Learning by Imitation in Theory, Field, and Lab*

PRELIMINARY AND INCOMPLETE. PLEASE DO NOT CITE!

Erik Mohlin† Robert Östling‡ Joseph Tao-yi Wang§

July 20, 2012

Abstract

We study how players learn to play the *l*owest *u*nique *p*ositive *i*nteger (LUPI) game using field data from a Swedish gambling company. Whoever chooses the lowest unique number wins a fixed prize. The unique symmetric equilibrium is completely mixed and is difficult to compute. Still, over the 49 days that the game was played, behavior came remarkably close to the equilibrium prediction. Due to the limited feedback, the large strategy space and the large number of players standard models of belief-based learning and reinforcement learning are unable to account for the observed learning pattern. We argue that a simple model of global proportional imitation (GPI) of previous winning numbers, and numbers that are close to winning numbers, can explain the data. We analyse the continuous time dynamic derived from the discrete time imitation process by means of stochastic approximation. It is shown that imitation of winning numbers gives rise to the replicator dynamic in LUPI, and that LUPI is strictly stable, implying convergence to the unique equilibrium from any interior initial condition. To corroborate our findings, we also analyze data from laboratory experiments. Imitation can explain the learning pattern in the laboratory data as well, but learning is much quicker in the laboratory than in the field.

JEL CLASSIFICATION: C72, C73, L83.
KEYWORDS: Learning; Imitation; Behavioral game theory; Stochastic approximation; Replicator dynamic; Stable game; Congestion game; Mixed equilibrium; Guessing game.

*We are grateful for comments from seminar audiences at Edinburgh, Lund, Oxford, and University College London. Parts of this paper were included in an earlier version of Östling et al. (2010). Erik Mohlin acknowledges financial support from the European Research Council, Grant no. 230251, Robert Östling acknowledges financial support from the Jan Wallander and Tom Hedelius Foundation, and Joseph Tao-yi Wang acknowledges support from the NSC of Taiwan.

†Department of Economics, University College London, Gower Street, London WC1E 6BT, United Kingdom. Telephone: +44 (0)20 7679 5485. Fax: +44 (0)207 916 2775. E-mail: e.mohlin@ucl.ac.uk.

‡Institute for International Economic Studies, Stockholm University, SE–106 91 Stockholm, Sweden. E-mail: robert.ostling@iies.su.se.

§Department of Economics, National Taiwan University, 21 Hsu-Chow Road, Taipei 100, Taiwan. E-mail: josephw@ntu.edu.tw.

1 Introduction

Learning by copying others is prevalent in the animal world as well as in human societies.[1] As children we imitate both parents and peers, but imitation is likely to place an important role also as adults when facing new problems in which the optimal course of action is not obvious (e.g. finding our way in a new country, buying new complex consumer products, learning which animals and plants that are dangerous). In a recent tournament organized by evolutionary biologists, learning algorithms heavily based on imitation proved to be most successful in solving a complex and dynamically changing multiarmed bandit problem (Rendell, Boyd, Cownden, Enquist, Eriksson, Feldman, Fogarty, Ghirlanda, Lilicrap and Laland, 2010).[2] In many economically relevant contexts information about successful strategies and successful players is disproportionally available through direct observation, mass media, or information services provided by various consulting firms. For example, there is usually plenty of information available about the relatively few succesful entrepreneurs, whereas there is much less information about the majority of entrepreneurs that failed (or did not even get started). This suggests that we primarily learn from others' success rather than their mistakes.

In this paper, we study how players learn to play a novel game in the field. The game has simple and clear rules, yet the theoretical prediction for the game is a complicated unique mixed strategy equilibrium which is very difficult to compute. Most likely no player could figure out the equilibrium and therefore had to resort to some other heuristic to guide their behavior. The players receive very limited feedback, thereby restricting the set of applicable learning rules considerably. We show that a simple imitate-the-best learning model can parsimoniously rationalize some striking patterns in the data; in particular how players so quickly can come close to equilibrium play. To derive theoretical predictions we have to make some simplifications. To corroborate the model, and our interpretation of the data, we therefore analyze data from laboratory experiments that implemented the theoretical assumptions of the of the field game much more closely. We obtain very similar results for the field and lab.

[1] See for example Laland (2001) for a discussion of imitation in the animal world. We are not aware of similar surveys for human behavior, but see Section III in Armstrong and Huck (2010) for a survey of some of the relevant research in economics.

[2] Copying the successful behavior of others appears innocuous when solving individual decision problems, but it is less clear how useful imitation is as a learning rule in strategic settings. Duersch, Oechssler and Schipper (2010) have shown that a simple "imitate-the-best" learning rule cannot be beaten by any other type of learning rule (including rational and forward-looking behavior) in most symmetric two-player games. In other games, such as rock-papers-scissors, however, imitating players can easily be exploited by the opponent. LUPI has a structure that is similar to that of rock-papers-scissors. Still, the large number of strategies and payoffs in practice makes it very difficult to exploit a population of imitators in LUPI.

Studying how people learn to play games using field data is a challenging task. Game-theoretic predictions often depend critically on the rules and payoffs of the game which are often hard to observe in the field. In addition, to study learning we would ideally like to follow inexperienced players that gradually gain experience of a novel game. In this paper we exploit data from a Swedish lottery game which we think comes relatively close to these ideal conditions.

In this lottery, players simultaneously choose positive integers from 1 to K. The winner is the player who chooses the lowest number that nobody else picked. This game is called the LUPI game, because the *l*owest *u*nique *p*ositive *i*nteger wins. In the field $K = 99999$ and about 50,000 numbers were played in LUPI (on average) on 49 consecutive days. Östling, Wang, Chou and Camerer (2011) have previously used the data from this game to study how players play in the first round and whether play converges to equilibrium. They concluded that play did not converge exactly to equilibrium in 49 days, but that behavior quickly comes surprisingly close to the equilibrium prediction.

Explaining rapid convergence in the LUPI game is demanding for traditional models of learning. In particular it is difficult to explain how players can learn to play close to equilibrium in only 49 rounds (and even quicker in the laboratory). In the lab, players only received information about the winning number and their own payoff. In the field it was possible to obtain more information with some effort, but laboratory results confirm that this was not essential for the learning process. The information about the previous winner allows players to infer the payoffs of all strategies, but they cannot infer the behavior of their opponents. Learning based on reinforcement of chosen strategies is far too slow because players win very rarely (and hence, their strategies are rarely reinforced). The leading example of belief-based learning, fictitious play (see for example, Fudenberg and Levine (1998)), is also unable to explain learning in this context. Fictitious play assumes that players best respond to the average of the past empirical distributions, but in our case players do not have such information (at least not in the lab). Hybrid models like EWA (Camerer and Ho, 1999, Ho, Camerer and Chong, 2007) require the same information as fictitious play and therefore do not fit any better in this information environment. The myopic best response dynamic, originating with Cournot, postulates that players best respond to the behavior in the previous period. This is something that players could possibly do in the field (but not the lab) since a website provided information about the lowest unchosen number in the previous round. Still this explanation will not work directly in practice even in the field, since the lowest unchosen number was typically *above* the winning number (in 43 of 49 days).[3] There is one form of belief-based learning that

[3] Also, a complicated collusive strategy of choosing both the last winning number (to knock out the winner) *and the lowest unchosen number could work.*

could possibly be used by players with our limited feedback: Players enter the game with a prior about what strategy opponents' use, and after each round they update their belief in response to information about the winning number. In Appendix C we discuss such a model of belief-based learning and argue that, even under a charitable assumption about the prior players use, it is difficult to capture basic features of the data

Explaining learning in the LUPI game therefore requires a model that (1) does not involve best responses to the full empirical distribution, (2) does not exclusively rely on information about a player's own payoff, and (3) is only based on information about the structure of the game, a player's own payoff, and the winning numbers. An appealing alternative which satisfies these three criteria is a simple imitate-the-best learning model in which all players imitate a window of numbers around the previous winning number.[4] Many models of imitation assume pair-wise imitation according to which an individual who considers revising her strategy samples one other individual from the population. In contrast, we assume that each revising individual observes the payoffs of all other individuals. Moreover we assume that players accumulate propensities to play a particular action in response to how often that action, or similar actions, has won in the past. The propensities are transformed into a mixed strategy via a simple proportional rule. We call this *global (cumulative) proportional imitation (GPI).* This simple model can parsimoniously rationalize some striking patterns in the data, in particular how players so quickly can come close to equilibrium play by reacting only to winning numbers.

We analyse the discrete time stochastic GPI process and show that it behaves approximately like the replicator dynamic in LUPI. We show that LUPI is a strictly stable game (Hofbauer and Sandholm, 2009) which implies that learning according to GPI almost surely will converge to the unique Nash equilibrium of the LUPI game.[5] Furthermore we provide two generalizations of the GPI process, one whose deterministic approximation coincides with the replicator dynamic in all games, and one which does not.

Although the field data is in many ways suitable for testing game-theoretical predictions, not all details of the field game are captured by our theoretical model. Moreover, we have not gained access to panel data, so we cannot follow individual behavior over time.

[4] Since players' payoffs are symmetric, imitating winning numbers is psychologically similar to counterfactual "fictive" reinforcement of unchosen numbers. In fact, in explaining learning in weak-link games, Roth (1995) note that reinforcement according to chosen strategies fits very poorly, so he substitutes a different a model based on imitating the most successful players (pp. 38–39). Similarly, Roth and Erev (1995) model "public announcements" in proposer competition ultimatum games ("market games") as reinforcing the winning bid (p.191). Thus, in a sense our model of imitation is generalization of reinforcement learning, where strategies are reinforced with what other players earned as well.

[5] In stable stable games the set of symmetric Nash equilibria is convex and globally attracting under the Braun-von Neumann-Nash, the logit best response dynamic, the best response dynamics, and "nearby" related dynamics. In strictly stable games there is a unique Nash equilibrium. The replicator dynamic converges to the equilibrium from any interior initial condition in stable games. For more on these concepts see Hofbauer and Sandholm (2009) and Sandholm (2011), chapter 7.

We therefore also analyze data from laboratory experiments conducted by Östling et al. (2011) that are closer to our theoretical assumptions and allows us to track individual decisions over time. The laboratory also allows us to completely control the feedback that players receive, which is important in order to distinguish different learning models.

There is a long-standing discussion about the interpretation of mixed strategy Nash equilibrium. In one interpretation, players deliberately randomize according to the mixed strategy equilibrium probabilities. In laboratory experiments, aggregate behavior is often in line with the mixed strategy equilibrium prediction, but players typically fail to randomize at the individual level (Camerer, 2003, Chapter 3). Mixed equilibria need not be understood as deliberate randomization of rational players, however, but can also be seen as the end state of an evolutionary or learning process. John Nash proposed this "mass-action" interpretation of equilibrium already in his dissertation (Nash, 1950). The mass-action interpretation of equilibrium has inspired a large theoretical literature about evolution and learning in games, as well as a smaller experimental literature.[6] This paper contributes to the previous experimental literature by studying how a large population of players learn to play a mixed equilibrium in the field. In particular, the large number of players gives enough statistical power to study the rate of learning across the time series in a game in which the structure does not vary, which most other field studies cannot do. For example, several studies have used field data from tennis and soccer to test mixed-strategy equilibrium predictions (Walker and Wooders, 2001, Chiappori, Levitt and Groseclose, 2002, Palacios-Huerta, 2003 and Hsu, Huang and Tang, 2007). These studies use highly experienced players and the studies on soccer pool data across substantial spans of time to test the mixed equilibrium prediction powerfully. They do not study how players learn to play a mixed equilibrium within their samples.[7]

There is a large theoretical literature on imitation and the resulting evolutionary dynamics. Binmore, Samuelson and Vaughan (1995) and Binmore and Samuelson (1997) study stochastic evolutionary processes which can be interpreted as noisy pair-wise imi-

[6] See Weibull (1995), Fudenberg and Levine (1998), Young (2004), and Sandholm (2011) for surveys of the theoretical literature and Camerer (2003, Chapter 6) and Erev and Haruvy (2008) for overviews of the experimental literature.

[7] Chiappori et al. (2002) provide some suggestive evidence about learning by noting that among the kickers with the most experience in their sample (those with eight or more kicks) only one out of nine fails a randomness test at the 10% level. Hsu et al. (2007) also find suggestive evidence that junior tennis players behave somewhat differently from adult players. However, these are crude tests for learning effects compared to our data, which compare a much larger sample of choices with day-by-day comparisons. There is also a field study of randomization in gambling choices that are not strategic, and learning is not measured (Sundali and Croson, 2006). Ockenfels and Roth (2004) discuss an interesting natural experiment measuring learning about prices for a surprising new product. They studied prices for retailer-copied "Iraq most wanted cards" originally produced by the US Department of Defense to help soldiers identify high-value targets during the Iraq war. Retailers quickly copied the cards and offered decks for $5.95; but retailer decks were also traded at a much higher "buy-it-now" (BIN) prices on eBay. They found that BIN prices converged to the retail price in about 30 days.

tation based on dissatisfaction and aspiration levels. The noise is exogenous and does not vanish in the large population limit, and the process can be approximated by the replicator dynamic over finite time horizons. Björnerstedt and Weibull (1996) and Weibull (1995, section 4.4), study a population of infinitely lived agents who revise strategies asynchronously at a Poisson rate by drawing one other individual from the population to potentially imitate. The authors assume that the population is large and hence study deterministic dynamics directly. In particular, the population process is described by the replicator dynamic if (1) each revising individual draws her potential role model randomly and the revision rate is linearly decreasing in the payoff to the current strategy, (2) each revising individual receives a noisy signal uniformly distributed about a randomly drawn peer's payoff and imitates if and only if the other individual earns more, or, (3) each revising individual draws another individual with probability linearly increasing in the potential role model's payoffs. In contrast, our model has synchronous rather than asynchronous revisions (in our model everyone revises at the same time) and global rather than pair-wise imitation. Furthermore, we study a stochastic approximation process with decreasing step size where noise vanishes asymptotically, rather than increasing the population size.

Schlag (1998) studies pairwise imitation rules in the context of multi-armed bandits. A revising (new-born) individual observes one other individual's payoff and action from the previous period, and compares it with its own (or predecessor's) payoff in the previous period. Schlag shows that the optimal policy is to imitate with a probability that is proportional to the positive part of the payoff difference between the other individual and oneself (or predecessor). Moreover, the behavior of a population that follows the optimal rule can be approximated by the replicator dynamic. Schlag (1999) extends the analysis to allow sampling of two, rather than one, individuals' behavior and payoffs in the previous period. In contrast, our model is based on global rather than pairwise imitation, and in our model imitation is cumulative in that it responds to winners in all previous period.

Vega-Redondo (1997) provides a finite population stochastic evolutionary model of a Cournot market where firms adapt their output choices synchronously in discrete time. Adaptation takes the form of imitation of one of the strategies that earned the highest payoff in the previous period. In addition exogenous experimentation occurs with probability ε. In the small noise limit the unique stochastically stable state corresponds to the competitive equilibrium outcome.[8] There is a smaller experimental literature, which

[8] Like in our model there is synchronous revision and imitation of winners, but our model is cumulative. We use stochastic approximation techniques to derive and study a deterministic process whereas Vega-Redondo (1997) analyses the limiting distribution of the stochastic process. In the terminology of Binmore and Samuelson (1994) we study the medium and long run while Vega-Redondo (1997) studies the ultra-long run.

has focused on learning by imitation in Cournot oligopolies, e.g. Apesteguia, Huck and Oechssler (2007), who compare the imitation procedures studied by Schlag and Vega-Redondo.

Östling et al. (2011) analyze the LUPI game using the same data as in this paper. They derive theoretical equilibrium predictions using the theory of Poisson games which assumes that the number of players is Poisson distributed and test whether behavior in the field and lab converges to equilibrium. They do not study theoretical convergence and stability properties, and they do not analyse empiricaly how players learn to play close to the equilibrium. One minor contribution of the current paper is that we extend the analysis of Östling et al. (2011) to compute the fixed N-player symmetric equilibrium of the LUPI game for a very large number of players.

There are some other papers that have studied LUPI's close market analogue the lowest unique bid auction (LUBA). Gallice (2009), Raviv and Virag (2009) and Houba, van der Laan and Veldhuizen (2010) analyze LUBA theoretically, whereas Eichberger and Vinogradov (2009) and Rapoport, Otsubo, Kim and Stein (2009) also study lowest unique bid auctions empirically. Radicchi, Baronchelli and Amaral (2012) study a special version of the LUBA in which players are allowed to adjust their bids for a period of time before the winner is determined. Pigolotti, Bernhardsson, Juul, Galster and Vivo (2012) build on the analysis of Östling et al. (2011) to provide a partial analysis of the Poisson Nash equilibrium of LUBA. None of these papers characterize or prove uniqueness of the Nash equilibrium, and they do not study stability properties. Also, none of these papers study how players learn to play equilibrium. In a companion paper, Mohlin, Östling and Wang (2012), we prove that the LUBA is a stable game with a unique equilibrium, which is attracting for a large class of evolutionary/learning dynamics. The unique equilibrium has a convex support which includes the lowest bid. We provide an algorithm for computing all such candidate equilibria, and for determining the unique equilibrium. Because of the close similarity between LUPI and LUBA we expect the same kind of learning process to be relevant in LUBA as well.

Although LUPI was not designed to be an exact model of a particular economic game, it does combine some strategic features of interesting naturally-occurring games, such as choices of traffic routes and research topics, buyers and sellers choosing among multiple markets,[9] or patent races (in which being first matters). Moreover, an interesting theoretical aspect of the LUPI game, which may explain why imitation of previous winning numbers is so effective, is that probability matching, or all players picking each number with probability that is exactly proportional to the number of times that number is played,

[9]Note, however, that LUPI is not a congestion game as defined by Rosenthal (1973) since the payoff from choosing a particular number does not only depend on how many other players that picked that number, but also on how many that picked lower numbers.

happens to be an equilibrium. Both humans and animals have been shown to engage in probability matching also in settings where this is not optimal, but in the LUPI game probability matching happens to be the rational thing to do in equilibrium.[10]

2 The LUPI Game

In the LUPI game, $N \geq 3$ players choose integers from 1 to K simultaneously, and the lowest unique number wins. The winner earns a payoff of 1, while all others earn 0. If there is no lowest unique number then there is no winner, and everyone earns zero.

