# Learning by Similarity-weighted Imitation in Winner-takes-all Games

Erik Mohlin[a], Robert Östling[b], Joseph Tao-yi Wang[c]

[a] *Department of Economics, Lund University. Address: Tycho Brahes väg 1, SE-220 07 Lund, Sweden. E-mail: erik.mohlin@nek.lu.se.*

[b] *Department of Economics, Stockholm School of Economics. Address: P.O. Box 6501, SE-113 83 Stockholm, Sweden. E-mail: robert.ostling@hhs.se.*

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

## Abstract

We study a simple model of similarity-based global cumulative imitation in symmetric games with large and ordered strategy sets and a salient winning player. We show that the learning model explains behavior well in both field and laboratory data from one such "winner-takes-all" game: the lowest unique positive integer game in which the player that chose the lowest number not chosen by anyone else wins a fixed prize. We corroborate this finding in three other winner-takes-all games and discuss under what conditions the model may be applicable beyond this class of games. Theoretically, we show that global cumulative imitation without similarity weighting results in a version of the replicator dynamic in winner-takes-all games.

*JEL codes:* C72, C73, L83.

*Key words:* Learning, imitation, behavioral game theory, evolutionary game theory, stochastic approximation, replicator dynamic, similarity-based reasoning, beauty contest, lowest unique positive integer game, mixed equilibrium.

## 1. Introduction

People learn both from their own and from other's experiences. Imitative learning is often argued to be particularly prevalent in complex situations or when information about the rules of the game is limited (see e.g. Apesteguia et al., 2007, Alós-Ferrer & Weidenholzer, 2014 and Friedman et al., 2015). Although there is evidence that people do learn by imitation in games, relatively little is known about when they imitate, and exactly how they imitate. What kind of imitation that is used matters for the trajectory and rest points of the learning dynamics (Vega-Redondo, 1997, Schlag, 1998, Alos-Ferrer, 2004, Huck et al., 1999) and for the speed of convergence (Roth, 1995). The degree to which a learner relies on imitation, and what form of imitation the player employs, is likely to depend both on the game played and on the feedback available to players.

In this paper, we study a particular type of imitation that we hypothesize is relevant in contexts where information about the most successful player is readily available and salient, while other information that can guide the learning process is scarce. Specifically, we focus on symmetric games with large and ordered strategy spaces in which there is at most one winning player that earns a fixed prize, whereas all other players earn nothing. We call such games *winner-takes-all (WTA)* games. One prominent example belonging to this class of games is the beauty contest game (Nagel, 1995). An intuitively appealing learning heuristic in such games is to imitate strategies that are similar to previous winning strategies.

We are grateful for comments from Ingela Alger, Alan Beggs, Ken Binmore, Colin Camerer, Vincent Crawford, Ido Erev, David Gill, Yuval Heller, and Peyton Young, as well as seminar audiences at the Universities of Edinburgh, Essex, Lund, Oxford, Warwick, and St Andrews, University College London, the 4th World Congress of the Game Theory Society in Istanbul, the 67th European Meeting of the Econometric Society, and the 8th Nordic Conference on Behavioral and Experimental Economics. Kristaps Dzonsons (k-Consulting) and Kaidi Sun provided excellent research assistance. Erik Mohlin acknowledges financial support from the European Research Council (230251), Jan Wallander and Tom Hedelius Foundation (P2016-0079:1), and the Swedish Research Council (2015-01751). 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.

We specify a learning model according to which players react only to feedback about the winning strategies in WTA games. Propensities to play a particular action are updated cumulatively, in response to how often that action, or similar actions, won in the past. We call the resulting learning model similarity-weighted *global cumulative imitation (GCI)*. We analyze the discrete time stochastic GCI process without similarity-weighting in WTA games and show that, asymptotically, it can be approximated by the *replicator dynamic multiplied by the expected number of players*. It is known that any Nash equilibrium is a rest point of the replicator dynamic, and if the replicator converges to an interior state, then it is a Nash equilibrium.

Our proposed learning model was initially designed by combining features from existing learning models to understand learning in a particular WTA game, the *lowest unique positive integer* (LUPI) game. As shown by Östling et al. (2011), players quickly learn to play close to the complicated mixed strategy equilibrium of the LUPI game. In this paper, we show that traditional learning models cannot explain rapid convergence in the LUPI game, but that convergence can be explained by similarity-weighted GCI in both the field and laboratory. Studying the LUPI game has the additional advantage that we can study strategic learning in the field, which is rarely possible (a recent exception is Doraszelski et al., 2018).

To test whether similarity-weighted GCI learning can explain behavior in other games than LUPI, we test the model's out-of-sample performance in three additional WTA games: the *second lowest unique positive integer (SLUPI)* game, the *center-most unique positive integer (CUPI)* game and a variant of the *beauty contest (pmBC)* game. We find our learning model explains behavior at least as well in SLUPI and CUPI as in the LUPI game we designed similarity-weighted GCI for. In the pmBC, convergence to equilibrium is very rapid and our learning model only helps to explain behavior during the first few rounds of play.

Our main reason for focusing on WTA games is that once we move beyond the class of WTA games, GCI learning can take several different forms depending on how propensities are updated when several players earn positive payoffs, i.e., when there is not just a single winner that earns a fixed prize. By focusing on WTA games we avoid this multiplicity of possible specifications. A further reason to focus on WTA games is that the GCI learning process induces a dynamic that can be approximated by the replicator dynamic in WTA games, something that provides us with some theoretical predictions for the long-run behavior of the learning process. In the Appendix we relax the restriction to WTA games and define four different versions GCI learning for general games, distinguished by whether players respond to all payoffs proportionally or only on the highest payoff, and by whether players react to how many players that choose a given strategy. Only one of these combinations will induce a dynamic that can be approximated by the replicator dynamic in general games. The other combinations do not yield known or tractable dynamics.

WTA games are primarily useful as a vehicle for studying global imitation, but may also have some economic interest. For example, the LUPI game is reminiscent of situations in which firms compete to develop new products or researchers compete to develop new scientific ideas: there is both an incentive to find the most profitable product or interesting idea and an incentive to avoid congestion and develop a unique product or idea. Nevertheless, most games of economic interest do not fall within the confines of WTA games. In the final section of the paper, we therefore analyze how the different versions of similarity-weighted GCI performs empirically in two games that resemble WTA games: a Tullock contest and an all-pay auction. We show that players not only imitate the winning player in these games and that imitation of all players proportional to the payoffs received explain behavior better. We therefore conjecture that games beyond the class of WTA games are only conducive to imitation of the highest-earning player when feedback about the winning player is salient and the rules of the game does not make it apparent that solely imitating the winner is suboptimal.

Our proposed learning model assembles parts of pre-existing models. It is most closely related to Sarin & Vahid (2004), Roth (1995) and Roth & Erev (1995). In order to explain quick learning in weak-link games, Sarin & Vahid (2004) add similarity-weighted learning to the reinforcement learning model of Cross (1973), whereas Roth (1995) substitutes reinforcement learning (formally equivalent to the model of Harley, 1981) with a model based on imitating the most successful (highest earning) players (pp. 38–39). Similarly, Roth & Erev (1995) model "public announcements" in proposer competition ultimatum games ("market games") as reinforcing the winning bid (p. 191). Relatedly, Duffy & Feltovich (1999) study whether feedback about one other randomly chosen pair of players affects learning in ultimatum and best-shot games. Whereas the propensity to generalize stimuli according to similarity has been studied relatively little in strategic interactions, it is well-established in non-strategic settings (see e.g. Shepard, 1987). In contrast to most existing models that assume pair-wise imitation, we assume each revising individual imitates globally. Global imitation results in the replicator dynamic multiplied by the expected number of players in WTA games. In contrast, it is well-known that reinforcement learning (and pair-wise cumulative imitation) is generally approximated by the replicator dynamic without the expected number of players as an added multiplicative factor (c.f. Börgers & Sarin, 1997, Hopkins, 2002, and Björnerstedt & Weibull, 1996).

This paper is also related to unpublished work by Christensen et al. (2009) who study learning in LUPI laboratory experiments with rich feedback, but they do not study imitation and find that reinforcement learning performs worse than fictitious play. They also report field data from LUPI's close market analogue the lowest unique bid auction (LUBA), but their data do not allow them to study learning. The latter is also true for other papers that study LUBA, e.g. Raviv & Virag (2009), Houba et al. (2011), Pigolotti et al. (2012), Costa-Gomes & Shimoji (2014) and Mohlin et al. (2015).

The rest of the paper is organized as follows. Section 2 describes the equilibrium of WTA games we study and study GCI learning using stochastic approximation. Section 3 describes and analyzes the field and lab LUPI game. Section 4 analyzes the WTA games that were designed to assess the out-of-sample explanatory power of our model, and shows that other learning models cannot easily explain the data. Section 5 discusses how similarity-weighted GCI performs in some other games and 6 concludes the paper. A number of appendices provide additional results as well as proofs of theoretical results.

## 2. Theoretical Framework

### *2.1. WTA games*

We restrict attention to winner-takes-all (WTA) games in which $N$ players simultaneously choose integers from 1 to $K$. The number of players can be fixed or variable. The pure strategy space is denoted $S = \{1, 2, ..., K\}$ and the mixed strategy space is the $(K-1)$-dimensional simplex $\Delta$. A WTA game is defined by a mapping $k^* : S^N \to S \cup \{\emptyset\}$ that determines the winning number. All players earn zero except the player(s) choosing $k^*(s)$. If only one player chooses $k^*(s)$, that player earns 1, whereas one player is randomly selected to receive 1 if more than one player choose the winning number. If $k^*(s) = \emptyset$, all players earn zero.

### *2.1.1. Equilibrium in the LUPI Game*

In the LUPI game, the lowest uniquely chosen number wins. Let $U(s)$ denote the set of uniquely chosen numbers under strategy profile $s$,

$$U(s) = \{s_j \in \{s_1, s_2, ..., s_N\} \text{ s.t. } s_j \neq s_l \text{ for all } s_l \in \{s_1, s_2, ..., s_N\} \text{ with } l \neq j\}.$$

Then the winning number $k^*(s)$ is given by

$$k^*(s) = \begin{cases} \min_{s_i \in U(s)} s_i & \text{if } |U(s)| \neq 0, \\ \emptyset & \text{if } |U(s)| = 0. \end{cases}$$

Since a unique number cannot be chosen by more than one player, the payoff is

$$u_{s_i}(s) = u(s_i, s_{-i}) = \begin{cases} 1 & \text{if } s_i = k^*(s), \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

Primarily for tractability, we focus on the case then the number of players $N$ is uncertain and Poisson distributed with mean $n$. Let $p$ denote the population average strategy, i.e. $p_k$ is the probability that a randomly chosen player picks the pure strategy $k$. Östling et al. (2011) show the expected payoff to a player putting all probability on strategy $k$ given the population average strategy $p$ is

$$\pi_k(p) = e^{-np_k} \prod_{i=1}^{k-1} \left(1 - np_i e^{-np_i}\right).$$

Östling et al. (2011) show that the LUPI game with a Poisson distributed number of players has a unique (symmetric) Nash equilibrium, which is completely mixed. The equilibrium with 53,783 players (the average number of daily choices in the field) is shown by the dashed line in Figures 3a and 3b below. Östling et al. (2011) and Mohlin et al. (2015) show that the Nash equilibrium with Poission-distributed population uncertainty is a close approximation to the Nash equilibrium with a fixed number of players. Proposition A1 in Appendix A shows another equilibrium property, namely that the probability that $k$ is the winning number is proportional to the probability that $k$ is played.

*2.1.2. Equilibrium in CUPI, SLUPI and pmBC*

In the CUPI game, the winner is the uniquely chosen number in $U(s)$ closest to $(1+K)/2$. If two uniquely chosen numbers are equally close, the higher of the two numbers wins. In other words, the CUPI game is simply the LUPI game with a re-shuffled strategy space.

In the SLUPI game, the winning number $k^*(s)$ is the second lowest chosen number among the set of uniquely chosen numbers $U(s)$. If $|U(s)| < 2$, there is no winner, i.e. $k^*(s) = \emptyset$. SLUPI has $K$ symmetric pure strategy Nash equilibria in which all players choose the same number, but there is no symmetric mixed strategy Nash equilibrium. To see why there is no symmetric mixed strategy equilibrium, note that the lowest number in the support of such an equilibrium is guaranteed not to win. For the expected payoff to be the same for all numbers in the equilibrium support, higher numbers in the equilibrium support must be guaranteed not to win. This can only happen if the equilibrium consists of two numbers, but in that case the expected payoff from playing some other number would be positive.

In a pmBC game (Nagel, 1995, Ho et al., 1998), the winning number $k^*(s)$ is the chosen integer closest to $p$ times the median guess plus a constant $m$. (This scalar $p$ is not to be confused with the population average strategy $p$ defined above.) If more than one player picks $k^*(s)$ one of them is selected at random to be the winner. The unique Nash equilibrium is that all players choose the integer closest to $m/(1-p)$. In our laboratory experiment, $p = 0.3$ and $m = 5$, so the equilibrium is that all players choose number 7.

*2.2. Global Cumulative Imitation (GCI)*

We now define the GCI learning model for WTA games. Time is discrete and in each period $t \in \mathbb{N}$, $N$ individuals from a population are randomly drawn to play a symmetric game. The expected payoff to strategy $k$ under the population average strategy $p$ is denoted $\pi_k(p)$. Note that since the winning prize is normalized to 1 the expected payoff of strategy $k$ is equal to the probability of winning when using strategy $k$.

A learning procedure can be described by an *updating rule* that specifies how the attractions of different actions are modified, or reinforced, in response to experience, and a *choice rule* that specifies how the attractions of different actions are transformed into mixed strategies which then generate actual choices.

*Updating rule.* Let $A_k(t)$ denote the attraction of strategy $k$ at the beginning of period $t$. During period $t$, actions are chosen and attractions are then updated according to

$$A_k(t+1) = A_k(t) + r_k(t), \tag{2}$$

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

To capture global imitation we assume that the reinforcement factor is equal to 1 for the winning strategy and zero for all other strategies, i.e.

$$r_k(t) = \begin{cases} 1 & \text{if } k = k^*(s), \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

*Choice rule.* Consider an individual who uses the mixed strategy $\sigma(t)$ that puts weight $\sigma_k(t)$ on strategy $k$. Attractions are transformed into choice by the following power function (Luce, 1959),

$$\sigma_k(t) = \frac{A_k(t)^{\lambda}}{\sum_{j=1}^{K} A_j(t)^{\lambda}}. \tag{4}$$

Note that $\lambda = 0$ means uniform randomization and $\lambda \to \infty$ means playing only the strategy with the highest attraction. As pointed out by Roth & Erev (1995), this simple choice rule together with accumulating attractions has the realistic implication that the learning curve flattens over time.

In WTA there is a most one winner every period, so at most one action is reinforced every period, Since we consider games with large, ordered strategy sets, reinforcing only the winning number would result in a learning process that is slow and tightly clustered on previous winners. Therefore, we follow Sarin & 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 & Vahid (2004). This function implies that strategies close to previous winners are reinforced and that the magnitude of reinforcement decreases linearly with distance from the previous winner.

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

$$\eta_k(k^*) = \frac{\max\left\{0, 1 - \frac{|k^*-k|}{W}\right\}}{\sum_{i=0}^{K} \max\left\{0, 1 - \frac{|k^*-i|}{W}\right\}}. \tag{5}$$

We set $r_k(t) = \eta_k(k_t^*)$, where $k_t^*$ is the winning number in $t$ (if there is no winning number, reinforcements are zero). The similarity window is shown in Figure 1 for $k^* = 10$ and $W = 3$. Note that the similarity weights are normalized so that they sum to one.

[INSERT FIGURE 1 HERE]
**Figure 1. Bartlett similarity window** ($k^* = 10$, $W = 3$).

*2.3. Stochastic Approximation of GCI*

In this subsection, we study GCI learning in WTA games using stochastic approximation techniques. We derive analytical results for GCI without similarity-weighted imitation and under the assumption that $\lambda = 1$. Later we briefly discuss similarity-weighted GCI.

*2.3.1. Deriving the Perturbed Replicator Dynamic as Approximation*

The updating and choice rules described in the previous section together define a stochastic process on the set of mixed strategies (i.e. the probability simplex). Since new reinforcements are added to old attractions, the relative importance of new reinforcements will decrease over time. This means that the stochastic process moves with smaller and smaller steps. Under certain conditions, the stochastic process will eventually behave approximately like a deterministic process. By finding an expression for this deterministic process, and studying its convergence properties, we are able to infer convergence properties of the original stochastic process.

Recall that $p$ denotes the population average strategy. To simplify the exposition we assume that all individuals have the same initial attractions, so that all individuals play the same strategy, i.e. we assume

$$p_k(t) = \frac{A_k(t)^{\lambda}}{\sum_{j=1}^{K} A_j(t)^{\lambda}}.$$

As we demonstrate in Appendix A, this assumption can be relaxed, to allow individual $i$ to follow strategy $\sigma^i$ and letting $p$ be the average strategy in the population. The reason why this can be done is that all players asymptotically play according to the same strategy because all individuals reinforce the same strategy in all periods and initial attractions are therefore washed out asymptotically.

In order to apply the relevant stochastic approximation techniques, we need reinforcements to be strictly positive. We do this by adding a constant $c > 0$ to all reinforcements (c.f. Gale et al., 1995). Thus, in the context of stochastic approximation, reinforcements are defined as follows

$$r_k^c(t) = \begin{cases} 1 + c & \text{if } k = k^*(s), \\ c & \text{otherwise.} \end{cases} \tag{6}$$

The addition of the constant $c$ can be viewed as a way to represent noise in the perception of payoffs. The constant $c$ must be strictly positive for the stochastic approximation argument to go through, but can be made arbitrarily small (see Appendix A, remark A1).

The stochastic process moves in discrete time. In order to be able to compare it with a deterministic process that moves in continuous time, we consider the interpolation of the stochastic process. The following proposition ties together the interpolated process with a deterministic process. For definitions and explanations of the key concepts from the theory of stochastic approximation, see Appendix A.

**Proposition 1.** *Consider the class of WTA games with a Poisson distributed number of players. Define the continuous time interpolated stochastic GCI process $\tilde{p} : \mathbb{R}_+ \to \Delta$ 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, every $\omega$-limit set of $\tilde{p}$ is a compact invariant set $A \subset \Delta$ that admits no proper attractor, under the flow $\Phi$ induced by the following continuous time deterministic GCI dynamic*

$$\dot{p}_k = np_k \left( \pi_k (p) - \sum_{j=1}^{K} p_j \pi_j (p) \right) + c (1 - Kp_k) . \tag{7}$$

Equation (7) is the replicator dynamic (Taylor & Jonker, 1978) multiplied by $n$ plus a noise term due to the addition of the constant $c$ to all reinforcements. The replicator dynamic is arguably the most well studied deterministic dynamic within evolutionary game theory (Weibull, 1995). Börgers & Sarin (1997) and Hopkins (2002) use stochastic approximation to derive the replicator dynamic, *without* the multiple $n$, from reinforcement learning with decreasing step-size. Björnerstedt & Weibull (1996) (see also Weibull, 1995, Section 4.4) derive the replicator dynamic (without the multiple $n$) from learning by pairwise imitation in the large population limit learning (see also Binmore et al., 1995, and Schlag, 1998). Similarly, we can define pair-wise (cumulative) imitation, and obtain the replicator dynamic (without the multiple $n$) in the limit as step size decreases.[1] Thus we have found that global imitation leads to a faster learning process, and hence potentially faster convergence, than either reinforcement learning or pairwise cumulative imitation.

**Remark 1.** *Proposition 1 concerns games with a Poisson-distributed number of players. If the number of players is fixed and equal to $N$, then we will still obtain the same expression for the continuous time deterministic GCI dynamic (with $N$ in place of $n$) in the limit as $c \to 0$. This follows from propositions B1 and B2 in Appendix B.*

### *2.3.2. Implications for Dynamics and Convergence of GCI*

The fact that the GCI learning process is approximated by the pertrubed replicator dynamic (7) allows us to draw some conclusions about the likely behavior of the GCI learning process itself. It is well known that any Nash equilibrium is a rest point of the (unperturbed) replicator dynamic, and if the replicator dynamic converges to an interior state (a mixed strategy), then it is a Nash equilibrium (Weibull, 1995). Thus in WTA games we may, roughly speaking, expect that the GCI learning process has at least one rest point in the vicinity of a Nash equilibrium. For the LUPI and CUPI games, there is a unique interior rest point, so we may further expect that if the GCI learning process settles down, then it will settle down on an approximate Nash equilibrium. Our next results shows how these rough statements can be made precise and verified for the case of the LUPI and CUPI games.

**Proposition 2.** *Consider the LUPI and CUPI games. There is some $\bar{c}$ such that if $c < \bar{c}$ then the following holds.*

1. *The perturbed replicator dynamic (7) has a unique interior rest point $p^{c*}$.*
2. *If the stochastic GCI process converges to an interior point, then it converges to the unique interior rest point $p^{c*}$ of the perturbed replicator dynamic.*
3. *The stochastic GCI process almost surely does not converge to a point on the boundary, i.e. for all $k$, $\Pr (\lim_{t\to\infty} p_k (t) = 0) = 0$.*

In other words, Proposition 2, establishes that for small enough noise levels the perturbed replicator dynamic (7) has a unique interior rest point (part 1 of Proposition 2). Thus, if the GCI-process converges to an interior point, then it converges to the unique interior rest point of the perturbed replicator dynamic (part 2). In addition to the unique interior rest point, the unperturbed replicator dynamic has rest points on the boundary of the simplex. However, it can be shown that the stochastic GCI process almost surely does not converge to the boundary (part 3). Thus, we know that if the stochastic GCI process converges

[1] We can define pair-wise (cumulative) imitation for a setting with decreasing step-size as follows. As before we assume that strictly positive initial attractors $\{A_i (1)\}_{i=1}^{K}$ are exogenously given, let attractions be updated according to (2), and let the choice rule be (4). We define reinforcement factors in a different way than before. In every period each agent draws one other player as role model and reinforces the action taken by that role model with the payoff earned by the role model. For the same reasons as before we add a constant $c$ to all payoffs. Since the probability of that an action $k$ wins is independent of the total number of players that are realised in a given period, the expected reinforcement is $\frac{1}{n}\mathbb{E}[r_k (t) | \mathcal{F}_t] = p_k (t) \pi_k (p (t)) + \frac{c}{n}$. Plugging this into equation (A3) in Appendix A gives us the replicator dynamic (without a multiple $n$) plus a noise term.

to a point, then it must converge to the unique interior rest point of the perturbed replicator dynamic (7), which as $c \to 0$, moves arbitrarily close to the Nash equilibrium. Our empirical results suggests that learning converges to a point in the simplex. Proposition 2 then implies that if subjects learn by GCI, we should see convergence to the equilibrium (or a $c$-perturbed version thereof).

