it occurs more than half of the time).[18] This explains why large groups exhibit a considerable larger amount of sluggishness than small ones. In particular, large groups are less likely to coordinate on an SSE, which requires a 100% change in the group withdrawal rate each time the announcement changes.

In Table 6, we present the outcomes from simulations of the Individual Evolutionary Learning model of Arifovic & Ledyard (2012) that reproduces group size differences in the context of $r = 1.54$ but does not when $r = 1.33$, in line with our experimental results.[19] Those group size differences arise from the computation of the foregone payoff of each strategy 'wait' or 'run' that takes into account several periods (at least 2) and includes the possibility that

[18] To achieve a 10% change in smaller groups, it is sufficient that only one person changes their behavior. For a given probability of change, this is more likely to occur than at least 10 persons changing their behavior in a group of 100. This also illustrates the importance of group size and power in this environment

[19] The details of the algorithm are given in Appendix B.



Notes: We compute the absolute relative change in strategies as $\mid e_t - e_{t-1} \mid \times 100$, where e_t is the proportion of withdrawals in period t. This represents the fraction of participants who change strategies between two periods in large groups (blue histogram) and in small groups (red bars) in all treatments with $r = 1.54$. All numbers are expressed in percentage points (p.p.).

Figure 8: Distribution of the absolute relative changes in the withdrawal rates in treatments with $r = 1.54$

the player is pivotal. Indeed, each player computes the foregone payoff of the 'run' strategy in each past period by taking into account the number of others withdrawing and performing a 'what-if' analysis, i.e., counting herself among the withdrawals had she instead waited in that period. We conjecture that this class of explanation plays an important role in the observed group size differences. The design of an experiment to explicitly test for behavior in the presence of various market powers is left for future research given the scale of the present study, but is certainly of interest.

Of course there may be further explanations behind our results. For the reasons explained in Footnote 14, we did not measure risk aversion or numeracy[20]

[20] Kleber et al. (2013) show that less numerate people tend to think in absolute terms while more numerate people are sensitive to proportions. Hence, some participants in our experiment may have misperveiced probabilities from a binominal distribution and found more likely that 30 people out of a hundred choose to 'run' than 3 out of 10. If a large

Payoff structure	$r = 1.54$	$r = 1.33$
group size		
Small (10)	63.9% 'run', 36.1% 'wait'	95.7% 'wait',4.32% 'run'
Large (100)	96% 'run', 4% 'wait'	100% 'wait'

Table 6: Frequency of each equilibrium in the IEL simulations

Notes: 10,000 Monte Carlo simulations. Group size differences are robust to any value of the probability of experimentation $P^{ex} \in (0, 0.5)$ and any size of the individual strategy pool $J \geq 20$ and to any value of $m \geq 2$ and $\beta > 2$ (i.e., players have to take several past periods into account to evaluate the foregone utilities of each strategy that is realistic given that the whole history was available to them in the experiment; and the selection pressure has to be strong enough, which is realistic given there are only two simple strategies and the payoff from each of them is clearly visible to the subjects in the experiment). The initial share of withdrawals is taken from the experimental data.

of the participants, but our randomization procedure over the experimental groups and the two locations should eliminate variation in subjects' intrinsic characteristics as a cause of the systematic treatment differences. However, there is evidence that people might display different risk attitudes depending on the situation they face. Shimizu & Udagawa (2011) investigated the question with an Asian disease problem by varying the contextual group size (i.e., the number of people affected by the disease). The authors found that subjects become more risk-averse and tend to choose more often the safer option when their decision affects a larger crowd. This result, initially obtained with students, has been replicated on a larger scale using the general population. Furthermore, the authors observe that subjects are more likely to hide less socially desirable actions in a large group than in a small one, even though subjects are grouped anonymously.

part of the participants were subject to this bias, this effect could explain why a large group generates more strategic uncertainty than a small group. We do not have data to back up such a claim.

In the context of our experiment, those results may help explain why, in the riskier environment $r = 1.54$, we find coordination on the safest but Pareto-inferior strategy 'run' in all large groups, while small groups display more heterogeneous outcomes, including a substantial share of Pareto-optimal 'wait' equilibrium.

5 Conclusion

This paper reports on a large-scale laboratory experiment that investigates how group size affects coordination in a bank-run game with two pure-strategy equilibria and sunspot equilibria. We consider 'small' groups of 10 players, where the experimental literature usually defines this size as large, and our large groups involve 80+ players. The main result is that group size does matter and 10 is not large enough, contrary to what theory predicts and many experiments consider.

As soon as the environment is risky enough and the safe, risk-dominant 'run' strategy is relatively attractive compared to the payoff-dominant but risky 'wait' strategy, we observe group size difference. Small groups may coordinate on the efficient 'wait' outcome, or even on the volatile sunspot equilibrium, while large groups invariably and quickly converge to the inefficient 'run' equilibrium. We never observe any instance of coordination on sunspots in large groups, while we only report a few in small groups, which suggests that coordination on sunspots is fragile.

On the experimental front, our experiment sheds some light on the external validity of lab experiments. While there are obvious reasons why small groups are usually considered in the literature, it may not be such an innocuous assumption. Experimental results may not necessarily hold in larger groups, even when theory predicts so. This is an important message given that large groups are obviously more representative of the real world that group experiments aim to study than a handful of individuals. This suggests that we need to be cautious when interpreting the results from experiments in situations that would involve a mass of people in the real world, such as market experiments. Of course, we are not claiming that results from small groups do not give useful indications of treatment effects, but whether they would survive in a large crowd is not a trivial question.

This brings us to a couple of remarks on the theory front. First, group size can be used as an equilibrium refinement device, as our results suggest that large crowds lead to simpler pictures than small groups, even in the presence of multiple equilibria. If such simpler behavior is due to the disappearance, or at least the mitigation, of strategic behavior in a large crowd, our results bear potentially important implications for macro-finance models, where a continuum of agents is typically assumed to populate the model. In particular, if coordination on SSE is hard to sustain in a simple lab environment, their empirical relevance might have been overstated. We believe that more research is needed in this direction given the large existing literature on the topic.

Additionally, our results on group size effects are relevant for policy design. If it is particularly difficult to coordinate beliefs and related decisions in a large crowd even with clear messages in a stylized lab environment, our results may have implications for the effectiveness of communication policies. Policy makers may also want to alter the incentives additionally to giving clear messages to have an effect on agents' beliefs and subsequent actions. Note, however, that we have considered purely random announcements and subjects were aware of that, while real-world communication is rather a public signal correlated with the fundamentals, which may help make it a coordination device. Given the large literature on the effects of news and the empirical relevance of communication policies in macro-finance, this question also deserves more research.

Furthermore, our results on group size effects potentially bear implications for market design. Atomistic markets populated by a large amount of individual participants or large organizations may not necessarily be welfare-enhancing compared to small groups where each individual actor is endowed with some market power, especially when welfare depends on the ability of agents to coordinate on riskier but payoff-dominant actions.

While we understand the technical obstacles to large-scale lab experiments, we believe that our clear-cut and innovative results and the abundance of potential implications that they entail call for more research in this area.