There is a population of agents. In every period a number of players is drawn from the population to play the game. The number of players is either fixed or Poisson distributed. The fixed N case is discussed in detail in Appendix A. In the main text we focus on the Poisson case. Östling et al. (2011) show that the LUPI game with a Poisson distributed number of players has a unique (symmetric) equilibrium, which is completely mixed.[11] They characterize the equilibrium and compute it. The Poisson equilibrium with 53783 players is illustrated in Figure 1 As expected, higher numbers have lower probability. Less intuitively, the probability is a slowly decreasing concave function up until a threshold (at roughly number 5500) where the probability decreases dramatically and the function becomes convex. The probability remains strictly positive over the whole strategy space.

We will use the following notation: The pure strategy space is $S_i = \{1, 2, ..., K\}$, and the mixed strategy space is the $(K-1)$-dimensional simplex Δ_i. Let $k^*(s)$ denote the winning number under strategy profile $s \in \times_{i=1,..,N} S_i$,

$$k^*(s) = \min_{j \in \{k \in \{1,...,N\}: s_k \neq s_l, \forall l \neq k\}} s_j.$$

The payoff to a player playing strategy $k \in S_i$ under strategy profile s is

$$u_k(s) = \begin{cases} 1 & \text{if } k = k^*(s), \\ 0 & \text{otherwise.} \end{cases}$$

Let p denote the population average strategy,

Let p denote the population average strategy, i.e. p_i is the probability that a randomly chosen player will pick pure strategy i. Randomly drawing players from the population,

[10] See Vulkan (2000) for a survey of the evidence regarding probability matching.

[11] Note that when there is uncertainty about the number of players, one cannot define asymmetric equilibrium based on player identification, since players do not know who will participate. Instead, one can at most define heterogeneous equilibrium where different groups of players play according to their "types". However, as shown below, any heterogeneous equilibrium is outcome equivalent to a homogeneous equilibrium with a representative type playing the population average strategy.

who then draw actions according to their mixed strategies, is equivalent to drawing actions directly according to the population average strategy p. Randomly drawing players from the population, who then draw actions according to their mixed strategies, is equivalent to drawing actions directly according to the population average strategy p.[12] Let $X(i)$ be the *total* number of players who choose strategy i, viewed as a random variable, not conditional on the realized number of players. For a given fixed player who is drawn to play the game, let $Y(i)$ be the number of *other* players who pick i, again viewed as a random variable, not conditional on the realized number of players. In the Poisson game $\Pr(Y(i)) = \Pr(X(i))$ due to the "environmental equivalence"-property of Poisson games (Myerson, 1998), but in the fixed N game this is not the case. The expected payoff to a player putting all probability on strategy k given the population average strategy p is

$$\begin{aligned}
\pi_k(p) &= \Pr(Y(k) = 0) \prod_{i=1}^{k-1} \Pr(Y(i) \neq 1) \\
&= \Pr(X(k) = 0) \prod_{i=1}^{k-1} \Pr(X(i) \neq 1) \\
&= e^{-np_k} \prod_{i=1}^{k-1} \left(1 - np_i e^{-np_i}\right).
\end{aligned}$$

The simplicity of this expression is due to the special properties of Poisson games, in particular the 'independent actions property'. Let $\pi(p) = (\pi_1(p), ..., \pi_K(p))'$ be the column vector of payoffs at state p. The probability that number k is the winning number is

$$\begin{aligned}
\Pr(k = k^*(s) | p) &= \Pr(X(k) = 1) \prod_{i=1}^{k-1} \Pr(X(i) \neq 1) \\
&= np_k e^{-np_k} \prod_{i=1}^{k-1} \left(1 - np_i e^{-np_i}\right) \\
&= np_k \pi_k(p).
\end{aligned}$$

2.1 Strict Stability

A game is strictly stable (Hofbauer and Sandholm (2009)) if

$$(p - q)' (\pi(p) - \pi(q)) < 0, \tag{1}$$

[12] To see this, let each individual in the population be a different type $\theta \in \Theta$, putting weight $\sigma(k|\theta)$ on pure strategy k, so that the population share of each type t is $1/|\Theta|$, as defined on pp. 4–5 in Östling et al. (2011). The number of players (across all types) who choose action k is Poisson with mean $n \sum_{\theta \in \Theta} (1/|\Theta|) \sigma(k|\theta) = np_k$.

for all $p, q \in \Delta$, such that $p \neq q$. The following characterization is useful: A game with a C^1 payoff function is strictly stable if and only if its associated Jacobian, denoted $D\pi(p)$, is negative definite with respect to the tangent space. With K strategies the tangent space is $\mathbb{R}_0^K = \{v \in \mathbb{R}^K : \sum_i v_i = 0\}$ so the LUPI game is strictly stable if and only if $v' D\pi(p) v < 0$ for all $v \in \mathbb{R}_0^K$, $v \neq \mathbf{0}$, and all $p \in \Delta$. Similarly p^* is (locally) asymptotically stable if and only if $v' D\pi(p^*) v < 0$ for all $v \in \mathbb{R}_0^K$, $v \neq \mathbf{0}$, and all p in a neighborhood of p^*.

In the Poisson case we have

$$\frac{\partial \pi_k(p)}{\partial p_j} = \begin{cases} n\frac{(np_j-1)e^{-np_j}}{\left(1-np_je^{-np_j}\right)} \prod_{i\in\{1,\ldots,k-1\}} \left(1-np_ie^{-np_i}\right)e^{-np_k} & \text{if } j<k \\ -n \prod_{i\in\{1,\ldots,k-1\}} \left(1-np_ie^{-np_i}\right)e^{-np_k} & \text{if } j=k \\ 0 & \text{if } j>k \end{cases},$$

so the $(n \times n)$ Jacobian can be written

$$D\pi(p) = n \begin{pmatrix} -\pi_1 & 0 & 0 & \cdots & \cdots & 0 \\ z_1\pi_2 & -\pi_2 & 0 & \cdots & \cdots & 0 \\ z_1\pi_3 & z_2\pi_3 & -\pi_3 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ z_1\pi_K & z_2\pi_K & z_3\pi_K & \cdots & \cdots & -\pi_K \end{pmatrix},$$

where

$$z_i = \frac{np_i - 1}{e^{np_i} - np_i}.$$

In a triangular matrix the eigenvalues correspond to the diagonal entries. Hence for interior p all eigenvalues $\lambda_1, \ldots\lambda_K$ of $D\pi(p)$ are strictly negative. We suspect that the following result is already known, but since we have not found it in the literature we provide a proof of it.

Lemma 1 *A real square matrix with distinct eigenvalues is negative definite if and only if all its eigenvalues are negative.*

Proof. A $K \times K$ matrix A is diagonalizable if its eigenvalues are distinct: There is a matrix P such that $P^{-1}AP = P'AP = \Lambda$, where $\Lambda = diag(\lambda_1, \ldots\lambda_K)$. The columns of P are corresponding eigenvectors $w_1 \ldots w_K$, which are linearly independent. Since A and the eigenvalues are real P can be taken to be real. Moreover, by linear independence the eigenvectors are mutually orthogonal, so if all eigenvectors are normalized to length one then we have $w_i \cdot w_j = 1$ for $i = j$ and $w_i \cdot w_j = 0$ for $i \neq j$. It follows that P is an orthogonal matrix; $P'P = I = P^{-1}P$. Since the inverse is unique $P^{-1} = P'$. Thus $P^{-1}AP = P'AP = \Lambda$. For any $v \in \mathbb{R}^K$ let $y = P'v = P^{-1}v$, implying $v = Py$. It follows that

$$v'Av = y'(P'AP)\, y = y_1^2\lambda_1 + ... + y_K^2\lambda_K.$$

Since the columns of P are linearly independent $v \neq \mathbf{0}$ implies $y \neq \mathbf{0}$. Hence $v'Av$ is negative for all $v \neq \mathbf{0}$ if and only if all eigenvalues are negative. Q.E.D.

If a matrix is negative definite w.r.t. $\mathbb{R}^K$ then it is also negative definite w.r.t. the tangent space $\mathbb{R}_0^K$. Hence $D\pi(p)$ is negative definite w.r.t. the tangent space at any state p where all strategies earn distinct payoffs. This is the case for all but a measure zero set of states:

Lemma 2 *The Jacobian $D\pi(p)$ is negative definite w.r.t. the tangent space almost everywhere.*

Proof. Suppose $l > k$. It can be verified that the condition $\pi_k(p) = \pi_l(p)$ is equivalent to

$$\frac{e^{-np_k}}{(1 - np_k e^{-np_k})} = \prod_{i=k+1}^{l-1} \left(1 - np_i e^{-np_i}\right) e^{-np_l}.$$

The right hand side is strictly monotonic in p_i so for any choice of $p_{k+1}, ..., p_l$ this equation determines the choice of p_k that yields $\pi_k(p) = \pi_l(p)$. Hence the set of states at which two strategies earn the same has measure zero in Δ. Since there are finitely many strategies the set of states at which two or more strategies earn the same also has measure zero. Q.E.D.

We can use this to prove, by a continuity argument, that the Jacobian must in fact be negative definite everywhere on Δ.

Lemma 3 *The Jacobian $D\pi(p)$ is negative definite w.r.t. the tangent space for all $p \in \Delta$.*

Proof. Consider a point $p^0 \neq p^*$ at which two or more actions earn the same payoffs. Let $B_\delta(p^0)$ denote the δ-radius closed ball around p^0. Since all payoffs are strictly positive there is a δ and an ε such that for any $p \in B_\delta(p^0) \cap \Delta$, $\min_i \pi_i(p) > \varepsilon$. If all payoffs are distinct at a point p then the Jacobian can be diagonalized by an orthogonal matrix, as described in the proof of lemma 1. Moreover, since P spans $\mathbb{R}^K$ we are able to choose v in a way that guarantees

$$v'Av = y'(P'AP)\, y = y_1^2\lambda_1 + ... + y_K^2\lambda_K < -\varepsilon.$$

Hence, for any $\eta \leq \delta$ there is some $p \in B_\eta(p^0) \cap \Delta$ such that $v' D\pi(p) v < -\varepsilon$. Since $v' D\pi(p) v$ is continuous in p we must have $v' D\pi(p^0) v < 0$. Q.E.D.

From the above lemmata we have the following result, which is essential for our subsequent proof that learning by imitation converges in LUPI.

Proposition 1 *The Poisson LUPI game is strictly stable.*

2.2 Probability Matching

We prove a property of equilibrium that may provide an intuition for why imitation of winners leads to equilibrium in LUPI. Let w_k be the probability that number k wins the game. From above we know $w_k = np_k\pi_k$. Since it is always possible that no number is chosen uniquely, the w_k's will not sum up to one, i.e. $\sum w_k < 1$. Note that the payoff π_k is the probability that one player wins by playing k while all other players play according to the mixed strategy p.

Proposition 2 *Consider the Poisson LUPI game and suppose that p has full support. There is probability matching, $p_k = w_k / \sum_j w_j$ for all k, if and only if p is the symmetric Nash equilibrium.*

Proof. Suppose that p is the symmetric Nash equilibrium. Since p is an equilibrium, all $\pi_k = \pi^*$ for all k, so we have

$$w_k = np_k\pi^*. \tag{2}$$

Summing both sides of (2) over k gives

$$\sum w_k = n\pi^* \sum p_k = n\pi^*.$$

Dividing the LHS of (2) with $\sum w_k$ and the RHS with $n\pi^*$ gives $p_k = w_k / \sum w_k$.

To prove the other direction, suppose p is a mixed strategy with full support that satisfies $p_k = w_k / \sum_j w_j$. Since $w_k = np_k\pi_k$ we have

$$p_k = \frac{np_k\pi_k}{\sum_j w_j},$$

or equivalently $\pi_k = \sum_j w_j / n$. Since the RHS is the same for all k it must be a mixed strategy equilibrium. Q.E.D.

This result suggests that players might converge to equilibrium by simply choosing numbers in proportion to how often those numbers have won in the past.

2.3 Feedback

All players in the population are informed about the winner in each period. In the lab, as well as in our theoretical model, this is the only information players receive. How might the players adjust their behavior in response to such limited feedback? We will assume that players learn by imitation and adjust their choices in the direction the previous winning number. Figure 2 provides some evidence that this is what happens empirically. Standard reinforcement learning assumes that players reinforce their chosen strategies in proportion to the payoff they have earned. Since most players never win such a learning model would predict almost no changes in behavior over the 49 days. It is also possible that players follow some kind of belief-based learning procedure. However, prominent examples of belief-based learning are unable to explain the learning in this context. Fictitious play assumes that the players best respond to the average of the past empirical distributions, but in our case players do not have such information. The myopic best response dynamic, originating with Cournot, postulates that the players best respond to the behavior in the previous period. The required information is absent in the lab.[13] There is one form of belief-based learning that could possibly be used by players with our limited feedback: Players enter the game with a prior about what strategy opponents' use, and after each round they update their belief in response to information about the winning number. We study such a model of belief learning in Appendix C.

3 Learning Theory

3.1 Global Proportional Imitation (GPI)

A learning procedure can be described by (1) an *updating rule* that specifies how propensities for different actions are modified in response to experience, and (2) a *choice rule* that specifies how propensities for actions are transformed into actual choices. The updating rule must respect the feedback that players receive. In each period $t \in \mathbb{N}$, n individuals from a population are randomly drawn to play an n-player game (n can be fixed or variable). After each match all individuals receive the same feedback. In this respect our imitation model utilizes global information. We consider two different assumptions regarding the feedback.

Information assumption A: For each action, all individuals are able to infer the payoff earned by those who played that action in the preceding round of the game.

[13] This is something that players could possibly do in the field since a website provided information about the lowest unchosen number in the previous round. Still this explanation will not work in practice even in the field, since the lowest unchosen number was typically *above* the winning number (in 43 of 49 days).

Information assumption B: For each action, all individuals are able to infer the payoff earned by those who played that action in the preceding round of the game, and are able to infer the number of players who played it.

In the field and lab LUPI games only assumption A is satisfied. Individuals receive information about the winning number and are therefore able to infer the payoff of those who played a particular action in the preceding round of the game (1 for the winning number and 0 for all other numbers).

The two information assumptions A and B lead to two different imitation processes:

Global proportional imitation in proportion to payoffs of chosen strategies (GPI-A).

Global proportional imitation in proportion to payoffs of chosen strategies and number of players choosing these strategies (GPI-B).

The former imitation process (GPI-A) may be relevant in settings where players receive information about the payoffs earned from all the different strategies that where chosen by other players, but receive no information about strategies that were not chosen by anyone. This is the model we will use to explain behavior in LUPI. In a sense such a model of imitation is generalization of reinforcement learning, where strategies are reinforced with what other players earned as well. The latter imitation process (GPI-B) may be relevant in environments where players receive information about both payoffs earned and the number of players choosing different pure strategies.

3.1.1 Updating rule

Attractions: Let $A_k(t)$ denote the attraction of strategy k in period t. Attractions are updated according to

$$A_k(t) = A_k(t-1) + r_k(t). \tag{3}$$

where $r_k(t)$ is the reinforcement of attraction k in period t.[14] Strictly positive initial attractors $\{A_i(0)\}_{i=1}^{K}$ are given exogenously. Let

$$\rho_k(p) = \mathbb{E}[r_k(s)].$$

Reinforcements: For our analysis of learning by imitation in the LUPI game we assume that each number is reinforced by the payoff earned by those who picked that number. Thus a winning number is reinforced by one and all other numbers are reinforced by zero. For technical reasons we also assume that all numbers are reinforced by a small positive constant $c \in \mathbb{R}_+$. Recall that $u_k(t)$ is the payoff to a player who plays k under strategy

[14] More complicated variants of this function are possible, e.g., weighting lagged attractions by a "forgetting" or decay factor. However, it is difficult to identify parameters in these more complicated functions and we have therefore chosen to keep the model as simple as possible.

profile $s(t)$. Formally we may define

$$r_k^A(t) = u_k(t) + c. \tag{4}$$

Such reinforcements can be calculated under information assumption A. Alternatively, under the slightly more demanding information assumption B, let $m_k(t)$ be the number of players picking k at time t, and define

$$r_k^B(t) = m_k(t)(u_k(t) + c). \tag{5}$$

In order to apply the stochastic approximation techniques below we need reinforcements to be strictly positive. We do this by adding the constant c. It turns out that with specification (4) the constant will give rise to a noise term in the resulting ODE, like the one used by Gale, Binmore and Samuelson (1995). In contrast, with specification (5) the constant drops out of the resulting ODE. In the former case we may still ignore the constant by letting $c \to 0$ after the stochastic approximation.

3.1.2 Choice rule

Consider an individual who uses the mixed strategy $\sigma(t)$, that puts weight $\sigma_k(t)$ on strategy k. According to the *proportional rule*, attractions are transformed into choice by the following formula

$$\sigma_k(t+1) = \frac{A_k(t)}{\sum_{j=1}^{K} A_j(t)}. \tag{6}$$

The power function

$$\sigma_k(t+1) = \frac{A_k(t)^\lambda}{\sum_{j=1}^{K} A_j(t)^\lambda}, \tag{7}$$

is a generalization of the proportional rule. Note that $\lambda = 0$ means uniform randomization and $\lambda \to \infty$ means playing only the strategy with the highest attraction. We derive analytical results for the case of $\lambda = 1$. We will refer to the combination of the choice rule (6) or (7) together with updating rule (3) as *global proportional imitation (GPI).* With reinforcements (4) we obtain *GPI-A* and with reinforcements (5) we have *GPI-B.*

3.2 Stochastic Approximation of GPI

We derive analytical results for GPI under the assumption $\lambda = 1$. Consider one individual in a large population of players. The player's strategy is σ and the population average strategy is p. Since we consider a large population the population average strategy and the average strategy faced by a single player can be treated as identical. We have:

$$\sigma_k(t+1) - \sigma_k(t) = \frac{A_k(t-1) + r_k(t) - \sigma_k(t)\sum_{j=1}^K [A_j(t-1) + r_j(t)]}{\sum_{j=1}^K [A_j(t-1) + r_j(t)]}$$
$$= \frac{r_k(t) - \sigma_k(t)\sum_{j=1}^K r_j(t)}{\sum_{j=1}^K [A_j(t-1) + r_j(t)]}.$$

Note that the denominator is $O(t)$ since

$$tc \leq \sum_{j=1}^{K} (A_j(t-1) + u_j(t) + c) \leq t(1+c).$$

Since the step size is identical for all individuals the influence of initial attractors vanishes asymptotically. Hence we can write

$$p_k(t+1) - p_k(t) = O\left(\frac{1}{t}\right)\left(r_k(t) - p_k(t)\sum_{j=1}^{K} r_j(t)\right) + O\left(\frac{1}{t^2}\right).$$

This describes how p evolves as a stochastic process with decreasing step size.