The results in Proposition 2 do not preclude the theoretical possibility that the stochastic GCI-process could converge to something else than a point, e.g. a periodic orbit. In order to check whether this possibility can be ignored, we simulated the learning process for the LUPI game. We used the lab parameters $K = 99$ and $n = 26.9$, and randomly drew 100 different initial conditions. For each initial condition, we ran the process for 10 million rounds. The simulated distribution is virtually indistinguishable from the equilibrium distribution except for the numbers 11-14, where some minor deviations occur. This is illustrated in Figure C1 in Appendix C. It strongly indicates global convergence of the stochastic GCI process in LUPI.

### *2.4. Similarity-Weighted GCI*

In Appendix B we show that similarity-weighted GCI in WTA games does not result in the replicator dynamic (Proposition B3) and that the Nash equilibrium is not a rest point (not even as noise vanishes). We therefore instead simulated the similarity-weighted GCI process in the LUPI game to examine whether it converges, and to check how the limit point differs from the Nash equilibrium. We use the laboratory parameters, $K = 99$ and $n = 26.9$, and randomly draw 100 different initial conditions. For each initial condition and windows sizes $W = 3$ and $W = 6$, we simulate the learning model for $100,000$ rounds. Figure B1 shows the resulting distribution of end states averaged over the 100 initial conditions. The process does seem to converge, but as expected it does not converge to the Nash equilibrium. For the smaller window size, $W = 3$, the end state is very close to equilibrium, whereas it is a bit further away from equilibrium for the larger window size $W = 6$. In our estimations below we will see that the best-fitting window size is $W = 5$. Thus at least for the lab parameters, adding similarity weights does not seem to affect the qualitative insights gained from the model without similarity window.

### *2.5. GCI for General Games*

In WTA games there is no difference between imitating only the highest earners, and imitating everyone in proportion to their earnings. This is due to the fact that in every round, at most one person earns more than zero. For the same reasons, there is also no difference between imitation which is solely based on payoffs, and imitation which is sensitive both to payoffs and to how often actions are played.

To extend the application of the GCI model beyond WTA games, we need to calculate expected reinforcement more generally. This requires us to make two distinctions. First, imitation may or may not be responsive to the number of people who play different strategies, so we distinguish *frequency-dependent (FD)* and *frequency-independent (FI)* versions of GCI. For simplicity, we assume a multiplicative interaction between payoffs and frequencies, i.e. reinforcement in the frequency-dependent model depends on the total payoff of all players that picked an action. Second, imitation may be exclusively focused on emulating the winning action, i.e. the action that obtained the highest payoff, or be responsive to payoff-differences in a proportional way, so we differentiate between *winner–takes-all imitation (W)* and *payoff-proportional imitation (P)*. In total, we propose the following four members of the GCI family: *PFI*, *PFD*, *WFD*, and *WFI*.

In Appendix B, we discuss these different versions of GCI in greater detail. In particular, we show that in winner-takes-all games, they all coincide if the number of players is Poisson distributed. Furthermore, we show that, in general, it is *only* the payoff-proportional and frequency-dependent version (PFD) of GCI that induces the replicator dynamic multiplied by the expected number of players as its associated continuous time dynamic. PFD can be used in information environments where there is population-wide information available about both payoffs and frequencies of different actions.

In Appendix B we also generalize the similarity-weighted GCI model beyond WTA games. We show that if reinforcements are payoff-proportional and frequency-dependent, then the similarity-weighted GCI induces a relatively tractable deterministic dynamic, namely the replicator dynamic for similarity- and frequency-weighted payoffs, multiplied by the number of players.

## 3. The Field and Lab LUPI Data

The field version of LUPI was introduced by the government-owned Swedish gambling monopoly Svenska Spel on the 29th of January 2007 and subsequently played daily. We utilize daily aggregate choice data from Östling et al. (2011) for the first seven weeks of the game. In the field version of LUPI,

$K = 99,999$ and each player had to pay 10 SEK (approximately 1 euro) for each bet. The total number of bets for each player was restricted to six. It was possible for players to let a computer choose random numbers for them (and such choices cannot be disentangled from the rest). The winner was guaranteed to win at least $100,000$ SEK, but there were also smaller second and third prizes (of $1,000$ SEK and 20 SEK) for being close to the winning number. In the unlikely event that no number was uniquely chosen, the tie-breaking rule stated that the prize would be split among all players choosing the least frequently chosen number. The tie-breaking rule was never implemented, because there were at least $1,298$ uniquely chosen numbers in every round of the game.

Players could access the full distribution of previous choices through the company web site only in the form of raw text files, so few likely looked at it. Information about winning numbers was much more readily available on the web site and in a daily evening TV show, as well as at many outlets of the gambling company, making it the most commonly encountered feedback.

The rules of the field LUPI game differs from the theoretical assumptions in several respects, perhaps most notably the tie-breaking rule, the existence of prizes for being close to the winning number, and the possibility to play up to six numbers. Because simplifications are necessary to analyze the equilibrium of the game, Östling et al. (2011) conducted laboratory experiments that follow the theory much more closely. Their experiment consisted of 49 rounds in each session. Each player was allowed to choose only one number by themselves, there was only one prize of $7 per round, and if there was no unique number, nobody won. Crucially, the only feedback that players received after each round was the winning number. Players were allowed to choose integers between 1 and 99. The number of players in each round was drawn from a distribution with mean 26.9. Due to a mistake in the implementation of the experiment, the distribution of players followed an unknown distribution with a variance that was lower than a Poisson distribution in three out of four sessions. In contrast, the empirical variance of the number of players in the field data is too large to be consistent with a Poisson distribution. A more detailed description of both the field data and laboratory experiments can be found in Östling et al. (2011) and its Online Appendix.

### *3.1. Descriptive Statistics*

The top panel of Table 1 reports summary statistics for the field game averaged over seven days. The last column displays the corresponding statistics that would result from equilibrium play.

In the first week, behavior in the field is quite far from equilibrium: the average chosen number is far above the equilibrium prediction, and both the median chosen number and the average winning number are below what it should be in equilibrium. However, behavior changes rapidly and moves towards equilibrium over time. For example, both average winning numbers and the average numbers played in later rounds are similar to the equilibrium prediction. The median chosen number is much lower than the average number, which is boost by some playing very high numbers, but their difference decreases over time. Table 1 also shows the fraction of all choices that is correctly predicted by the equilibrium prediction, or the "hit rate" (c.f. Selten, 1991).[2] The hit rate increases from about 0.45 in the first week to 0.73 in the last week. The full empirical distribution, displayed in Figure 3 below, also shows clear movement towards equilibrium. However, Östling et al. (2011) can reject the hypothesis that behavior in the last week is in equilibrium.

[2] When defining the hit rate, we treat the mixed strategy equilibrium prediction as a deterministic prediction that a fraction $p(k)$ of all players will play number $k$. The hit rate is then formally defined by $\sum_{k=1}^{K} \min\{p(k), f(k)\}$, where $f(k)$ is the fraction of players that played $k$. The resulting number lies between 0 and 1, but even if all players individually play the equilibrium mixed strategy, the empirical distribution will deviate from the equilibrium prediction distribution and the hit rate will be below 1. Simulations show that the expected hit rate if all players play according to the equilibrium strategy is around 0.87 in the field and 0.74 in the lab game.

**Table 1. Field and lab descriptive statistics by round**

| | All | 1-7 | 8-14 | 15-21 | 22-28 | 29-35 | 36-42 | 43-49 | Eq. |
|---|---|---|---|---|---|---|---|---|---|
| Field | | | | | | | | | |
| # Bets | 53783 | 57017 | 54955 | 52552 | 50471 | 57997 | 55583 | 47907 | 53783 |
| Avg. number | 2835 | 4512 | 2963 | 2479 | 2294 | 2396 | 2718 | 2484 | 2595† |
| Median number | 1675 | 1203 | 1552 | 1669 | 1604 | 1699 | 2057 | 1936 | 2542 |
| Avg. winner | 2095 | 1159 | 1906 | 2212 | 1818 | 2720 | 2867 | 1982 | 2595† |
| Hit rate | 0.64 | 0.45 | 0.59 | 0.65 | 0.65 | 0.68 | 0.73 | 0.73 | 0.87 |
| | | | | | | | | | |
| Laboratory | | | | | | | | | |
| Avg. number | 5.96 | 8.56 | 5.24 | 5.45 | 5.57 | 5.45 | 5.59 | 5.84 | 5.22† |
| Median number | 4.65 | 6.14 | 4.00 | 4.57 | 4.14 | 4.29 | 4.43 | 5.00 | 5.00 |
| Avg. winner | 5.63 | 8.00 | 5.00 | 5.22 | 6.00 | 5.19 | 5.81 | 4.12 | 5.22† |
| Below 20 (%) | 98.02 | 93.94 | 99.10 | 98.45 | 98.60 | 98.85 | 98.79 | 98.42 | 100.00 |
| Hit rate | 0.70 | 0.63 | 0.69 | 0.68 | 0.71 | 0.72 | 0.74 | 0.73 | 0.74 |

†In equilibrium, the distribution of winning and chosen numbers is identical in the the LUPI game.

The bottom panel of Table 1 shows descriptive statistics for the laboratory experiment. We only report data for subjects that were randomly selected to participate in each round. Those not selected were still required to submit a number, but these choices were not incentivized. 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 the median number chosen corresponds closely to the equilibrium after the first seven rounds. The hit rate increases from 0.63 during the first seven rounds to very close to the theoretical maximum in the last 14 rounds. The overwhelming impression from the bottom panel of Table 1 is that convergence (close) to equilibrium is very rapid despite receiving feedback only about the winning number. As an additional indication of equilibrium convergence, Figure D1 in Appendix D shows a close correspondence between the distribution of chosen and winning numbers in all sessions from period 25 and onwards. In equilibrium, these two distributions should coincide (which is shown in Proposition A1 in Appendix A).

Before turning to estimation of the learning model, we analyze whether there is any direct evidence that players imitate previous winning numbers. Figure 2 provides some suggestive evidence that this is indeed the case in the field. Figure 2 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 $t$. In other words, the average number played generally moves in the same direction as winning numbers in the preceding periods.

[INSERT FIGURE 2 HERE]

**Figure 2. Winning and chosen numbers in the field LUPI game.**

The difference between the winning numbers at time $t$ and time $t-1$ compared to the difference between the average chosen number at time $t+1$ and time $t$.

In the laboratory we can regress individual choices on previous winning numbers. However, because choices are likely to be correlated within sessions and we only have data from four sessions, the standard errors must be interpreted with caution. Table 2 displays the results from an OLS regression predicting changes in guesses with lagged differences between winning numbers. Comparing the first 14 rounds with the last 14 rounds, the estimated coefficients are very similar, but the explanatory power of past winning numbers is much higher in the early rounds ($R^2$ is 0.026 in the first 14 rounds and 0.003 in the last 14 rounds). Figure D2 in Appendix D illustrates the co-movement of average guesses and previous winning numbers graphically.

**Table 2. Laboratory panel data OLS regression**

| Dependent variable: $t$ guess minus $t-1$ guess | | | |
|---|---|---|---|
| | All periods | 1–14 | 36–49 |
| $t-1$ winner minus $t-2$ winner | 0.154$^{***}$ | 0.147$^{***}$ | 0.172$^{**}$ |
| | (0.04) | (0.04) | (0.07) |
| $t-2$ winner minus $t-3$ winner | 0.082$^{*}$ | 0.089 | 0.169$^{*}$ |
| | (0.04) | (0.05) | (0.08) |
| $t-3$ winner minus $t-4$ winner | 0.047 | 0.069 | 0.078 |
| | (0.03) | (0.04) | (0.07) |
| Observations | 5662 | 1216 | 1710 |
| $R^2$ | 0.009 | 0.026 | 0.003 |

Standard errors within parentheses are clustered on individuals.
Constant included in all regressions.

*3.2. Estimation Procedure*

The similarity-weighted GCI learning model has two free parameters: the size of the similarity window, $W$, and the precision of the choice function, $\lambda$. In our baseline estimations, we fix $\lambda = 1$ and estimate $W$. When estimating the model, we also need to make assumptions about the choice probabilities in the first period, as well as the initial sum of attractions.

We use the empirical frequencies to create choice probabilities (same for all agents) for the first period. Given these probabilities and $\lambda$, we determine $A(0)$ so that equation (4) gives the actual choice probabilities $\sigma_k(1)$. Because the power choice function is invariant to scaling, the level of attractions is indeterminate. In our baseline estimations, we scale attractions so that they sum to one, i.e., $A_0 \equiv \sum_{k=1}^{K} A_k(0) = 1$. Reinforcement factors are scaled to sum to one in each period, so the first period choice probabilities carry the same weight as each of the following periods of reinforcement. The reinforcement factors $r_k\,(t)$ depend on the winning number in $t$. For the empirical estimation of the learning model, we use the actual winning numbers.

Given the history of previous winning numbers, the learning model provides a prediction for the current period. We estimate the model by minimizing the squared difference between the predicted distribution of choices and the actual distribution of choices. Summing over all periods from the second period and onwards gives us the sum of squared deviations (SSD). Formally, let $p_k\,(W,\lambda)\,(t)$ be the learning model prediction for period $t$ under parameters $W$ and $\lambda$ and let $\sigma_k\,(t)$ be the actual proportion of players who choose action $k$ in period $t$. SSD is defined as

$$SSD\,(W,\lambda) = \sum_{t=2}^{T} \left(p_k\,(W,\lambda)\,(t) - \sigma_k\,(t)\right)^2 ,$$

where $T$ is the total number of periods. We search for the parameter values that minimize $SSD\,(W,\lambda)$ summed over sessions.[3] We obtain standard errors for the laboratory games through the following boostrapping procedure. We sample sessions with replacement and estimate the model for each possible permutation of sessions ($4^4 = 256$ permutations for the laboratory LUPI and $6^6 = 46,656$ for the other WTA games). The standard error correspond to the standard deviation of the estimated parameters in the boostrapped samples.

In order to measure goodness-of-fit and compare different models, we perform leave-one-out cross-validation (see e.g. Hastie et al., 2009). For each laboratory session, we use the remaining sessions to estimate the parameters of the model. We then obtain the model prediction for these parameters and calculate SSD for the excluded session. We do this for each session and sum across sessions to obtain an out-of-sample SSD which we use as the measure of goodness-of-fit. To assess the goodness-of-fit for the equilibrium prediction or the learning model with fixed parameters, we simply compute SSD without cross-validation as no parameters are estimated in these cases.

[3] We focus on minimizing SSD rather than maximizing the likelihood function to facilitate comparison with equilibrium. The equilibrium prediction is numerically zero for most numbers and the likelihood of the equilibrium prediction will therefore always be zero.

**Table 3. Estimation of learning model (all data; $\lambda = 1$)**

| | $A_0 = 0.5$ | | $A_0 = 1$ | | $A_0 = 2$ | | $A_0 = 4$ | |
|---|---|---|---|---|---|---|---|---|
| | $W$ | $SSD$ | $W$ | $SSD$ | $W$ | $SSD$ | $W$ | $SSD$ |
| *Field* (Equilibrium $SSD$ $= 0.0090$) | | | | | | | | |
| Actual | 2117 | 0.0051 | 1999* | 0.0044 | 1369 | 0.0039 | 1190 | 0.0042 |
| Uniform | 1978 | 0.0065 | 1392 | 0.0065 | 1318 | 0.0066 | 1179 | 0.0069 |
| *Laboratory period 1-7* (Equilibrium $SSD = 1.27$) | | | | | | | | |
| Actual | 8 | 1.20 | 6* | 1.21 | 6 | 1.25 | 6 | 1.39 |
| | (0.76) | | (0.72) | | (0.18) | | (0.52) | |
| Uniform | 8 | 1.25 | 6 | 1.32 | 6 | 1.45 | 6 | 1.70 |
| | (0.63) | | (0.70) | | (0.14) | | (0.61) | |
| *Laboratory period 1-14* (Equilibrium $SSD = 2.77$) | | | | | | | | |
| Actual | 6 | 2.82 | 6* | 2.82 | 5 | 2.89 | 5 | 3.08 |
| | (0.29) | | (0.24) | | (0.26) | | (0.52) | |
| Uniform | 6 | 2.91 | 6 | 3.03 | 5 | 3.20 | 4 | 3.58 |
| | (0.35) | | (0.40) | | (0.44) | | (0.61) | |
| *Laboratory period 1-49* (Equilibrium $SSD = 8.54$) | | | | | | | | |
| Actual | 5 | 8.88 | 5* | 8.79 | 4 | 8.80 | 4 | 9.03 |
| | (0.20) | | (0.48) | | (0.26) | | (0.46) | |
| Uniform | 5 | 9.01 | 5 | 9.04 | 4 | 9.27 | 4 | 9.80 |
| | (0.24) | | (0.49) | | (0.26) | | (0.49) | |

Comparison of the cross-validated fit ($SSD$) and estimated window sizes ($W$) of similarity-weighted GCI with $\lambda = 1$. Initial attractions are either actual choices in the first period or uniform distribution. Bootstrapped standard errors within parentheses. Baseline estimates marked with asterisks.

### *3.3. Estimation Results*

For the field data, we search over all window sizes of the Bartlett similarity window between 500 and 2500. (We also verified that smaller/larger windows did not improve the fit.) We find a best-fitting window of 1999, or 3996 numbers in addition to the winning number are reinforced (as long as the winning number is above 1998). The sum of squared deviations between predicted and empirical frequencies is 0.0044, compared to 0.0090 for the equilibrium prediction. This is reported in the top panel of Table 3. The estimated window size is also sensitive to the assumption about initial choice probabilities and attractions. To see this, the top panel of Table 3 shows that the best-fitting window size is smaller if the initial choice probabilities are uniform, but it is also smaller the more weight is given to initial attractions.

Figure 3 displays the daily predicted densities of the learning model for numbers up to 6000 along with the data and equilibrium starting from the second day. To make the figures readable, the data has been smoothed using moving averages (over 201 numbers). The vertical dotted lines show the previous winning number. The main feature of learning is that the frequency of very low numbers shrinks and the gap between the predicted frequency of numbers between 2000 and 5000 is gradually filled in.

[INSERT FIGURE 3A HERE]

**Figure 3A. Learning model, equilibrium and data for day 2-25 in the field LUPI game**

Daily empirical densities (bars), estimated learning model (solid lines), equilibrium (dashed line), and the winning number in the previous period (dotted lines) for the field LUPI game day 2-25. Estimated values $W = 1999$, and $\lambda = 1$. To improve readability the empirical densities have been smoothed with a moving average over 201 numbers

[INSERT FIGURE 3B HERE]

**Figure 3B. Learning model, equilibrium and data for day 26-49 in the field LUPI game**
Daily empirical densities (bars), estimated learning model (solid lines), equilibrium (dashed line), and the winning number in the previous period (dotted lines) for the field LUPI game day 26-49. Estimated values $W = 1999$, and $\lambda = 1$. To improve readability the empirical densities have been smoothed with a moving average over 201 numbers

We can also estimate the model by fitting both $W$ and $\lambda$. To do this, we let $W$ vary from 100 and 2500 and determine the best-fitting value of $\lambda$ through interval search for each window size (we let $\lambda$ vary between 0.005 and 2). The best-fitting parameters are $W = 1310$ and $\lambda = 0.81$. The sum of squared deviations is 0.0043, so letting $\lambda$ vary does not seem to improve the fit of the learning model to any particular extent. Moreover, the sum of squared deviations is relatively flat with respect to $W$ and $\lambda$ when both parameters increase proportionally, so it is challenging to identify both parameters (see Figure D3 in Appendix D). A higher window size $W$ combined with higher response sensitivity $\lambda$ generates a very similar sum of squared deviations (since a higher $W$ is generating a wider spread of responses and a higher $\lambda$ is tightening the response).

When estimating the model for the pooled laboratory sessions, the resulting window size is 5 (bottom panel of Table 3). The sum of squared deviations is 8.79, which is close to the accuracy of the equilibrium prediction (8.54). Indeed, players in the laboratory seem to learn to play the game more quickly than in the field, so there is less learning to be explained by the learning model. The difference between the learning model and equilibrium is consequently larger in early rounds. If only the seven first rounds are used to estimate the learning model, the best-fitting window size is 6 and the sum of squared deviations is 1.21, which is slightly better than the equilibrium fit of 1.27.

The bottom panel of Table 3 also shows the estimated window sizes for different initial choice probabilities and weights on initial attractions. The estimated window size is typically smaller when the initial attractions are scaled up. It is clear that our model works best in the initial rounds of play (when most of the learning takes place). Figures D4 to D7 in Appendix D therefore show the prediction of the learning model along with the data and equilibrium prediction for rounds 2–6 for each session separately. We have also estimated the model allowing $\lambda$ to vary using the first 7 rounds in the laboratory. Like in the field, this only improves fit marginally and the best-fitting $\lambda$ is close to 1. Moreover, like with the field data, it is challeging to estimate both parameters of the model because the sum of squared deviations is relatively flat with respect to $W$ and $\lambda$ (see Figure D8 in Appendix D).

Our learning model assumes a triangular similarity window. To investigate if this is supported by the data, we back out the implied reinforcement factors directly from the data. To do this we assume attractions are updated according to equation (2) and that attractions are transformed into mixed strategies according to equation (4) with $\lambda = 1$. Using the empirical distribution in a period as a measure of the mixed strategy played in that period, and assuming that initial attractions sum to one, we can solve for the implied reinforcement of each action in each period. More precisely the empirical estimate of the reinforcement factor of number $k$ in period $t$ is

$$\widehat{r}_k\left(t\right) = \left[\widehat{p}_k\left(t+1\right) - \widehat{p}_k\left(t\right)\right]\left(t+1\right) + \widehat{p}_k\left(t\right),$$

where $\widehat{p}_k\left(t\right)$ is the empirical frequency with which number $k$ is played in $t$. Note that this estimation strategy does *not* assume reinforcement factors to be similarity-weighted. Although reinforcement factors are non-negative in the learning model, estimated reinforcement factors may be negative if a number is played less than in the previous period.

For the field data, Figure 4 shows the estimated reinforcement factors close to the winning number, averaged over days 2 to 49. The reinforcement factor for the winning number (estimated to be about 0.007) is excluded in order to enhance the readability of the figure. The black line in Figure 4 shows a moving average (over 201 numbers) of the estimated reinforcement factors, which are symmetric around the winning number and could be quite closely approximated by a Bartlett similarity window of about 1000. In a finite sample the empirical frequencies with which a number is played may diverge from the theoretical distribution implied by the attractions. For this reason the empirical estimate of reinforcements may sometimes be negative. Note that the variance of reinforcement factors is larger for numbers far below the winning number, likely due to fewer data when the winning number is below 1000. It may appear surprising that the structurally estimated window size is so much larger than what is suggested by the estimated reinforcements in Figure 4. However, Figure 4 only shows changes close to the winning number, whereas the learning model also needs to explain the "baseline" level of choices. Moreover, if we restrict the similarity window to be 1000, then the sum of squared deviations is 0.0046, i.e. only a slightly worse fit.

[INSERT FIGURE 4 HERE]

**Figure 4. Estimated reinforcement factors in the field LUPI game**

The winning number is excluded. Black solid line represents a moving average over 201 numbers.