References

Arifovic, J., Evans, G. W. & Kostyshyna, O. (2020), 'Are sunspots learnable? An experimental investigation in a simple macroeconomic model', *Journal of Economic Dynamics and Control* **110**(C).

Arifovic, J., Hommes, C. & Salle, I. (2019), 'Learning to believe in simple equilibria in a complex olg economy - evidence from the lab', *Journal of Economic Theory* **183**, 106–182.

Arifovic, J. & Jiang, J. H. (2019), 'Strategic uncertainty and the power of extrinsic signals – evidence from an experimental study of bank runs', *Journal of Economic Behavior & Organization* **167**, 1–17.

Arifovic, J., Jiang, J. H. & Xu, Y. (2013), 'Experimental evidence of bank runs as pure coordination failures', *Journal of Economic Dynamics and Control* **37**(12), 2446–2465.

Arifovic, J. & Ledyard, J. (2012), 'Individual evolutionary learning, other-regarding preferences, and the voluntary contributions mechanism', *Journal of Public Economics* **96**(9), 808–823.

Azariadis, C. (1981), 'Self-fulfilling prophecies', *Journal of Economic Theory* **25**(3), 380–396.

Bossaerts, P. & Plott, C. (2004), 'Basic principles of asset pricing theory: Evidence from large-scale experimental financial markets', *Review of Finance* **8**, 135–169.

Carlsson, H. & van Damme, E. (1993), 'Global games and equilibrium selection', *Econometrica* **61**(5), 989–1018.

Cass, D. & Shell, K. (1983), 'Do sunspots matter?', *Journal of Political Economy* **91**(2), 193–227.

Diamond, D. W. & Dybvig, P. H. (1983), 'Bank runs, deposit insurance, and liquidity', *Journal of Political Economy* **91**(3), 401–419.

Diederich, J., Goeschl, T. & Waichman, I. (2016), 'Group size and the (in)efficiency of pure public good provision', *European Economic Review* **85**, 272–287.

Duffy, J. (2016), 'Macroeconomics: A Survey of Laboratory Research', *in* J. H. Kagel & A. E. Roth, eds, *The Handbook of Experimental Economics, Volume 2*, Introductory Chapters, Princeton University Press.

Duffy, J. & Fisher, E. O. (2005), 'Sunspots in the laboratory', *American Economic Review* **95**, 510–529.

Duffy, J. & Tavits, M. (2008), 'Beliefs and voting decisions: A test of the pivotal voter model', *American Journal of Political Science* **52**(3), 603–618.

Duffy, J. & Xie, H. (2016), 'Group size and cooperation among strangers', *Journal of Economic Behavior & Organization* **126**, 55–74.

Evans, G. W. & Honkapohja, S. (2001), *Learning and Expectations in Macroeconomics*, Princeton University Press.

Farmer, R. (1999), *Macroeconomics of Self-fulfilling Prophecies*, Vol. 1, 2 edn, The MIT Press.

Fehr, D., Heinemann, F. & Llorente-Saguer, A. (2018), 'The power of sunspots: an experimental analysis', *Journal of Monetary Economics* **in press**.

Foster, D. & Young, P. (1990), 'Stochastic evolutionary game dynamics', *Theoretical Population Biology* **38**(2), 219–232.

Gracia-Lázaro, C., Ferrer, A., Ruiz, G., Tarancón, A., Cuesta, J. A., Sánchez, A. & Moreno, Y. (2012), 'Heterogeneous networks do not promote cooperation when humans play a Prisoner's Dilemma', **109**(32), 12922–12926.

Harsanyi, J. C. & Selten, R. (1988), *A general theory of equilibrium selection in games*, Vol. 1 of *MIT Press Books*, The MIT Press.

Heinemann, F., Nagel, R. & Ockenfels, P. (2009), 'Measuring strategic uncertainty in coordination games', *The Review of Economic Studies* **76**(1), 181–221.

Isaac, R., Walker, J. M. & Williams, A. W. (1994), 'Group size and the voluntary provision of public goods: Experimental evidence utilizing large groups', *Journal of Public Economics* **54**(1), 1–36.

Kandori, M., Mailath, G. J. & Rob, R. (1993), 'Learning, mutation, and long run equilibria in games', *Econometrica* **61**(1), 29–56.

Kim, Y. (1996), 'Equilibrium selection in n-person coordination games', *Games and Economic Behavior* **15**(2), 203–227.

Kleber, J., Dickert, S., Peters, E. & Florack, A. (2013), 'Same number, different meanings: How numeracy influences the importance of numbers for pro-social behavior', *Journal of Experimental Social Psychology* **49**(4), 699–705.

Ledyard, J. O. (1995), 'Public goods: a survey of experimental research', *in* J. Kagel & A. Roth, eds, *Handbook of experimental economics*, Princeton University Press, Princeton, NJ, pp. 111–194.

Marimon, R., Spear, S. E. & Sunder, S. (1993), 'Expectationally driven market volatility: an experimental study', *Journal of Economic Theory* **61**(1), 74–103.

Matsui, A. & Matsuyama, K. (1995), 'An approach to equilibrium selection', *Journal of Economic Theory* **65**(2), 415–434.

Monderer, D. & Shapley, L. S. (1996), 'Potential games', *Games and Economic Behavior* **14**(1), 124–143.

Peck, J. & Shell, K. (1991), 'Market uncertainty: correlated and sunspot equilibria in imperfectly competitive economies', *The Review of Economic Studies* **58**(5), 1011–1029.

Riedl, A., Rohde, I. M. T. & Strobel, M. (2016), 'Efficient coordination in weakest-link games', *Review of Economic Studies* **83**(2), 737–767.

Sally, D. (1995), 'Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992', *Rationality and Society* **7**(1), 58–92.

Shimizu, K. & Udagawa, D. (2011), 'A re-examination of the effect of contextual group size on people's attitude to risk', *Judgment and Decision Making* **6**(2), 156.

Temzelides, T. (1997), 'Evolution, coordination, and banking panics', *Journal of Monetary Economics* **40**(1), 163–183.
URL: *http://www.sciencedirect.com/science/article/pii/S0304393297000330*

Van Huyck, J., Battalio, R. & Beil, R. (1990), 'Tacit coordination games, strategic uncertainty, and coordination failure', *American Economic Review* **80**(1), 234–248.

Weber, R. A. (2006), 'Managing growth to achieve efficient coordination in large groups', *American Economic Review* **96**(1), 114–126.

Weimann, J., Brosig-Koch, J., Heinrich, T., Hennig-Schmidt, H. & Keser, C. (2019), 'Public good provision by large groups - the logic of collective action revisited', *European Economic Review* **118**, 348–363.

Williams, A. W. & Walker, J. M. (1993), 'Computerized laboratory exercises for microeconomics education: Three applications motivated by experimental economics', *Journal of Economic Education* **24**, 291–315.