We borrow the following notation and definitions from Benaïm (1999). Consider a metric space (X, d) and a semi-flow $\Phi : \mathbb{R}_+ \times X \to X$ induced by a vector field F on X. A point $x \in X$ is an equilibrium if $\Phi_t(x) = x$ for all t. A point $x^* \in X$ is an ω-limit point of x if $x^* = \lim_{t_k \to \infty} \Phi_{t_k}(x)$ for some sequence $t_k \to \infty$. The ω-limit set of x, denoted $\omega(x)$, is the set of ω-limit points of x. The definition of an ω-limit can be extended to a discrete time system. A set $A \subseteq X$ is invariant if $\Phi_t(A) = A$ for all $t \in \mathbb{R}$. For $\delta > 0$, and $T > 0$, a (δ, T)-*pseudo-orbit* from $a \in X$ to $b \in X$ is a finite sequence of partial trajectories $\{\Phi_t(y_i) : 0 \leq t \leq t_i\}_{i=0,\ldots,k-1}$, with $t_i \geq T$, such that $d(y_0, a) < \delta$, $d\left(\Phi_{t_j}(y_j), y_{j+1}\right) < \delta$ for $j = 0, ..., k-1$, and $y_k = b$. A point $a \in X$ is chain recurrent if there is a (δ, T)-pseudo-orbit from a to a for every $\delta > 0$, and $T > 0$. Let $\Lambda \subseteq X$ be a non-empty invariant set. Φ is called chain recurrent on Λ if every point $x \in \Lambda$ is a chain recurrent point for $\Phi|\Lambda$, the restriction of Φ to Λ. A compact invariant set on which Φ is chain recurrent is called an internally chain recurrent set. A subset $A \subseteq X$ is an attractor for Φ if (i) A is non-empty, compact and invariant, and (ii) A has a (so-called fundamental) neighborhood $U \subseteq X$ such that $\lim_{t\to\infty} d(\Phi_t, A) \to 0$ uniformly in $x \in U$ (distance between Φ_t and closest point in A). An attractor A is a proper attractor if it contains no proper subset that is an attractor.[15]

[15] Our stochastic process is a Robbins-Monro algorithm (Robbins and Monro (1951)). The ODE method originates with Ljung (1977). For a book-lenght treatment of the theory of stochastic approximation see Benveniste, Priouret and Métivier (1990).

We have:

Proposition 3 *Define the continuous time interpolated stochastic GPI process* $\tilde{p} : \mathbb{R}_+ \to \mathbb{R}^m$ *by*

$$\tilde{p}(t+s) = p(t) + s\frac{p(t+1) - p(t)}{1/(t+1)},$$

for all $n \in \mathbb{N}$ *and* $0 \leq s \leq 1/(t+1)$. *With probability 1, the* ω*-limit set of* $\tilde{p}$, $\omega(\tilde{p})$, *is connected and internally chain recurrent for the flow* Φ *induced by the continuous time deterministic GPI dynamic*

$$\dot{p}_k = \rho_k(p) - p_k \sum_{j=1}^{K} \rho_j(p) \tag{8}$$

Equivalently; With probability 1, $\omega(\tilde{p})$ *is a compact invariant set for* Φ, *and* $\Phi|\Lambda$ *admits no proper attractor.*

Proof. We can write $p(t+1) - p(t) = \gamma(t+1)(F(p(t)) + U(t+1) + b(t+1))$, where $\gamma(t+1) = O(1/t)$, $F(p(t))$ equals the right hand side of (8),

$$U(t+1) = r_k(t) - \rho_k(p(t)) - p_k(t) \sum_{j=1}^{K} (r_j(t) - \rho_j(p(t))),$$

and $b(t+1) = O(1/t)$. We have $\mathbb{E}[U(t+1)|p(t)] = 0$, and $\sup_t \mathbb{E}\left[\|U(t+1)\|^2\right] \leq C$ for some constant C. Moreover, $\lim_{t\to\infty} \gamma(t) = 0$, $\sum_{t=1}^{\infty} \gamma(t) = \infty$, and $\sum_{t=1}^{\infty} (\gamma(t))^2 < \infty$. Also F is a bounded locally Lipschitz vector field. Propositions 4.1 and 4.2, with remarks 4.3 and 4.5 in Benaïm (1999) imply that with probability 1 the interpolated process $\tilde{p}$ is an asymptotic pseudotrajectory of the flow Φ induced by F. Since $\{\tilde{p}(t) : t \geq 0\}$ is precompact, the desired result follows from Benaïm's theorem 5.7 and proposition 5.3. Q.E.D.

In other words, the realization of $p(t)$ almost surely converges to a compact invariant set that admits no proper attractor under the flow induced by the GPI dynamic (8).

The next step is to calculate the expected reinforcement under reinforcement rules (4) and (5).

Lemma 4 *Consider a symmetric fixed* N*-player game and* $c > -\min_{s\notin S} u(s_i, s_{-i})$. ***(a)*** *The expected reinforcement (4) is*

$$\rho_k^A(p) = Np_k \sum_{j=0}^{N-1} \frac{1}{j+1} \Pr(Y(k) = j)\, \mathbb{E}_p[u(k, s_{-i}) + c|Y(k) = j]. \tag{9}$$

(b) *The expected reinforcement (5) is*

$$\rho_k^B(p) = Np_k(\pi_k(p) + c). \tag{10}$$

Proof. (a) Since $\mathbb{E}_p[r_k(s)|X(k)=0] = 0$ we have

$$\begin{aligned}
\rho_k^A(p) &= \mathbb{E}_p\left[r_k^A(s)\right] \\
&= \sum_{j=1}^{N} \Pr(X(k)=j)\, \mathbb{E}_p\left[r_k^A(s)\,|X(k)=j\right] \\
&= \sum_{j=1}^{N} \Pr(X(k)=j)\, \mathbb{E}_p\left[u(k,s_{-i}) + c|Y(k)=j-1\right] \\
&= \sum_{j=0}^{N-1} \Pr(X(k)=j+1)\, \mathbb{E}_p\left[u(k,s_{-i}) + c|Y(k)=j\right].
\end{aligned} \tag{11}$$

We need to translate from $\Pr(X(k)=j+1)$ to $\Pr(Y(k)=j)$. Simply use

$$\begin{aligned}
\Pr(Y(k)=j) &= \binom{N-1}{j} p_k^j (1-p_k)^{N-1-j} \\
&= \frac{(n-1)!}{j!\,(n-1-j)!} p_k^j (1-p_k)^{N-1-j},
\end{aligned}$$

and

$$\begin{aligned}
\Pr(X(k)=j+1) &= \binom{N}{j+1} p_k^{j+1} (1-p_k)^{N-(j+1)} \\
&= \frac{N!}{(j+1)!\,(N-(j+1))!} p_k^{j+1} (1-p_k)^{N-j-1} \\
&= \frac{Np_k}{j+1} \frac{(N-1)!}{j!\,(N-j-1)!} p_k^j (1-p_k)^{N-j-1} \\
&= \frac{Np_k}{j+1} \Pr(Y_i(k)=j).
\end{aligned}$$

Plugging this into (11) finishes the proof.

(b) Similar to (a). Q.E.D.

By plugging the relevant expression for expected reinforcement into the general stochastic approximation result (8), we obtain the ODE for the corresponding version of GPI. Reinforcement of the form *(5)*, with expected reinforcement (10), results in the replicator dynamic, scaled by the number of players:

Proposition 4 *In a symmetric fixed N-player game, the GPI continuous time dynamic* *with reinforcement (5) is*

$$\dot{p}_k = Np_k \left(\pi_k(p) - \sum_{j=1}^{K} p_j \pi_j(p) \right).$$

It follows that, the realization of the stochastic GPI-B process almost surely converges to a compact invariant set that admits no proper attractor under the flow induced by the replicator dynamic (14). The following result, adapted from Hopkins and Posch (2005), rules out convergence to non-Nash rest points in general under the dynamic with reinforcements defined by (5). In LUPI the same result holds also for reinforcements defined by (4).

Proposition 5 *If p' is a rest point of the replicator dynamic (14) and not a Nash equilibrium, then for the stochastic imitation process $p(t)$ defined by the choice rule (6) and updating rule (3) with reinforcement rule (5),* $\Pr\left(\lim_{t\to\infty} p(t) = p'\right) = 0$.

Proof. The proof of proposition 3 in Hopkins and Posch (2005) can be adapted. Q.E.D.

Reinforcement of the form *(4)*, with expected reinforcement (9), gives rise to a quite complicated deterministic dynamic. We therefore restrict attention to symmetric two-player games where payoffs are represented by a matrix $A = [a_{ij}]$. This special case has been the focus of much of evolutionary game theory.

Proposition 6 *In a 2-player game, where payoffs are represented by a matrix $A = [a_{ij}]$, the GPI continuous time dynamic with reinforcement (4) is*

$$\dot{p}_k = 2V_k^{RD}(p) + V_k^{DD}(p), \tag{12}$$

where

$$V_k^{RD}(p) = p_k \left(\pi_k(p) - \sum_{j=1}^{K} p_j \pi_j(p) \right),$$

and

$$V_k^{DD}(p) = p_k \left(p_k \left(-\left(a_{kk} + c\right)\right) - \sum_{j=1}^{K} p_j^2 \left(-\left(a_{jj} + c\right)\right) \right).$$

Proof. With $N = 2$ the expression (9) becomes

$$\begin{aligned}
\rho_k(p) &= p_k \begin{pmatrix} 2\Pr(Y_i(k)=0)\,\mathbb{E}_p[u(k,s_{-i})+c|Y_i(k)=0] \\ +\Pr(Y_i(k)=1)\,\mathbb{E}_p[u(k,s_{-i})+c|Y_i(k)=1] \end{pmatrix} \\
&= p_k\left(\pi_k(p)+c+\Pr(Y_i(k)=0)\,\mathbb{E}_p[u(k,s_{-i})+c|Y_i(k)=0]\right) \\
&= p_k\left(\pi_k(p)+c+\sum_{i\neq k} p_i(a_{k,i}+c)\right) \\
&= p_k\left(2(\pi_k(p)+c)-p_k(a_{kk}+c)\right).
\end{aligned}$$

Plugging this into the stochastic approximation result (8), and rearranging we obtain (12). Q.E.D.

Define the *diagonal dynamic* as

$$\dot{p}_k = V_k^{DD}(p)\,. \tag{13}$$

Clearly, pure Nash equilibria are rest points of (13) but mixed Nash equilibria are generally not. However in constant sum games a mixed equilibrium p satisfies $V^{DD}(p) = 0$ if and only if all strategies in the support of p have equal weight. Conversely, if a completely mixed equilibrium p puts equal weight on all strategies then the game must be constant sum in order to satisfy $V^{DD}(p) = 0$. It turns out that even in these games, the dynamic (12) is quite different from that of other well-studied dynamics in the literature. In what follows we therefore restrict attention to games with a completely mixed equilibrium that puts equal probability on all strategies. In such games we have

$$\dot{p}_k = -\left(\frac{C}{2}+c\right)p_k\left(p_k-\sum_{j=1}^{K}p_j^2\right),$$

where $C = a_{ij} + a_{ji}$ is the constant sum. We can use the same Lyapunov function as for the replicator dynamic. With its help we prove:

Proposition 7 *Consider a C-constant sum game with a unique interior equilibrium that puts equal weight on all strategies. The GPI2 dynamic (12) diverges towards the boundary.*

Proof. The games we consider are null stable (Hofbauer and Sandholm (2009)). In null stable games the solutions to replicator dynamic are closed orbits. It is therefore sufficient to show that the diagonal dynamic (13) diverges. With the Kullback-Leibler relative entropy function (see below, equation (15)) we have, since $p_k^* = 1/K$,

$$\dot{H}_{p^*}(p) = \left(\frac{C}{2} + c\right) \sum_k p_k^* \left(p_k - \sum_{j=1}^K p_j^2\right) = \left(\frac{C}{2} + c\right) \left(\frac{1}{K} - \sum_{j=1}^K p_j^2\right)$$

The function $\sum_{j=1}^K p_j^2$ attains its maximum $1/K$ when $p_j = 1/K$ for all j. Hence, since $c > -\min_{s \notin S} u(s_i, s_{-i})$, we have $\dot{H}_{p^*}(p) \geq 0$ with equality only at $p = p^*$, implying that the GPI dynamic (12) diverges from p^*. Q.E.D.

Example 1 *The above proposition implies that we have divergence from the Nash equilibrium towards the boundary in RPS under GPI with reinforcements of the form (4). This stands in contrast to the replicator dynamic that results from reinforcements of the form (5). As is well-known, the replicator dynamic cycles in RPS. This is illustrated in figures 16a and 16b.*

3.3 Convergence of GPI in the Poisson LUPI game

In LUPI expected reinforcement is

$$\rho_k(p) = np_k \pi_k(p) + c,$$

so (8) takes the form

$$\dot{p}_k = np_k \left(\pi_k(p) - \sum_{j=1}^K p_j \pi_j(p)\right) + c(1 - Kp_k(t)).$$

This is the replicator dynamic multiplied by n plus a noise term due to the addition of c to reinforcements. In what follows we will ignore c. It must be strictly positive for the stochastic approximation argument to go through, but we are allowed to make it arbitrarily small.[16] Hence we will focus on

$$\dot{p}_k = np_k \left(\pi_k(p) - \sum_{j=1}^K p_j \pi_j(p)\right). \quad (14)$$

[16] If $c = 0$ then we face the problem that the stochastic sequence of reinforcements is not guaranteed to satisfy $\lim_{t\to\infty} \gamma(t) = 0$, $\sum_{t=1}^\infty \gamma(t) = \infty$, and $\sum_{t=1}^\infty (\gamma(t))^2 < \infty$. With $c = 0$ proposition 3 would continue to hold if almost surely $\lim_{t\to\infty} \gamma(t) = 0$, almost surely $\sum_{t=1}^\infty \gamma(t) = \infty$, and $\mathbb{E}\left[\sum_{t=1}^\infty (\gamma(t))^2\right] < \infty$. In LUPI these conditions hold if the probability of a tie is bounded away from zero. Unfortunately along trajectories towards the boundary, specifically towards monomorphic states this need not be the case.

It follows that in LUPI, the realization of the stochastic GPI process almost surely converges to a compact invariant set that admits no proper attractor under the flow induced by the replicator dynamic (14).

It is well known that the Kullback-Leibler relative entropy function

$$H_{p^*}(p) = \sum_i p_i^* \log\left(\frac{p_i^*}{p_i}\right), \tag{15}$$

is a Lyapunov function for the replicator dynamic (Weibull (1995), chapters 3 and 6, and Sandholm (2011) pp. 234-5). It satisfies $H_{p^*}(p) \geq 0$ with equality if and only if $p = p^*$, and along any solution trajectory p it holds that $\dot{H}_{p^*}(p) = -\sum_i p_i^* \frac{\dot{p}_i}{p_i}$. Hence, in LUPI we have $\dot{H}_{p^*}(p) = -n(p^* - p)' \pi(p)$. Consequently, in order to demonstrate convergence from any interior initial state we must show that

$$(p^* - p)' \pi(p) = \sum_k (p_k^* - p_k) \pi_k(p) > 0, \tag{16}$$

for all $p \in \Delta$, such that $p \neq p^*$. For p^* to be (locally) asymptotically stable it is sufficient that (16) holds for all p in a neighborhood of p^*. The condition (16) is satisfied by strictly stable games. By propositions 1, 3 and 5 (noting that the latter is applicable also for GPI-A in LUPI) we conclude:

Proposition 8 *The deterministic GPI-dynamic (14) converges to the unique Nash equilibrium from any interior initial condition. The stochastic GPI-process almost surely converges to the unique Nash equilibrium.*

3.4 Similarity-Based Imitation of Winners

Since the strategy set is so large in the LUPI field game, only reinforcing the previous winning number would result in a learning process that is too slow and too tightly clustered on previous winners. We therefore follow Sarin and Vahid (2004) by assuming that numbers that are similar to the winning number may also be reinforced. We use the triangular Bartlett similarity function used by Sarin and Vahid (2004), which puts reinforcement factors on strategies near the previous winner that declines linearly with distance from the previous winner. In Appendix B we consider two other forms of similarity-based imitation in LUPI: One imitates strategies that are similar to, but distinct from, the winning number. The other imitates strategies that are similar, but not above, the winning number.

Let W denote the size of the "similarity window" and define the similarity function

$$\gamma_k(j) = \frac{\max\left\{0, 1 - \frac{|j-k|}{W}\right\}}{\sum_{i=0}^{K} \max\left\{0, 1 - \frac{|j-i|}{W}\right\}}.$$

This is depicted in figure 3 for $j = 10$ and $W = 3$. Note that the similarity weights are scaled so that they sum to one. In general the normalized (Bartlett-)similarity-modified payoff to strategy k at time t, is

$$\tilde{u}_k(t) = \sum_{j=0}^{K} u_j(s(t))\,\gamma_k(j).$$

Recall $k^*(s) = \min_{j\in\{k\in N: s_k \neq s_l, \forall l\neq k\}} s_j$. Thus in LUPI this reduces to

$$\tilde{u}_k(t) = \gamma_k(k^*(s(t))).$$

Reinforcements take the form $\tilde{r}_k(t) = \tilde{u}_k(t) + c$. In LUPI this means that $\gamma_k(j)$ is the reinforcement of strategy k when strategy j wins. When $W = 1$ we have $\tilde{u}_k(t) = u_k(t)$. Thus, the stochastic approximation results derived above hold exactly when $W = 1$, and it might be hoped that they will also hold for sufficiently low $W > 1$.