Figure 5 shows the reinforcement factors in the lab estimated using the exact same procedure. The top panel of Figure 5 reports the estimated reinforcement factors for all periods in the laboratory, and the results suggest that only the winning number (and the numbers immediately below and above) are reinforced. During the first 14 rounds, however, the window seems to be slightly larger, as shown by the middle panel. However, "reinforcing" the previous winning number might be a statistical artefact: the number that wins is typically picked less than average in that period, so reversion to the mean implies that it will be guessed more often in the next period. The bottom panel of Figure 5 therefore shows the estimated reinforcement factors from a simulation of equilibrium play with $n = 26.9$. Comparing the real and simulated data in Figure 5 suggests that players indeed imitate numbers that are similar to previous winning numbers, but it is not clear to what extent they imitate the exact winning number.

[INSERT FIGURE 5 HERE]

**Figure 5. Estimated reinforcement factors in the laboratory LUPI game**

Top panel: Average over periods 1-49. Middle panel: Average over periods 1-14. Bottom panel: Average of 1000 simulations of 49 rounds of play.

## 4. Other WTA Games and Learning Models

Similarity-weighted GCI seems to be able to capture how players in both the field and the laboratory learn to play the LUPI game. However, the learning model was developed after observing Östling et al's (2011) LUPI data, which raises worries that the model is only suited to explain learning in this particular game. Therefore, we conducted new experiments with SLUPI, CUPI and pmBC. We choose games with relatively complex rules so that it would not be transparent to calculate best responses. We made no changes to the similarity-weighted GCI model after observing the results from these games.

### *4.1. Experimental Design*

Experiments were run at the Taiwan Social Sciences Experimental Laboratory (TASSEL), National Taiwan University in Taipei, Taiwan, during June 23-27, 2014 and July 23-30, 2018. We conducted six sessions with 29 to 31 players in each session. A total of 179 subjects participated. In each session, all subjects actively participated in 20 rounds of each of the three games. The order of the games varied across sessions: CUPI-pmBC-SLUPI in the first session (June 23, 2014), pmBC-CUPI-SLUPI in the second (June 25, 2014), SLUPI-pmBC-CUPI in the third session (June 27, 2014), SLUPI-CUPI-pmBC in the fourth session (July 23, 2018), pmBC-SLUPI-CUPI in the fifth session (July 25, 2018) and CUPI-SLUPI-pmBC in the sixth session (July 30, 2018). The prize to the winner in each round was NT$200 (approximately US$7). Each subject was informed, immediately after each round, what the winning number was (in case there was a winning number), whether they had won in that particular round, and their payoff so far during the experiment. There were no practice rounds. All sessions lasted for less than 125 minutes, and the subjects received a show-up fee of NT$100 in the 2014 sessions and NT$200 in the 2018 sessions in addition to earnings from the experiment (which averaged NT$385, ranging from NT$0 to NT$1400). Experimental instructions translated from Chinese are available in Appendix E. The experiments were conducted using the experimental software zTree (Fischbacher, 2007) and subjects were recruited using the TASSEL website.

*4.2. Descriptive Statistics*

Figure 6 shows how subjects played in the first and last five rounds in the three different games. The black lines show the mixed equilibrium of the CUPI game (with 30 players). Since there is no obvious theoretical benchmark for the SLUPI game, we instead simulate 20 rounds of the similarity-weighted GCI 100,000 times and show the average prediction for the last round. In this simulation, we set $\lambda = 1$ and use the best-fitting window size for the first 20 rounds of the LUPI laboratory experiment ($W = 5$). The initial attractions were uniform.

[INSERT FIGURE 6 HERE]

**Figure 6. Data and theoretical benchmark for SLUPI, pmBC and CUPI**

Empirical densities shown as bars. The solid lines show the equilibrium prediction for CUPI and the simulated similarity-based GCI model for the SLUPI game (period 20 prediction averaged over 100,000 simulations with $W = 5$ and $\lambda = 1$).

It is clear from Figure 6 that players learn to play close to the theoretical benchmark in all three games. The learning pattern is particularly striking in the pmBC game: in the first period, 10% play the equilibrium strategy, which increases to 59% in round 5 and 92% in round 10. In the CUPI game, subjects primarily learn not to play 50 so much – in the first round 30 percent of all subjects play 50 – and there are fewer guesses far from 50. In the SLUPI game, it is less clear how behavior changes over time, but it is clear that there are fewer very high choices in the later periods.

**Table 4. Panel data OLS regression in SLUPI, pmBC, and CUPI**

| Dependent variable: $t$ mean guess minus $t-1$ mean guess | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | SLUPI | | pmBC | | CUPI | | CUPI (trans.) | |
| | 1–20 | 1–5 | 1–20 | 1–5 | 1–20 | 1–5 | 1–20 | 1–5 |
| Change in t–1 | 0.130*** | 0.258*** | 4.616 | 1.568*** | -0.013 | -0.042 | 0.014* | 0.162** |
| | (0.04) | (0.07) | (3.25) | (0.30) | (0.01) | (0.05) | (0.01) | (0.05) |
| Change in t–2 | -0.000 | | -0.053 | | -0.001 | | 0.017* | |
| | (0.01) | | (0.82) | | (0.01) | | (0.01) | |
| Change in t–3 | 0.006 | | -0.384 | | -0.005 | | 0.006 | |
| | (0.01) | | (0.21) | | (0.01) | | (0.01) | |
| Observations | 2,864 | 537 | 2,864 | 537 | 2,864 | 537 | 2,864 | 537 |
| $R^2$ | 0.046 | 0.033 | 0.002 | 0.034 | 0.002 | 0.003 | 0.004 | 0.049 |

Standard errors within parentheses are clustered at the individual level. Constant included in all regressions.

To investigate whether subjects adjust their choices in response to past winners, we run the same kind of regression as we did for the LUPI lab data: OLS regressions with changes in average guesses as the dependent variable, and lagged differences between winning numbers as independent variables. Because we only have data from six sessions, the standard errors in these regressions must be interpreted with caution. In LUPI, pmBC and SLUPI, it is clear that the prediction of similarity-weighted GCI is that lagged differences between winning numbers should be positively related to differences in average guesses. In CUPI, however, it is possible that players instead imitate numbers that are similar in terms of distance to the center rather than similar in terms of actual numbers. Therefore, we also report the results after transforming the strategy space. In this transformation, we re-order the strategy space by distance to the center so that 50 is mapped to 1, 51 to 2, 49 to 3, 52 to 4 and so on. The reason we use this asymmetric transformation rather than simply using the distance to 50 is that our tie-breaking rule is slightly asymmetric; if two numbers are uniquely chosen then the higher number wins. The regression results are reported in Table 4.

In SLUPI and pmBC, it is clear that guesses move in the same direction as the winning number in the previous round during the first five rounds. After the initial five rounds, this tendency is less clear, especially in the pmBC game where players learn to play equilibrium very quickly. In the CUPI, subjects seem to imitate based on the transformed strategy set rather than actual numbers. In the remainder, we therefore report CUPI results with the transformed strategy space. Again, the tendency to imitate is strongest during the first five rounds. It is primarily during these first periods that we should expect our model to predict well, because learning slows down after the initial periods. The relationship between winning and chosen numbers in pmBC, CUPI and SLUPI is illustrated in Figure D9 in Appendix D.

We also estimate the reinforcement factors following the same procedure as in LUPI. The result when all periods are included is shown in Figure 7. Since there is most clear evidence of imitation in early rounds, Figure D10 in Appendix D reports the corresponding estimation when restricting the attention to periods 1-5 only. Figure 7 indicates that there is a triangular singularity window in both SLUPI and CUPI. As Figure D10 reveals, however, this is less clear in early rounds – players seem to avoid imitating the exact winning number from the previous round. In pmBC, players are predominantly playing the previous winning number which is due to the fact that most players play equilibrium after the fifth round. When restricting the attention to the first five periods, estimated reinforcement has a triangular shape, although it is clear that players primarily imitate the winning number and numbers below the winning number. One likely explanation for players choosing numbers below the previous winning number in pmBC is that they understand that the best response is below the previous winning number. In fact, in beauty contests with simpler rules, for example guess half-the-average, many players best respond to the previous winning number and belief-based models may outperform imitation learning models. In fact, in Nagel's (1995) half-the-average beauty contest treatment, very few players imitated winning numbers and a majority approximately best-responded to the previous winning number (which correspond to "L1" in her paper).

[INSERT FIGURE 7 HERE]

**Figure 7. Estimated reinforcement factors for SLUPI, pmBC and CUPI**

*4.3. Estimation Results*

The results in the previous section suggest that the similarity-weighted GCI model might be able to explain the learning pattern observed in the data. To verify this, we set $\lambda = 1$ and fix the window size at $W = 5$, which is the best-fitting window size for the first 20 periods in the laboratory LUPI game. As in our baseline estimation for the LUPI game, we burn in attractions using first-period choices and set the sum of initial attractions to 1. The results are displayed in Table 5 (third row). Table 5 also reports the results for estimating the GCI model separately for each game (first row), estimating the GCI model without similarity, i.e. with $W = 1$ (fourth row). Furthermore, Table 5 reports the fit of the equilibrium prediction for LUPI, CUPI and pmBC, as well as the fit of a dummy model that assumes that current period's play equal the actual distribution of choices in the previous period. In addition, Table 5 reports estimation results based on the reinforcement learning model discussed in the next subsection.

**Table 5. Estimation results for LUPI and other WTA games**

| | LUPI | | SLUPI | | pmBC | | CUPI | |
|---|---|---|---|---|---|---|---|---|
| | $W$ | $SSD$ | $W$ | $SSD$ | $W$ | $SSD$ | $W$ | $SSD$ |
| GCI Estimation | 5 | 4.11 | 5 | 5.74 | 1 | 10.32 | 5 | 5.05 |
| | (0.44) | | (1.46) | | (0.00) | | (0.48) | |
| RF Estimation | 3 | 7.01 | 1 | 10.22 | 1 | 54.54 | 1 | 8.37 |
| | (0.40) | | (1.43) | | (0.00) | | (0.31) | |
| GCI $W = 5$ | 5 | 4.09 | 5 | 5.71 | 5 | 61.26 | 5 | 5.01 |
| GCI $W = 1$ | 1 | 12.99 | 1 | 16.43 | 1 | 10.32 | 1 | 17.92 |
| Equilibrium | | 4.12 | | N/A | | 14.18 | | 5.75 |
| Dummy | | 5.95 | | 6.94 | | 3.06 | | 6.02 |

Comparison of the cross-validated fit ($SSD$) and estimated window sizes ($W$) of similarity-weighted GCI and reinforcement learning, both with $\lambda = 1$, for periods 1-20 in LUPI and the other WTA games. Numbers within parentheses are bootstrapped standard errors. The table also report the fit of GCI using the best-fitting window for the LUPI data ($W = 5$) and without a similarity window ($W = 1$). The last two rows shows the fit of equilibrium and a dummy model in which players play the actual distribution in the previous period.

Table 5 shows that the window size estimated using the LUPI data is identical to the best-fitting window size in both SLUPI and CUPI. In both these games, the fit is considerably poorer without the similarity-weighted window, indicating that similarity is important to explain the speed of learning in these games. The learning model seems to improve a little over the equilibrium prediction for the CUPI game, but not to any large extent. The fit of the learning model is better than the dummy model in all games but pmBC. In the pmBC game, the window estimated using the LUPI data provides a poor fit and the best-fitting window is 1. This is primarily because many players play the equilibrium number in later rounds. If the model is estimated using only the first five periods, the best-fitting window size for pmBC is $W = 3$.[4] Comparing the SSD scores divided by the number of sessions (four in LUPI, six in SLUPI and CUPI) shows that similarity-weighted GCI learning performs no worse in the new games SLUPI and CUPI than in LUPI, the game for which it was initially created.

### *4.4. Alternative Learning Models*

Apart from showing that the LUPI data is consistent with our proposed learning model, it can also be noted that the data cannot easily be explained by existing learning models. The leading example of belief-based learning, fictitious play (see e.g. Fudenberg & Levine, 1998), is not applicable in the feedback environment we study. Standard fictitious play assumes that players best respond to the average of the past empirical distributions, but in the laboratory experiments, players only received information about the winning number and their own payoff. In the field LUPI game it was possible to obtain more information with some effort, but the laboratory results suggest that this was not essential for the learning process. We nevertheless estimate a fictitious play model using the field data and find that the fit is poorer than the imitation-based model. These results are relegated to Appendix F. A particular variant of fictitious play posits that players estimate their best responses by keeping track of forgone payoffs. Again, this information is not available to our subjects, since the forgone payoff associated with actions below the winning number depends on the (unknown) number of other players choosing that number. Hybrid models like EWA (Camerer & Ho, 1999, Ho et al., 2007) require the same information as fictitious play and are therefore also not applicable in this context. Action sampling learning and impulse matching learning (Chmura et al., 2012) as well as myopic best response (Cournot) dynamic suffers from similar problems. In the field, players could possibly best respond to behavior in the previous round, but because the lowest unchosen number was above the winning number during 43 out of 49 days, best-responding players would be indifferent about what to play during most rounds. These observations also apply to SLUPI and CUPI, whereas there are several possible learning models that can explain learning in the pmBC.

One may postulate players could potentially adopt a more general form of belief-based learning with our limited feedback: players enter the game with a prior about what strategy opponents' use, and update their beliefs after each round in response to information about the winning number. In Appendix F, we discuss this possibility and argue that it requires strained assumptions about the prior distribution, as well as a high degree of forgetfulness about experiences from previous rounds of play, in order to explain the data.

Learning based on reinforcement of chosen actions (e.g. Cross, 1973, Arthur, 1993, and Roth & Erev, 1995) *is* consistent with the feedback that our subjects receive in all games we study. However, reinforcement learning is too slow to explain learning in the field game, because only 49 players win and only these players would change their behavior. As shown by Sarin & Vahid (2004), reinforcement learning is quicker if players update strategies that are similar to previous successful strategies. To see whether similarity-weighted reinforcement learning can explain behavior in the laboratory, we compare similarity-weighted GCI with similarity-weighted reinforcement learning. We use the reinforcement learning model of Roth & Erev (1995) since this model is structurally very similar to GCI – the only difference is that under reinforcement learning only actions that one has taken oneself are reinforced.

Table 5 compares the similarity-weighted GCI model with similarity-weighted reinforcement learning. It is clear that GCI results in a better fit than reinforcement learning. In fact, the dummy model also outperforms reinforcement learning. Table 5 also shows that reinforcement learning fits better without a similarity window (i.e. $W = 1$) in all games but LUPI. However, if the model is estimated using only data from the first five periods, reinforcement learning fits better with a similarity window than without one, in all games except pmBC.[5]

[4] Using period 1-5 data only, the sum of squared deviations is 3.60, which is better than equilibrium ($SSD = 12.16$) but worse than the dummy model that assumes that players play like the previous period ($SSD = 2.06$).

[5] The best-fitting window for the reinforcement learning model when only using five periods of data is 3 in LUPI and

One worry when comparing learning models is that player heterogeneity might bias estimates (Wilcox, 2006). We have therefore also repeated the estimations in Table 5 by fitting individual-specific similarity windows. In this estimation, we estimate the best-fitting window size $W_i$ separately for each subject $i$ by minimizing the sum of squared deviations between the learning model's prediction and the subject's choices. The average estimated window sizes are similar to those reported in Table 5, and GCI always has a better fit than reinforcement learning.

## 5. Other Games

In this section, we study similarity-weighted GCI in two games that are not WTA games, but resemble them in that there is a salient winning player. As discussed in Section 2.5, non-WTA games raise the question whether players primarily imitate the strategy that yielded the highest payoff or imitate in proportion to the payoff of all strategies.

The first game we consider is the Tullock contest. In a Tullock contest, a prize is awarded to one player and the probability of winning is proportional to the player's share of total effort. Provided that players do not provide higher effort than the value of the prize, the winning player is also the player who receives the highest payoff. In contrast to WTA games, however, the non-winning players earn different payoffs depending on their strategy choices and players may react to the payoffs of non-winning players as well. To investigate this possibility, we analyze data from Mago et al's (2016) four-player Tullock contest experiment. In the experiment, the same group of players play a contest game with a prize of 80 for 20 periods. Players could choose integers between 0 and 80. We focus on the two treatments with complete feedback about all players' strategy choices and payoffs ($N = 120$). The grey bars in Figure 8 shows average play across all sessions during the last five rounds of the experiment. It is immediately apparent from the data that the fraction of players choosing to play 0 or 5 is large. Very low numbers rarely win (numbers below 6 won in 28 out of 600 cases) and imitation based only on previous winning strategies can therefore not explain the popularity of these numbers. We therefore estimate the version of GCI that assumes that players imitate in proportion to the total payoff received by each strategy (PFD-GCI). The black line in Figure 8 show the average model prediction for the last five periods.

[INSERT FIGURE 8 HERE]

**Figure 8. Data and learning model for the Tullock contest**

Bars show empirical densities and solid lines the similarity-weighted PFD-GCI learning model for periods 15-20 in the Tullock contest. A payoff of 80 has been added to each players' payoff to make reinforcements positive. Initial attractions are based on first-period choices and scaled to sum to 320 (which is the total payoff in a round if all players play 0). Estimated window size is $W = 1$ and precision $\lambda = 0.76$.

The best-fitting window size is 1 ($\lambda = 0.76$), so similarity does not play a role, which is related to the fact that players predominantly play numbers divisible by 5 in this experiment. It is also clear from Figure 8 that the model somewhat underpredicts the popularity of number 5 (which rarely wins), and somewhat overpredicts some intermediate numbers (that win much more often). Figure D11 shows that the model appears to capture the speed and direction of learning during most sessions.

The second game we consider is Gneezy and Smorodinsky's (2006) all-pay auction experiment with 4 to 12 players. In their experiment, players played for 10 round and could bid any amount for a prize worth 100. The prize is allocated to the highest bidder, all players pay their bid, and ties are broken randomly. Except in a few cases where bids exceed 100, the winning player earns the highest payoff. However, just like in a contest, non-winning players that play 0 earns the second-highest payoff. Gneezy & Smorodinsky (2006) report that play largely follows a bimodal distribution: some play numbers close to 100 whereas a majority of players play numbers close to 0. Just like in the Tullock contest, low numbers rarely win (bids below 10 never won in the experiment) and imitation of only previous winning bids cannot explain why so many players play low numbers. We therefore estimated the PFD-GCI for this experiment as well. The data and model prediction for the last five rounds are shown in Figure 9.

CUPI, and 5 in SLUPI. Both similarity-weighted GCI and the dummy model fit better than reinforcement learning also when only using data from the first five periods.

[INSERT FIGURE 9 HERE]

**Figure 9. Data and learning model for the all-pay auction**

Bars show empirical densities and solid lines the similarity-weighted PFD-GCI learning model for periods 5-10 in the all-pay auction. A payoff of 105 has been added to each players' payoff to make reinforcements positive. Choices have been rounded off to the nearest integer. Initial attractions are based on first-period choices and scaled to sum to 105 multiplied by the by the number of players in the session. Estimated window size is $W = 1$ and precision $\lambda = 0.93$.

The best-fitting window size is 1 ($\lambda = 0.93$), so there is no evidence of similarity-weighted imitation in this game either. Figure D12 shows that the model tracks the speed and direction of learning in all sessions, but does not capture all period-to-period adjustments in the data. We also estimated PFI-GCI, which assumes that players imitate in proportion to the average payoff of each strategy. This alternative model resulted in a very similar fit in both the Tullock contest and the all-pay auction.

A common feature of contests and all-pay auctions is that it is obvious how to increase the chance of winning by bidding higher, whereas in order to learn to play optimally players have to trade-off a higher probability of winning against the potential costs associated with not winning. A similar observation applies to auctions. For example, it is clear that you increase your chance of winning by rasing your bid in a first-price auction, whereas it is much more difficult to learn how to optimally shade your bid. Because many players are likely to understand that it is not always optimal to imitate previous winning strategies, it is likely that players will not just imitate previous winning bids.

The analysis in this section suggests that payoff-proportional GCI provides a better account of learning in two non-WTA games than imitation of winners only. A more in-depth analysis is required to investigate whether GCI learning also explains the data better than other learning models such as reinforcement and belief-based learning models.

## 6. Concluding Remarks

Different strategic environments evoke different learning heuristics and no model of learning is likely to explain learning across all strategic settings. In this paper we studied similarity-weighted GCI learning and showed it can explain rapid movement toward equilibrium in the LUPI game as well as in some other WTA games. We also carried out some analyses of GCI learning beyond the class of WTA games. We found that players in both a Tullock contest and an all-pay auction not only respond to the strategy choice of the highest-earning player and that behavior in these games is better explained by a version of GCI that assumes players imitate all strategies proportionally to the total or average payoff of each strategy. These results suggest that players generally do not blindly imitate salient winning strategies. However, if the rules of the game are complex, like the WTA games studied in this paper, players may not understand that imitation of the winning strategy is suboptimal and they may therefore respond to salient information about previous winning strategies.

Another reason why players may imitate a salient winning player is because the rules of the games are not disclosed to players. This type of environment has been little studied in the laboratory (recent exceptions are Friedman et al., 2015, Nax et al., 2016 and Oechssler et al., 2016). Outside the laboratory, however, the rules of the strategic interaction are often largely unknown, whereas information about others' successful behavior is abundantly available through the Internet and mass media. For example, stories about the relatively small number of successful entrepreneurs are widely circulated, whereas much less information is available about the majority of entrepreneurs that failed, or did not even get started. Other commonly studied learning models may not even be applicable in such environmentss

Two ingredients of the similarity-weighted GCI model are key to its ability to explain the speed of learning in the LUPI data. The first ingredient is that imitation is global, i.e. players imitate all players' strategy choices. This is crucial for explaining rapid learning in the LUPI game and the other WTA games we study – pairwise cumulative imitation or reinforcement learning based only on own experience would imply too slow learning.

The second ingredient of our learning model is that players imitate numbers that are similar to winning numbers. In our model, similarity is operationalized as a triangular window around the previous winning number, but we also test this assumption by estimating similarity weights directly from the data. In our estimation of similarity weights, we assume choice probabilities are given by the ratio of attractions and that attractions are updated by simply adding reinforcement factors. These two assumptions are common features of many learning models, so a similar estimation procedure may prove useful in future research to elicit similarity weights.

Our direct estimation of similarity weights reveal that people's similarity-weighted reasoning generally appears to be more sophisticated than a simple triangular window. In the laboratory experiments, there is some indication that players avoid exactly the winning number in the unique positive integer games, whereas the similarity window is asymmetric in the beauty contest game. Another sign of more sophisticated similarity-weighted reasoning is that players in the CUPI game imitate numbers based on strategic similarity rather than numeric similarity. Moreover, we find no evidence of similarity-weighted reasoning in all-pay auctions and the Tullock contest games. In Tullock contest game players rather seem to categorize strategy sets by mainly considering numbers divisible by five. We leave it to future research to develop a more general understanding of how players mentally process large strategy sets. For some attempts in this direction see Jehiel (2005), Jehiel & Samet (2007), and Mohlin (2014).

## References

Alos-Ferrer, C. (2004), 'Cournot versus walras in dynamic oligopolies with memory', *International Journal of Industrial Organization* **22**(2), 193–217.

Alós-Ferrer, C. & Weidenholzer, S. (2014), 'Imitation and the role of information in overcoming coordination failures', *Games and Economic Behavior* **87**, 397–411.