Appendix A Equilibrium selection: technical appendix

This section collects the equilibrium outcomes from a wide range of selection criteria in coordination games. For the purpose of our experiment, we present the generalization to an $n > 2$-player game. We refer the reader to Kim (1996, Sec. 3–4) for the details of the derivation of the five criteria below with n players. For the sake of clarity, we adopt the same notations. Within the context of our game, the 'L'-equilibrium is the sub-optimal 'run'-equilibrium, and the Pareto-optimal equilibrium 'H' corresponds to 'wait'. The payoff of 'waiting' when k players wait and, hence, $n-k$ players withdraw is given by:

$$\pi_k^H = \max\left(0, R\frac{n - r(n-k)}{k}\right) \tag{3}$$

where the subscript denotes the total number of players adopting the strategy 'H'/'wait'.[21] Similarly, the payoff of 'withdrawing' when k players withdraw and $n-k$ wait is given by:

$$\pi_k^L = \min\left(r, \frac{n}{k}\right) \tag{4}$$

In particular, we have the payoff from 'waiting' when all other players withdraw: $\pi_1^H = 0$ if $r \geq \frac{n}{n-1}$, and $R[n-r(n-1)]$ otherwise (i.e., $\pi_1^H = \max\left(0, R[n - r(n-1)]\right)$). The payoff of 'withdrawing' when all other players are waiting is $\pi_1^L = r$. The payoff of withdrawing when all players do so as well is given by $\pi_n^L = 1$, and the payoff of waiting when all players do so is given by $\pi_n^H = R$, with $R = 2$ in our experiment.

We note that the first two criteria hereafter refer to a static selection process, while the last three correspond to a dynamic form of selection. In the generalization to n-player games, it is assumed that every player is repeatedly and randomly matched to play the game with $n-1$ other players, out of a finite or infinite population of $N >> n$ players. To stay close to the design of the experiment, we can think of the limit case in which $N \to \infty$ or arbitrary large, we are interested in the effects of n, the size of the game, on the equilibrium selection outcome.

[21]Note that this notation might differ from the notation in the main text. The reason is that we chose to use the number of withdrawals as reference for subjects following previous experimental studies (Arifovic & Jiang 2019), whereas Kim (1996) measures the number of people choosing 'wait'.

A.1 Risk-dominance (Harsanyi & Selten 1988)

We use the infinite N-case in Kim (1996, Eq. 2) to derive the function Φ that maps the share $y \in (0,1)$ of players choosing the strategy 'wait' onto the payoff difference function $\Phi(y)$:

$$\Phi(y) = \sum_{k=1}^{n} \binom{n-1}{k-1} y^{k-1}(1-y)^{n-k}\phi_k \tag{5}$$

with $\phi_k \equiv \pi_k^H - \pi_{n-k+1}^L$, as given in Eq. (3) and (4). In particular, we have $\Phi(0) < 0 < \Phi(1)$. Figure 9a represents the function Φ for each of our four treatments. Note that the function is not monotone due to the non-linear payoff functions.

Kim (1996, Eq. 14) states that if the following relation is true:

$$\Phi\left(1-\mu_n\right) > 0, \text{ with } (1-\mu_n) \equiv \frac{\phi_n}{\phi_n - \phi_1}, \tag{6}$$

(where $1-\mu_n$ measures the net gain from coordinating on H rather than on L), then the 'wait'-equilibrium risk-dominates the 'run'-equilibrium.

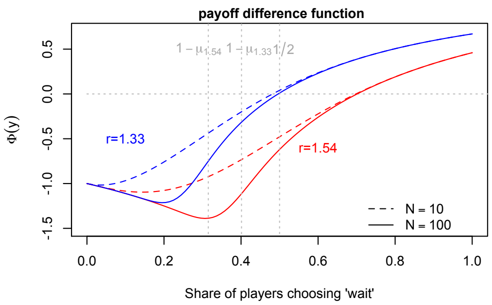
Within the bank-run game, we have:

$$1-\mu_{n,r} = \frac{\pi_n^H - \pi_1^L}{(\pi_n^H - \pi_1^L) + (\pi_n^L - \pi_1^H)} = 1 - \frac{1-\max(0, R[n-(n-1)r])}{R+1-r-\max(0, R[n-(n-1)r]} \tag{7}$$

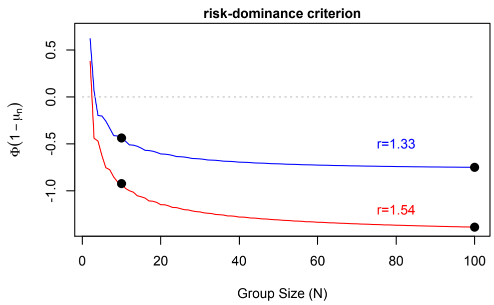
As clear from Eq. 7, $1-\mu_{n,r} = \frac{R-r}{R+1-r} \perp\!\!\!\perp n$ as soon as $r \geq \frac{n}{n-1}$, which is the case in all our experimental treatments. Yet, the group size n still influences the shape of the function Φ. Furthermore, $1-\mu_{n,r}$ is decreasing in $r \in (1, R)$ (we use $R = 2$ in our experiment). As $r \to R = 2$, $1-\mu_{n,r} \to 0$ and $\Phi(0) < 0$. Therefore, when r increases, Condition (6) is less likely to hold, and 'withdraw' becomes the risk-dominant equilibrium.

In the experimental treatments, $1-\mu_{n=10,r=1.54} = 1-\mu_{n=100,r=1.54} = 1-\mu_{1.54} = 0.315$, and $1-\mu_{n=10,r=1.33} = 1-\mu_{n=100,r=1.33} = 1-\mu_{1.33} = 0.4012$. Those thresholds are reported on Figure 9a. We have $\Phi(0.315) < 0$ and $\Phi(0.4012) < 0$, hence Condition (6) does not hold in any of our experimental treatments and *the criterion of risk-dominance always selects the 'run'-equilibrium.*

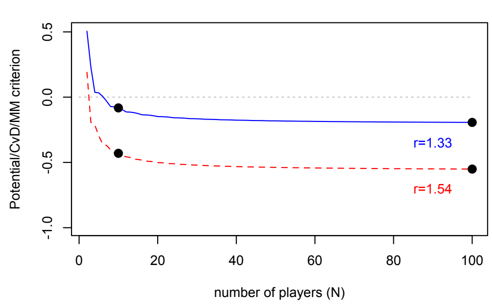
For the purpose of illustration, Figure 9b plots the values of Φ with $r = 1.33$ and $r = 1.54$ as a function of the group size n. Black dots represent the experimental treatments that we consider. The values of Φ, and hence the likelihood of 'wait' to risk-dominate, are decreasing in n. For fairly small groups ($n < 3$ with $r = 1.54$ and $n < 4$ with $r = 1.33$), the 'wait'-equilibrium



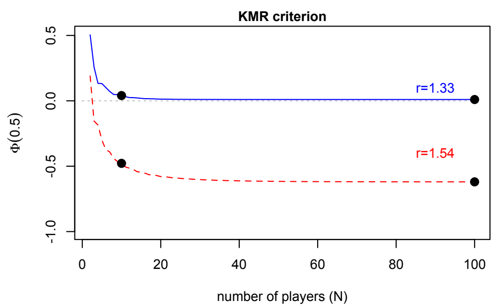
(a) Function Φ as a function of r and N



(b) Φ values as a function of r and N



(c) Potential of the game, CvD criterion of global perturbation and MM criterion



(d) KMR Darwinian criterion

Figure 9: Criteria of equilibrium selection

is risk-dominant. However, the 'run'-equilibrium becomes risk-dominant for any higher group sizes.