Under similarity-based imitation the expected reinforcement is

$$\tilde{\rho}_k(t) = \mathbb{E}[\tilde{r}_k(s)] = \sum_{j=0}^{K} \rho_j(p)\,\gamma_k(j) + c = \sum_{j=0}^{K} np_j\pi_j(p)\,\gamma_k(j) + c.$$

The expected similarity-modified payoff to strategy k under strategy profile p, is

$$\tilde{\pi}_k(p) = \mathbb{E}[\tilde{u}_k(s)] = \sum_{j=0}^{K} \mathbb{E}[u_j(s)]\,\gamma_k(j) = \sum_{j=0}^{K} \pi_j(p)\,\gamma_k(j). \tag{17}$$

Note that if p^* is an interior equilibrium so that $\pi_k(p^*) = \pi^*$ for all k then

$$\tilde{\pi}_k(p^*) = \sum_{j=0}^{K} \pi_j(p^*)\,\gamma_k(j) = \pi^* \sum_{j=0}^{K} \gamma_k(j) = \pi^*,$$

for all k. Hence p^* is an equilibrium for the game with similarity-modified payoffs as well.[17]

[17] Note that this hinges crucially on the normalization in the definition of $\gamma_k(j)$. The un-normalized Bartlett window is much less tractable.

Lemma 5 *The Nash equilibrium of the Poisson LUPI game is a Nash equilibrium of the Poisson LUPI game with γ-similarity-modified payoffs.*

Importantly this result does not rule out multiple equilibria.

We can analyze stability and convergence properties under the similarity based imitation process by simply applying the GPI dynamic (8) to a modified game in which expected payoffs are given by $\tilde{\pi}_k(p)$ rather than $\pi_k(p)$. Instead of studying the dynamic (14) we need to study

$$\dot{p}_k = np_k \left(\tilde{\pi}_k(p) - \sum_{i=1}^{K} p_i \tilde{\pi}_i(p) \right). \tag{18}$$

In order for GPI to converge to p^* from any interior initial conditions in the LUPI game with similarity-modified payoffs we need, in line with (16),

$$-\sum_k (p_k^* - p_k) \tilde{\pi}_k(p) = -\sum_k (p_k^* - p_k) \sum_{j=0}^{K} \gamma_k(j) \pi_j(p) < 0$$

for all $p \neq p^*$. This is something we would have to evaluate numerically.

Clearly, whether convergence is preserved is going from π to $\tilde{\pi}$ depends on how similar the modified payoffs are to the original payoffs.

Conjecture 2 *In the Poisson LUPI game with γ-similarity-modified payoffs, if W is sufficiently small then the GPI dynamic (14) convergences, from any interior initial state, to a unique connected and compact set, containing the symmetric Nash equilibrium of the game without similarity-modified payoffs.*

4 Estimation

Since we only have aggregate data for the field, we focus on representative agent learning models. Each player is assumed to play according to the same distribution $p_k(t)$ that is updated after each round. The model can also be thought of as a population learning model. In this interpretation, there are instead players of K types and the proportion of players of different types in the population, $p_k(t)$, changes after each round. Most likely there is some truth to both interpretation, but without individual level data it is impossible to empirically distinguish the two (in the field).[18]

[18]Some players in the lab do indeed choose the same number throughout the experiments, but other do not.We have tried to keep the learning model as simple as possible. It is possible to construct more complicated, potentially more realistic models, with more parameters, but it would most likely be difficult to identify models with more parameters (in particular with the laboratory data).

The learning model has two free parameters: the size of the similarity window, W, and the precision of the choice function, λ. We estimate the best-fitting values by minimizing the squared deviation between predicted choice densities and empirical densities summed over all numbers and rounds. Estimation of any learning model requires an assumption about the choice probabilities in the first period, $p_k(1)$. We use the empirical frequencies to create choice probabilities in the first period ("burning in"). Given these probabilities and λ, we determine $A(1)$ so that equation (7) gives the assumed choice probabilities $p_k(1)$. Since the power choice function is invariant to scaling, we determine the attractions in the first period so that they sum to one, i.e., $\sum_{k=1}^{K} A_k(1) = 1$. At the end of each period t, strategies are reinforced by a factor $\tilde{r}_k(t)$, which depends on the winning number in period t, as described above. For the empirical estimation of the learning model we use the actual winning numbers from the field.

5 The Field LUPI Game

The field version of LUPI, called Limbo, was introduced by the government-owned Swedish gambling monopoly Svenska Spel on the 29th of January 2007. Daily data were downloaded for the first seven weeks, ending on the 18th of March 2007. This section describes its essential elements; additional description of the game is available in Östling et al. (2011).

In Limbo, players chose an integer between 1 and 99,999. Each number bet costs 10 SEK (approximately 1 EURO). The game was played daily and the winning number was presented on TV in the evening and on the Internet. The winner received 18 percent of the total sum of bets, with the prize guaranteed to be at least 100,000 SEK (approximately 10,000 EURO). If no number was unique the prize was shared evenly among those who chose the smallest and least-frequently chosen number. There were also smaller second and third prizes (1000 SEK and 20 SEK) for being close to the winning number.

During the first three to four weeks, it was only possible to play the game at physical branches of Svenska Spel by filling out a form, but after that online play was also allowed. The form allowed players to bet on up to six numbers[19], to play the same numbers for up to 7 days in a row, or to let a computer choose random numbers for them (a "HuxFlux" option).

[19]The rule that players could only pick up to six numbers a day was enforced by the requirement that players had to use a "gambler's card" linked to their personal identification number when they played. Colluding in LUPI can conceivably increase the probability of winning but would require a remarkable degree of coordination across a large syndicate, and is also risky if others might be colluding in a similar way since one has to be unique to win. Interestingly, the game was stopped on March 24th, one day after a newspaper article claimed that some players had colluded in the game, but it is unclear whether collusion actually occurred or how it could be detected.

Unfortunately, we have only gained access to aggregate daily frequencies, not to individual-level data. We also do not know how many players used the randomization HuxFlux option. However, because the operators told us how HuxFlux worked, we can estimate that approximately 19 percent of players were randomizing in the first week.[20]

Note that the theoretical analysis of the LUPI game in the previous sections differs from the field LUPI game in four ways. First, the theory used a tie-breaking rule in which nobody wins if there is no uniquely chosen number (to simplify expected payoff calculations). In the field game, however, players who tie by choosing the smallest and least-frequently chosen number share the prize. This is a minor difference because the probability that there is no unique number is very small and it never happened during the 49 days for which we have data. A second, more important, difference is that we assume that each player can only pick one number whereas players in the field are allowed to bet on up to six numbers. Third, we do not take the second and third prizes present in the field version into account, but this is unlikely to make a big difference given the strategic nature of the game. Finally, in the lab, the only feedback the players receive is information about the winning number in the previous period.

Nevertheless, these four differences between the game analyzed theoretically and the field game, are a motivation for running laboratory experiments with single bets, no opportunity for direct collusion, only a first prize, and feedback only about winning numbers, which match the game analyzed theoretically more closely.

5.1 Descriptive Statistics

Table 1 reports weekly summary statistics for the first 49 days of the game. The last column displays the corresponding statistics that would result from play according to the symmetric equilibrium (we use the Poisson-Nash equilibrium as an approximation in this case).

Overall, the average number of bets N was 53,783, but there was considerable variation over time. There is no apparent time trend in the number of participating players, but there is less participation on Sundays and Mondays.

Despite some differences between the simplified theory and the way the field lottery game was implemented, the average number chosen overall was 2835, which is close to the equilibrium prediction of 2595. Note that the median number chosen is much lower than the average number, which is due to some players playing very high numbers (but the difference between the average and median decreases over time). Winning numbers, and the lowest numbers not chosen by anyone, varied a lot over time. It is noteworthy

[20] In the first week, the randomizer chose numbers from 1 to 15,000 with equal probability. The drop in numbers just below and above 15,000 suggests the 19 percent figure.

that all the aggregate statistics in Table 1 move closer to the equilibrium predictions from the first to the last week. For example, the mean number chosen in the last week is 2484, compared to the prediction of 2595. In equilibrium essentially nobody should choose a number above 10,000. In the first week 12 percent chose these high numbers, but in the last week only 1 percent did.

Table 1. Field descriptive statistics by week

	All	W1	W2	W3	W4	W5	W6	W7	Eq.
# Bets	53783	57017	54955	52551	50471	57997	55583	47907	53783
Avg. number	2835	4512	2963	2479	2294	2396	2718	2484	2595
Median number	1674	1203	1551	1668	1603	1698	2056	1935	2541
Avg. winner	2095	1159	1906	2212	1818	2720	2866	1982	2585
Lowest not played	3103	1745	2594	3169	3528	3498	3729	3462	4077
Below 100 (%)	6.08	15.16	6.33	5.76	4.82	4.29	3.01	3.19	2.02
Below 1000 (%)	32.31	44.91	33.03	32.42	32.43	31.84	24.01	27.52	20.05
Below 5000 (%)	92.52	78.75	89.28	95.10	96.09	96.22	95.76	96.44	93.34
Below 10000 (%)	96.63	88.07	96.66	97.86	98.33	98.27	98.39	98.81	100.00

Another noteworthy finding is that the average winning number is close to the median number during most weeks. This suggests that players change strategies in the direction of previous winners. To see this more clearly, figure 2 displays the effect of winning numbers on chosen numbers in the field. It shows how the difference between the winning number at time t and the winning number at time $t-1$ closely matches the difference between the average chosen number at time $t+1$ and the average chosen number at time t. As a result, figure 4 shows how the relationship between the median number chosen on day t is related to the median of the winning numbers from day 1 until $t-1$.

5.2 Results

Estimation of the learning model for the field data is quite computationally demanding so we had to rely on a relatively coarse grid search to find best-fitting parameters. The sum of squared deviations with respect to both W and λ appears to be relatively smooth and convex, so it is likely that we have find the best-fitting values. The estimated values for the field data are $W = 344$ and $\lambda = 0.0085$. Recall that a window size of 344 implies that a window of 687 numbers are reinforced.

To see how the learning model fits the data, Figure 5 displays the average weekly predicted densities of the learning model for numbers up to 6000 (along with the data and equilibrium). The main feature of learning is that the number of low numbers shrinks and the gap between the predicted frequency of numbers between 2000 and 5000 is gradually filled in. Figure 6 shows summary statistics week-by-week in a boxplot.

Östling et al. (2011) can reject the hypothesis that the behavior in week seven is in equilibrium. Still, over the 49 days there is clear movement towards equilibrium. Figure 7 displays a formal measure of this: The proportion of the empirical distribution that is below the theoretical equilibrium distribution increases over time, from 0.5 in the first week to 0.8 in the last week, not very far from the theoretical maximum at 1.0

6 The Laboratory LUPI Game

The field LUPI game does not exactly match the theoretical assumptions used to generate the equilibrium prediction. In the field data, some choices were made by a random number generator, some players might have chosen multiple numbers or colluded and there were multiple prizes. In addition, in the field we did not have full control over the feedback, and we could not follow individuals over time. We therefore analyze the laboratory data from Östling et al. (2011) that follows the theoretical assumptions much more closely, with one exception. Östling et al. (2011) implemented the games with population uncertainty so that the number of participants varied from round to round. We analyze this data as if there was a Poission distributed number of players, but we have already seen that this does not matter much for the equilibrium prediction.

Their experiment consisted of 49 rounds in each session and the prize to the winner in each round was $7. The number of players in each round was drawn from a distribution with mean 26.9.[21] In the lab, each player was allowed to choose only one number, they could not use a random number generator (as in the field game), there was only one prize per round, and if there was no unique number nobody won. The only feedback that players received after each round was the winning number.

In particular, at the beginning of each session, the experimenter first explained the rules of the LUPI game. The instructions were based on a version of the lottery form for the field game translated from Swedish to English (see Östling et al. (2011)). Subjects

[21] In three of the four sessions, subjects were told the mean number of players, and that the number varied from round to round, but did not know the distribution (in order to match the field situation in which players were very unlikely to know the total number playing each day). Due to a technical error, in these three sessions, the variance was lower than the Poisson variance (7.2 to 8.6 rather than 26.9). However, this mistake is likely to have little effect on behavior because subjects did not know the total number of players in each round. In the last session, the number of players in each round was drawn from a Poisson distribution with mean 26.9 and the subjects were informed about this. Furthermore, the data from the true Poisson session and the lower-variance sessions look statistically similar so we pool them in our analysis.

were then given the option of leaving the experiment, but none of the recruited subjects chose to leave.

In three of the four sessions, subjects were told that the experiment would end at a predetermined, but non-disclosed time to avoid an end-game effect. Subjects were also told that participation was randomly determined at the beginning of each round, with 26.9 subjects participating on average. Subjects in the fourth session were explicitly told there were 49 rounds, and the number of players was drawn from a Poisson distribution. They were also shown in the instructions a graph showing a distribution function for a Poisson distribution with mean 26.9.

In the beginning of each round, subjects were informed whether they would actively participate in the current round (i.e., if they had a chance to win). They were required to submit a number in each round, even if they were not selected to participate.

When all subjects had submitted their chosen numbers, the lowest unique positive integer was determined. If there was a lowest unique positive integer, the winner earned \$7; if no number was unique, no subject won. Each subject was privately informed, immediately after each round, what the winning number was, whether they had won that particular round, and their payoff so far during the experiment. This procedure was repeated 49 times, with no practice rounds. All sessions lasted for less than an hour, and subjects received a show-up fee of \$8 or \$13 in addition to earnings from the experiment (which averaged \$8.60).

A more detailed description of the experiment can be found in Östling et al. (2011).

6.1 Lab Descriptive Statistics

We focus only on the choices from incentivized subjects that were selected to actively participate in the remainder of the paper. It is noteworthy, however, that the choices of participating and non-participating subjects did not significantly differ (p-value 0.16, Mann-Whitney). The choices from the session with the announced Poisson distribution and the pooled other three sessions do not significantly differ at the five percent level ($p = 0.59$, t-test with clustered standard errors). In the remainder of the paper we therefore pool all four sessions.

Table 2 shows some descriptive statistics for the participating subjects in the lab experiment. As in the field, some players in the first rounds tend to pick very high numbers (above 20) but the percentage shrinks to approximately 1 percent after the first seven rounds. Both the average and median number chosen corresponds closely to equilibrium after the first seven rounds. The average winning numbers are too high compared to equilibrium play, which is consistent with players picking very low numbers too much, creating non-uniqueness among those numbers so that unique numbers are unusually high. The overwhelming impression from Table 2 is that convergence (close) to equilibrium is very rapid despite receiving feedback only about the winning number.

Table 2. Laboratory descriptive statics

	All	1-7	8-14	15-21	22-28	29-35	36-42	43-49	Eq.
Avg. number	5.96	8.56	5.24	5.45	5.57	5.45	5.59	5.84	5.18
Median number	4.65	6.14	4.00	4.57	4.14	4.29	4.43	5.00	5.00
Avg. winner	5.95	8.44	5.34	5.28	6.60	5.51	5.87	4.58	5.18
Below 20 (%)	98.02	93.94	99.10	98.45	98.60	98.85	98.79	98.42	100.00

The movement towards equilibrium is clear when measuring the proportion of the empirical distribution below the empirical distributions, as displayed in figure 8(a)-(d).

Östling et al. (2011) reports the result from a post-experimental questionnaire. The most interesting finding from their analysis was that several subjects said that they responded to previous winning numbers. Figure 9(a)-(d) displays the effect of winning numbers on chosen numbers in the lab for sessions 1-4. The difference between the winning number at time t and the winning number at time $t-1$ closely matches the difference between the average chosen number at time $t+1$ and the average chosen number at time t. Furthermore, table 3 displays the results from an OLS regression with changes in average guesses as dependent variable, and lagged differences between winning numbers as independent variables. At least during the first two weeks lagged differences in winning numbers have a clear effect on average choices. The fact that the effect is weaker in later weeks is consistent with the GPI model since the decreasing step size implies that the influence of winning numbers grow smaller over time.

Table 3. Laboratory panel data OLS regression

Dependent variable: t mean guess – $t-1$ mean guess			
	All periods	1-14	15-49
$t-1$ winner – $t-2$ winner	0.154***	0.147***	0.168***
	(4.29)	(3.41)	(4.62)
$t-2$ winner – $t-3$ winner	0.082**	0.089*	0.055
	(2.01)	(1.79)	(1.05)
$t-3$ winner – $t-4$ winner	0.047*	0.069*	-0.027
	(1.42)	(1.74)	(-0.47)
Observations	5662	1216	4446
R^2	0.0086	0.0256	0.0036

Standard errors are clustered on individual level.

As a further illustration of the relationship between chosen numbers and winning numbers, Figure 10 displays the distribution of winnings number in all rounds in all sessions, and chosen numbers from period 25 and onwards. Recall from Proposition 2 that in equilibrium the choice probabilities should coincide with the probability that each number wins, and as can be seen from Figure 10 the correspondence is quite close.

6.2 Learning Results

Estimating the learning model for the laboratory data is much less computationally demanding than in the field and we can therefore use a much finer grid to find the best-fitting parameters. As shown by Figure 11, however, the sum of the squared deviations for the laboratory data is relatively flat with respect to both W and λ when both parameters increases proportionally. For the lab data, W and λ largely play inverse roles. A higher window size W combined with higher response sensitivity λ generate very close squared deviations (since higher W is generating a wider spread of responses and higher λ is tightening the response). The higher W is, the higher is λ, but the overall fit is nearly unchanged as W varies between 3 and 12.

We divide the estimated window size from the field by 100 and fix $W = 3$. This resulted in an estimated λ for the laboratory data of 0.31. If we do not restrict $W = 3$, then the best-fitting W and λ for the laboratory data gives $W = 11$ and $\lambda = 1.84$.

As was discussed in the previous section, players in the laboratory seem to learn to play the game much quicker than in the field, so there is less learning to be explained by the learning model. The learning model can explain some of the ups and downs during the first 14 rounds in the laboratory, as well as the shrinking dispersion of numbers over time, but after the first few weeks there is no trend towards higher numbers as seen in the field data. Figure 13 displays box plots for the 14 first rounds in the four sessions. Note that the learning model predicts much more dispersion of numbers in the early rounds in the first session. This is explained by the fact that players played very high numbers in the first round in that session and that a very high number, 67, won in the fourth period. The imitation-based model is substantially affected by that outlying win.

7 Concluding Discussion

The current paper provides suggestive evidence that a simple imitation learning model can explain the quick movement toward equilibrium in both the field and lab.

References

Apesteguia, J., Huck, S. and Oechssler, J. (2007), 'Imitation–theory and experimental evidence', *Journal of Economic Theory* **136**, 217–235.

Armstrong, M. and Huck, S. (2010), 'Behavioral economics as applied to firms', *Competition Policy International* **6**(1), 3–45.