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

Arthur, W. B. (1993), 'On designing economic agents that behave like human agents', *Journal of Evolutionary Economics* **3**(1), 1–22.

Beggs, A. (2005), 'On the convergence of reinforcement learning', *Journal of Economic Theory* **122**(1), 1–36.

Benaïm, M. (1999), Dynamics of stochastic approximation algorithms, *in* J. Azéma, M. Émery, M. Ledoux & 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. & Métivier, M. (1990), *Adaptive algorithms and stochastic approximations*, Springer-Verlag New York, Inc., New York, USA.

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

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

Börgers, T. & Sarin, R. (1997), 'Learning through reinforcement and replicator dynamics', *Journal of Economic Theory* **77**(1), 1–14.

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

Chmura, T., Goerg, S. J. & Selten, R. (2012), 'Learning in experimental 2x2 games', *Games and Economic Behavior* **76**(1), 44–73.

Christensen, E. N., De Wachter, S. & Norman, T. (2009), Nash equilibrium and learning in minbid games. Mimeo.

Costa-Gomes, M. A. & Shimoji, M. (2014), 'Theoretical approaches to lowest unique bid auctions', *Journal of Mathematical Economics* **52**, 16–24.

Cross, J. G. (1973), 'A stochastic learning model of economic behavior', *The Quarterly Journal of Economics* **87**(2), 239–266.

Doraszelski, U., Lewis, G. & Pakes, A. (2018), 'Just starting out: Learning and equilibrium in a new market', *American Economic Review* **108**(3), 565–615.

Duffy, J. & Feltovich, N. (1999), 'Does observation of others affect learning in strategic environments? an experimental study', *International Journal of Game Theory* **28**(1), 131–152.

Fischbacher, U. (2007), 'z-tree: Zürich toolbox for readymade economic experiments', *Experimental Economics* **10**(2), 171–178.

Friedman, D., Huck, S., Oprea, R. & Weidenholzer, S. (2015), 'From imitation to collusion: Long-run learning in a low-information environment', *Journal of Economic Theory* **155**, 185–205.

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

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

Gneezy, U. & Smorodinsky, R. (2006), 'All-pay auctions–an experimental study', *Journal of Economic Behavior and Organization* **61**, 255–275.

Harley, C. B. (1981), 'Learning the evolutionarily stable strategy', *Journal of Theoretical Biology* **89**(4), 611–633.

Hastie, T., Tibshirani, R. & Friedman, J. (2009), *The Elements of Statistical Learning Theory*, Springer, New York.

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

Ho, T.-H., Camerer, C. & Weigelt, K. (1998), 'Iterated dominance and iterated best response in experimental" p-beauty contests"', *American Economic Review* **88**, 947–969.

Hopkins, E. (2002), 'Two competing models of how people learn in games', *Econometrica* **70**(6), 2141–2166.

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

Houba, H., Laan, D. & Veldhuizen, D. (2011), 'Endogenous entry in lowest-unique sealed-bid auctions', *Theory and Decision* **71**(2), 269–295.

Huck, S., Normann, H.-T. & Oechssler, J. (1999), 'Learning in cournot oligopoly - an experiment', *Economic Journal* **109**(3), C80–C95.

Jehiel, P. (2005), 'Analogy-based expectation equilibrium', *Journal of Economic Theory* **123**, 81–104.

Jehiel, P. & Samet, D. (2007), 'Valuation equilibrium', *Theoretical Economics* **2**, 163–185.

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

Luce, R. D. (1959), *Individual Choice Behavior: A Theoretical Analysis*, Wiley, New York.

Mago, S. D., Samak, A. C. & Sheremeta, R. M. (2016), 'Facing your opponents: Social identification and information feedback in contests', *Journal of Conflict Resolution* **60**, 459–481.

Mohlin, E. (2014), 'Optimal categorization', *Journal of Economic Theory* **152**, 356–381.

Mohlin, E., Östling, R. & Wang, J. T.-y. (2015), 'Lowest unique bid auctions with population uncertainty', *Economics Letters* **134**, 53–57.

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

Nagel, R. (1995), 'Unraveling in guessing games: An experimental study', *American Economic Review* **85**(5), 1313–1326.

Nax, H. H., Burton-Chellew, M. N., West, S. A. & Young, H. P. (2016), 'Learning in a black box', *Journal of Economic Behavior and Organization* **127**, 1–15.

Oechssler, J., Roomets, A., & Roth, S. (2016), ‘From imitation to collusion: a replication’, *Journal of the Economic Science Association* **2**, 13–21.

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

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

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

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

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

Roth, A. & 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.

Sarin, R. & 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.

Selten, R. (1991), ‘Properties of a measure of predictive success’, *Mathematical Social Sciences* **21**, 153–167.

Shepard, R. N. (1987), ‘Toward a universal law of generalization for psychological science’, *Science* **237**(4820), 1317–1323.

Taylor, P. D. & Jonker, L. (1978), ‘Evolutionarily stable strategies and game dynamics’, *Mathematical Biosciences* **40**, 145–156.

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

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

Wilcox, N. T. (2006), ‘Theories of learning in games and heterogeneity bias’, *Econometrica* **74**(5), 1271–1292.

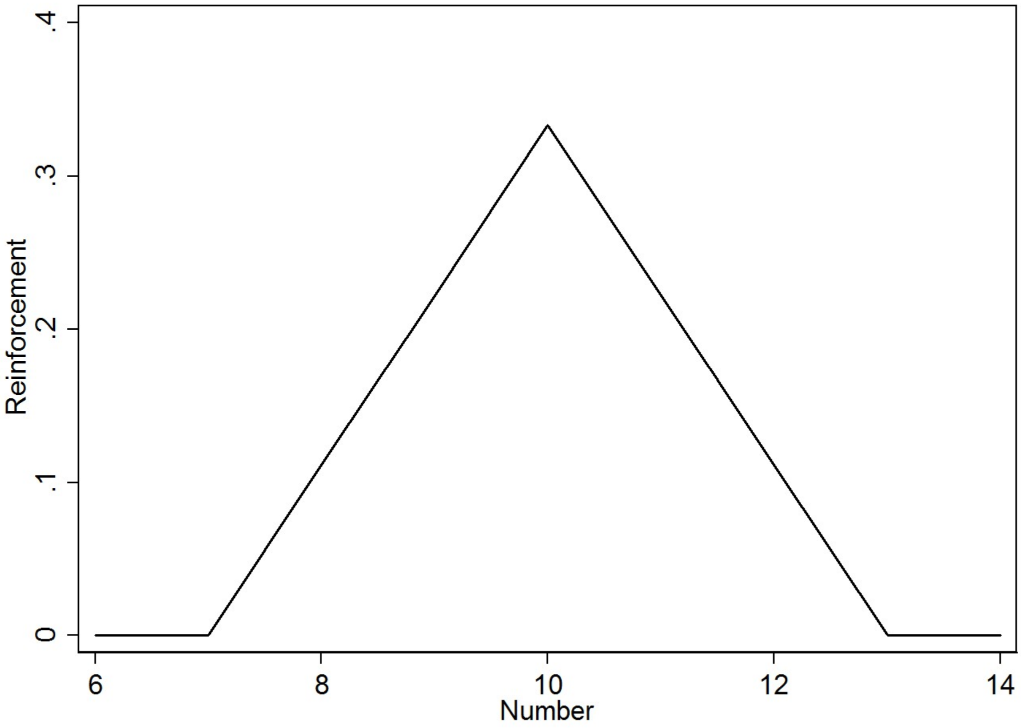


**Figure 1**

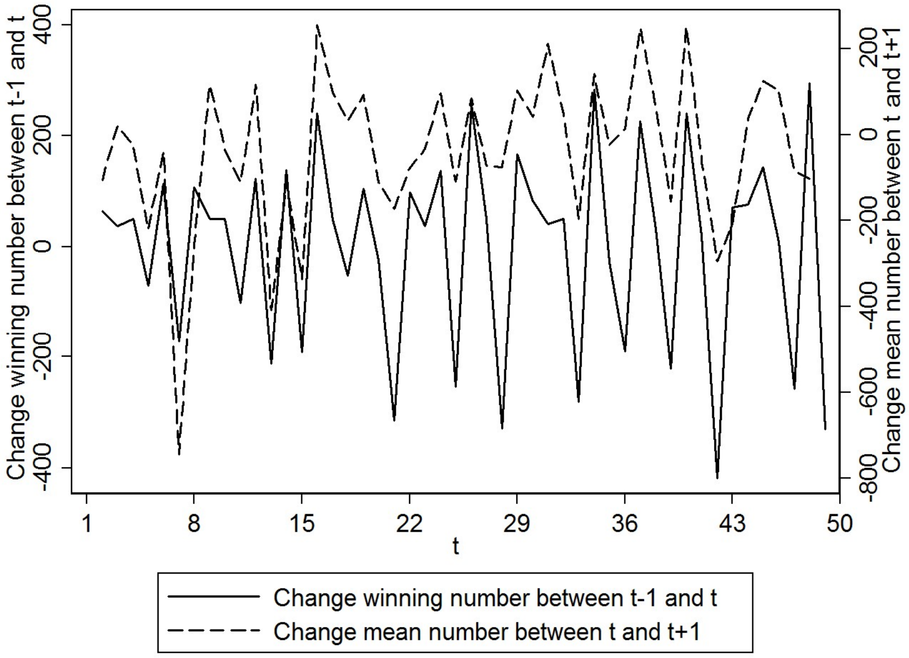


**Figure 2**

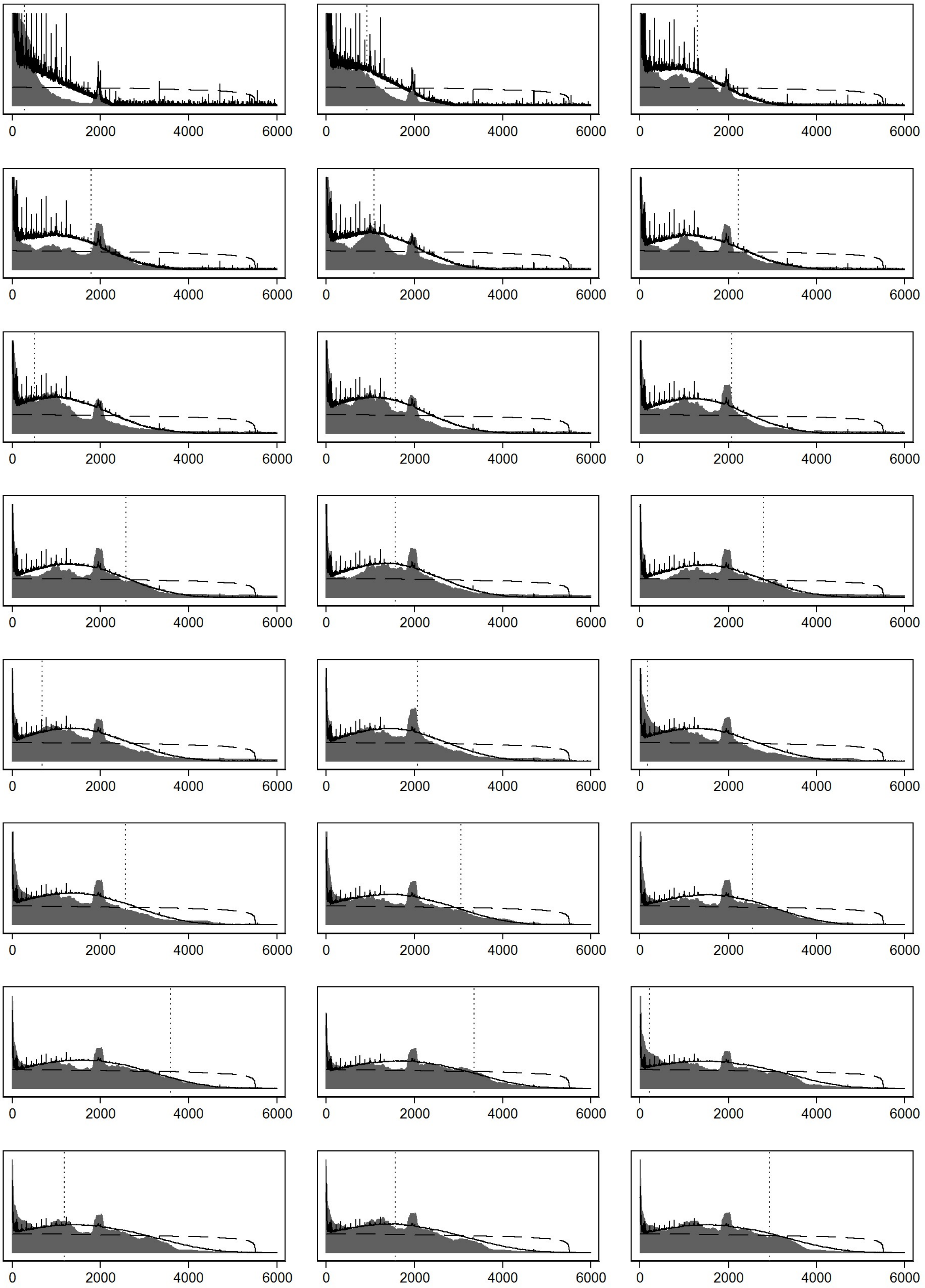


**Figure 3A**

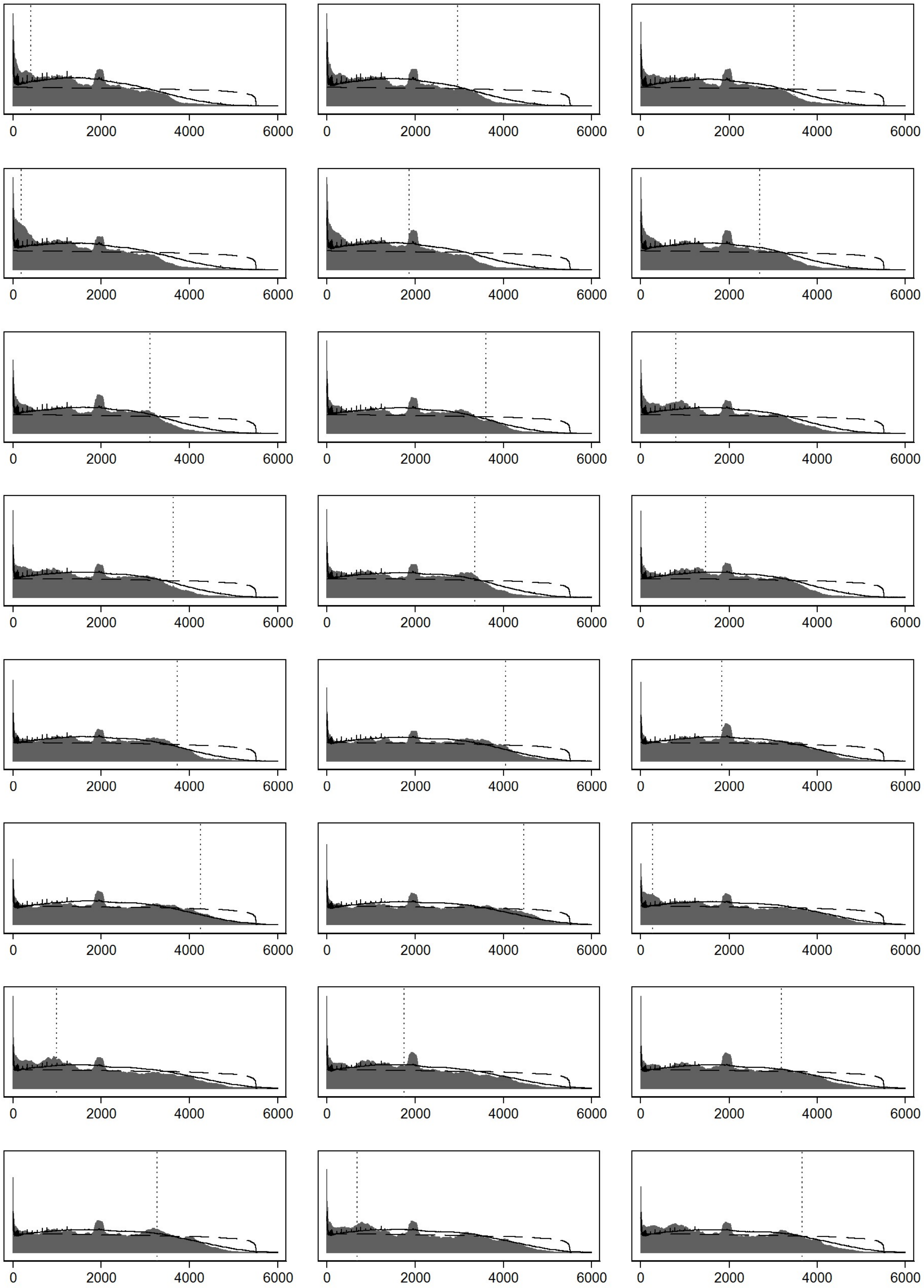

0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000
0 2000 4000 6000

**Figure 3B**

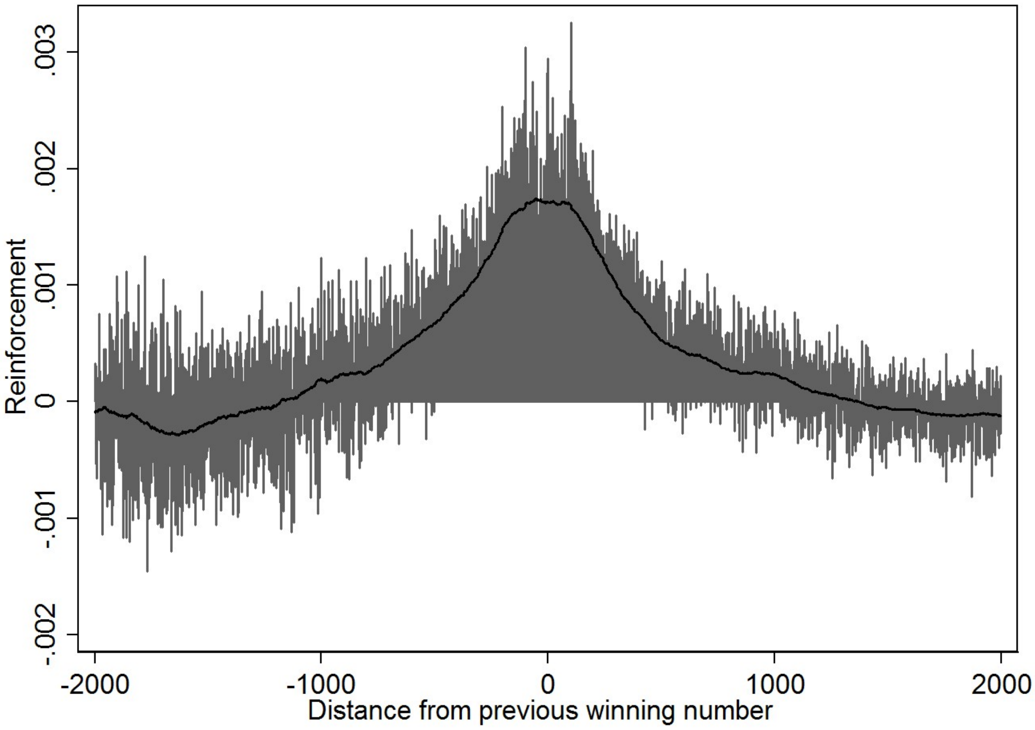

.003
.002
.001
0
-.001
-.002
Reinforcement
-2000
-1000
0
1000
2000
Distance from previous winning number

**Figure 4**

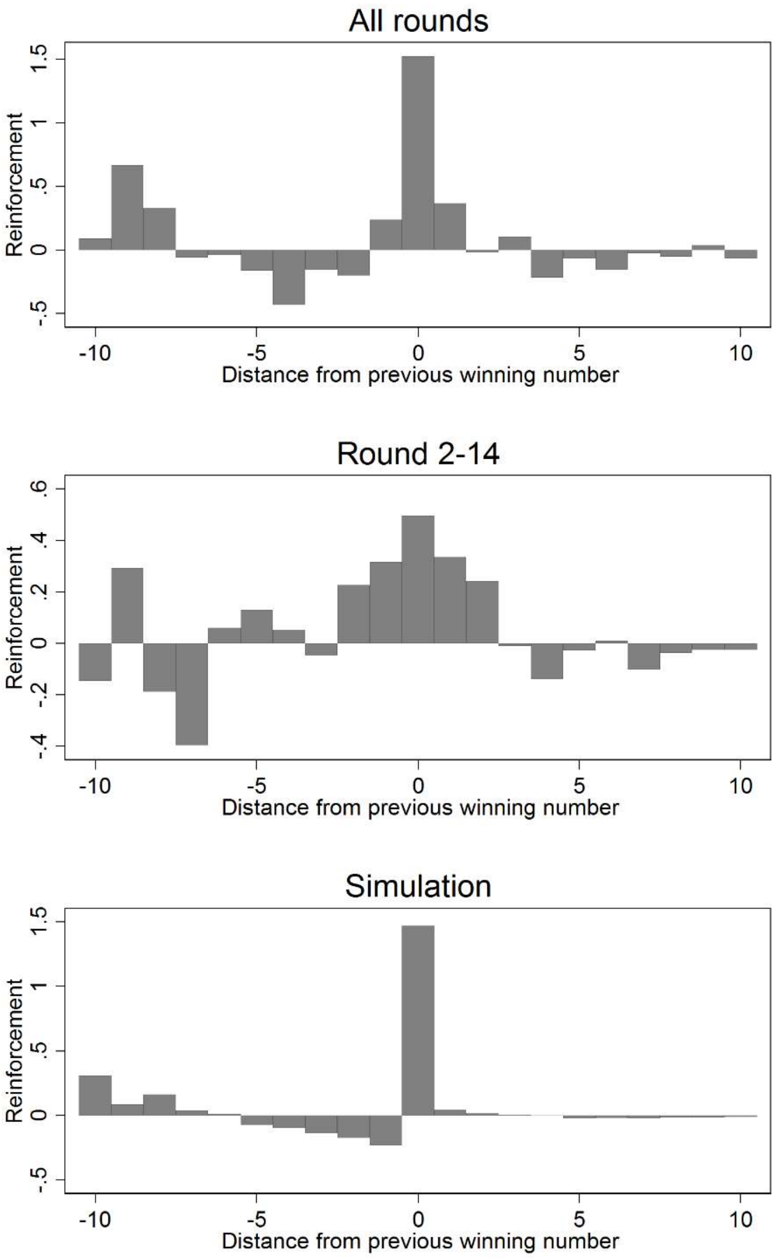

All rounds
Reinforcement
-.5
0
.5
1
1.5
-10
-5
0
5
10
Distance from previous winning number
Round 2-14
Reinforcement
-.4
-.2
0
.2
.4
.6
-10
-5
0
5
10
Distance from previous winning number
Simulation
Reinforcement
-.5
0
.5
1
1.5
-10
-5
0
5
10
Distance from previous winning number

**Figure 5**

.