A.2 Maximization of the potential (Monderer & Shapley 1996)

Another existing equilibrium refinement concept is the maximization of the potential of a game. Monderer & Shapley (1996) show that any congestion game has a potential function and provide a way of constructing this potential. It is easy to see that our bank-run game is a congestion game, in which the payoff of the players only depend on the number of players choosing a given action from a finite set of actions. For our game, the potential function is defined by the following function (isomorph up to a constant):

$$P(A) = \sum_{l=1}^{e_1} \min\left\{r, \frac{n}{l}\right\} + \sum_{l=1}^{e_2} \max\left\{0, \frac{n-(n-l)r}{l} R\right\}, \tag{8}$$

where A denotes an action profile of the players with $a_i \in \{\text{withdraw,wait}\}$, e_1 (e_2) is the number of agents withdrawing (waiting) according to A with $e_1 + e_2 = n$. This function is indeed a potential function: if we fix all but 1 player's action and look at how the value of the potential changes by changing the last player's action, we find that this change is exactly the change in the utility of this last player. Thus, $P(A)$ is a potential function of the game. Note that this function is not differentiable everywhere, thus maximization is not straightforward. However what is most interesting to us is to look at the two pure-strategy equilibria and calculate the potential value in these equilibria. The equilibrium with the highest potential is the selected one. The potential of the 'run'-equilibrium is:

$$P(e_1 = n) = \sum_{l=1}^{\lfloor n/r \rfloor} r + \sum_{l=\lceil n/r \rceil}^{n} \frac{n}{l} = \sum_{k=1}^{n} \pi_k^L, \tag{9}$$

and the potential of the 'wait'-equilibrium is:

$$P(e_2 = n) = \sum_{l=1}^{\lfloor n-n/r \rfloor} 0 + \sum_{l=\lceil n-n/r \rceil}^{n} \frac{n-(n-l)r}{l} R = \sum_{k=1}^{n} \pi_k^H \tag{10}$$

Therefore, if

$$P(e_2 = n) - P(e_1 = n) > 0 \tag{11}$$

the 'wait'-equilibrium is selected.

Some algebra shows that (9) is increasing in r, while (10) is decreasing in r (fixing R and n). For (10) this is straightforward: as r increases, we have fewer nonzero elements in the sum, and each element becomes smaller. As for (9), we fix $r' > r$. Then we have $P(e_1 = n) = \sum_{l=1}^{\lfloor n/r' \rfloor} r' + \sum_{l=\lceil n/r' \rceil}^{n/r} \frac{n}{l} + \sum_{l=\lceil n/r \rceil}^{n} \frac{n}{l}$. Since $r' > r$ the number of elements in the first two expressions is exactly the same as in (9), but all of these elements are larger for r' than in (9) with r. Hence, (9) is increasing in r.

Thus, for all group sizes, there is a threshold r^* for which (9) and (10) are equal. For higher short-term rates, Condition (11) is less likely to hold and the 'run'-equilibrium is more likely to be selected. For smaller r-values, Condition (11) is more likely to hold and the 'wait'-equilibrium to be selected.

The threshold r^* is also group size dependent: the higher N, the smaller the region where the 'wait'-equilibrium is chosen. To see that, Figure 9c displays the LHS of (11) for $n = 2, ..., 100$ for $r = 1.54$ and $r = 1.33$. The black dots correspond to the experimental treatments. In fact, small population sizes play a role: with $r = 1.54$, 'wait' is only selected if $n = 2$, and for $r = 1.33$, 'wait' is selected for $n < 7$. However, the threshold values r^* become essentially independent from n as the group size becomes large enough.

From Figure 9c, it is clear that, in neither of our treatments does the Condition (11) hold as the values of the function are all negative. Hence, in our game, *the maximum potential criterion selects the 'run'-equilibrium in all our experimental treatments.*

A.3 Global payoff uncertainty (Carlsson & van Damme 1993)

In the global payoff uncertainty approach developed by Carlsson & van Damme (1993), each player observes the payoff matrix with some (small) noise before selecting a strategy. This refinement approach (denoted by CvD hereafter) states that an equilibrium is selected if it is robust with respect to global perturbation. In a 2-player game, Carlsson & van Damme (1993) show that the iterated dominance principle selects the risk-dominant equilibrium.

Following Kim (1996), in our n-player game, the 'wait'-equilibrium is selected by the CvD criterion if:

$$\frac{1}{n}\sum_{k=1}^{n}[\pi_k^H - \pi_k^L] > 0 \Leftrightarrow \sum_{k=1}^{n}\pi_k^H > \sum_{k=1}^{n}\pi_k^L \tag{12}$$

By noticing that

$$\sum_{k=1}^{n}\pi_k^L = \sum_{l=1}^{\lfloor n/r \rfloor} r + \sum_{l=\lceil n/r \rceil}^{n} \frac{n}{l} \tag{13}$$

and

$$\sum_{k=1}^{n}\pi_k^H = \sum_{l=1}^{\lfloor n-n/r \rfloor} 0 + \sum_{l=\lceil n-n/r \rceil}^{n} \frac{n-(n-l)r}{l} R \tag{14}$$

it is easy to see that Condition (12) is the same as Condition (11) discussed above. Hence, in our n-player game, the CvD criterion still selects the risk-dominant equilibrium, and *the 'run'-equilibrium is always selected.*

A.4 Dynamic random matching (Matsui & Matsuyama 1995)

Matsui & Matsuyama (1995) consider a dynamic random matching framework, in which players are rational and maximize their expected future discounted payoff, but cannot switch strategies at will due to frictions. A selected equilibrium is called uniquely absorbing. As shown in Kim (1996), in an n-player game, the 'wait'-equilibrium is selected if Condition (12) holds. Therefore, in our n-players framework, the equivalence between risk-dominance and equilibrium selection in Matsui & Matsuyama (1995) survives, and this criterion *selects the 'run'-equilibrium in all our treatments.*

A.5 The evolutionary criterion of Kandori et al. (1993) (KMR hereafter)

Kandori et al. (1993) consider an evolutionary criterion in discrete time with a finite population size, guided by the Darwinian principle that the strategy with the highest payoff spreads out in the population of the strategies at the expense of the least-performing one. This environment corresponds to the limiting case of Matsui & Matsuyama (1995) where there is no friction and players are myopic, i.e., the best-response myopic dynamics. In this case, each player adopts a best-response against the current configuration of the society as a whole: given a proportion y_t of players committed to 'wait', a player chooses the 'wait' strategy if $\Phi(y_t) > 0$ and the 'run'-strategy if $\Phi(y_t) < 0$ (and is indifferent between the two in case of strict equality).