Benaïm, M. (1999), Dynamics of stochastic approximation algorithms, *in* J. Azéma, M. Émery, M. Ledoux and M. Yor, eds, 'Séminaire de Probabilités XXXIII', Vol. 1709 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin/Heidelberg, pp. 1–68.

Benveniste, A., Priouret, P. and Métivier, M. (1990), *Adaptive algorithms and stochastic approximations*, Springer-Verlag New York, Inc., New York, USA.

Binmore, K. G., Samuelson, L. and Vaughan, R. (1995), 'Musical chairs: Modeling noisy evolution', *Games and Economic Behavior* **11**(1), 1–35.

Binmore, K. and Samuelson, L. (1994), 'An economist's perspective on the evolution of norms', *Journal of Institutional and Theoretical Economics* **150/1**, 45–63.

Binmore, K. and Samuelson, L. (1997), 'Muddling through: Noisy equilibrium selection', *Journal of Economic Theory* **74**(2), 235–265.

Björnerstedt, J. and Weibull, J. (1996), Nash equilibrium and evolution by imitation, *in* K. J. Arrow, E. Colombatto, M. Perlman and C. Schmidt, eds, 'The Rational Foundations of Economic Behaviour', MacMillan, London, pp. 155–171.

Camerer, C. F. (2003), *Behavioral Game Theory*, Princeton University Press, Princeton.

Camerer, C. F. and Ho, T. H. (1999), 'Experience-weighted attraction learning in normal form games', *Econometrica* **67**(4), 827–874.

Chiappori, P. A., Levitt, S. D. and Groseclose, T. (2002), 'Testing mixed strategy equilibrium when players are heterogeneous: The case of penalty kicks', *American Economic Review* **92**(4), 1138–1151.

Dasgupta, P. and Maskin, E. (1986), 'The existence of equilibrium in discontinuous economic games, i: Theory', *Review of Economic Studies* **53**(1), 1–26.

Duersch, P., Oechssler, J. and Schipper, B. C. (2010), Unbeatable imitation. Mimeo.

Eichberger, J. and Vinogradov, D. (2009), Least unmatched price auctions: A first approach. Mimeo.

Erev, I. and Haruvy, E. (2008), Learning and the economics of small decisions. Mimeo.

Fudenberg, D. and Levine, D. K. (1998), *The Theory of Learning in Games*, MIT Press.

Gale, D., Binmore, K. G. and Samuelson, L. (1995), 'Learing to be imperfect', *Games and Economic Behavior* **8**, 56–90.

Gallice, A. (2009), Lowest unique bid auctions with signals. Mimeo.

Ho, T. H., Camerer, C. F. and Chong, J.-K. (2007), 'Self-tuning experience weighted attraction learning in games', *Journal of Economic Theory* **133**(1), 177–198.

Hofbauer, J. and Sandholm, W. H. (2009), 'Stable games and their dynamics', *Journal of Economic Theory* **144**(4), 1665–1693.

Hopkins, E. and Posch, M. (2005), 'Attainability of boundary points under reinforcement learning', *Games and Economic Behavior* **53**(1), 110–125.

Houba, H., van der Laan, D. and Veldhuizen, D. (2010), 'Endogenous entry in lowest-unique sealed-bid auctions', *Theory and Decision* p. forthcoming.

Hsu, S.-H., Huang, C.-Y. and Tang, C.-T. (2007), 'Minimax play at wimbledon: Comment', *American Economic Review* **97**(1), 517–523.

Laland, K. N. (2001), Imitation, social learning, and preparedness as mechanisms of bounded rationality, *in* G. Gigerenzer and R. Selten, eds, 'Bounded Rationality', MIT Press, Boston, chapter 13, pp. 233–247.

Ljung, L. (1977), 'Analysis of recursive stochastic algorithms', *IEEE Trans. Automatic Control* **22**, 551–575.

Mohlin, E., Östling, R. and Wang, J. T.-y. (2012), The lowest unique bid auction: Equilibirum and stability properties. Mimeo.

Myerson, R. B. (1998), 'Population uncertainty and poisson games', *International Journal of Game Theory* **27**, 375–392.

Nash, J. (1950), *Non-cooperative Games*, Princeton University.

Ockenfels, A. and Roth, A. E. (2004), Convergence of prices for a new commodity: 'iraq most wanted' cards on ebay. Mimeo.

Östling, R., Wang, J. T.-y., Chou, E. Y. and Camerer, C. F. (2011), 'Testing game theory in the field: Swedish lupi lottery games', *American Economic Journal: Microeconomics* **3**(3), 1–33.

Palacios-Huerta, I. (2003), 'Professionals play minimax', *Review of Economic Studies* **70**(2), 395–415.

Pigolotti, S., Bernhardsson, S., Juul, J., Galster, G. and Vivo, P. (2012), 'Equilibrium strategy and population-size effects in lowest unique bid auctions', *Physical Review Letters* **108**, 088701.

Radicchi, F., Baronchelli, A. and Amaral, L. A. N. (2012), 'Rationality, irrationality and escalating behavior in lowest unique bid auctions', *PLoS One* **7**(1), e29910.

Rapoport, A., Otsubo, H., Kim, B. and Stein, W. E. (2009), Unique bid auction games. Mimeo.

Raviv, Y. and Virag, G. (2009), 'Gambling by auctions', *International Journal of Industrial Organization* **27**, 369–378.

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., Fogarty, L., Ghirlanda, S., Lilicrap, T. and Laland, K. N. (2010), 'Why copy others? insights from the social learning strategies tournament', *Science* **328**, 208–213.

Robbins, H. and Monro, S. (1951), 'A stochastic approximation method', *Annals of Mathematical Statistics* **22**, 400–407.

Rosenthal, R. W. (1973), 'A class of games possessing pure-strategy nash equilibria', *International Journal of Game Theory* **2**(1), 65–67.

Roth, A. E. (1995), Introduction to experimental economics, *in* A. E. Roth and J. Kagel, eds, 'Handbook of Experimental Economics', Princeton University Press, Princeton, chapter 1, pp. 3–109.

Roth, A. and Erev, I. (1995), 'Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term', *Games and Economic Behavior* **8**(1), 164–212.

Sandholm, W. H. (2011), *Population Games and Evolutionary Dynamics*, MIT Press, Cambridge.

Sarin, R. and Vahid, F. (2004), 'Strategy similarity and coordination', *Economic Journal* **114**, 506–527.

Schlag, K. H. (1998), 'Why imitate, and if so, how?: A boundedly rational approach to multi-armed bandits', *Journal of Economic Theory* **78**(1), 130–156.

Schlag, K. H. (1999), 'Which one should i imitate?', *Journal of Mathematical Economics* **31**(4), 493–522.

Sundali, J. and Croson, R. (2006), 'Biases in casino betting: The hot hand and the gambler's fallacy', *Judgement and Decision Making* **1**(1), 1–12.

Vega-Redondo, F. (1997), 'The evolution of walrasian behavior', *Econometrica* **65**(2), 375–384.

Vulkan, N. (2000), 'An economist's perspective on probability matching', *Journal of Economic Surveys* **14**(1), 101–118.

Walker, M. and Wooders, J. (2001), 'Minimax play at wimbledon', *American Economic Review* **91**(5), 1521–1538.

Weibull, J. W. (1995), *Evolutionary Game Theory*, MIT Press.

Young, P. (2004), *Strategic Learning and Its Limits*, Oxford University Press.

Östling, R., Wang, J. T., Chou, E. and Camerer, C. (2010), Testing game theory in the field: Swedish lupi lottery games. Mimeo.

8 Appendix A: Fixed N Player LUPI

Suppose that in every period N players are drawn from the population. The probability that a randomly drawn individual will pick strategy k is p_k. From the point of view of any single player the expected payoff to strategy k, under population average strategy p, is

$$\pi_k(p) = \prod_{i=1}^{k-1} \Pr\left(Y(i) \neq 1 \middle| \bigwedge_{j<i} (Y(j) \neq 1) \right) \left(\Pr\left(Y(k) = 0 \middle| \bigwedge_{j<k} (Y(j) \neq 1) \right) \right).$$

If $p_i < 1$ for all i then $\pi_k(p) > 0$ for all k.

Note that ($s_1 = k \wedge k$ wins) and ($s_2 = k \wedge k$ wins) are independent events since number k only wins if there is a unique player picking it. In the fixed N LUPI game we therefore have

$$\begin{aligned}
\Pr(k = k^*(s) | p) &= \sum_{i=1}^{N} \Pr(s_i = k \wedge k \text{ wins}) \\
&= \sum_{i=1}^{N} \Pr(k \text{ wins } | s_i = k) \Pr(s_i = k) \\
&= \sum_{i=1}^{N} \pi_k(p) p_k \\
&= N p_k \pi_k(p)
\end{aligned}$$

We now prove that the fixed N player LUPI game has a unique symmetric equilibrium, which is completely mixed.

Proposition 9 *There is at least one symmetric equilibrium of the fixed N player LUPI game and any such equilibrium has full support.*

Proof. Existence of a symmetric equilibrium follows directly from Lemma 6 in Dasgupta and Maskin (1986). To see that a symmetric equilibrium must be in mixed strategies, note that if all players pick the same number and there are at least three players, any of the players would be guaranteed to win by deviating to any other number. (If $N = 2$, then both players playing 1 is a symmetric equilibrium.) Hence we have $p_i < 1$ for all i (and consequently $\pi_k(p) > 0$ for all k). To see that a symmetric equilibrium must have full support, first note that $p_1 > 0$, since otherwise picking 1 would guarantee a win. As inductive hypothesis suppose that $p_i > 0$ for all $i \leq k$. If $p_{k+1} = 0$ then

$$\begin{aligned}\frac{\pi_{k+1}(p)}{\pi_k(p)} &= \frac{\prod_{i=1}^{k} \Pr\left(Y(i) \neq 1 \middle| \bigwedge_{j<i}(Y(j) \neq 1)\right)}{\prod_{i=1}^{k-1} \Pr\left(Y(i) \neq 1 \middle| \bigwedge_{j<i}(Y(j) \neq 1)\right)} \\ &\times \frac{\Pr\left(Y(k+1) = 0 \middle| \bigwedge_{j<k+1}(Y(j) \neq 1)\right)}{\Pr\left(Y(k) = 0 \middle| \bigwedge_{j<k}(Y(j) \neq 1)\right)} \\ &= \frac{\Pr\left(Y(k) \neq 1 \middle| \bigwedge_{j<k}(Y(j) \neq 1)\right)}{\Pr\left(Y(k) = 0 \middle| \bigwedge_{j<k}(Y(j) \neq 1)\right)}.\end{aligned}$$

Since $p_k > 0$ this implies $\pi_{k+1} > \pi_k$, which cannot hold in an equilibrium with $p_k > 0$ and $p_{k+1} = 0$. Hence, we must have $p_{k+1} > 0$. By induction we have full support. Q.E.D.

Next we prove that when we consider completely mixed strategies (in line with the previous proposition) then the fixed N player LUPI game is strictly stable. This implies that there is a unique symmetric equilibrium (Hofbauer and Sandholm (2009)).

Proposition 10 *The fixed N player LUPI game has a unique symmetric equilibrium.*

Proof. Consider a focal player who plays strategy k. Let $E(M)$ be the event that up to $k-1$ there was no winning number and M players picked numbers below k, i.e.

$$E(M) = \left(\bigwedge_{j<k}(Y(j) \neq 1)\right) \wedge \left(\sum_{j=0}^{k-1} Y(j) = M\right).$$

We will not write down an explicit expression for $\Pr(E(M))$ but is clear that it will be independent of p_j for all $j \geq k$. Note that

$$\begin{aligned}\Pr\left(Y(k) = 0 \middle| \bigwedge_{j<k}(Y(j) \neq 1)\right) &= \frac{\Pr\left((Y(k) = 0) \wedge \left(\bigwedge_{j<k}(Y(j) \neq 1)\right)\right)}{\Pr\left(\bigwedge_{j<k}(Y(j) \neq 1)\right)} \\ &= \frac{\sum_{M=0}^{N-1} \Pr\left((Y(k) = 0) \wedge E(M)\right)}{\Pr\left(\bigwedge_{j<k}(Y(j) \neq 1)\right)} \\ &= \frac{\sum_{M=0}^{N-1} \left(\Pr\left(Y(k) = 0 | E(M)\right) \Pr\left(E(M)\right)\right)}{\Pr\left(\bigwedge_{j<k}(Y(j) \neq 1)\right)}.\end{aligned}$$

It follows that

$$\begin{aligned}\pi_k(p) &= \prod_{i=1}^{k-1} \Pr\left(Y(i) \neq 1 \,\middle|\, \bigwedge_{j<i}(Y(j) \neq 1)\right) \Pr\left(Y(k) = 0 \,\middle|\, \bigwedge_{j<k}(Y(j) \neq 1)\right) \\ &= \prod_{i=1}^{k-1} \Pr\left(Y(i) \neq 1 \,\middle|\, \bigwedge_{j<i}(Y(j) \neq 1)\right) \\ &\times \frac{\sum_{M=0}^{N-1} \left(\Pr\left(Y(k) = 0 | E(M)\right) \Pr\left(E(M)\right)\right)}{\Pr\left(\bigwedge_{j<k}(Y(j) \neq 1)\right)} \\ &= W \sum_{M=0}^{N-1} \left(\Pr\left(Y(k) = 0 | E(M)\right) \Pr\left(E(M)\right)\right),\end{aligned}$$

where

$$W = \frac{\prod_{i=1}^{k-1} \Pr\left(Y(i) \neq 1 \,\middle|\, \bigwedge_{j<i}(Y(j) \neq 1)\right)}{\Pr\left(\bigwedge_{j<k}(Y(j) \neq 1)\right)}.$$

Conditional on not having picked a number below k the probability of picking number k is

$$\tilde{p}_k = \frac{p_k}{1 - \sum_1^{k-1} p_i},$$

and

$$\Pr\left(Y\left(k\right) = 0 | E\left(M\right)\right) = \left(1 - \tilde{p}_k\right)^{N-1-M}.$$

Using this we can write

$$\pi_k(p) = W \sum_{M=0}^{N-1} \left(1 - \tilde{p}_k\right)^{N-1-M} \Pr\left(E(M)\right).$$

Since W and $\Pr\left(E\left(M\right)\right)$ are independent of p_j for all $j \geq k$ we have

$$\frac{\partial \pi_k\left(p\right)}{\partial p_j} = \begin{cases} \sum_{M=0}^{N-1} \left(1 - \tilde{p}_k\right)^{N-1-M} \frac{\partial\left(W \Pr(E(M))\right)}{\partial p_j} & \text{if } j < k \\ -\frac{\partial \tilde{p}_k}{\partial p_k} W \sum_{M=0}^{N-1} \Pr\left(E(M)\right)\left(N-1-M\right)\left(1 - \tilde{p}_k\right)^{N-2-M} & \text{if } j = k \\ -\frac{\partial \tilde{p}_k}{\partial p_j} W \sum_{M=0}^{N-1} \Pr\left(E(M)\right)\left(N-1-M\right)\left(1 - \tilde{p}_k\right)^{N-2-M} & \text{if } j > k \end{cases},$$

Since $\partial \tilde{p}_k / \partial p_j = 0$ for if $j > k$ the Jacobian is lower triangular. Since $\partial \tilde{p}_k / \partial p_k = 1/\left(1 - \sum_{i=1}^{k-1} p_i\right)$ the diagonal is negative for all p such that $p_i < 0$ for all i.[22] In LUPI $\pi_k\left(p\right) > 0$ for all k and interior p. Thus we can use the logic behind proposition 1 to

[22] Note

$$\tilde{p}_k = \frac{p_k}{1 - \sum_{i=1}^{k-1} p_i} = \frac{p_k}{\sum_k^K p_i}.$$

conclude that the Jacobian for the fixed N LUPI game is negative definite for all p such that $p_i < 1$ for all i. Since we know from proposition 9 that any symmetric equilibrium must have full support, and since any strictly stable game has a unique symmetric equilibrium (Hofbauer and Sandholm (2009)), the desired result follows. Q.E.D.

In order to compute a symmetric equilibrium of the fixed N player LUPI game, let q_k denote the conditional probability that a player picks number k conditional on not having guessed a lower number (so q_k depends on $p_1, p_2, ..., p_k$). This implies that $q_1 = p_1$, $q_2 = p_2/(1-q_1)$, $q_3 = p_3/\left[(1-q_1)(1-q_2)\right]$, $q_4 = p_4/\left[(1-q_1)(1-q_2)(1-q_3)\right]$, and so on.

For each number k, and each m with $0 \leq m < N$, we compute recursively the probability $l_{k,m}$ that the there is no winner below k and m *other* players have not guessed numbers below k:

$$l_{k+1,m} = \sum_{m' \geq m, m' \neq m+1} l_{k,m'} \binom{m'}{m} q_k^{(m'-m)} (1-q_k)^m .$$

Based on these probabilities, the probability of winning on each number can be written as

$$w_k = \begin{cases} \sum_{m=0}^{N-1} l_{k,m}(1-q_k)^m & \text{if } k < K, \\ l_{k,0} & \text{if } k = K. \end{cases}$$

As a simple example, consider the case when $N = 3$ and $K = 3$. First, by the assumption that there are 3 players, $l_{1,2} = 1$ and $l_{1,1} = l_{1,0} = 0$. For $k = 2$ and $k = 3$ the corresponding probabilities are

$$\begin{aligned} l_{2,2} &= \sum_{m' \geq m, m' \neq m+1} l_{k,m'} \binom{m'}{m} q_k^{(m'-m)} (1-q_k)^m = l_{1,2} (1-g_1)^2 = (1-q_1)^2 \\ l_{2,1} &= l_{1,1} (1-q_1) = 0 \\ l_{2,0} &= l_{1,0} + l_{1,2} q_1^2 (1-q_1)^0 = q_1^2 \\ l_{3,2} &= l_{2,2} (1-q_2)^2 = (1-q_1)^2 (1-q_2)^2 \\ l_{3,1} &= l_{2,1} (1-q_2) = 0 \\ l_{3,0} &= l_{2,0} + l_{2,2} q_2^2 = q_1^2 + (1-q_1)^2 q_2^2 \end{aligned}$$

We have

$$\frac{\partial}{\partial p_k}\left(\frac{p_k}{1-\sum_{i=1}^{k-1} p_i}\right) = \frac{1}{1-\sum_{i=1}^{k-1} p_i},$$

but

$$\frac{\partial}{\partial p_k}\left(\frac{p_k}{\sum_k^K p_i}\right) = \frac{\sum_k^K p_i - p_k}{\left(\sum_k^K p_i\right)^2}.$$

With the latter expression the Jacobian would not be lower triangular (but still have a negative diagonal). Presumably the resulting Jacobian would still be negative definite with respect to the tangent space.