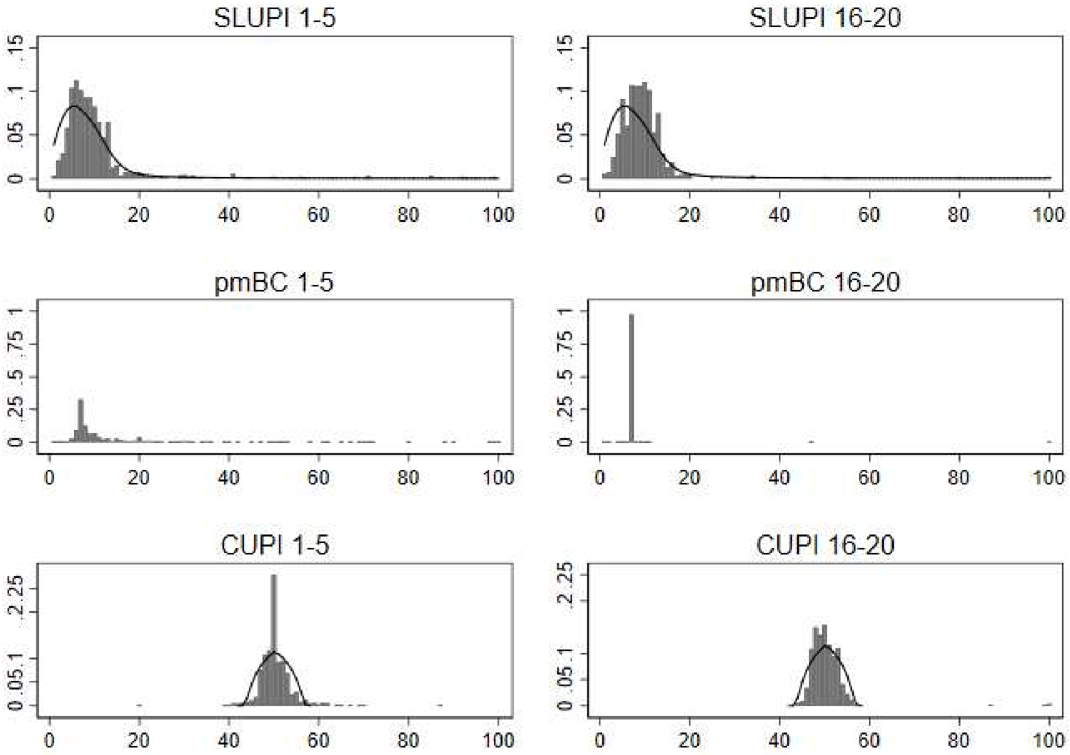

SLUPI 1-5
SLUPI 16-20
pmBC 1-5
pmBC 16-20
CUPI 1-5
CUPI 16-20
0
.05
.1
.15
0
.25
.5
.75
1
0
.05
.1
.2
.25
0
20
40
60
80
100

**Figure 6**

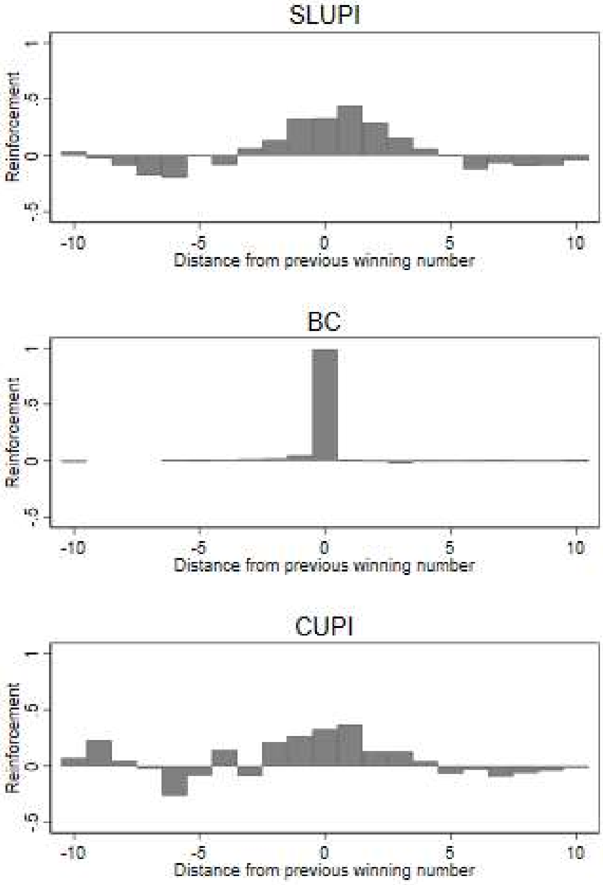


**Figure 7**

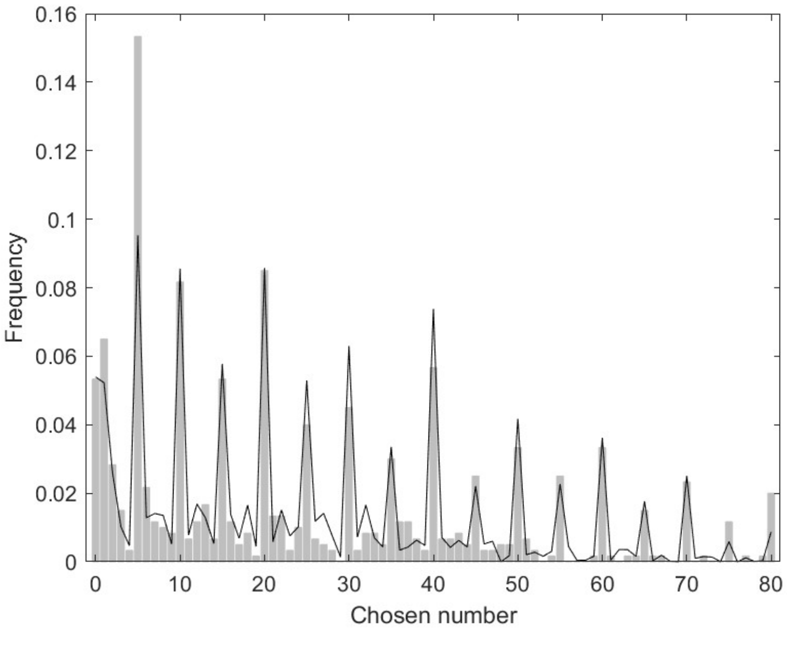


**Figure 8**

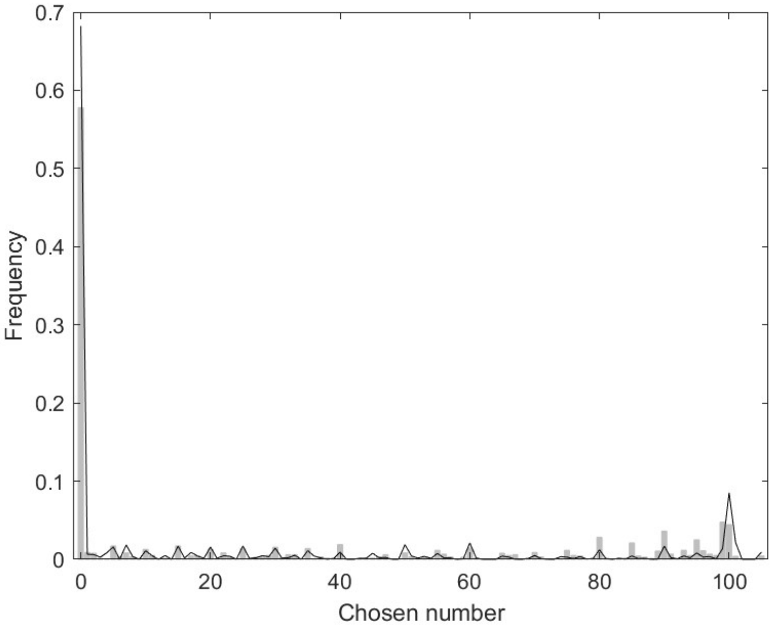


**Figure 9**

## Online Appendix

## Appendix A: Additional Results and Proofs

*Probability Matching in LUPI*

Proposition A1 shows that, in equilibrium, the probability a number is played is proportional to the probability that number wins. The probability that number $k$ wins in the LUPI game is $w_k(p) = Np_k\pi_k(p)$, where $N$ can be either fixed or Poisson-distributed.

**Proposition A1.** *Consider the 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 $p$ has full support $\pi_k = \pi^*$ for all $k$ we have

$$w_k = Np_k\pi^*. \tag{A1}$$

Summing both sides of (A1) over $k$ gives

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

Dividing the left-hand side of (A1) with $\sum w_k$ and the right-hand side with $N\pi^*$ gives $p_k = w_k / \sum w_k$.

To prove the other direction, suppose that $p$ 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 $\pi_k = \sum_j w_j / N$. Because the right-hand side is the same for all $k$, it must be a mixed strategy equilibrium. ■

We can use this property to verify whether the equilibrium with population uncertainty is a good approximation of the equilibrium with a fixed number of players. We simulated a fixed number of players ($N = 53,783$) playing according to the equilibrium with population uncertainty about 750 million times, and find the resulting distribution of winning numbers is practially indistinguishable from the equilibrium disribution. This strongly suggests the equilibrium with Poisson-distributed population uncertainty is a very good approximation of the fixed-$N$ equilibrium when the number of players is large.

*Concepts and Notation*

We borrow the following notation and definitions from Benaïm (1999): Consider a metric space $(X, d)$ (in our case it is the simplex $\Delta$ and Euclidean distance) 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 a rest point (an equilibrium in Benaïm's terminology) if $\Phi_t(x) = x$ for all $t$. A point $x^* \in X$ is an $\omega$-limit point of $x$ if $x^* = \lim_{t_k \to \infty} \Phi_{t_k}(x)$ for some sequence $t_k \to \infty$. Intuitively, an $\omega$-limit point of $x$ is a point to which the semi-flow $\Phi_t(x)$ always returns. The $\omega$-limit set of $x$, denoted $\omega(x)$, is the set of $\omega$-limit points of $x$. The definition of an $\omega$-limit can be extended to a discrete time system. A set $A \subseteq X$ is invariant if $\Phi_t(A) = A$ for all $t \in \mathbb{R}$. A subset $A \subseteq X$ is an attractor for $\Phi$ if (i) $A$ is non-empty, compact and invariant, and (ii) $A$ has a neighborhood $U \subseteq X$ such that $\lim_{t\to\infty} d(\Phi_t, A) \to 0$ uniformly in $x \in U$ (the distance between $\Phi_t$ and the closest point in $A$). An attractor $A$ is a proper attractor if it contains no proper subset that is an attractor.

For $\delta > 0$, and $T > 0$, a $(\delta, T)$-*pseudo-orbit* from $a \in X$ to $b \in X$ is a finite sequence of partial trajectories $\{\Phi_t(y_i) : 0 \le t \le t_i\}_{i=0,...,k-1}$, with $t_i \ge T$, 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 $(\delta, T)$-pseudo-orbit from $a$ to $a$ for every $\delta > 0$, and $T > 0$. Let $\Lambda \subseteq X$ be a non-empty invariant set. $\Phi$ is called chain recurrent on $\Lambda$ if every point $x \in \Lambda$ is a chain recurrent point for $\Phi|\Lambda$, the restriction of $\Phi$ to $\Lambda$. A compact invariant set on which $\Phi$ is chain recurrent is called an internally chain recurrent set.

The study of this kind of stochastic processes was initiated by Robbins & Monro (1951). The ODE method originates with Ljung (1977). For a book-length treatment of the theory of stochastic approximation, see Benveniste et al. (1990).

*Proof of Proposition 1*

We being by deriving an expressions for the law of motion of $p(t)$.

$$\begin{aligned} p_k(t+1) - p_k(t) &= \frac{A_k(t+1)}{\sum_{j=1}^K A_j(t+1)} - \frac{A_k(t)}{\sum_{j=1}^K A_j(t)} \\ &= \frac{A_k(t) + r_k(t)}{\sum_{j=1}^K (A_j(t) + r_j(t))} - p_k(t) \\ &= \frac{A_k(t) + r_k(t) - p_k(t) \sum_{j=1}^K (A_j(t) + r_j(t))}{\sum_{j=1}^K (A_j(t) + r_j(t))} \\ &= \frac{r_k(t) - p_k(t) \sum_{j=1}^K r_j(t)}{\sum_{j=1}^K A_j(t+1)}. \end{aligned} \tag{A2}$$

This formulation makes it clear that $p(t)$ is a process with decreasing step size since $c > 0$ ensures that the sum of reinforcements grows without bound. Note that the stochastic nature of the process is due to randomness of the reinforcement terms $\{r_j(t)\}_{j=1}^K$, which also enter into $\sum_{j=1}^K A_j(t+1)$, by the definition of the updating rule (2).

Let $(\Omega, \mathcal{F}, \mu)$ be a probability space and $\{\mathcal{F}_t\}$ a filtration such that $\mathcal{F}_t$ is a sigma-algebra that represents the history of the system up until the beginning of period $t$. The process $p$ is adapted to $\{\mathcal{F}_t\}$. We can write

$$p(t+1) - p(t) = \gamma(t+1)(F(t) + U(t+1)),$$

where the step size is

$$\gamma(t+1) = \frac{1}{\sum_{j=1}^K A_j(t+1)},$$

the expected motion is

$$F(t) = \mathbb{E}[r_k(t) | \mathcal{F}_t] - p_k(t) \sum_{j=1}^K \mathbb{E}[r_j(t) | \mathcal{F}_t],$$

and $U(t+1)$ is a stochastic process adapted to $\{\mathcal{F}_t\}$;

$$U(t+1) = r_k(t) - \mathbb{E}[r_k(t) | \mathcal{F}_t] - p_k(t) \sum_{j=1}^K (r_j(t) - \mathbb{E}[r_j(t) | \mathcal{F}_t]).$$

We write $\gamma(t+1)$ and $U(t+1)$ but $F(t)$ because the former two terms depend on events that take place after the beginning of period $t$, whereas the latter term only depends on the attractions at the beginning of period $t$.

The stochastic process moves in discrete time. In order to be able to compare it with a deterministic process that moves in continuous time, we consider the interpolation of the stochastic process. As defined in the main text the continuous time interpolated stochastic GCI process $\tilde{p} : \mathbb{R}_+ \to \mathbb{R}^m$ is

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

for all $n \in \mathbb{N}$ and $0 \leq s \leq 1/(t+1)$.

Note that $\mathbb{E}[U(t+1) | \mathcal{F}_t] = 0$, and $\sup_t \mathbb{E}\left[\|U(t+1)\|^2 | \mathcal{F}_t\right] \leq C$ for some constant $C$. Moreover, for any realization $\lim_{t\to\infty} \gamma(t) = 0$, $\sum_{t=1}^\infty \gamma(t) = \infty$, 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 remark 4.3 in Benaïm (1999) imply that with probability 1, the interpolated process $\tilde{p}$ is an asymptotic pseudotrajectory of the flow $\Phi$ induced by $F$. Since $\{\tilde{p}(t) : t \geq 0\}$ is precompact, we obtain the following result from Benaïm's Theorem 5.7 and Proposition 5.3.

*With probability 1, every $\omega$-limit set of $\tilde{p}$ is a compact invariant set $\Lambda$ for the flow $\Phi$ induced by the continuous time deterministic GCI dynamic*

$$\dot{p}_k = \mathbb{E}[r_k(t) | \mathcal{F}_t] - p_k(t) \sum_{j=1}^K \mathbb{E}[r_j(t) | \mathcal{F}_t], \tag{A3}$$

*and $\Phi|\Lambda$, the restriction of $\Phi$ to $\Lambda$, admits no proper attractor.*

The next step is to calculate the expected reinforcement. Using our specification of reinforcements (6) with the added term $c$, it is easy to find that

$$\mathbb{E}\left[r_k^c(t)\,|\mathcal{F}_t\right] = np_k(t)\,\pi_k(p(t)) + c,$$

where $\pi_k$ is the expected payoff, i.e. the probability of winning, when playing $k$ with probability one. By plugging this into the general stochastic approximation result (A3) and suppressing the reference to $t$, we obtain the desired result.

**Remark A1.** *If $c = 0$ then we face the problem that the step size $\gamma(t) = 1/\sum_{i=1}^K r_i(t)$ 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 1 would continue to hold if almost surely $\lim_{t\to\infty}\gamma(t) = 0$, almost surely $\sum_{t=1}^\infty \gamma(t) = \infty$, and $\mathbb{E}\left[\sum_{t=1}^\infty \gamma(t)^2\right] < \infty$. 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.*

*Proposition 1 with Heterogenous Initial Attractions*

We may relax the assumption that all individuals have the same initial attractions. Then, we have to distinguish the strategy of individual $i$, denoted $\sigma^i$, from the average strategy in the population. Suppose that there $M$ individuals in the population from which players are drawn. (Think of $M$ as being arbitrarily large but finite.) The average strategy in the population is

$$p = \frac{1}{M}\sum_{i=1}^M \sigma^i.$$

We have

$$\begin{aligned}\sigma_k^i(t+1) - \sigma_k^i(t) &= \frac{r_k(t) + \sigma_k^i(t)\sum_{j=1}^K r_j(t)}{\sum_{j=1}^K\left(A_j^i(t) + r_j(t)\right)} \\ &= \frac{r_k(t) + \sigma_k^i(t)\sum_{j=1}^K r_j(t)}{\sum_{j=1}^K A_j^i(1) + \sum_{j=1}^K\left(\sum_{\tau=1}^t r_j(\tau)\right)} \\ &= \frac{r_k(t) + \sigma_k^i(t)\sum_{j=1}^K r_j(t)}{\sum_{j=1}^K\left(\sum_{\tau=1}^t r_j(\tau)\right)} + O\left(\frac{1}{\left(\sum_{j=1}^K\left(\sum_{\tau=1}^t r_j(\tau)\right)\right)^2}\right).\end{aligned}$$

Next, use this to find

$$\begin{aligned}&p_k(t+1) - p_k(t) \\ &= \frac{1}{M}\sum_{i=1}^M \frac{r_k(t) + \sigma_k^i(t)\sum_{j=1}^K r_j(t)}{\sum_{j=1}^K\left(A_j^i(t) + r_j(t)\right)} \\ &= \frac{r_k(t) + \left(\frac{1}{M}\sum_{i=1}^M \sigma_k^i(t)\right)\sum_{j=1}^K r_j(t)}{\sum_{j=1}^K\left(\sum_{\tau=1}^t r_j(\tau)\right)} + O\left(\frac{1}{\left(\sum_{j=1}^K\left(\sum_{\tau=1}^t r_j(\tau)\right)\right)^2}\right) \\ &= \frac{1}{\sum_{j=1}^K\left(\sum_{\tau=1}^t r_j(\tau)\right)}\left(r_k(t) + p_k(t)\sum_{j=1}^K r_j(t) + O\left(\frac{1}{\sum_{j=1}^K\left(\sum_{\tau=1}^t r_j(\tau)\right)}\right)\right).\end{aligned}$$

We can write

$$p(t+1) - p(t) = \gamma(t+1)\left(F(t) + U(t+1) + b(t+1)\right),$$

where $F(t)$ and $U(t+1)$ are defined as before, the step size is slightly modified (initial attractions are removed),

$$\gamma(t+1) = \frac{1}{\sum_{j=1}^K\left(\sum_{\tau=1}^t r_j(\tau)\right)},$$

and the new term is -

$$b(t+1) = O\left(\frac{1}{\sum_{j=1}^{K}\left(\sum_{\tau=1}^{t} r_j(\tau)\right)}\right).$$

(We write $\gamma(t+1)$, $U(t+1)$, and $b(t+1)$, but $F(t)$, because the former three terms depend on events that take place after the beginning of period $t$ whereas the latter term only depends on the attractions at the beginning of period $t$.) Note that $\lim_{t\to\infty} b(t) = 0$. With the added help of remark 4.5 in Benaïm (1999), the proof of Proposition 1 can be used again.

*Proof of Proposition 2*

In proving this result we refer to the LUPI game. The result holds for the CUPI game as well since it's strategy space is merely a permutation of the strategy space of the LUPI game.

We start by noting that the dynamic (7) can be rewritten as follows

$$\dot{p}_k = np_k\left(\pi_k^c(p) - \sum_{j=1}^{K} p_j\left(\pi_j^c(p)\right)\right), \tag{A4}$$

where, noting that $\lim_{p_i\to\infty}\frac{c}{np_i} = +\infty$, we define

$$\pi_i^c(p) = \begin{cases} \pi_i(p) + \frac{c}{np_i} & \text{if } p_i > 0 \\ +\infty & \text{if } p_i = 0 \end{cases}.$$

We may consider an auxiliary *perturbed LUPI game* with expected payoffs $\pi_i^c(p)$ rather than $\pi_i(p)$ for all $i$. Hence, the perturbed replicator dynamic for a LUPI game with a Poisson distributed number of players game can be interpreted as the unperturbed replicator dynamic for the perturbed LUPI game with a Poisson distributed number of players. It is immediate that (7) has a rest point $p^{c*}$ at which $\pi_i(p) + \frac{c}{np_i} = \pi_i(p^{c*})$ for all $i$. As $c \to 0$, this rest point converges to the Nash equilibrium of the unperturbed game.

*Part 1*

We show that the perturbed replicator dynamic (7) has a unique interior rest point $p^{c*}$, by showing that the auxiliary perturbed LUPI game with a Poisson distributed number of players has a unique symmetric interior equilibrium $p^{c*}$.

Existence follows from Myerson (1998). Full support is ensured by the fact that $\lim_{p_i\to\infty}\left(\pi_i(p) + \frac{c}{np_i}\right) = +\infty$, while $\pi_i(p) + \frac{c}{np_i}$ is finite on the interior. In equilibrium, the expected payoff is the same for each action, so

$$\begin{aligned}\pi_{k+1}^c(p) &= e^{-np_{k+1}}\prod_{i=1}^{k}\left(1 - np_i e^{-np_i}\right) + \frac{c}{np_{k+1}} \\ &= e^{-np_k}\prod_{i=1}^{k-1}\left(1 - np_i e^{-np_i}\right) + \frac{c}{np_k} = \pi_k^c(p),\end{aligned}$$

or equivalently,

$$\frac{e^{np_{k+1}}}{e^{np_k}} = e^{np_{k+1}}\frac{\frac{c}{n}\left(\frac{1}{p_{k+1}} - \frac{1}{p_k}\right)}{\prod_{i=1}^{k-1}\left(1 - np_i e^{-np_i}\right)} + \left(1 - np_k e^{-np_k}\right).$$

Taking logarithms on both sides

$$p_{k+1} - p_k = \frac{1}{n}\ln\left(e^{np_{k+1}}\frac{\frac{c}{n}\left(\frac{1}{p_{k+1}} - \frac{1}{p_k}\right)}{\prod_{i=1}^{k-1}\left(1 - np_i e^{-np_i}\right)} + \left(1 - np_k e^{-np_k}\right)\right). \tag{A5}$$

Note that as $c \to 0$, the left-hand side approaches $\frac{1}{n} \ln (1 - np_k e^{-np_k})$. Since $(1 - np_k e^{-np_k}) \in (0,1)$ for all $p \in int\,(\Delta)$, there is some $c\,(k)$ such that if $c < c\,(k)$, then we have $\frac{1}{n} \ln (1 - np_k e^{-np_k}) < 0$ for the equilibrium $p$. This implies that $p_{k+1} < p_k$. We can establish such a bound $c\,(k)$ for each $k$. Let $\bar{c} = \min_k c\,(k)$, so that if $c < \bar{c}$ then $p_{k+1} < p_k$ for all $k$. For every candidate equilibrium value of $p_1$ the relationship (A5) recursively determines all equilibrium probabilities. Since the probabilities sum to one and since $p_{k+1} < p_k$ for all $k$, there is a unique equilibrium.