Under this evolutionary process, a selected equilibrium is the long-run equilibrium of the game. Following Kim (1996), the 'wait'-equilibrium is selected if:

$$\Phi\left(\frac{1}{2}\right) = \sum_{k=1}^{n} \binom{n-1}{k-1} \left(\frac{1}{2}\right)^{n-1} \phi_k > 0 \tag{15}$$

Along the same way of reasoning as in Section A.2, as r increases, 'withdraw' is more likely to be selected, and the effect of n is milder as soon as the population size is large enough.

Looking at Figure 9a for 0.5 on the x-axis, Condition (15) holds when $r = 1.33$ and $N = 10$ or 100 ($\Phi(0.5) > 0$), but does not for $r = 1.54$ and $n = 10$ or 100 ($\Phi(0.5) < 0$). Figure 9d displays the values of (15) for the two chosen levels of r and n varying from 2 to 100, with the dots corresponding to our treatments. For $r = 1.54$, (15) is positive and the 'wait'-equilibrium prevails only for $n = 2$. For $r = 1.33$, the values of (15) are all positive.

Independently from r, these values become essentially flat once $n > 20$.

Hence, *'wait' is the long-run equilibrium if* $r = 1.33$, independently from the group size n, and *'run' is the long-run equilibrium if* $r = 1.54$ as soon as $n > 2$.

A.6 Evolutionary dynamics à la Foster & Young (1990)

The selected equilibrium is the stochastically stable equilibrium. The share of players committed to the 'wait'-strategy in time t, y_t, evolves according to a deterministic replicator dynamics as: $dy_t = y_t(1 - y_t)\Phi(y_t)dt$. The equilibrium is selected by minimizing the corresponding potential function U defined by:

$$U(y) = -\int_0^y x(1-x)\Phi(x)dx \tag{16}$$

with $y \in (0, 1)$ indicating as above the share of players choosing to wait. Kim (1996, p. 223) shows that if $U(1) > 0$ (resp. $U(1) < 0$), then the 'run'-equilibrium (resp. the 'wait' equilibrium) is stochastically stable. With our payoff functions, with $r = 1.54$, we have $U(1) = 0.076 > 0$ for $n = 10$ and $U(1) = 0.102 > 0$ for $n = 100$, and with $r = 1.33$, we have $U(1) = 0.008 > 0$ with $n = 10$ and $U(1) = 0.028 > 0$ with $n = 100$. Hence, *the 'run'-equilibrium is always selected.*

Appendix B Individual Evolutionary Learning

The Individual Evolutionary Learning algorithm

Initialization ($t = 1$)

1. Set the parameter values: number of agents $N = \{10, 100\}$, length of the simulation $T = 50$, short- and long-run interest rate $r = \{1.33, 1.54\}$ and $R = 2$, number of strategies per agent $J = 40$, probability of experimentation $p^{\text{ex}} = 10\%$, intensity of choice $\beta = 5$, memory $m = 15$, initial probability of choosing 'run' $p_1^{\text{run}} = 0.3$.
2. Create a population of N agents, each endowed with a pool of J strategies $\{s_{j,i,1}\}_{\{i=1,...,N;\ j=1,...,J\}}$ as follow: each component $s_{j,i,1}$ takes the value '0' (i.e., 'wait') with probability $1 - p_1^{\text{run}}$, and '1' (i.e., 'run') otherwise.
3. For each agent i, select randomly with uniform probability over J a strategy $j^\star$ and select 'wait' if $s_{j^\star,i,1} = 0$ and 'run' if $s_{j^\star,i,1} = 1$.
4. Compute the total number of withdrawals e_1.
5. For each agent i, compute the number $e_{-i,1}$ of other agents that chose to withdraw as: if agent i withdraws in 1 ($s_{j^\star,i,1} = 1$), $e_{-i,1} = e_1 - 1$ and $e_{-i,1} = e_1$ otherwise.

Execution (for each period $t = 2, ..., T$):

6. *Experimentation*: For each agent i and each strategy j, flip each component $s_{j,i,t}$ from 0 to 1 or 1 to 0 with probability p^{ex}; leave unchanged otherwise.
7. *Computation of the foregone payoff* for each agent i:
   (a) for the strategy 'wait': $UWait_{i,t} = \sum_{\tau=t-1}^{t-m} \max\left(0, \frac{N-re_{-i,t}}{N-e_{-i,t}} R\right)$,
   (b) for the strategy 'run': $URun_{i,t} = \sum_{\tau=t-1}^{t-m} \min\left(r, \frac{N}{e_{-i,t}+1}\right)$.
8. For each agent i, compute the relative fitness of each strategy as:
   - $p_{i,t}^{\text{wait}} \equiv \frac{\exp(\beta UWait_{i,t})}{\exp(\beta UWait_{i,t}) + \exp(\beta URun_{i,t})}$ for strategy 'wait',
   - $p_t^{\text{run}} \equiv \frac{\exp(\beta URun_{i,t})}{\exp(\beta UWait_{i,t}) + \exp(\beta URun_{i,t})}$ for strategy 'run', where $p_t^{\text{wait}} + p_t^{\text{run}} = 1$.
9. *Reproduction* for each agent i and each strategy $j \in J$: with probability $p_{i,t}^{\text{wait}}$, set $s_{j,i,t} = 0$ (i.e., wait), otherwise set $s_{j,i,t} = 1$ and 'run'.
10. *Selection* for each agent i, randomly with uniform probability over J, draw a strategy $j^\star$ and select 'wait' if $s_{j^\star,i,t} = 0$ and 'run' if $s_{j^\star,i,t} = 1$.
11. Compute the total number of withdrawals e_t.
12. For each agent i, compute the number $e_{-i,t}$ of other agents that chose to withdraw: if agent i withdraws in t ($s_{j^\star,i,t} = 1$), $e_{-i,t} = e_t - 1$ and $e_{-i,t} = e_t$ otherwise.

Appendix C Experimental instructions

Below are the experimental instructions presented to participants. The treatment-specific information is denoted by italics and put in brackets. The numbers are calculated for the specific group sizes by the program, denoted by *XX*. All the instructions and screens were translated to Spanish as well, and participants had the opportunity to choose their language at the beginning of the experiment.

INSTRUCTIONS

Today you will participate in an experiment in economic decision making. You will be paid for your participation. There is a participation fee of 5 euros. The additional amount of cash that you earn will depend upon your decisions and the decisions of other participants. You will be earning experimental currency. At the end of the experiment, you will be paid in euros at the exchange rate of 4 experimental currency units = 1 euro.

Since your earnings depend on the decisions that you will make during the experiment, it is important to understand the instructions. Read them carefully. If you have any questions, raise your hand and an experimenter will come to your desk and provide answers.