The win probabilities are given by

$$
\begin{aligned}
w_1 &= l_{1,2}(1-q_1)^2 = (1-q_1)^2 \\
w_2 &= l_{2,0} + l_{2,1}(1-q_2) + l_{2,2}(1-q_2)^2 = q_1^2 + (1-q_1)^2(1-q_2)^2 \\
w_3 &= l_{3,0} = q_1^2 + (1-q_1)^2 q_2^2
\end{aligned}
$$

Setting these probabilities equal gives $q_1 = 2\sqrt{3}-3$ and $q_2 = 1/2$. This implies that $p_1 = 2\sqrt{3}-3 = 0.4641$, $p_2 = (1-2\sqrt{3}+3)/2 = 0.26795$ and $p_3 = 1-2\sqrt{3}+3-(1-2\sqrt{3}+3)/2 = 0.26795$.

In order to compute the equilibrium for larger N and K, we make use of the fact that $w_k = \sum \pi_k/N$ and $\sum \pi_k \in (0,1)$. We start by guessing $\sum \pi_k$, so we can use the expression w_k to solve recursively for all $p_1, ..., p_K$. We search over $\sum \pi_k$ until these probabilities sum to one.

The solid lines in Figure 14 display the symmetric equilibrium for $N = 27$, $N = 100$ and $N = 400$ with $K = 99$ in all figures. After a certain threshold (which is increasing in N), the computed probabilities becomes numerically indistinguishable from zero. The value of K does therefore not affect the numerically computed equilibrium as long as K is sufficiently high. We have not discovered any other symmetric equilibria than those shown in Figure 14, so we conjecture that the LUPI game has a unique symmetric equilibrium. In addition, the equilibrium is practically indistinguishable from the unique equilibrium when the number of players is assumed to be Poisson distributed (see Östling, Wang, Chou and Camerer, 2010 for the full analysis of this case). This is shown by the dashed lines in Figure 14 which correspond to the Possion-Nash equilibrium for the same (average) number of players.

We have not (yet) been able to calculate the equilibrium with a fixed number of players when $N = 53783$ (as in the field data). The Poisson-Nash equilibrium can be used as an approximation of the symmetric equilibrium with a fixed number of players. It was displayed in Figure 1.

9 Appendix B: Further Models of Similarity-Based Imitation

We consider three kinds of similarity based imitation.

1. Imitation of strategies that are similar to the winner
2. Imitation of strategies that are similar to, but distinct from, the winner
3. Imitation of strategies that are similar to, but not higher than, the winner

The first one was discussed in the main text and is the basis of our estimated model. The second and third serves a more conceptual purpose. They are discussed in this appendix

9.1 Imitation of strategies that are similar to, but distinct from, the winner

Given the nature of LUPI it could seem reasonable to pick strategies that are close to previous winners, but not to choose exactly the number that has won. Define

$$\phi_k(j) = \begin{cases} 0 & \text{if } k = j \\ \frac{\max\left\{0,1-\frac{|j-k|}{W}\right\}}{-1+\sum_{i=0}^{K}\max\left\{0,1-\frac{|j-i|}{W}\right\}} & \text{otherwise} \end{cases},$$

This similarity function is displayed in Figure 15b. The similarity-modified payoff to strategy k under strategy profile s, is

$$\check{u}_k(s) = \sum_{j=0}^{K} u_j(s)\,\phi_k(j) = \phi_k(k^*(s)).$$

The expected reinforcement is now

$$\check{\rho}_k(t) = \mathbb{E}[\check{r}_k(t)] = \sum_{j=0}^{K} \rho_j(p)\,\phi_k(j) + c = \sum_{j=0}^{K} np_j\pi_j(p)\,\phi_k(j) + c.$$

and the expected similarity-modified payoff is

$$\check{\pi}_k(p) = \sum_{j=0}^{K} \pi_j(p)\,\phi_k(j).$$

By the same argument as before we have:

Lemma 6 *The Nash equilibrium of the Poisson LUPI game is a Nash equilibrium of the Poisson LUPI game with ϕ-similarity-modified payoffs.*

As above, convergence properties have to be verified numerically.

9.2 Imitation of strategies that are similar to, but not higher than, the winner

Given the nature of LUPI it could seem reasonable to pick strategies that are close to previous winners, but not higher than them. Define

$$\varphi_k(j) = \begin{cases} \max\left\{0, 1 - \frac{|j-k|}{W}\right\} & \text{if } k < j \\ 0 & \text{otherwise} \end{cases},$$

and

$$\mu_k(j) = \frac{\max\left\{0, 1 - \frac{|j-k|}{W}\right\}}{\sum_{i=0}^{K} \varphi_i(j)}.$$

The similarity function is graphed in Figure 15c. The resulting following normalized similarity-modified payoff to strategy k under strategy profile s, is

$$\acute{u}_k(s) = \sum_{j=0}^{K} u_j(s)\,\mu_k(j) = \mu_k(k^*(s)).$$

The expected reinforcement is

$$\acute{\rho}_k(t) = \mathbb{E}[\acute{r}_k(t)] = \sum_{j=0}^{K} \rho_j(p)\,\mu_k(j) + c = \sum_{j=0}^{K} np_j\pi_j(p)\,\mu_k(j) + c.$$

and the expected similarity-modified payoff is

$$\acute{\pi}_k(p) = \sum_{j=0}^{K} \pi_j(p)\,\mu_k(j).$$

By the same argument as before we have:

Lemma 7 *The Nash equilibrium of the Poisson LUPI game is a Nash equilibrium of the Poisson LUPI game with μ-similarity-modified payoffs.*

Note that

$$\begin{aligned}
\frac{\partial \acute{\pi}_k(p)}{\partial p_i} &= \sum_{j=i}^{k} \frac{\partial \pi_j(p)}{\partial p_i} \mu_k(j) \\
&= \begin{cases} \sum_{j=i}^{k} \frac{\partial \pi_j(p)}{\partial p_i} \mu_k(j) & \text{if } i < k \\ \frac{\partial \pi_k(p)}{\partial p_k} \mu_k(k) & \text{if } i = k \\ 0 & \text{if } i > k \end{cases} \\
&= \begin{cases} n \sum_{j=i}^{k} z_i \pi_j(p) \mu_k(j) & \text{if } i < k \\ -n\pi_k(p) \mu_k(k) & \text{if } i = k \\ 0 & \text{if } i > k \end{cases}
\end{aligned}$$

Thus the Jacobian $D\acute{\pi}(p)$ is a lower triangular matrix where the k^{th} diagonal entry is $-n\pi_k(p)\mu_k(k)$. Since $\mu_k(k)$ is independent of p the same argument can be used as in the case of no similarity adjustment of payoffs. Thus the Poisson LUPI game with μ-similarity-modified payoffs is strictly stable. This implies that it has a unique equilibrium. We have:

Proposition 11 *The μ-similarity-based deterministic GPI-dynamic converges to the unique Nash equilibrium from any interior initial condition. The μ-similarity-based stochastic GPI-process almost surely converges to the unique Nash equilibrium.*

10 Appendix C: Belief-Based Learning

In this section we briefly discuss whether belief-based learning can rationalize imitation of previous winning numbers. Suppose a player in the LUPI game uses the winning numbers to update her prior belief about the distribution of all players' play using Bayes' rule. The resulting beliefs depend critically upon the choice prior distribution. In order to simulate such belief-updating, we need a parametrized prior distribution that can be updated by winning numbers. Since we could not find a standard distribution that is flexible enough to capture the typical distribution of play in the data, we used the Poisson equilibrium distribution for different values of n. Recall that for low n this distribution is very steep (like an exponential distribution), while for high n it is very spread out and have the peculiar "concave-convex" shape. Since we simply use this as a parameterized prior distribution, n is simply a parameter of the distribution and should not be confused with the actual number of players in the game. Hereafter we therefore call this distribution parameter x instead.

For each allowed value of x, we calculate the probability that number k wins given that N players play according to the distribution. Let $w_x(k)$ be the probability that number k wins if everybody plays according to the equilibrium distribution with the distribution parameter equal to x. Let p_x^t be the agent's belief in period t that the parameter of the prior distribution is x. Beliefs are updated according to

$$p_x^{t+1} = \frac{w_x^t(k)\, p_x^t + \varepsilon}{\sum_y w_y^t(k)\, p_y^t + \varepsilon},$$

where k is the winning number in period t and where $\varepsilon = 10^{-7}$ for the lab estimations and $\varepsilon = 10^{-27}$ for the field estimations. The reason that we include ε is to ensure that all probabilities are positive – otherwise some probabilities will be numerically zero.

For the field data we allow $x = \{5, 10, 15, ..., 99995, 100000\}$ and set $N = 53783$. Based on the actual winning number, beliefs quickly come close the actual number of players in the field. In the last week, the agent believes that x is most likely to be around 57000. However, in earlier rounds, the agent believes that x is most likely to be higher than 57000. The reason is that in the early rounds "too low" numbers are winning (compared to if everybody played according to the equilibrium distribution with $N = 53783$), which is rationalized by believing that guesses are more spread out than they actually are. This implies that the best response is to pick 1 in all rounds.

We also simulated the same model for the lab data and it is clear that the model cannot rationalize imitation in the lab either. For the lab data we allowed x to take all integer values between 3 and 50. The best-response given updated belief is typically to play 1 (when x is believed to be higher than N) or to pick a number around 8 to 10 (when x is believed to be lower than N).

It is clear that belief-based learning with our particular choice of a parameterized distribution cannot rationalize imitation, neither in the lab or field. This conclusion depends critically on the prior distribution that was chosen. Most likely it is possible in theory to find a prior distribution that would do the trick. For example, if player's prior distribution is the actual distribution from the field for all days it might work. The crucial assumption seems to be the smoothness of the prior distribution. If a low number wins, guesses are believed to be more spread out than in equilibrium (with the correct number of players), which makes is optimal to pick 1. In the data, on the other hand, the relatively low winning numbers are to large extent instead due to the spikiness of the data.

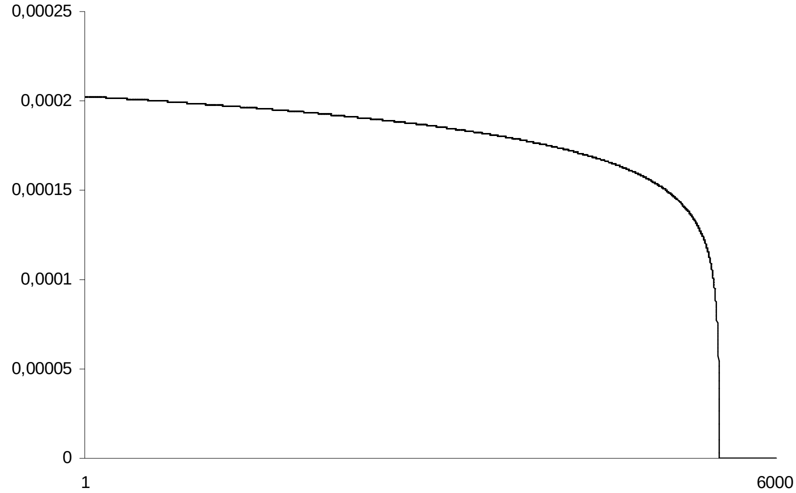


Figure 1. Poisson-Nash equilibrium for *n*=53793 and *K*=99999.

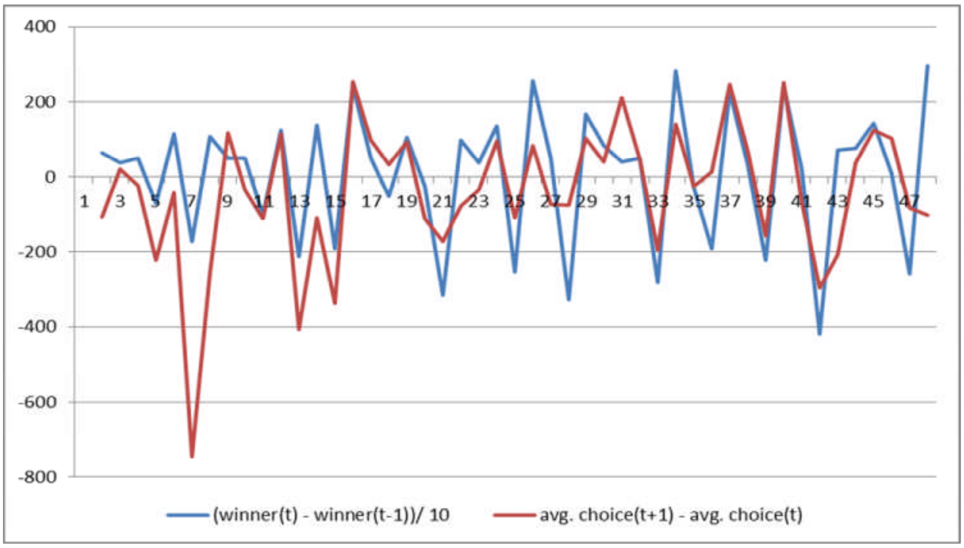


Figure 2. The effect of winning numbers on chosen numbers in the field. The difference between the winning numbers at time t and time t-1 compared with the difference between the average chosen number at time t+1 and time t.

(Winning numbers divided by 10 for comparability, since the variance of winning numbers is larger than variance of mean guesses.)

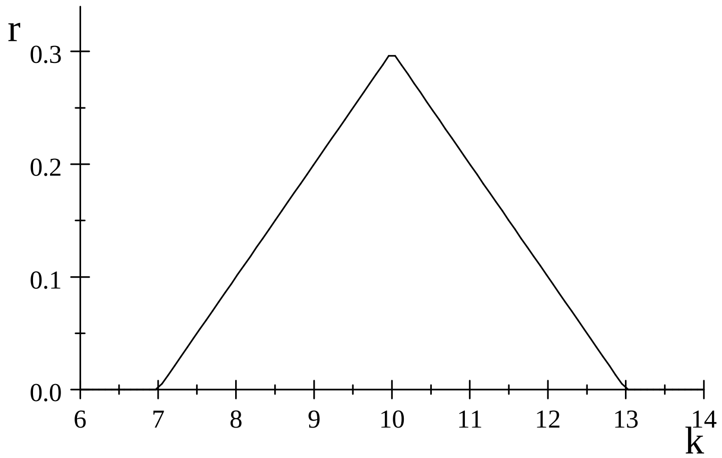


Figure 3. Bartlett similarity window (*k**=10, *W*=3).

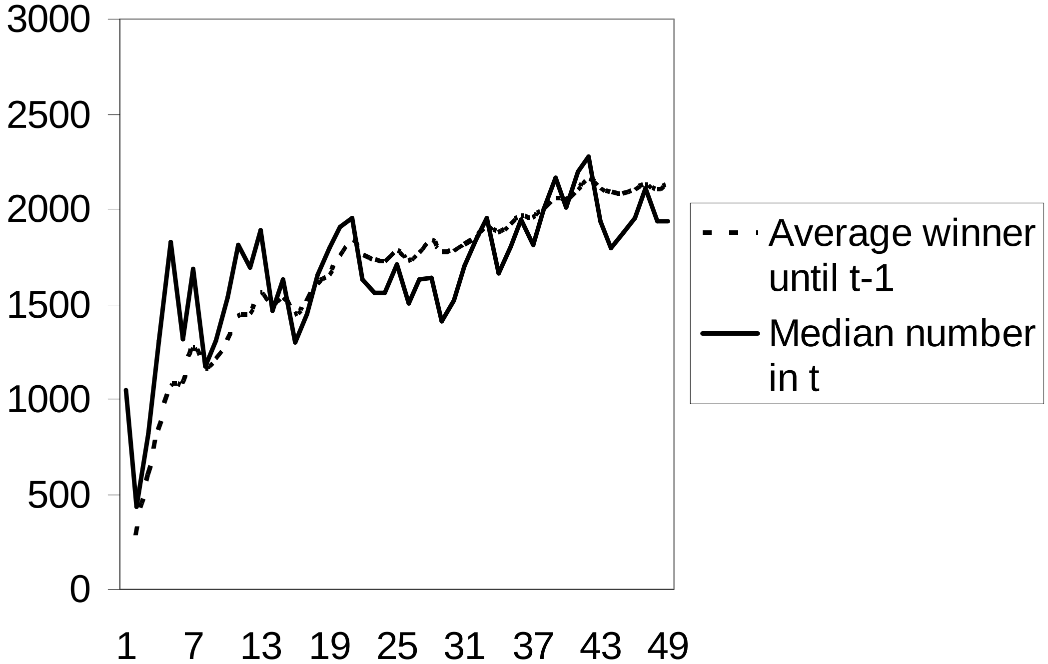


Figure 4. Median winner and median choices in the field

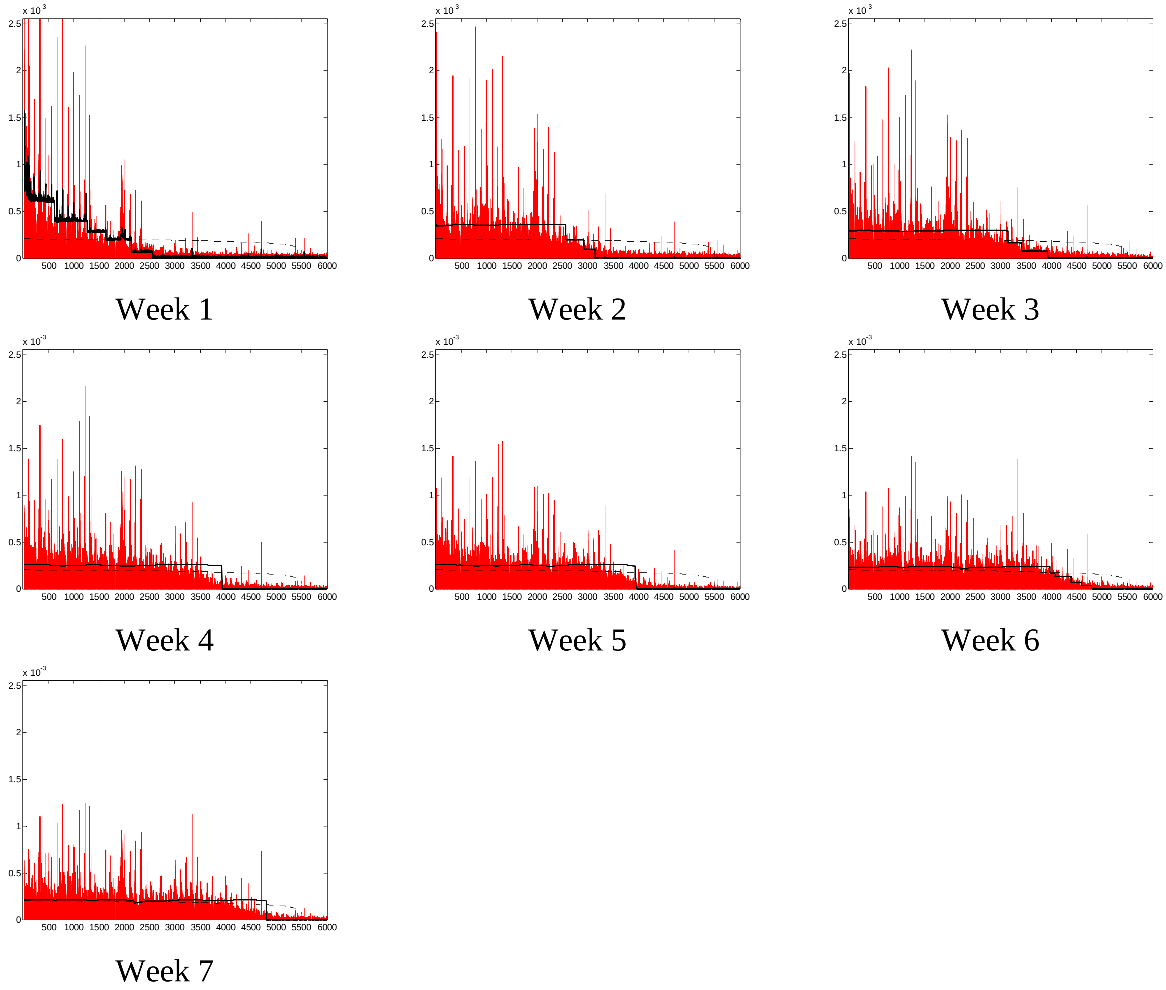


Figure 5. Average weekly empirical densities (bars), estimated learning model (lines) and Poisson-Nash equilibrium (dotted lines) for the field (W = 344, λ = 0.0085).
Note that the learning model fits extremely well in week 1 by construction because it was initialized using actual data from week 1.