*Part 2*

Proposition 1 implies that the realization of the stochastic GCI process almost surely converges to a compact invariant set that admits no proper attractor under the flow induced by the perturbed replicator dynamic (7). Part 1 implies that the only candidate rest point in the interior is the perturbed Nash equilibrium.

*Part 3*

To rule out convergence to the boundary, recall that the initial attractions are strictly positive. Since no boundary point is a Nash equilibrium, the proofs of Lemma 3 and Proposition 3 in Hopkins & Posch (2005) can be adapted; for instance one may consider the unperturbed dynamic in the perturbed game (defined by the perturbed payoffs $\pi^c$). If $p' \neq p^{c*}$, then $p'$ is not a Nash equilibrium of the perturbed game. If a point $p'$ is not a Nash equilibrium, then the Jacobian for the replicator dynamic, evaluated at $p'$, has at least one strictly positive eigenvalue. Hopkins & Posch (2005) show that this rules out convergence. For a related point, see Beggs (2005).

## Appendix B: A Family of GCI Models

As described in section 2.5, in order to be able to generalize the learning rule that we defined for WTA games, we define four different versions of GCI that happen to coincide in WTA games, but which may yield different predictions in other games.

Let $M_k(s)$ denote the set of players picking strategy $k$ under strategy profile $s$, and let $m_k(s) = |M_k(s)|$ be the number of players picking strategy $k$ under strategy profile $s$. The realized payoff to player $i$ unders trategy profile $s$ is $u_i(s, \omega)$, where $\omega$ denotes the realisation of a random variable. For example, in the beauty contest $\omega$ determines which player that wins the price in case of a tie. We assume that $\frac{1}{m_k(s(t))} \sum_{ieM_k(s(t))} u_i(s(t), \omega)$ is constant across realisations $\omega$, meaning that $\omega$ only affects the distribution of payoffs at a given strategy profile.

Under *payoff-proportional frequency-independent global cumulative imitation (PFI-GCI)* reinforcements are

$$r_k^{PFI}(t) = \begin{cases} \frac{1}{m_k(s(t))} \sum_{ieM_k(s(t))} u_i(s(t), \omega) + c & \text{if } M_k(s) \neq \varnothing, \\ c & \text{otherwise.} \end{cases} \tag{B1}$$

Here reinforcement of a strategy $k$ is proportional to the sum of payoffs earned by those playing that strategy. Such reinforcements can be calculated based only on information about the payoff that was received by actions that someone played. Alternatively, players may also have information about the number of players playing each strategy.

Under *payoff-proportional frequency-dependent global cumulative imitation (PFD-GCI)* reinforcements are

$$r_k^{PFD}(t) = \begin{cases} m_k(t) \sum_{ieM_k(s(t))} u_i(s(t), \omega) + m_k(t)\, c & \text{if } M_k(s) \neq \varnothing, \\ m_k(t)\, c & \text{otherwise.} \end{cases} \tag{B2}$$

Note that $r_k^{PFD}(t) = m_k(t)\, r_k^{PFI}(t)$. In the WTA games, subjects do not have any information about $m_k(t)$ unless $k$ is the winning number. However, if $c = 0$ then $m_k(t)\, c = 0$ so that $r_k^{PFD}(t) = 0$ for all $k$ other than the winning number. Thus, for $c = 0$ subjects in our WTA experiments could update attractions with reinforcements of the form $r_k^{PFD}(t)$.

Next consider imitation that only reinforces the winning actions – the highest earning action – by the highest amount earned. In line with Roth (1995), we define *winner-takes-all frequency-independent global cumulative imitation (WFD-GCI),*

$$r_k^{WFI}(t) = \begin{cases} u_i(s(t), \omega) + c & \text{if } i \in \max_i u_i(s(t), \omega), \\ 0 & \text{otherwise.} \end{cases} \tag{B3}$$

Roth does not explicitly add a constant $c$ but he assumes, equivalently, that all payoffs are strictly positive.

We also define a frequency-dependent version of winner-takes-all imitation (which is not mentioned in Roth, 1995). The set of highest earning players is $\arg\max_{i'} u_{i'}(s(t), \omega)$, and such a player earns $\max_{i'} u_{i'}(s(t), \omega)$. We define *winner-takes-all frequency-dependent global cumulative imitation (WFI-GCI) i*

$$r_k^{WFD}(t) = \begin{cases} \arg\max_{i'} u_{i'}(s(t), \omega)\, (u_i(s(t), \omega) + c) & \text{if } i \in \arg\max_{i'} u_{i'}(s(t), \omega) \text{ and } s_i = k, \\ 0 & \text{otherwise.} \end{cases} \tag{B4}$$

In the WTA games, if $c = 0$, then $r_k^{WFD}(t) = 0$ for all $k$ other than the winning number. Thus, for $c = 0$, subjects in our WTA experiments could update attractions with reinforcements of the form $r_k^{WFD}(t)$.

The following proposition relates the four different members of the GCI family in winner-takes-all games.

**Proposition B1.** *In WTA games*

$$\lim_{c \to 0} r_k^{PFI}(t) = \lim_{c \to 0} r_k^{PFD}(t) \lim_{c \to 0} = r_k^{WFI}(t) \lim_{c \to 0} = r_k^{WFD}(t) = \begin{cases} 1 & \textit{if } k = k^*(s(t)) \\ 0 & \textit{otherwise.} \end{cases} .$$

**Proof.** Follows from the fact that in WTA games, $m_k(t)\, u_i(s(t), \omega) = 1$ for winning $i$ and $k = s_i$ and $m_k(t)\, u_i(s(t), \omega) = 0$ for all other $k$ and $i$. ■

Proposition B1 implies that we are unable to distinguish the members of the GCI family in winner-takes-all games. However, in general, the different members of the GCI-family can be distinguished as they induce different dynamics. We can show that PFD induces a noisy replicator dynamic in all games.

**Proposition B2.** *Consider a symmetric game and assume that $c > \min_i u_i(s,\omega)$ for all $\omega$. In a fixed $N$-player game, the GCI continuous time dynamic with PFD-reinforcement (B2) is*

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

*In a Poisson $n$-player game, the GCI continuous time dynamic with PFD-reinforcement (B2) is*

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

**Proof.** Let $X_t(k)$ be the *total* number of players who are drawn to participate and choose strategy $k$ in period $t$. For a given focal individual who is drawn to play the game, let $Y_t(k)$ be the number of *other* players who pick $k$ in period $t$. In the Poisson game, the ex ante probability of $X_t(k) = m$ is equal to the probability that $Y_t(k) = m$ conditional on the focal individual being drawn to play. This is due to the *environmental equivalence*-property of Poisson games (Myerson, 1998). However in a game with a fixed number of $N$ players, this is not the case.

We now derive the expected reinforcement $\rho^{PFD}$. To simplify the exposition, we suppress the reference to $\mathcal{F}_t$. To simplify notation we write $\bar{u}_k(s(t)) = m_k(t) \sum_{ieM_k(s(t))} u_i(s(t), \omega)$. For both fixed and Poisson distributed number of players, we have

$$\begin{aligned}
&\mathbb{E}\left[r_k^{PFD}(t) | \mathcal{F}_t\right] \\
&= \sum_{j=1}^{N} \Pr(X(k) = j)\, \mathbb{E}\left[r_k^{PFD}(s) | X(k) = j\right] + \Pr(X(k) = 0)(c \cdot 0) \\
&= \sum_{j=1}^{N} \Pr(X(k) = j)\, \mathbb{E}\left[j \cdot (\bar{u}_k(s(t)) + c) | Y(k) = j - 1 \wedge X(k) = j\right] \\
&= \sum_{j=0}^{N-1} \Pr(X(k) = j+1)\, \mathbb{E}\left[(j+1)(\bar{u}_k(s(t)) + c) | Y(k) = j \wedge X(k) = j+1\right]. \qquad \text{(B5)}
\end{aligned}$$

For *fixed $N$-player games*, we need to translate from $\Pr(X(k) = j+1)$ to $\Pr(Y(k) = j)$. Use

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

to obtain

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

Plugging this into (B5) yields

$$\begin{aligned}
&\mathbb{E}\left[r_k^{PFD}(t) | \mathcal{F}_t\right] \\
&= \sum_{j=0}^{N-1} \frac{Np_k}{j+1} \Pr(Y_i(k) = j)\, \mathbb{E}\left[(j+1)(\bar{u}_k(s(t)) + c) | Y(k) = j \wedge X(k) = j+1\right] \\
&= Np_k \sum_{j=0}^{N-1} \Pr(Y_i(k) = j)\, \mathbb{E}\left[(\bar{u}_k(s(t)) + c) | Y(k) = j \wedge X(k) = j+1\right],
\end{aligned}$$

or

$$\mathbb{E}\left[r_k^{PFD}(t)\,|\mathcal{F}_t\right] = Np_k\left(\pi_k\left(p\left(t\right)\right)+c\right). \tag{B6}$$

Plugging (B6) into the general stochastic approximation result (A3) gives the desired result for fixed $N$-player games.

For *Poisson-distributed* $N$, we have

$$\Pr\left(X\left(k\right)=j+1\right) = \frac{e^{np_k}\left(np_k\right)^{j+1}}{(j+1)!} = \frac{np_k}{j+1}\frac{e^{np_k}\left(np_k\right)^{j}}{j!} = \frac{np_k}{j+1}\Pr\left(X\left(k\right)=j+1\right).$$

Plugging this into (B5) yields

$$\begin{aligned}
&\mathbb{E}\left[r_k^{PFD}(t)\,|\mathcal{F}_t\right] \\
&= \sum_{j=0}^{N-1}\frac{np_k}{j+1}\Pr\left(X\left(k\right)=j+1\right)\mathbb{E}\left[\left(j+1\right)\left(\bar{u}_k\left(s\left(t\right)\right)+c\right)|Y\left(k\right)=j\wedge X\left(k\right)=j+1\right] \\
&= np_k\sum_{j=0}^{N-1}\Pr\left(X\left(k\right)=j+1\right)\mathbb{E}\left[\left(\bar{u}_k\left(s\left(t\right)\right)+c\right)|Y\left(k\right)=j\wedge X\left(k\right)=j+1\right] \\
&= np_k\left(\pi_k\left(p\left(t\right)\right)+c\right).
\end{aligned}$$

Using this in the general stochastic approximation result (A3) gives the desired result for Poisson games. ■

The other three GCI models – PFI, WFI and WFD – do not generally lead to any version of the replicator dynamic. This can be verified by calculating the expected reinforcement and plugging it into equation (A3). The different models also differ in their informational requirements: WDI requires the least feedback, whereas PFD requires the most. Nevertheless, players could still use all four models in our experimental games although they only receive feedback about the action that obtained the highest payoff, i.e. the winner. Since players can infer the payoff of all other players (zero unless they win), they can use both winner-imitation and proportional imitation. Moreover, even though they only know the number of individuals who picked the winning action (one individual), they are still able to compute the product of payoff and the number of players for all actions (since it is zero for all non-winning actions). For this reason, they are able to use both frequency dependent and frequency independent imitation.

***Similarity-weighted GCI in WTA games***

We may add similarity-weights to each of the specifications of reinforcement defined above. With the similarity function (5) reinforcement factors in the WTA games are

$$\hat{r}_k\left(t\right) = \begin{cases} \eta_k\left(k^*\left(t\right)\right)+c & \text{if there is a winner, } k^*\left(t\right)\text{, in period } t, \\ c & \text{otherwise.} \end{cases} . \tag{B7}$$

**Proposition B3.** *In a LUPI game with a Poisson distributed number of players, the GCI continuous time dynamic with reinforcement (B7) is the following dynamic*

$$\dot{p}_k = n\left(\hat{\pi}_k\left(p\right)-p_k\sum_{j=1}^{K}\hat{\pi}_j\left(p\right)\right)+\left(1-K\right)c, \tag{B8}$$

*where* $\hat{\pi}_k$ *denotes the similarity- and frequency-weighted payoff*

$$\hat{\pi}_k\left(p\right) = \sum_{l=0}^{K}p_l\pi_l\left(p\right)\eta_k\left(l\right).$$

**Proof.** Expected reinforcement *(B7)* is

$$
\begin{aligned}
\mathbb{E}\left[r_k(t)\,|\mathcal{F}_t\right] &= \mathbb{E}\left[\eta_k\left(k^*\left(s(t)\right)\right)|\mathcal{F}_t\right] + c \\
&= \mathbb{E}\left[\left.\frac{\max\left\{0, 1-\frac{|k^*(s(t))-k|}{W}\right\}}{\sum_{i=0}^{K}\max\left\{0, 1-\frac{|k^*(s(t))-i|}{W}\right\}}\right|\mathcal{F}_t\right] + c \\
&= \sum_{l=0}^{K}\Pr\left(k^*\left(s(t)\right)=l|\mathcal{F}_t\right)\left(\frac{\max\left\{0, 1-\frac{|l-k|}{W}\right\}}{\sum_{i=0}^{K}\max\left\{0, 1-\frac{|l-i|}{W}\right\}}\right) + c \\
&= \sum_{l=0}^{K} np_l(t)\,\pi_l\left(p(t)\right)\eta_k(l) + c.
\end{aligned}
$$

By using the expression for $\mathbb{E}\left[r_k(t)\,|\mathcal{F}_t\right]$ in the general stochastic approximation result (A3) and suppressing the reference to $t$, we obtain the desired result. ■

Note that this is not the noisy replicator dynamic, hence we cannot be sure that the limiting behavior of (B8) is the same as that of (7). A rest point $\hat{p}$ of (B8) solves, for each $k$,

$$
\hat{p}_k = \frac{n\hat{\pi}_k(\hat{p}) + c}{n\sum_{j=1}^{K}\hat{\pi}_j(\hat{p}) + Kc}.
$$

We can verify that $p^*$ (the equilibrium of the game without similarity-adjustment) is not a rest point (when $c = 0$):

$$
\begin{aligned}
\frac{\hat{\pi}_k(p^*)}{\sum_{j=1}^{K}\hat{\pi}_j(p^*)} &= \frac{\sum_{l=0}^{K} np_l^*\pi_l(p^*)\,\eta_k(l)}{\sum_{l=0}^{K} np_l^*\pi_l(p^*)} \\
&= \frac{\sum_{l=0}^{K} np_l^*\pi^*\eta_k(l)}{\sum_{l=0}^{K} np_l^*\pi^*} \\
&= \frac{\sum_{l=0}^{K} p_l^*\eta_k(l)}{\sum_{l=0}^{K} p_l^*} \\
&= \sum_{l=0}^{K} p_l^*\eta_k(l) \\
&\neq p_k^*.
\end{aligned}
$$

Figure B1 shows the results from the simulations of similarity-weighted GCI in the LUPI games, described in section 2.4.

[INSERT FIGURE B1 HERE]

**Figure B1. Simulated similarity-weighted GCI process for the laboratory LUPI** game ($K = 99$ and $n = 26.9$)

The (blue) line corresponds to the equilibrium. The crosses indicate the average end state after 100,000 rounds of simulated play with 100 different initial conditions. The top panel shows similarity-weighted GCI for window size $W = 3$ and the bottom panel for $W = 6$. The noise parameter $c$ is set to 0.00001. The error bars show one standard deviation above/below the mean across the 100 simulations.

We study similarity-weighted GCI in the general case by restricting attention to payoff-proportional, frequency-dependent GCI:

$$
\hat{r}_k^{PFD}(t) = \sum_{l=1}^{K}\eta_k(l)\,r_l^{PFD}(t) = \begin{cases} \sum_{l=1}^{K}\eta_k(l)\,m_l(t)\left(u_{s_i(t)}\left(s(t)\right)+c\right) & \text{if } s_i(t)=k \text{ for some } i, \\ \sum_{l=1}^{K}\eta_k(l)\,m_k(t)\,c & \text{otherwise.} \end{cases} \tag{B9}
$$

For this kind of reinforcement the deterministic dynamic is a replicator dynamic for similarity- and frequency-weighted payoffs, plus a noise term.

**Proposition B4.** *Consider a symmetric game with an ordered strategy set $S = \{1, 2, ..., K\}$ for each player. Assume that $c > \min_i u_i(s, \omega)$ for all $\omega$. In a fixed $N$-player game, the GCI continuous time dynamic with similarity-weighted PFD-reinforcement (B9) is*

$$\dot{p}_k = N\left(\hat{\pi}_k(p) - p_k(t)\sum_{j=1}^{K}\hat{\pi}_j(p)\right) + cN\left(\sum_{l=1}^{K}p_l\eta_k(l) - p_k(t)\right).$$

*In a Poisson $n$-player game, the GCI continuous time dynamic with PFD-reinforcement (B2) is*

$$\dot{p}_k = n\left(\hat{\pi}_k(p) - p_k(t)\sum_{j=1}^{K}\hat{\pi}_j(p)\right) + cn\left(\sum_{l=1}^{K}p_l\eta_k(l) - p_k(t)\right).$$

*As in the main text $\hat{\pi}_k$ denotes the similarity- and frequency-weighted payoff*

$$\hat{\pi}_k(p) = \sum_{l=0}^{K}p_l\pi_l(p)\eta_k(l).$$

**Proof.** In the Poisson case expected reinforcement is,

$$\begin{aligned}
\mathbb{E}\left[\hat{r}_k^{PFD}(t)\,|\mathcal{F}_t\right] &= \sum_{l=1}^{K}\eta_k(l)\,\mathbb{E}\left[r_l^{PFD}(t)\,|\mathcal{F}_t\right]. \\
&= \sum_{l=1}^{K}\eta_k(l)\,np_l(\pi_l(p(t)) + c) \\
&= \left(\sum_{l=1}^{K}\eta_k(l)\,np_l\pi_l(p(t)) + \sum_{l=1}^{K}\eta_k(l)\,np_lc\right) \\
&= n\left(\hat{\pi}_k(p) + c\sum_{l=1}^{K}\eta_k(l)\,p_l\right)
\end{aligned}$$

Using this in (A3) yields

$$\begin{aligned}
\dot{p}_k &= n\left(\hat{\pi}_k(p) + c\sum_{l=1}^{K}\eta_k(l)\,p_l\right) - np_k(t)\sum_{j=1}^{K}\left(\hat{\pi}_j(p) + c\sum_{l=1}^{K}\eta_j(l)\,p_l\right) \\
&= n\left(\hat{\pi}_k(p) - p_k(t)\sum_{j=1}^{K}\hat{\pi}_j(p)\right) + cn\sum_{l=1}^{K}(\eta_k(l)\,p_l - p_k(t)\,p_l) \\
&= n\left(\hat{\pi}_k(p) - p_k(t)\sum_{j=1}^{K}\hat{\pi}_j(p)\right) + cn\left(\sum_{l=1}^{K}p_l\eta_k(l) - p_k(t)\right),
\end{aligned}$$

A similar result is obtained for the fixed $N$ case. ■

## Appendix C: Global Convergence

In the main text it is established that if the stochastic GCI-process converges to a point then it must converge to (a perturbed version of) the unique interior equilibrium. In order to establish that the process does indeed converge to this point and not to something else than a point – e.g. a periodic orbit – we simulated the learning process. As explained in Section 2.3.2 we used the lab parameters $K = 99$ and $n = 26.9$, and randomly drew 100 different initial conditions. For each initial condition, we ran the process for 10 million rounds. Figure C1 shows the resulting distribution at the end of these 10 million rounds, averaged over the 100 initial conditions.

[INSERT FIGURE C1 HERE]

**Figure C1. Simulated GCI process for the laboratory LUPI** game ($K = 99$ and $n = 26.9$) The (blue) line corresponds to the equilibrium. The crosses indicate the average end state after 10 million rounds of simulated play with 100 different initial conditions. The noise parameter $c$ is set to 0.00001. The error bars show one standard deviation above/below the mean across the 100 simulations.

## Appendix D: Additional Empirical Results

[INSERT FIGURE D1 HERE.]

**Figure D1. Distribution of chosen (thick solid line) and winning (thin solid line) numbers in all sessions from period 25 and onwards and equilibrium (dashed line).**

[INSERT FIGURE D2 HERE.]

**Figure D2. Correlation between winning and chosen numbers in laboratory LUPI game**

The difference between the winning numbers at time $t$and time $t-1$ (solid line) compared to the difference between the average chosen number at time $t+1$ and time $t$ (dashed line). Data from one periond in the first session excluded to make figure readable (winner was 67).

[INSERT FIGURE D3 HERE.]

**Figure D3. Fit of GCI learning model for field LUPI data for different values of $W$ and $\lambda$.**

This figure shows the sum of squared deviations between the field LUPI data and the GCI learning model for $W = 500, ..., 2500$ and $\lambda$ between 0.4 and 2.

[INSERT FIGURE D4 HERE.]

**Figure D4. Empirical densities (bars), estimated learning model (solid lines), Poisson-Nash equilibrium (dashed line), and winning numbers (dotted lines) for laboratory session 1, period 2-6.**

[INSERT FIGURE D5 HERE.]

**Figure D5. Empirical densities (bars), estimated learning model (solid lines), Poisson-Nash equilibrium (dashed line), and winning numbers (dotted lines) for laboratory session 2, period 2-6.**

[INSERT FIGURE D6 HERE.]

**Figure D6. Empirical densities (bars), estimated learning model (solid lines), Poisson-Nash equilibrium (dashed line), and winning numbers (dotted lines) for laboratory session 3, period 2-6.**

[INSERT FIGURE D7 HERE.]

**Figure D7 Empirical densities (bars), estimated learning model (solid lines), Poisson-Nash equilibrium (dashed line), and winning numbers (dotted lines) for laboratory session 4, period 2-6.**

[INSERT FIGURE D8 HERE.]

**Figure D8: Fit of GCI learning model for laboratory LUPI data for different values of $W$ and $\lambda$.**

[INSERT FIGURE D9 HERE.]

**Figure D9: The effect of winning numbers on chosen numbers in SLUPI, pmBC and CUPI.**

The difference between the winning numbers at time $t$ and time $t-1$ (solid line) compared to the difference between the average chosen number at time $t+1$ and time $t$ (dotted line). Winning numbers that change more than 10 numbers is shown as 10/-10 in graph. The strategy space in CUPI has been transformed as described in the main text.

[INSERT FIGURE D10 HERE.]

**Figure D10. Estimated reinforcement factors in SLUPI, pmBC and CUPI including only period 1-5**

[INSERT FIGURE D11 HERE.]

**Figure D11. Average effort (dotted lines) and average predicted choice from similarity-weighted PFD-GCI learning model (solid lines) in the Tullock contest.**

A payoff of 80 has been added to each players' payoff to make reinforcements positive. Initial attractions are based on first-period choices and scaled to sum to 320 (which is the total payoff in a round if all players play 0). Estimated window size is $W = 1$ and precision $\lambda = 0.76$.