Your Task

(You and the other ***XX participants*** *in the session, play together in a group of XX. / During the experiment, you will be matched with 9 other participants, and you will play together in a group of 10. The group composition will not change during the experiment.)* Each of you starts each period with 1 experimental euro (EE) deposited in the experimental bank. You must decide **whether to withdraw your 1 EE or wait** and leave it deposited in the bank. The bank promises to pay ***1.54 / 1.33*** EEs to each withdrawer. After the bank pays the withdrawers, the money that remains in the bank will be doubled, and the proceeds will be divided evenly among people who choose to wait. Note that if too many people desire to withdraw, the bank may not be able to fulfill the promise to pay ***1.54 / 1.33*** to each withdrawer. In that case, the bank will divide the XX EEs evenly among all withdrawers and those who choose to wait will get nothing.

Your payoff depends on your own decision and the decisions of the other XX people in the group. Specifically, how much you receive if you

make a withdrawal request or how much you earn by waiting depends on how many people in the group place withdrawal requests.
Note that you are not allowed to ask other participants what they will choose. You must **guess** what other people will do - how many of the other XX people will withdraw - and act accordingly. The graph below shows your payoff for withdrawing or waiting depending on the number of **other** withdrawers in your group. During the experiment you will see the same graph, and a calculator on your screen. The calculator returns your payoff for each action when you enter how many **other** participants hypothetically withdraw. Note that everybody earns about the same amount if exactly XX people decide to withdraw their money. Let's look at two examples:

Example 1.
Suppose 2 subjects choose to withdraw. If you choose to withdraw, your payoff is XX, and if you choose to wait, your payoff is XX.

Example 2.
Suppose XX subjects choose to withdraw. If you choose to withdraw, your payoff is XX, and if you choose to wait, your payoff is 0.

In each period you have **one** minute to make your decision. If you do not submit a decision on time, you do not earn any money for that period. Your last decision will be duplicated for the given period, and that will be taken into account for the others' earnings. (If you do not make a decision on time in the first period, your action will be randomly determined with equal probabilities.) A timer on the top left part of the screen will show you the remaining time for each period.

<u>Announcement</u>
In each period, an announcement will show up on the screen to forecast the number of withdrawers for this period.

The announcement will be either

- "The forecast is that XX or more people will choose to withdraw", or
- "The forecast is that XX or fewer people will choose to withdraw".

Everybody receives the **same** message. The announcements are randomly generated. *(There is a possibility of seeing either announcement, but the chance of seeing the same message that you saw in the previous period is* higher *than the chance of seeing a different announcement. / There is an equal chance of* *seeing either announcement in each period.)* These announcements are forecasts, which can be right or wrong. The experimenter does not know better than you how many people will choose to withdraw (or wait) in each period. The actual number of withdrawals is determined by the decisions of all participants. Your actual payoff depends only on your own choice and the choices of other participants.

Training Periods

This experimental session consists of 56 periods. The **first 6 periods** are training periods, and **do not count** towards your earnings. This is an opportunity for you to become familiar with the task you will perform during the experiment. During the training period, you **are *not* playing** with your peers from this experiment. Instead, you are playing with XX robot players, whose decisions were generated **before** the experiment. All of your peers are also playing with the **same** XX robot players. None of your decisions have an influence on the behavior of the robot players. After the 6 training periods, the formal experiment starts. There are no robots any more, and you will only play with other participants from this experimental session.

Earnings

We will pay you in cash at the end of the experiment based on the points you earned in the 50 periods. You earn 1 euro for each XX points you make plus an additional 5 euros of participation fee.

On the next screen you are asked to answer some understanding questions.

Control questions:

Before the experiment starts, please answer some questions. You can return to the instructions by clicking on the menu at the top of this page. If you need help, please raise your hand.

1. How many other participants are you playing with in the formal experimental periods?

2. In which period do you start playing with your fellow participants?

3. Do you know how your partners decide when you are making your decision?

4. Do the announcements have a direct effect on your payoffs?

    a. Yes, always.

b. No, it might only influence the decision of others that determines my payoff.

5. Suppose that all of your group-members decide to wait. What is your payoff if you wait as well?

6. Suppose that you do not make a decision on time, and your last decision was to withdraw. In this period, 8 other players decided to withdraw. What is your payoff for this period?

Post-Experimental Questionnaire:

Please answer the following questions seriously. **Your answers will help us understanding the findings of this study.** The questionnaire is anonymous. Unless otherwise specified, please answer the following questions on a five-point scale where **"1" indicates that you strongly disagree** with the statement, **"3" means neutral**, and **"5" means strongly agree.**

1. When I made my decision, I thought carefully about what the others were doing.

2. When I saw an announcement I tried to follow it.

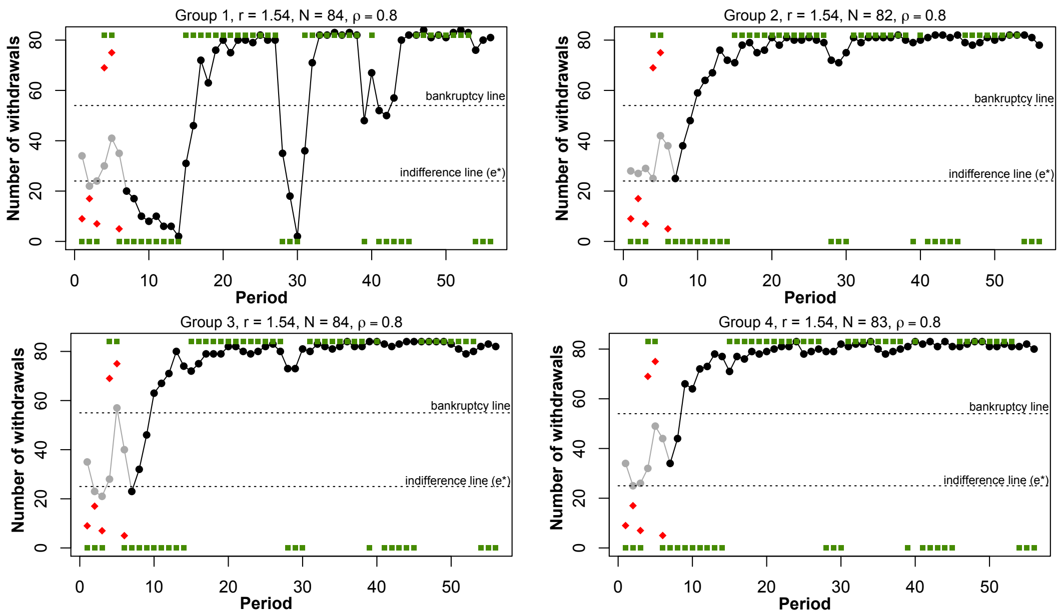
3. When I saw that the announcement changed, I did not follow immediately, but waited to see what others were doing.