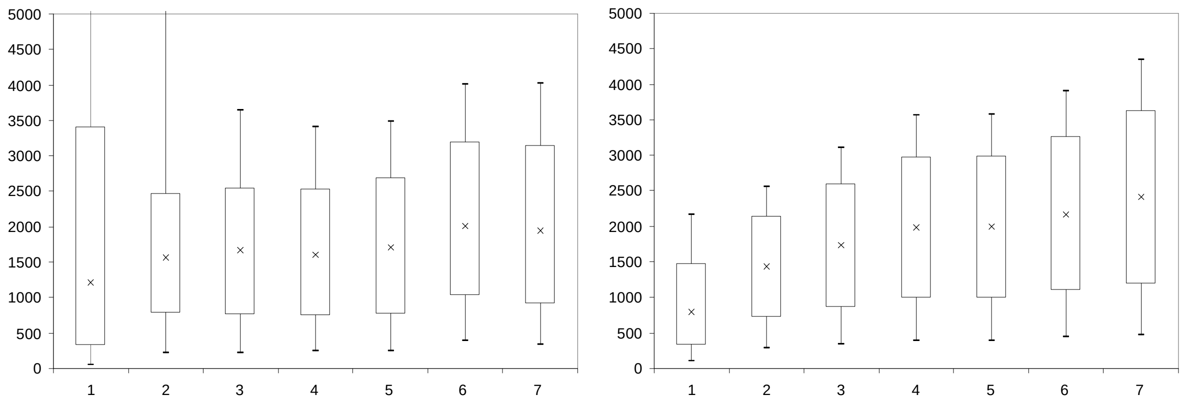


Figure 6. Weekly box plots of data (left) and estimated learning model (right) (10-25-50-75-90 percentile box plots, W = 344, λ = 0.0085).

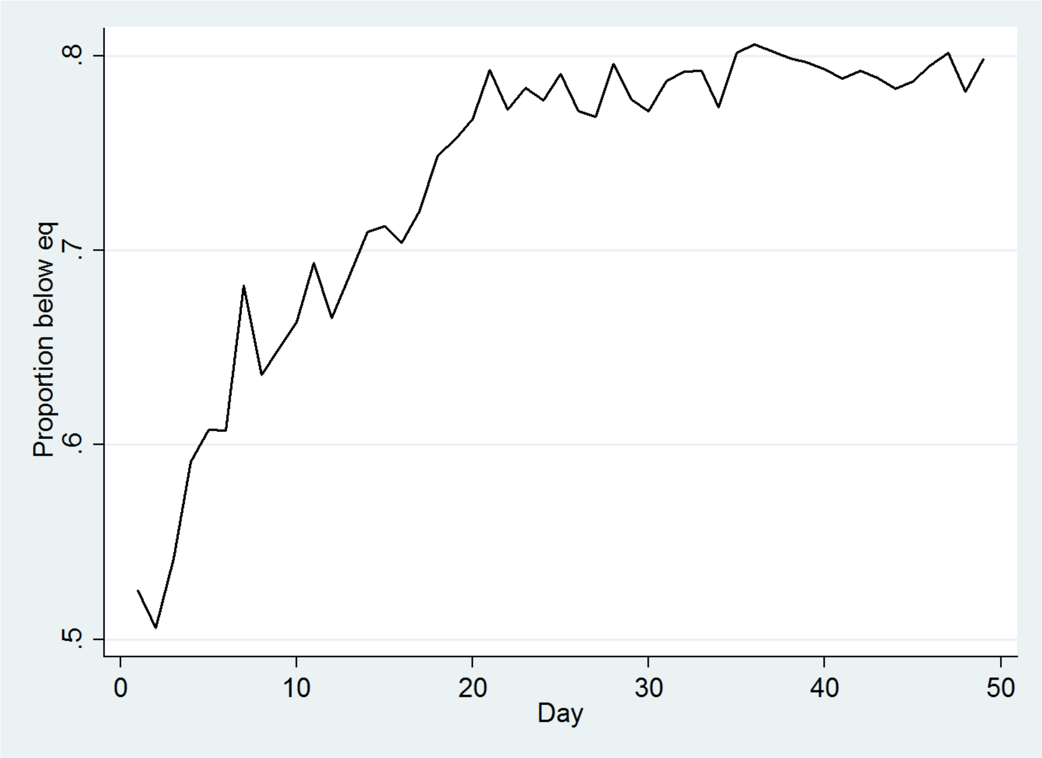
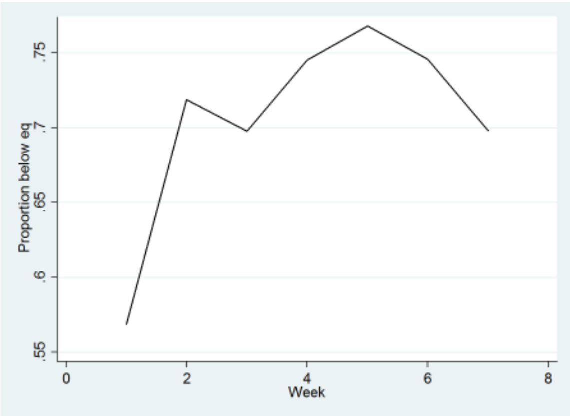
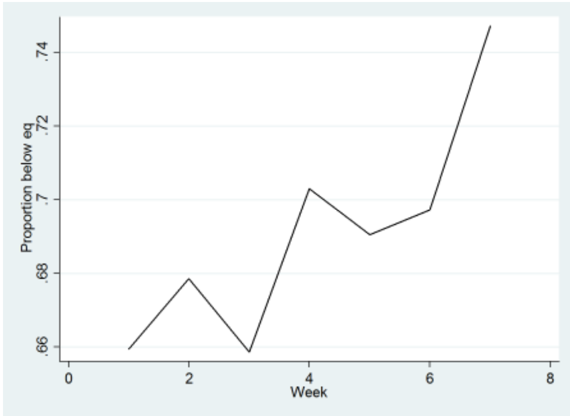


Figure 7. Evidence of movement towards equilibrium in the field. Daily values of the proportion of the empirical distribution that lies below the equilibrium distribution.

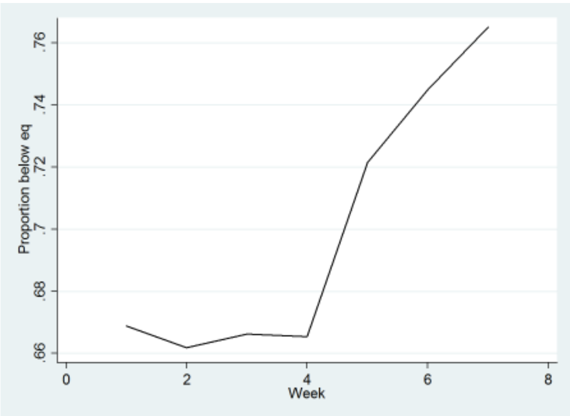
Figure 8. Evidence of movement towards equilibrium in the lab. Weekly values of the proportion of the empirical distribution that lies below the equilibrium distribution.



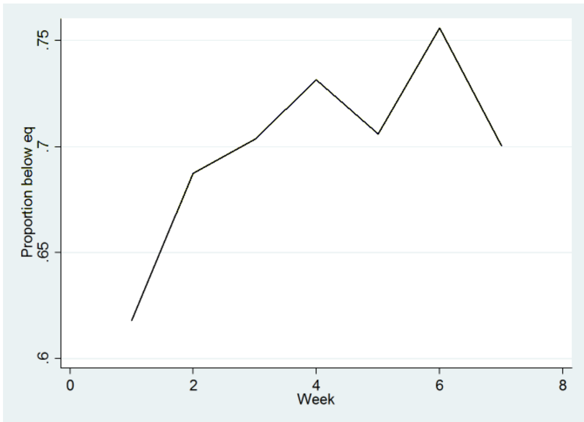
(a) Session 1.



(b) Session 2.

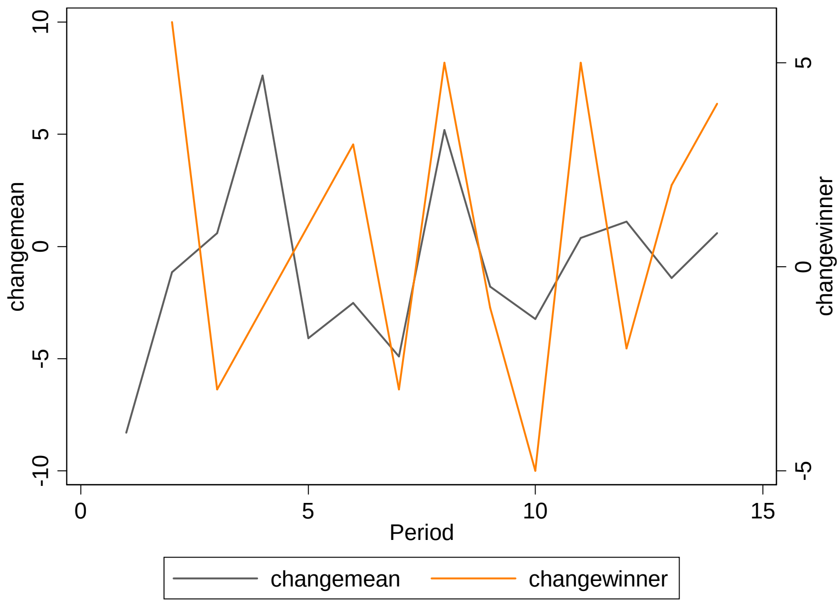


(c) Session 3.

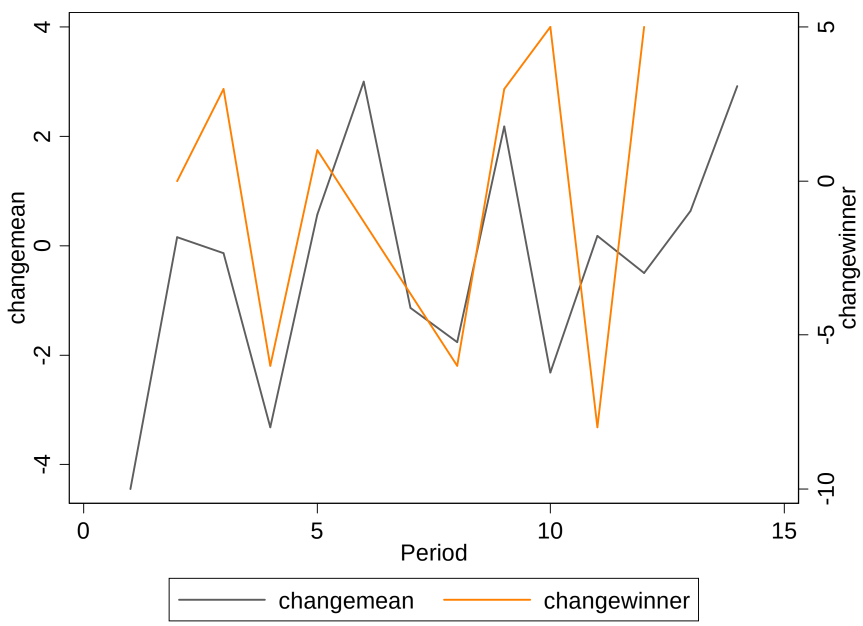


(d) Session 4.

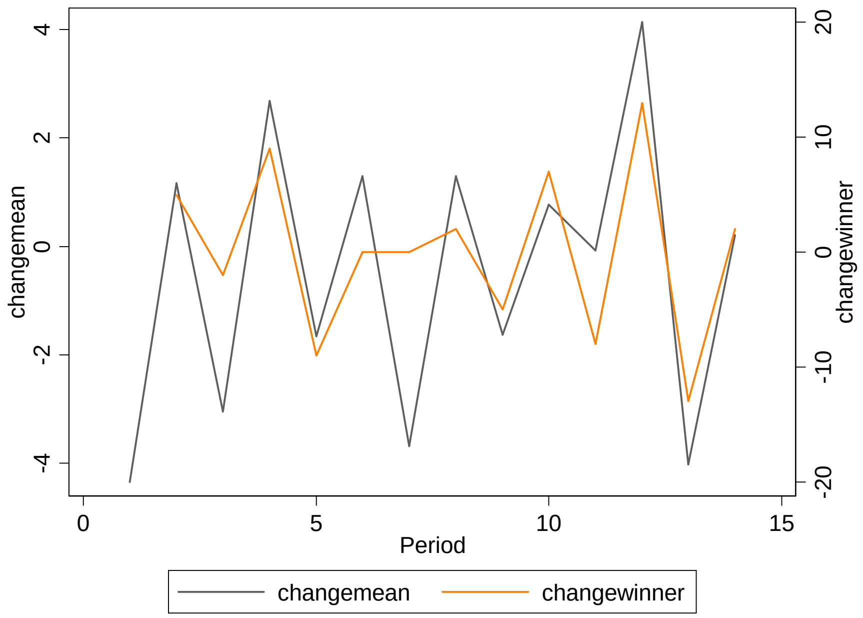
Figure 9. The effect of winning numbers on chosen numbers in the lab.



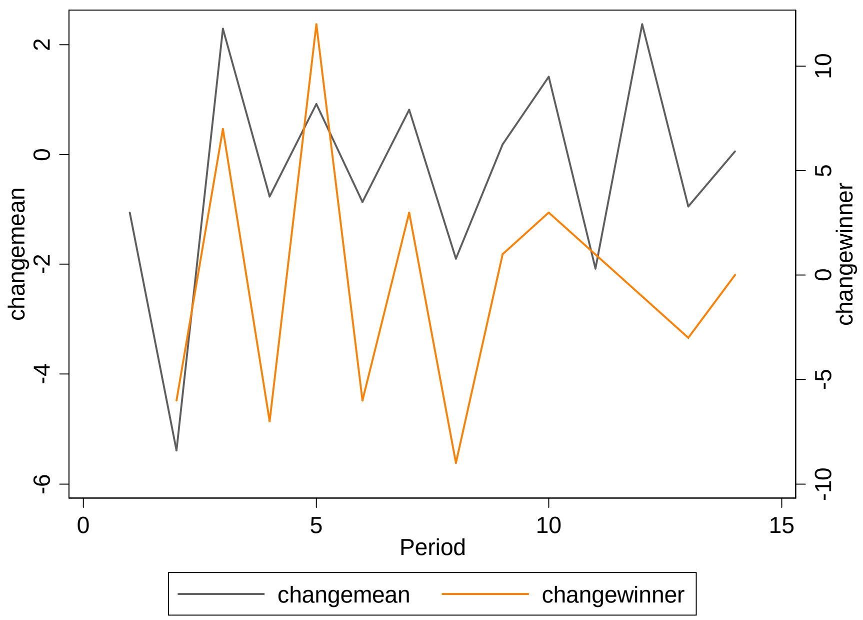
(9a) The difference between the winning numbers at time t and time t-1 compared with the difference between the average chosen number at time t+1 and time t (session 1, winning number of 67 excluded)



(9b) The difference between the winning numbers at time t and time t-1 compared with the difference between the average chosen number at time t+1 and time t (session 2)



(9c) The difference between the winning numbers at time t and time t-1 compared with the difference between the average chosen number at time t+1 and time t (session 3)



(9d) The difference between the winning numbers at time t and time t-1 compared with the difference between the average chosen number at time t+1 and time t (session 4)

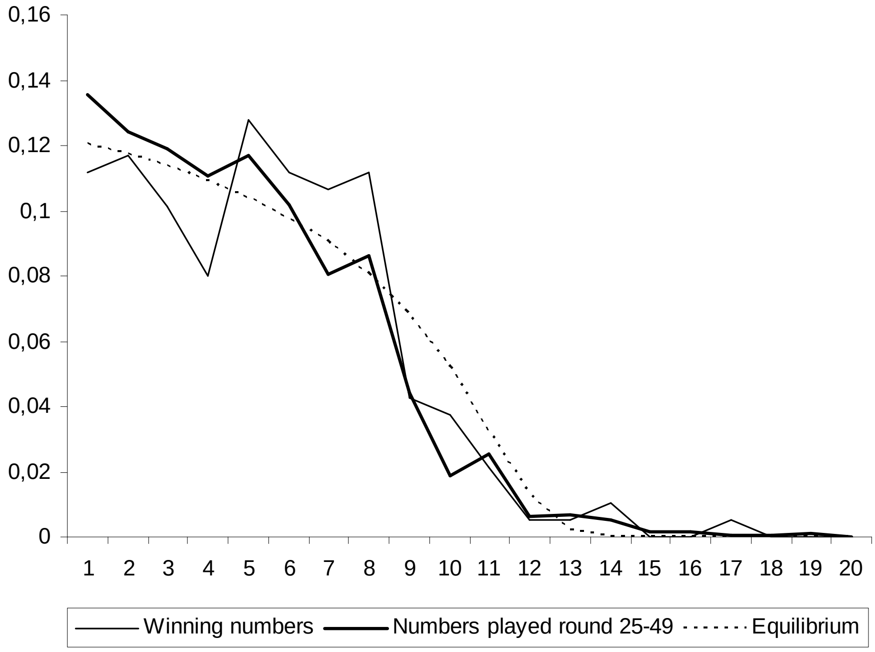


Figure 10. Distribution of winning numbers (all rounds) and numbers played in the last seven rounds.