[INSERT FIGURE D12 HERE.]

**Figure D12. Average bid (dotted lines) and average predicted choice from similarity-weighted PFD-GCI learning model (solid lines) in the all-pay auction.**
A payoff of 105 has been added to each players' payoff to make reinforcements positive. Choices have been rounded off to the nearest integer. Initial attractions are based on first-period choices and scaled to sum to 105 multiplied by the number of players in the session. Estimated window size is $W = 1$ and precision $\lambda = 0.93$.

## Appendix E: Experimental Instructions

*Experimental Payment*

At the end of the experiment, you will receive a show-up fee of NT$100, and whatever amount of Experimental Standard Currency (ESC) you earned in the experiment converted into NT dollars. The amount you will receive, which will be different for each participant, depends on your decisions, the decisions of others, and chance. All earnings are paid in private and you are not obligated to tell others how much you have earned. Note: The exchange rate for Experimental Standard Currency and NT dollars is 1:1 (1 ESC = NT$1).

Note: Please do not talk during the experiment. Raise your hand if you have any questions; the experimenter will come to you and answer them.

*Instructions for Part I*

Part I consists of 20 rounds. In each round, everyone has to choose a whole number between 1 and 100. Whoever chooses the second-lowest, uniquely chosen number wins. For example, if the chosen numbers are (in order) 1, 1, 1, 2, 3, 3, 4, 5, 5, 5, 6, 7, 7, the unique numbers are 2, 4, 6. The second lowest among them is 4, so whoever chose 4 is the winner of this round. If there is no second-lowest unique number, nobody wins this round.

Raise your hand if you have any questions; the experimenter will come to you and answer them.

Now we will start Part I and there will be 20 rounds. All of the Experimental Standard Currency (ESC) you earn in these rounds will be converted into NT dollars according to the 1:1 exchange rate and given to you. So please chose carefully when making your decisions.

*Instructions for Part II*

Part II also consists of 20 rounds. In each round, everyone has to choose a whole number between 1 and 100. The computer will then calculate the median of all chosen numbers. Whoever chooses closest to "(median)x0.3+5" wins. For example, if there are three participants and they choose 1, 2, and 3. The median is 2, and 2x0.3+5=5.6. Among 1, 2, and 3, the closest number to 5.6 is 3, so whoever chose 3 is the winner of this round. If there are two or more people who choose the closest number, the computer will randomly choose one of them to be the winner.

Raise your hand if you have any questions; the experimenter will come to you and answer them.

Now we will start Part II and there will be 20 rounds. All of the Experimental Standard Currency (ESC) you earn in these rounds will be converted into NT dollars according to the 1:1 exchange rate and given to you. So please chose carefully when making your decisions.

*Instructions for Part III*

Part III consists of 20 rounds. In each round, everyone has to choose a whole number between 1 and 100. Whoever chooses closest to 50, uniquely chosen number wins. If there are two numbers of the same distance to 50, the larger number wins. For example, you win if there are two or more who choose 50 and you uniquely choose 51. If there are two or more who choose 50 and 51, we will have to check (in order) if anyone uniquely chose 49, 52, 48, etc.

[INSERT FIGURE E1 HERE]

**Figure E1. Figure included in the CUPI game instructions.**

If no number is uniquely chosen, nobody wins in this round.

Raise your hand if you have any questions; the experimenter will come to you and answer them.

Now we will start Part III and there will be 20 rounds. All of the Experimental Standard Currency (ESC) you earn in these rounds will be converted into NT dollars according to the 1:1 exchange rate and given to you. So please chose carefully when making your decisions.

## Appendix F: Belief-based Learning

In this section, we briefly discuss whether fictitious play and Bayesian belief-based learning can rationalize behavior in the field LUPI game.

*Bayesian Belief-based Learning*

Suppose that a player of the LUPI game uses previous winning numbers to update her prior belief about the distribution of all players' play using Bayes' rule. The resulting posterior would depend critically upon the prior distribution. The fact that a particular number wins in a round is informative about the probability that the winning number was chosen, but says very little about the likelihood that other numbers were chosen – lower numbers than the winning number could either have been chosen a lot or not chosen at all. Allowing a completely flexible Dirichlet prior with $K$ parameters would both be computationally infeasible and result in very slow learning. Therefore, we instead pick a particular parameterized prior distribution and assume that the player updates her beliefs about the parameter of that distribution. Since we could not find a standard distribution that is flexible enough to capture the patterns seen in the data, we used the Nash equilibrium distribution with different values of $n$. For low $n$, this distribution is steep, while for high $n$ it is spread out and has 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. To avoid confusion, we hereafter instead call this distribution parameter $x$. Figure F1 illustrates this distribution for some different values of $x$.

[INSERT FIGURE F1 HERE]

**Figure F1. The Poisson Nash-equilibrium distribution for different values of the parameter x.**

In order to simulate belief-based learning using this particular distribution, we first calculate the probability that number $k$ wins if all players play according to the prior distribution for each value of $x$. We assume that all individuals share the same prior. Let $w_x(k)$ be the probability that number $k$ wins if $Poisson(n)$ players play according to the equilibrium distribution with the distribution parameter equal to $x$. Let $b_x(t) \in [0,1]$ be the agent's belief in period $t$ that the parameter of the prior distribution is $x$. Beliefs are updated according to

$$b_x(t+1) = \frac{w_k(x) b_x(t) + \varepsilon}{\sum_y [w_k(y) b_y(t) + \varepsilon]},$$

where $k$ is the winning number in period $t$. If $\varepsilon = 0$, this is equivalent to standard Bayesian updating, whereas $\varepsilon > 0$ implies that there is some noise in the updating process. This noise term is required to ensure that all probabilities are positive – otherwise some probabilities will be rounded off to zero.

We have estimated this belief-based learning model for the field data using the actual winning numbers and setting $n = 53,783$ and $K = 99,999$. We allowed $x \in \{1,2,3,...,99999\}$ and assumed a uniform prior over $x$, i.e. $b_x(0) = 1/99999$ for all $x$. We first set $\varepsilon$ to $10^{-20}$. Figure F2 shows the value of $x$ that results in the highest value of $w_x(k)$ along with the winning numbers in the field. As is clear from Figure F2, the most likely $x$ closely follows the winning number. The reason is that the most likely value of $x$ when $k$ wins is such that the equilibrium distribution "drops" to zero just around $k$. The best-response to this distribution would be to play just above $k$ in the next round. However, belief-learners also take winning numbers from previous rounds into account. Number 280 wins in the first day, and beliefs in the second day are therefore centered around $x = 1731$. The best-response to this belief is to play 281. On the second day, number 922 wins, which is extremely unlikely if players play according to a distribution with $x = 1731$. As shown by Figure F3, the agent therefore starts believing that $x$ is around $60,000$ from the third day and onwards, i.e. close to the actual number of players in the field. The reason is that a low number could win either if the distribution happens to drop at the right place, or when the distribution is very spread out. In the last week, beliefs are centered around $x = 57,000$. Since the agent believes that $x$ is higher than the number of players, guesses are believed to be more spread out than they actually are and the best response is to pick 1 from the third round and onwards.

[INSERT FIGURE F2 HERE]

**Figure F2. Winning numbers in the field (solid lines) along with the most likely value of x given that all players play according to prior distribution.**

[INSERT FIGURE F3 HERE]

**Figure F3. Evolution of posterior beliefs about parameter x.**

It is clear that belief-based learning with our particular choice of a parameterized distribution cannot rationalize imitative behavior in the field. Interestingly, however, the model can rationalize imitative behavior for higher values of the noise parameter. A high epsilon essentially implies a higher degree of forgetting and, consequently, that the experience of the last round is relatively more important. For example, if we set $\varepsilon = 10^{-10}$, the peak of the agent's posterior corresponds to the most likely $x$ in each period shown in Figure F2. The best-response to these beliefs is to pick a number slightly above the previous winning numbers during most of the rounds.

*Fictitious Play*

In our laboratory LUPI experiments, players only observed previous winning numbers. In the field game, however, it was possible to do so with some effort (by downloading and processing raw text files from the gambling company's website). Although we strongly suspect that not many players did this, we cannot rule it out. We therefore also estimate a fictitious play learning model in which players form beliefs about which numbers that will be chosen based on the past empirical distribution, and noisily best respond to those beliefs.

In this model, the perceived probability that number $k$ is chosen in period $t+1$ is given by

$$b_k\left(t+1\right) = \sum_{s=1}^{t} \frac{\widehat{p}_k\left(s\right)}{t},$$

where $\widehat{p}_k\left(t\right)$ is the empirical frequency with which number $k$ was played in $t$. For these beliefs, we calculate the expected payoff of each number assuming that the number of players are Poisson distributed. These expected payoffs are transformed into choice probabilities using the same power function (4) as in the estimation of the other learning models. Choices in the first period are assumed be identical to the actual distribution of play. The resulting model only has one free parameter, the precision parameter $\lambda$.

The best-fitting lambda is $\lambda = 0.0036$ and the SSD is 0.0075. The fit is considerably poorer than the imitation learning model which has a SSD of 0.0044 in our baseline estimation. One caveat is that player heterogeneity might bias estimates in favor of imitation learning, as discussed by Wilcox (2006). Since we do not have individual-level data for the field LUPI game, it is difficult to correct for this potential bias. Figure F4 shows the median chosen number in the field together with the predicted median choice according to the estimated fictitious play and GCI model. Although it is clear that fictitious play predicts the upward drift in choices in the field data, fictitious play seems to be too rapid and predicts too high numbers. Figure F5 shows that the fictitious play model also seems to underpredict the fraction of low numbers that are played – since numbers below 100 are very common in the data, the expected payoff of playing low numbers is low, and fictitious play therefore predicts that low numbers are played with low probability. Although the fit of the fictitious play model might be improved, for example by assuming that there is a constant inflow of new players with uniform priors, we believe that fictitious play is a less convincing explanation for several other reasons: 1) few players probably accessed the complete distribution, 2) calculating expected payoffs given the empirical distribution is very complicated, and 3) learning in the laboratory is very rapid despite the fact that only feedback about winning numbers is available.

[INSERT FIGURE F4 HERE]

**Figure F4. Actual and predicted number chosen in the field LUPI game (period 2-49)**

The solid line shows the actual median played, the dashed line the predicted median from the baseline GCI baseline estimation, and the dotted line the predicted median according to the estimated fictitious play learning model.

[INSERT FIGURE F5 HERE]

**Figure F5. Actual and predicted numbers below 100 in the field LUPI game (period 2-49)**
The solid line shows the actual fraction of numbers below 100, the dashed line the predicted fraction of numbers below 100 from the baseline GCI baseline estimation, and the dotted line the corresponding prediction of the estimated fictitious play learning model.

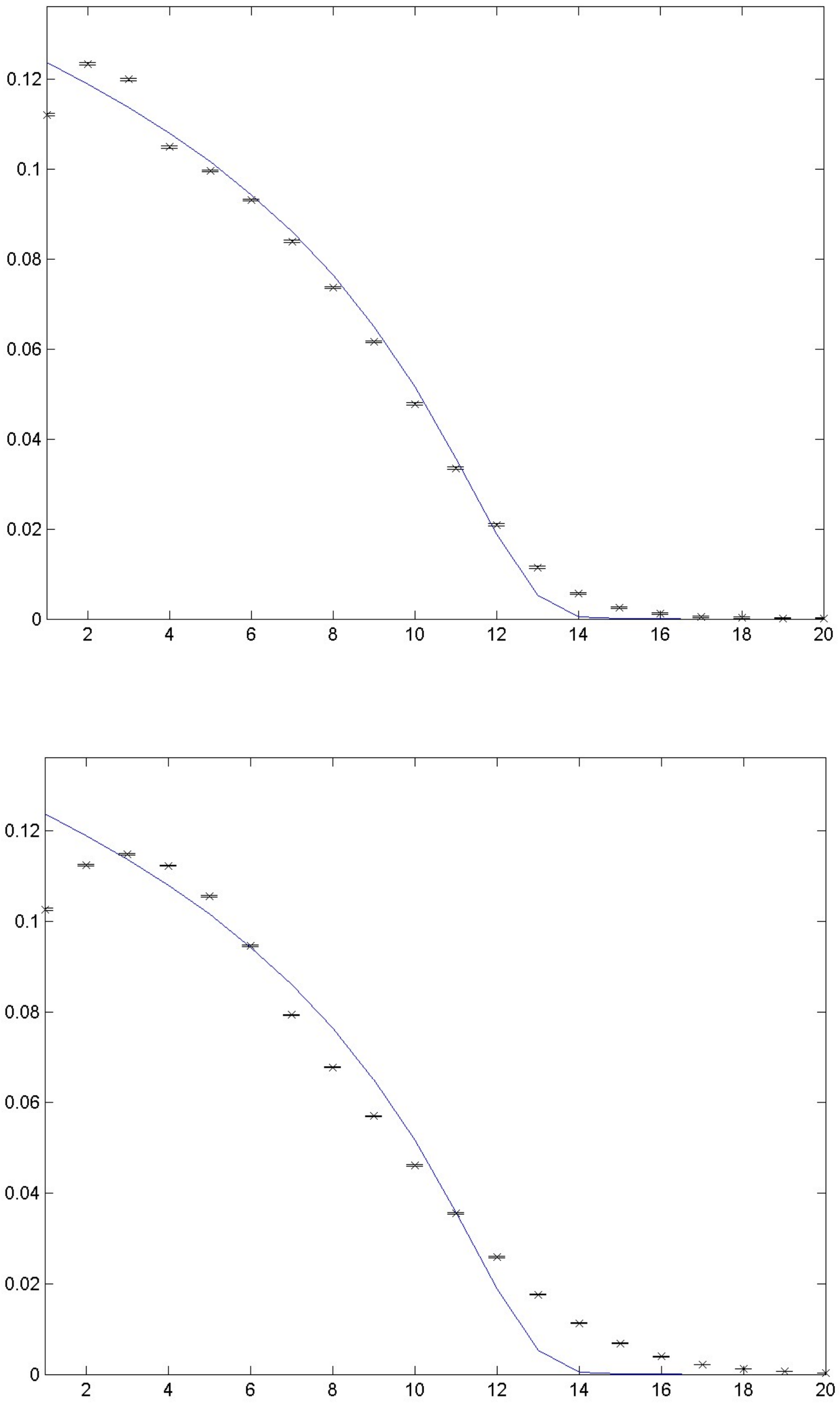

0.12
0.1
0.08
0.06
0.04
0.02
0
2
4
6
8
10
12
14
16
18
20
0.12
0.1
0.08
0.06
0.04
0.02
0
2
4
6
8
10
12
14
16
18
20

**Figure B1**

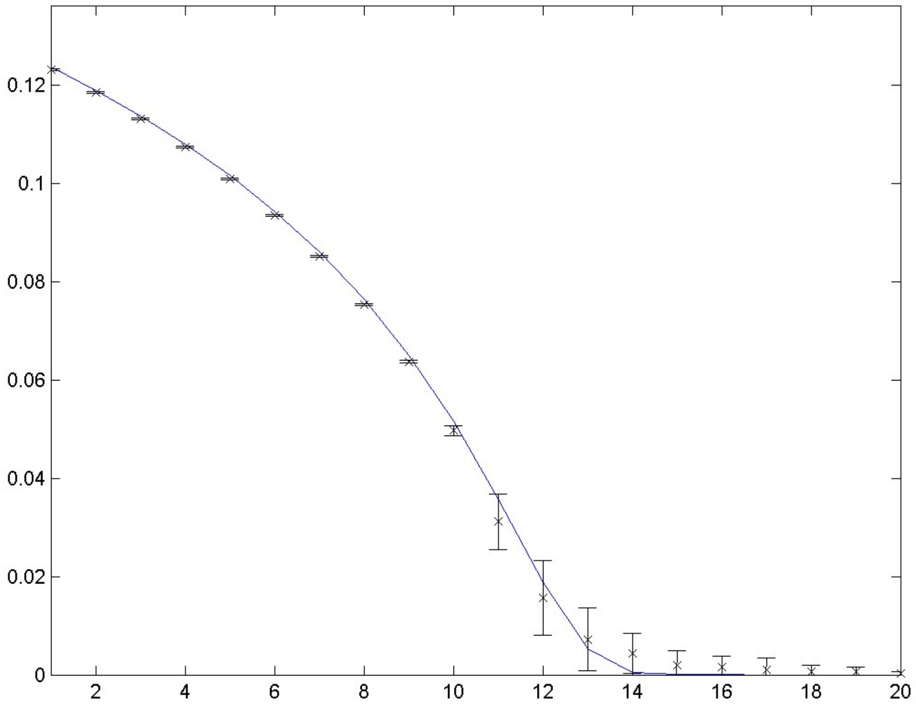


**Figure C1**

.