4. I found safer not to follow the announcement.

5. I found it difficult to think about what others would do in a group with XX other people.

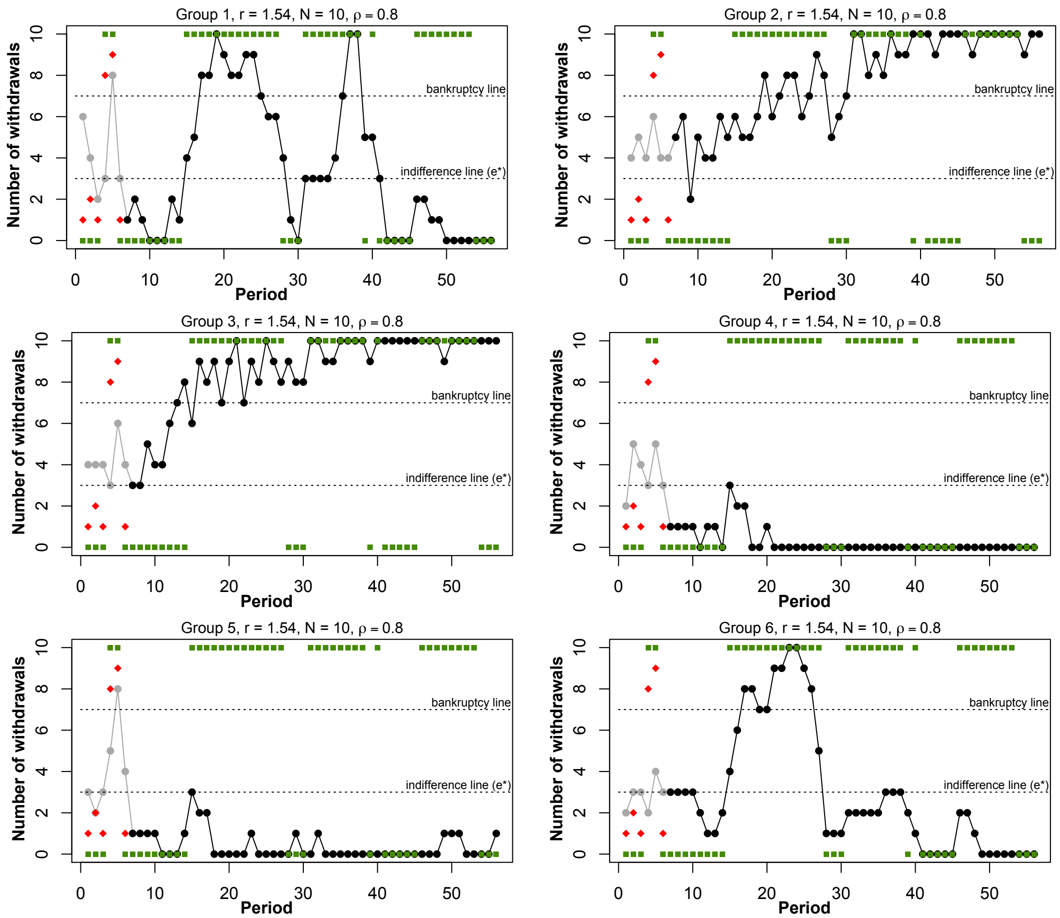
6. If you followed a specific decision rule, please explain it here.

Appendix D Additional figures



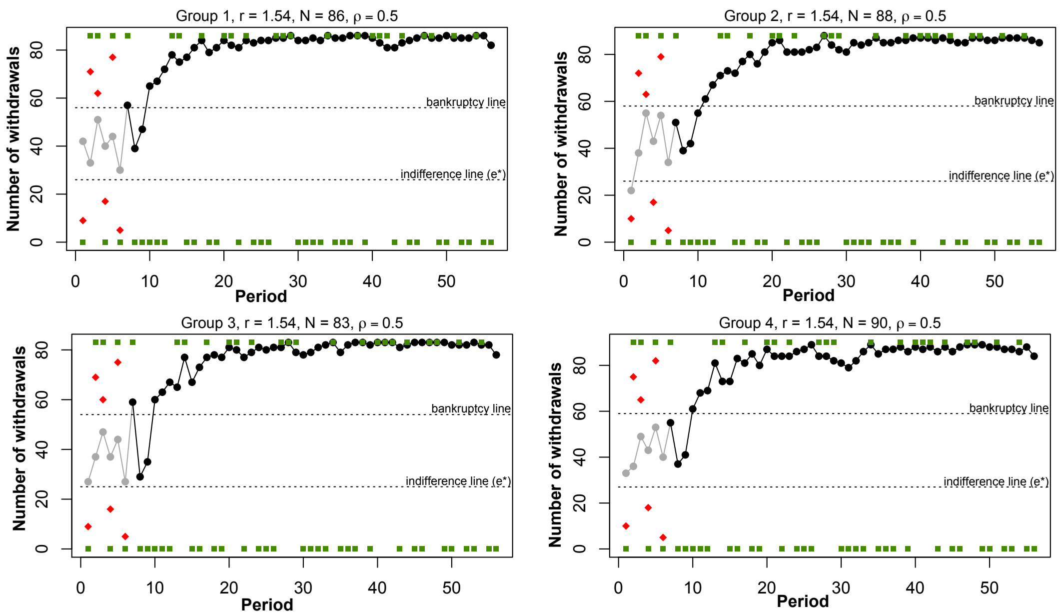
Notes: See explanation under Figure 3 and in the main text.

Figure 10: Withdrawals in large groups with $r = 1.54$ and $\rho = 0.8$



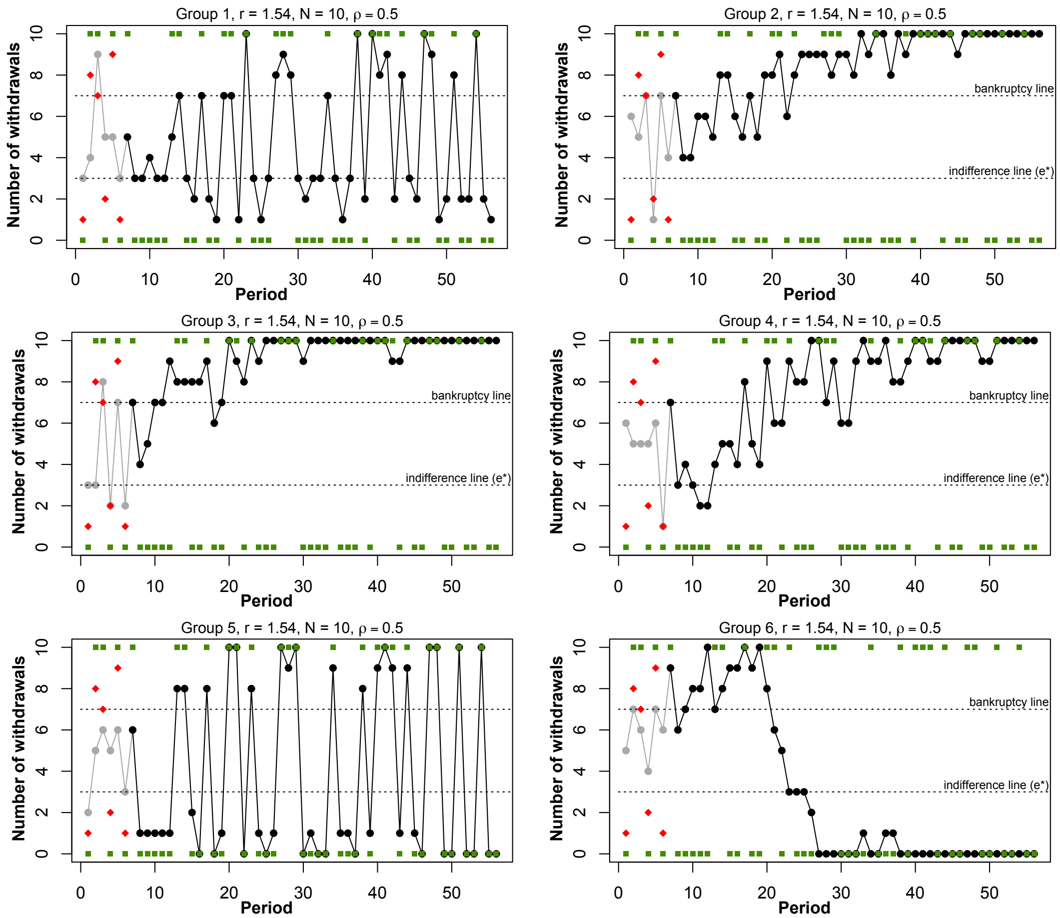
Notes: See explanation under Figure 3 and in the main text.