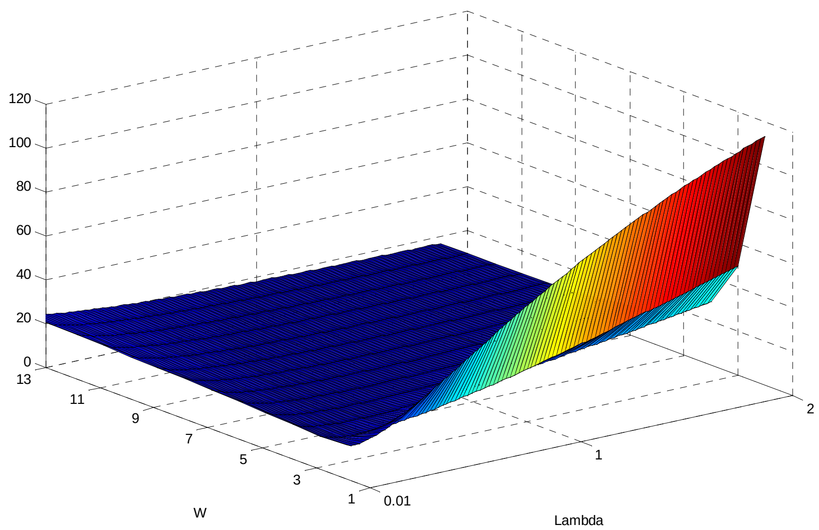
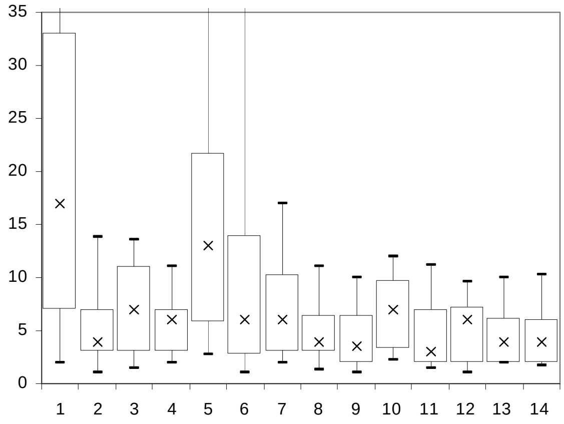
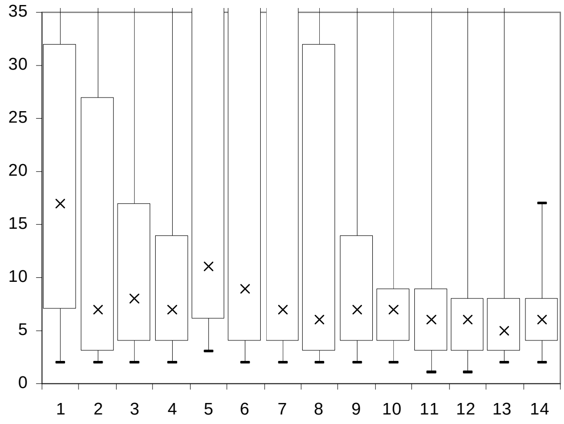


Figure 11. Sum of squared deviation for learning model in the laboratory (W = 1,...,13, λ = 0.01,...,2).

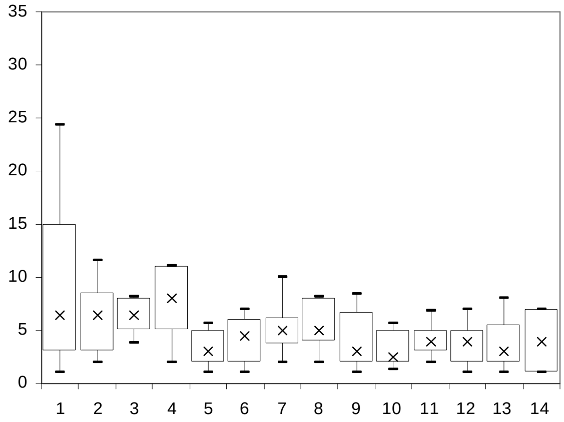
Figure 12: TBA



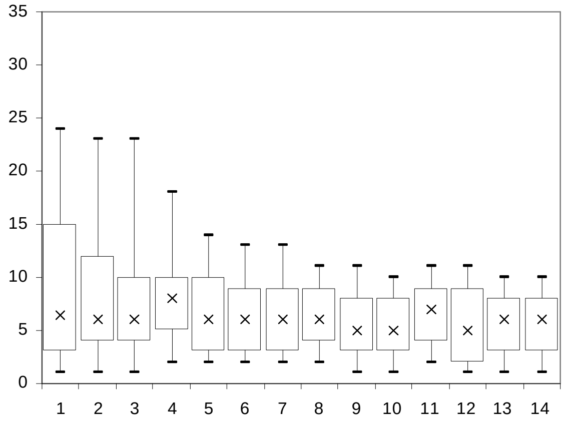
Data (Session 1)



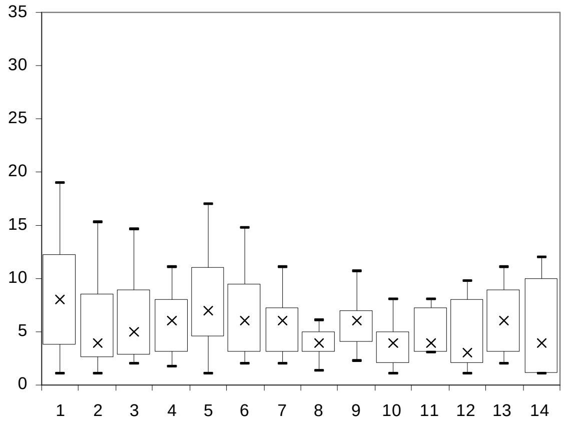
Learning model (Session 1)



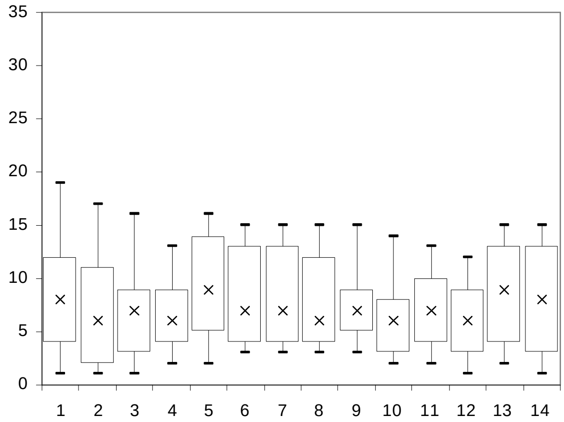
Data (Session 2)



Learning model (Session 2)



Data (Session 3)



Learning model (Session 3)

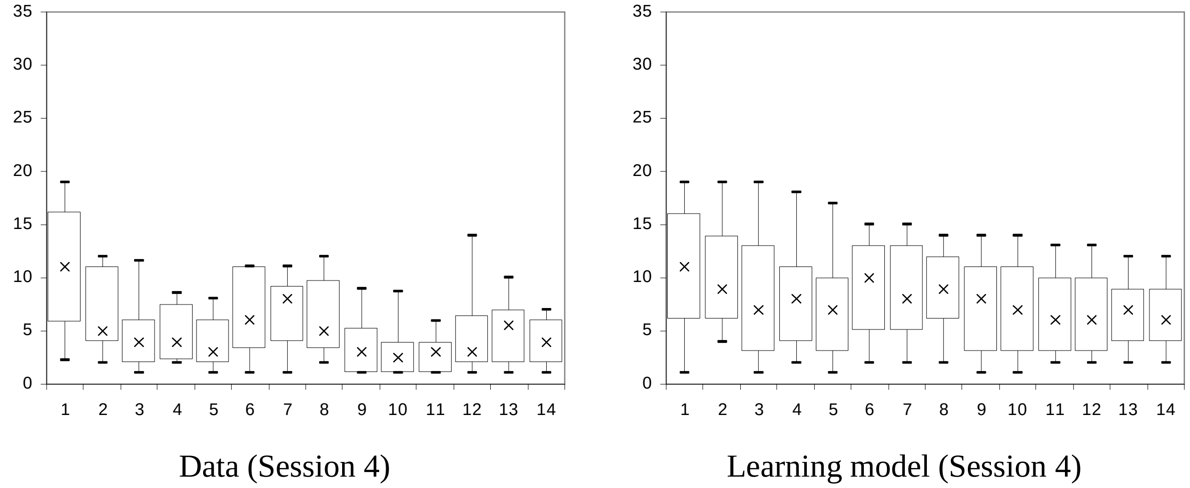


Figure 13. Box plots of data (left) and estimated learning model (right) for round 1-14 in the four laboratory sessions (10-25-50-75-90 percentile box plots, W = 3, λ = 0.31).

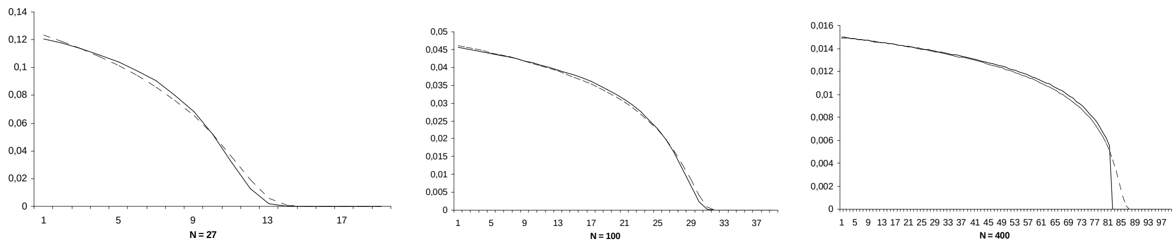
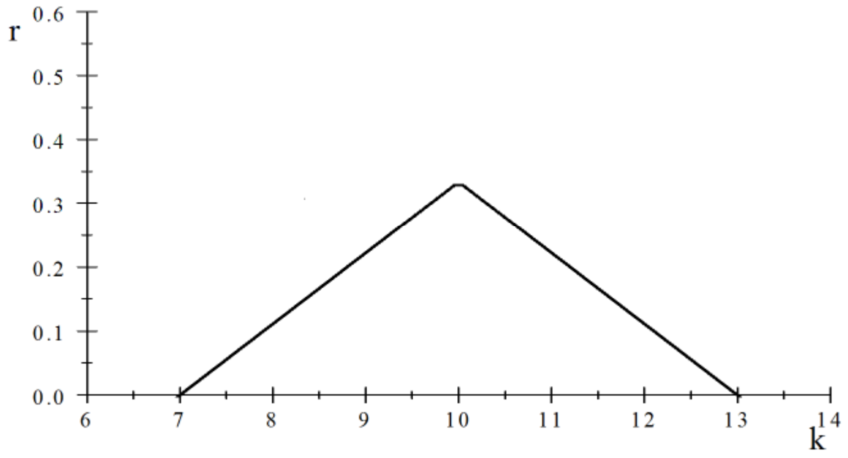


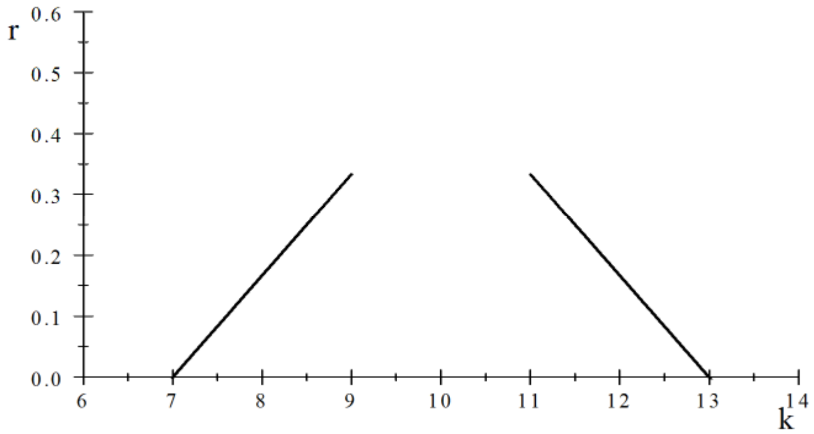
Figure 14. Symmetric Nash equilibrium and Poisson equilibrium for *N*=27, *N*=100 and *N*=400.

Figure 15 a-c. Similarity windows

a) Bartlett similarity window (j=10, W=3)



b) Strategies that are similar to, but distinct from the winning strategy (j=10, W=3).



c) Strategies that are similar to, but not higher than the winning strategy (j=10, W=3).

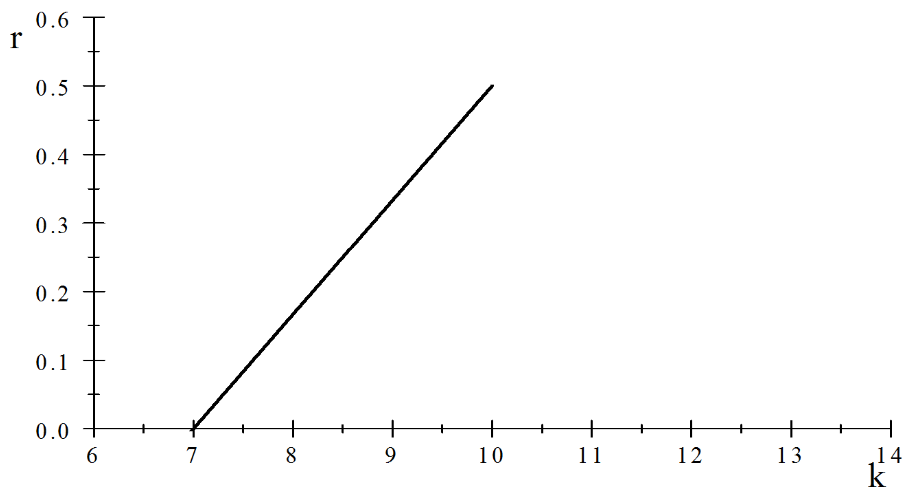
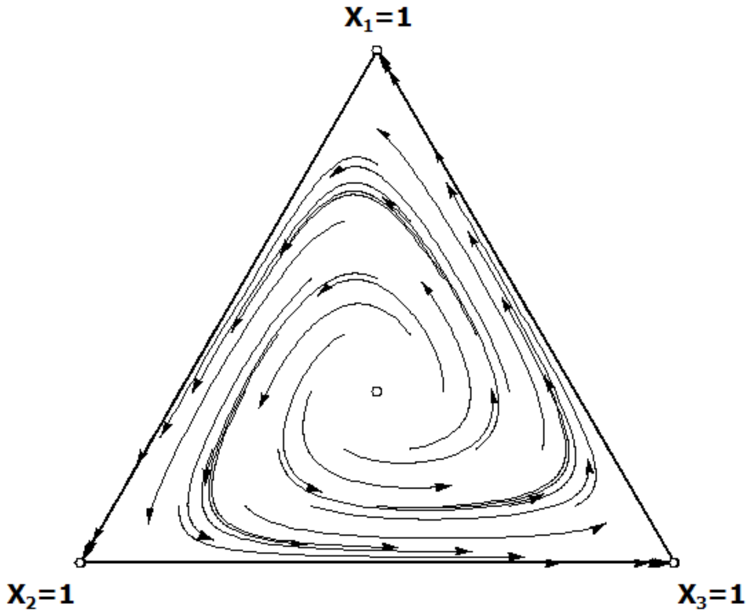
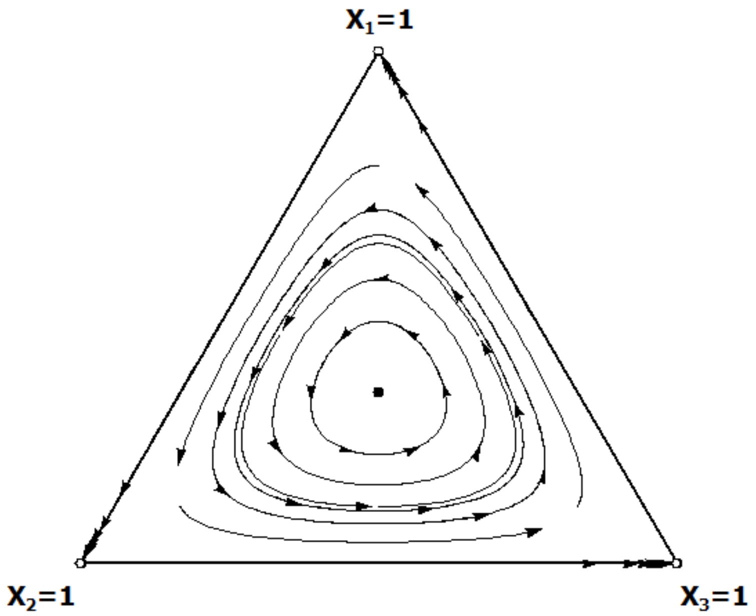


Figure 16 a-b. Phase diagrams for two versions of GPI in Rock Paper Scissors



a) GPI with imitation in proportion to payoffs only, in RPS.



b) GPI with imitation in proportion to payoffs and number of players, in RPS.