Figure 11: Withdrawals in small groups ($n = 10$) with $r = 1.54$ and $\rho = 0.8$



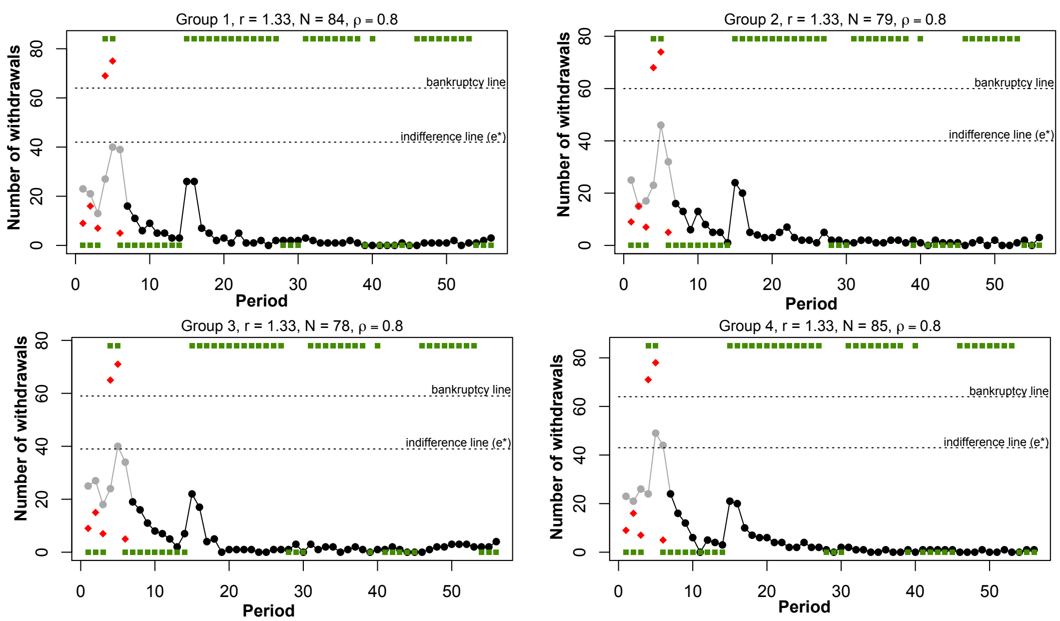
Notes: See explanation under Figure 3 and in the main text.

Figure 12: Withdrawals in large groups with $r = 1.54$ and $\rho = 0.5$ (no persistence)



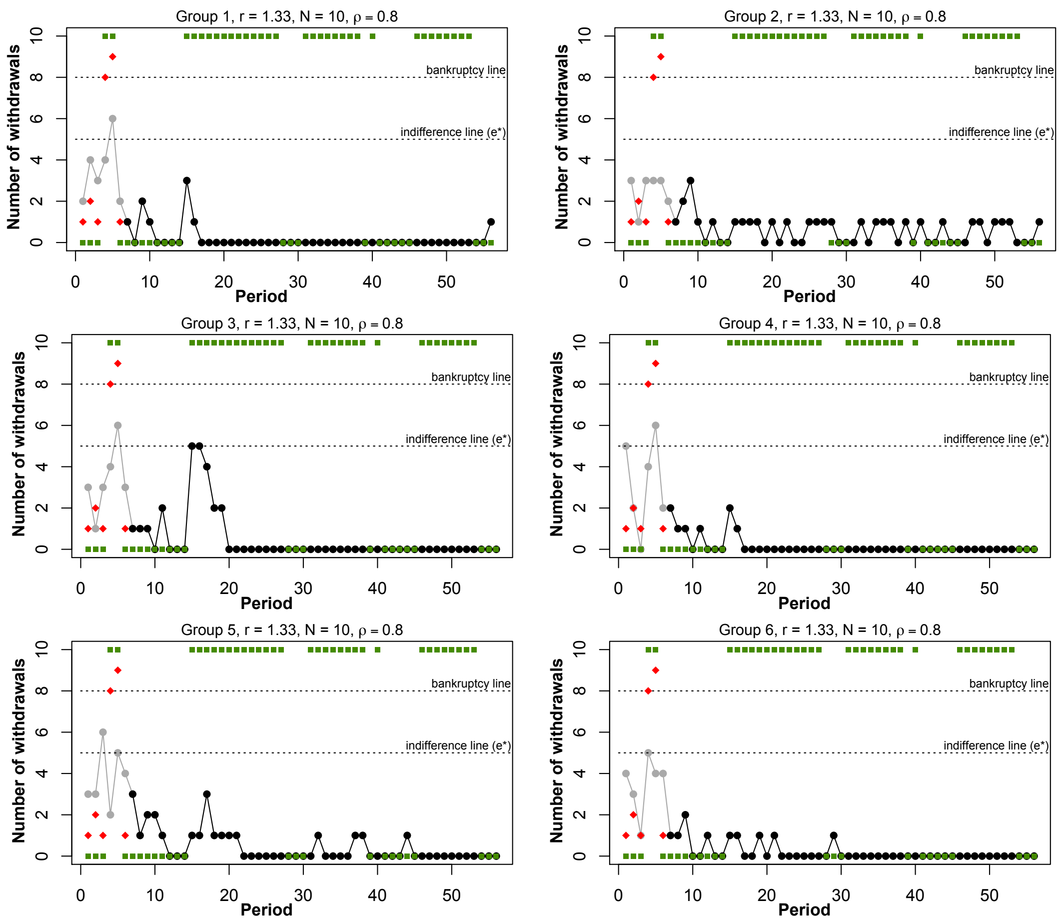
Notes: See explanation under Figure 3 and in the main text.

Figure 13: Withdrawals in small groups ($n = 10$) with $r = 1.54$ and $\rho = 0.5$ (no persistence)



Notes: See explanation under Figure 3 and in the main text.

Figure 14: Withdrawals in large groups with $r = 1.33$



Notes: See explanation under Figure 3 and in the main text.

Figure 15: Withdrawals in small groups ($n = 10$) with $r = 1.33$