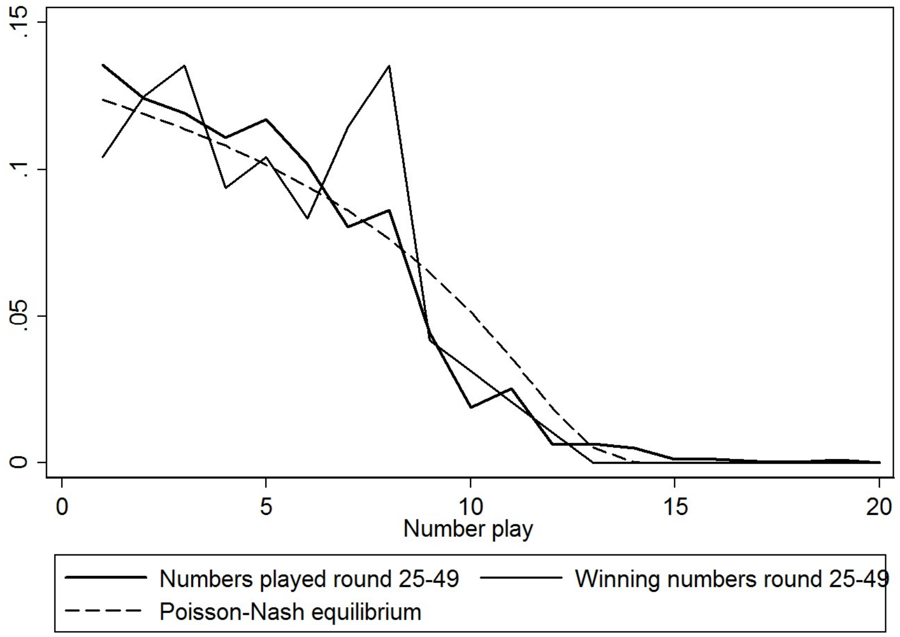


**Figure D1**

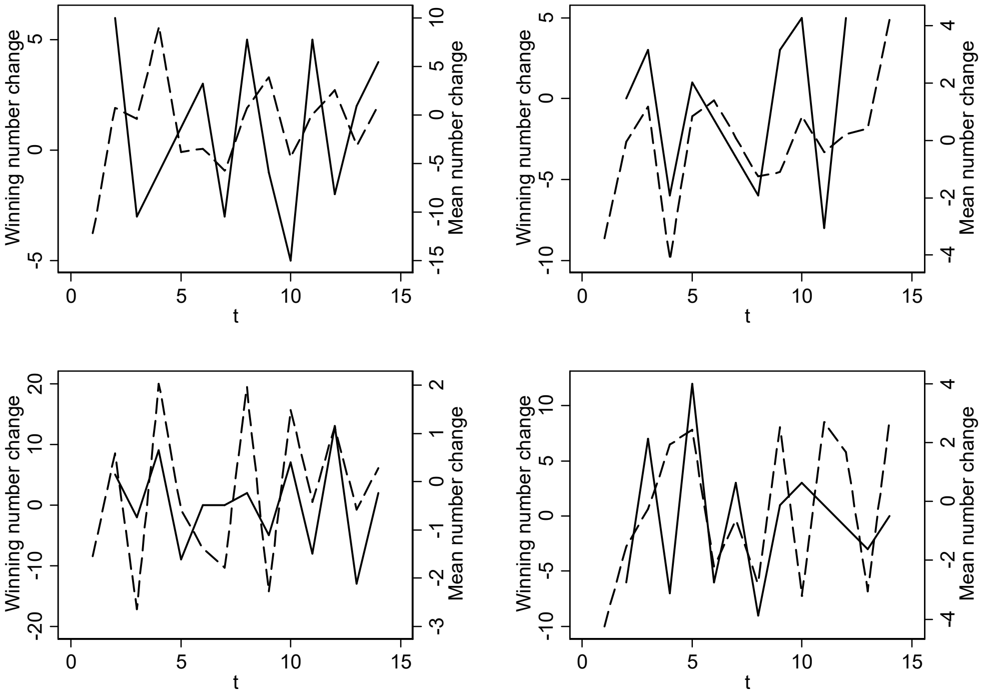


**Figure D2**

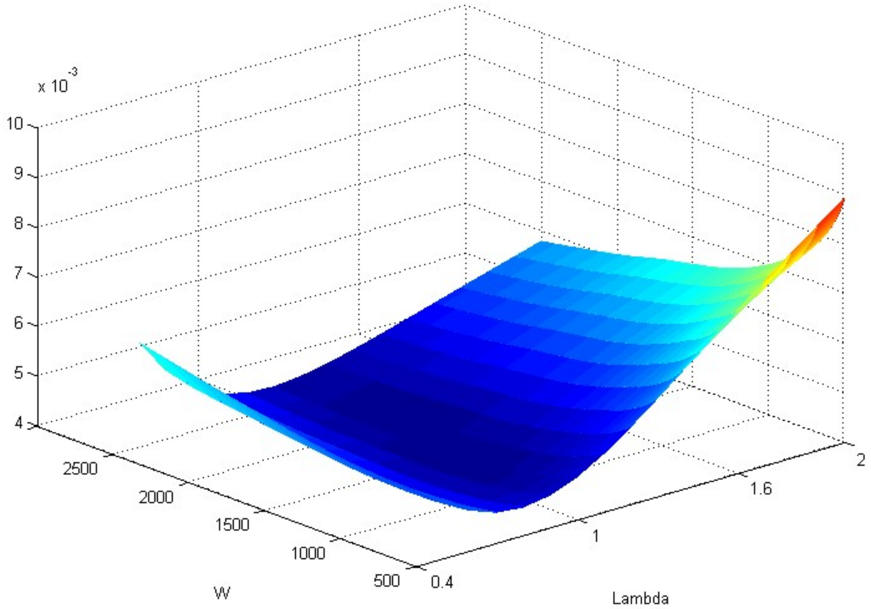


**Figure D3**

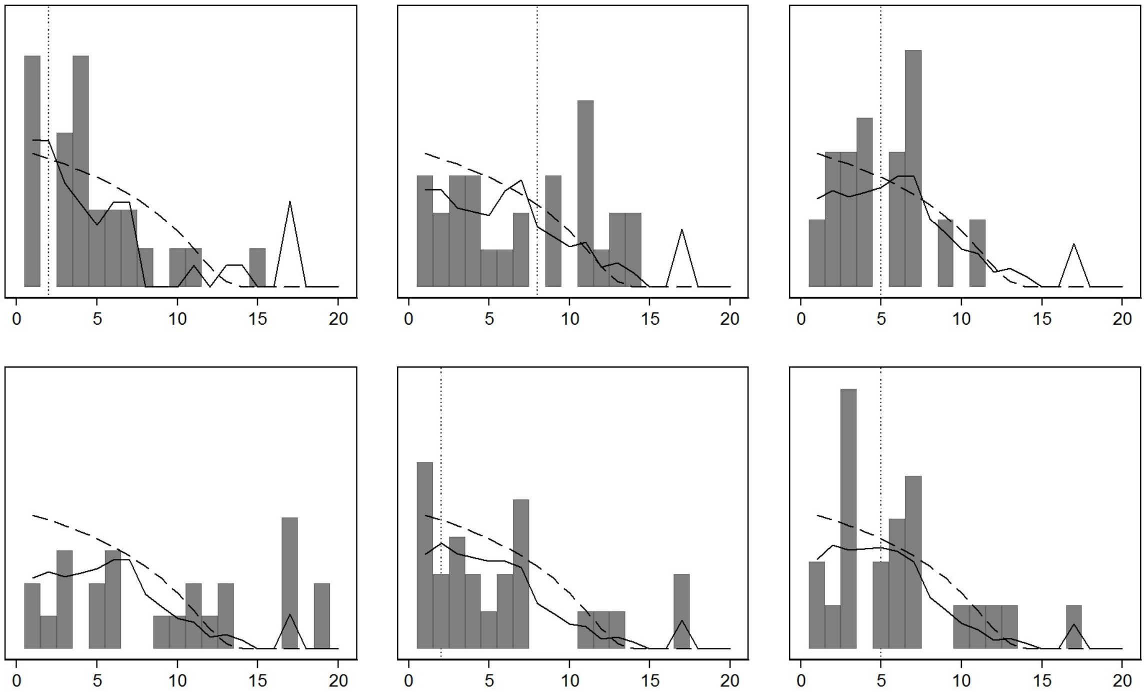


**Figure D4**

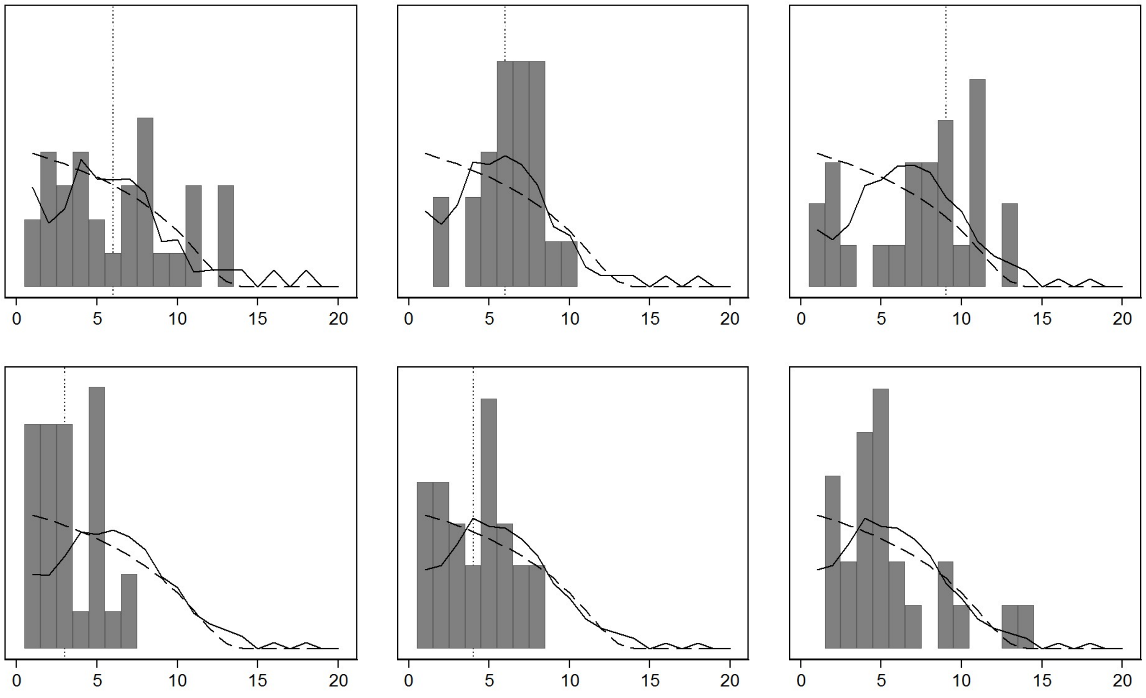


**Figure D5**

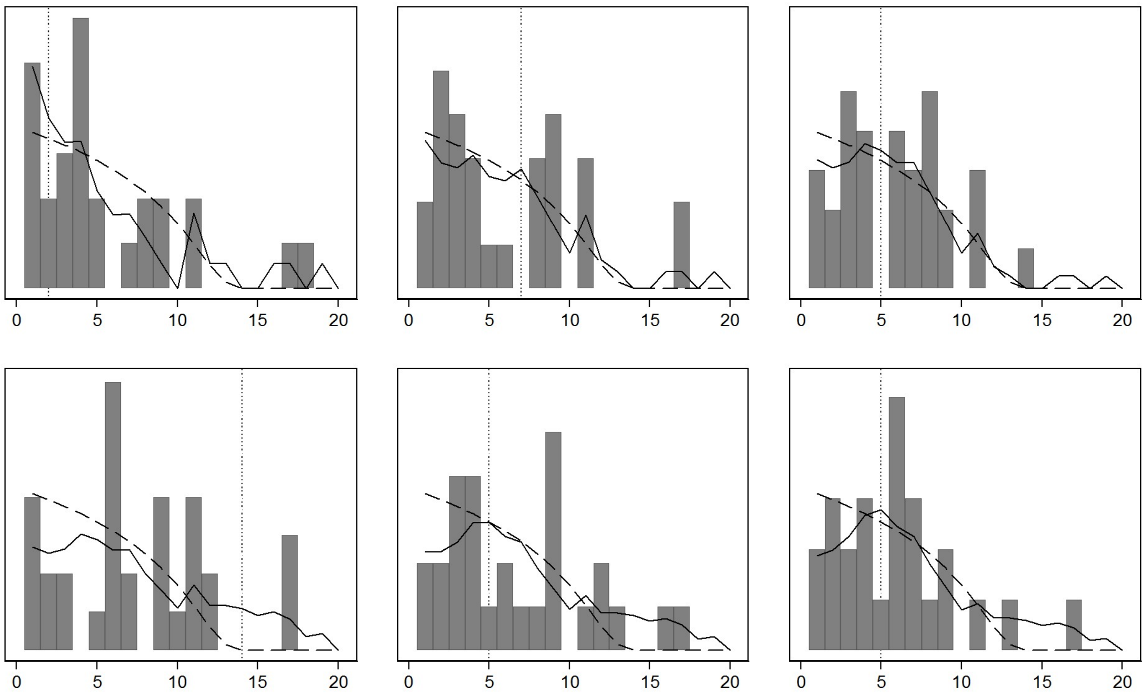


**Figure D6**

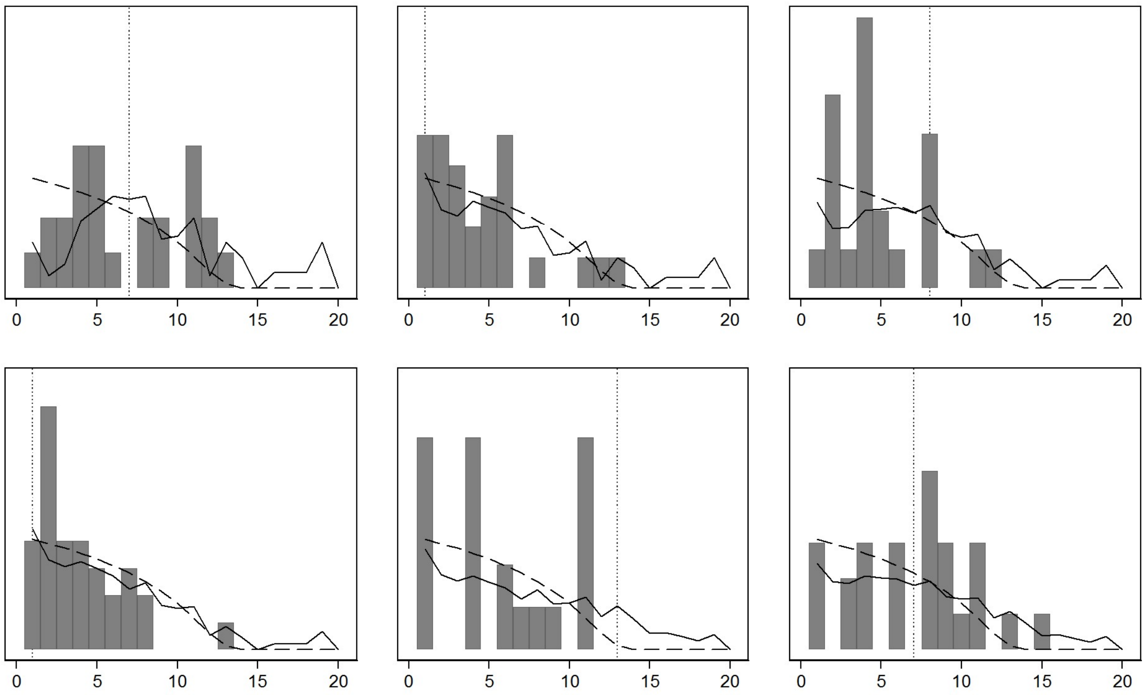


**Figure D7**

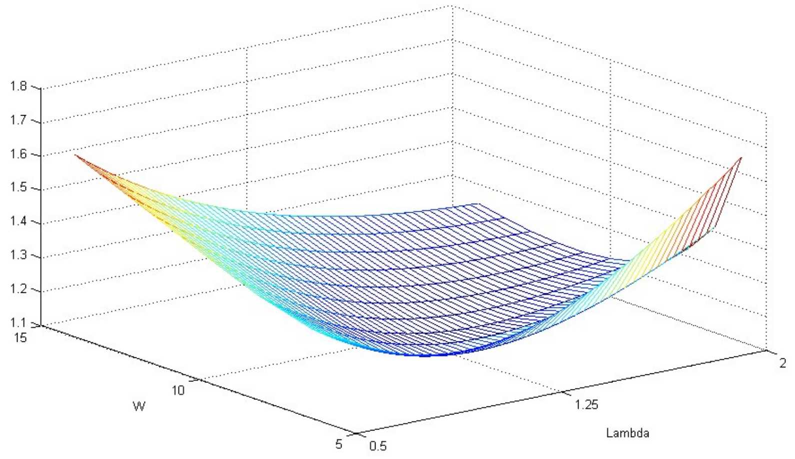


**Figure D8**

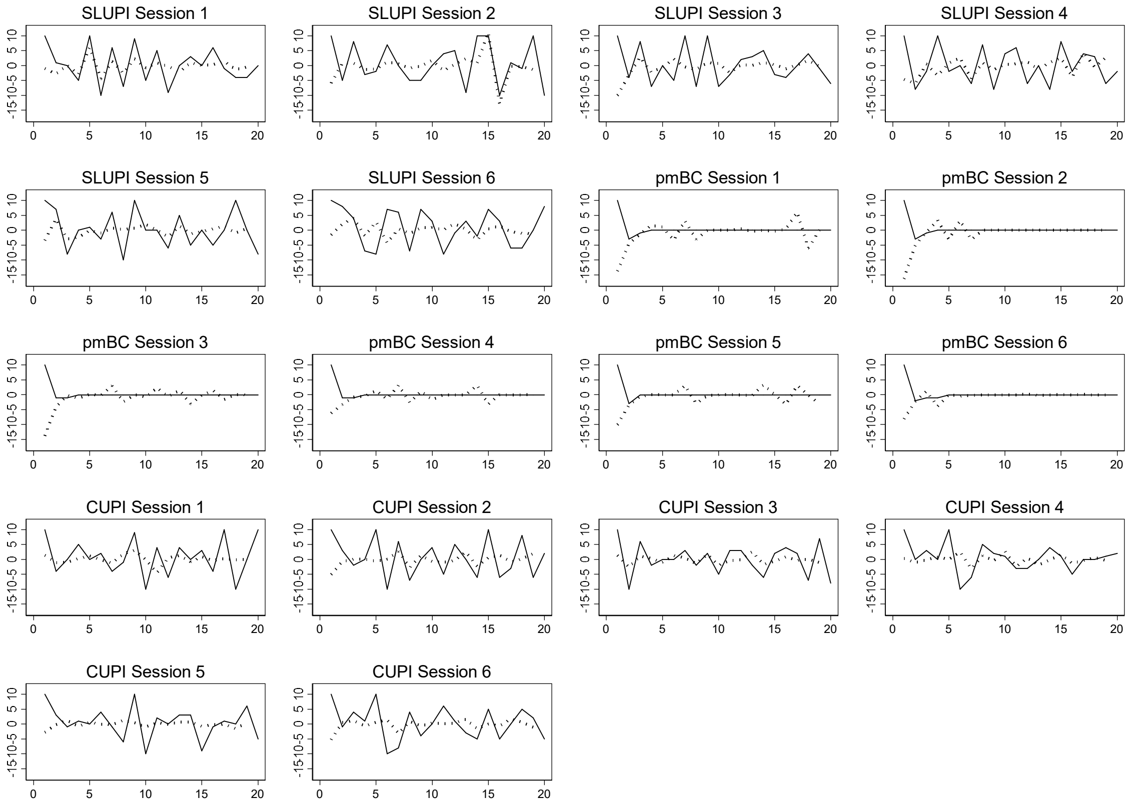
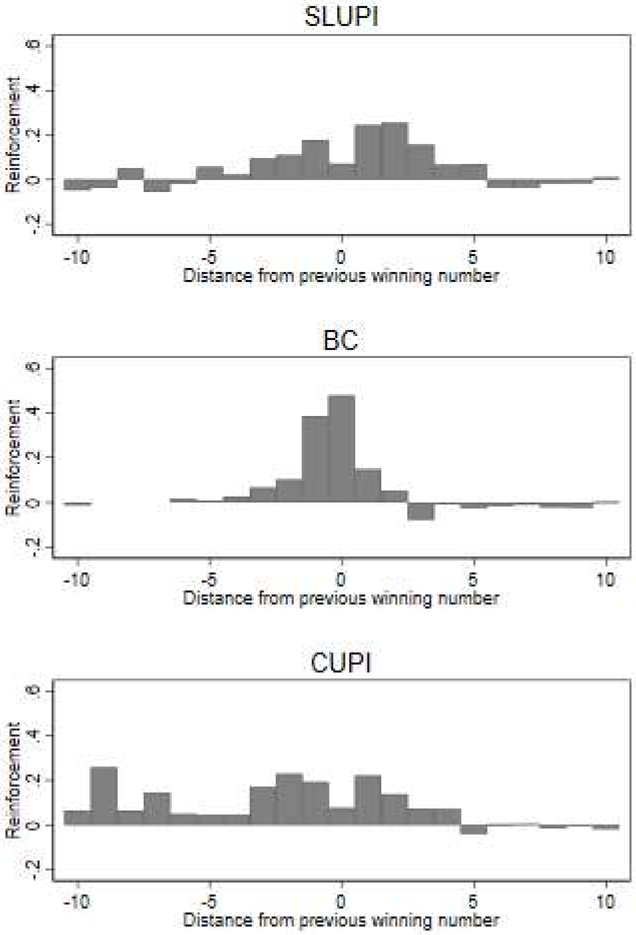


**Figure D9**


SLUPI

BC

CUPI

**Figure D10**

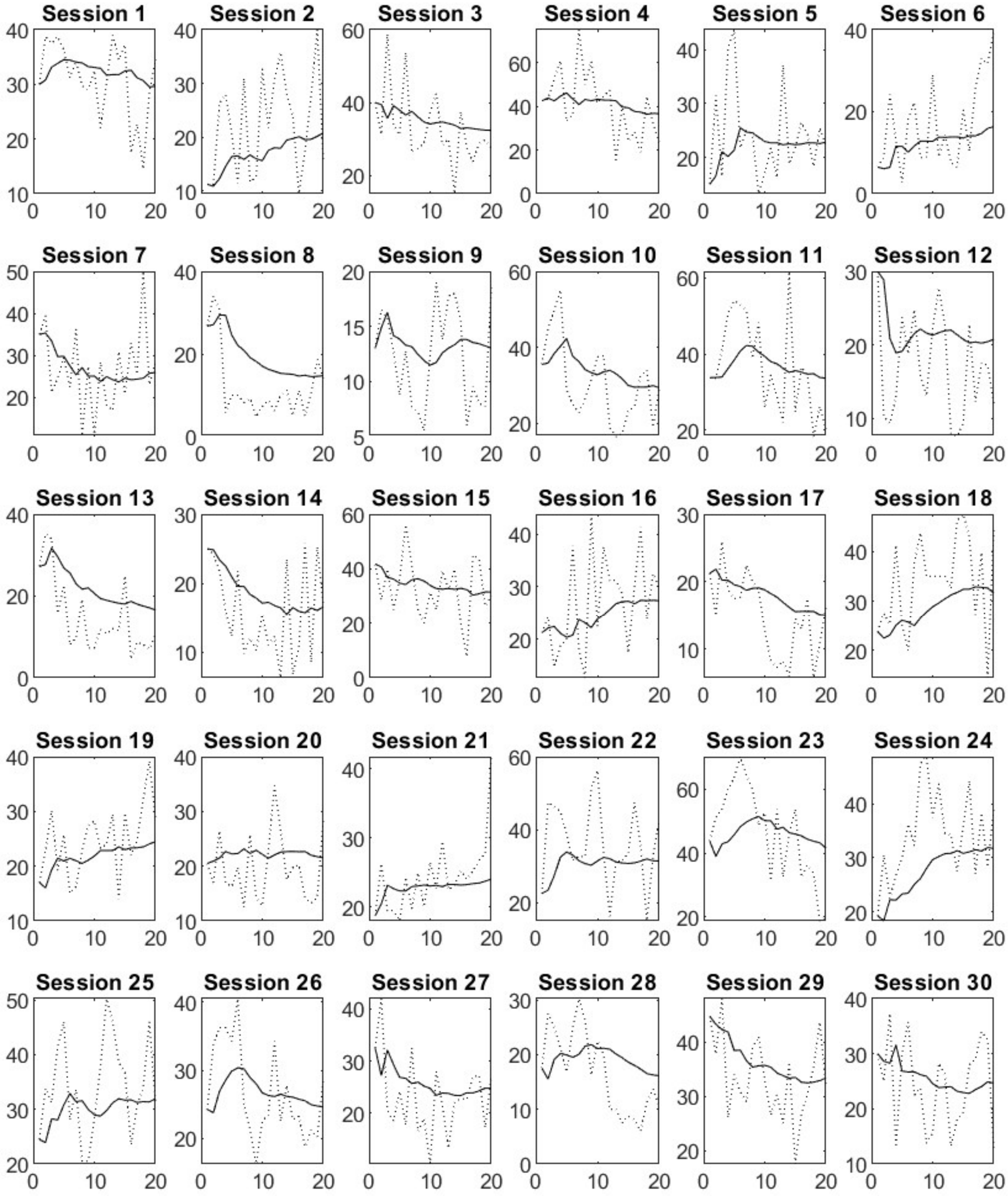

Session 1
Session 2
Session 3
Session 4
Session 5
Session 6
Session 7
Session 8
Session 9
Session 10
Session 11
Session 12
Session 13
Session 14
Session 15
Session 16
Session 17
Session 18
Session 19
Session 20
Session 21
Session 22
Session 23
Session 24
Session 25
Session 26
Session 27
Session 28
Session 29
Session 30
0
10
20

**Figure D11**

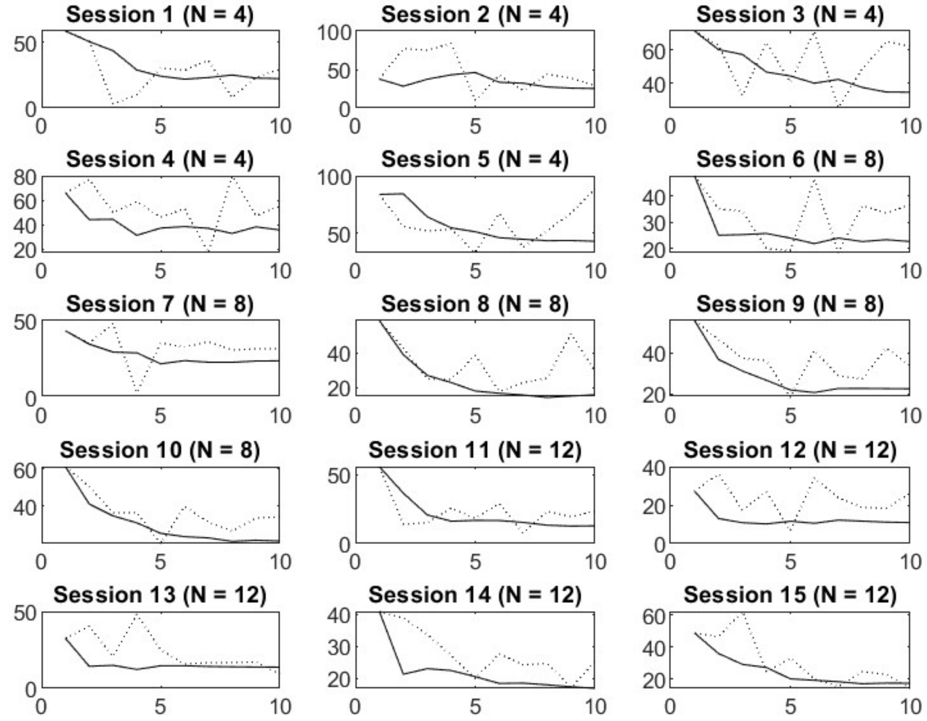

Session 1 (N = 4)
Session 2 (N = 4)
Session 3 (N = 4)
Session 4 (N = 4)
Session 5 (N = 4)
Session 6 (N = 8)
Session 7 (N = 8)
Session 8 (N = 8)
Session 9 (N = 8)
Session 10 (N = 8)
Session 11 (N = 12)
Session 12 (N = 12)
Session 13 (N = 12)
Session 14 (N = 12)
Session 15 (N = 12)
0
5
10

**Figure D12**

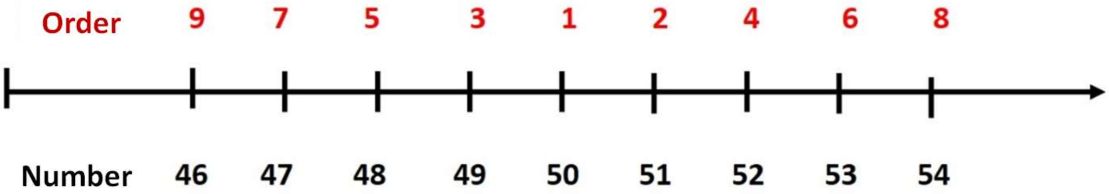


**Figure E1**

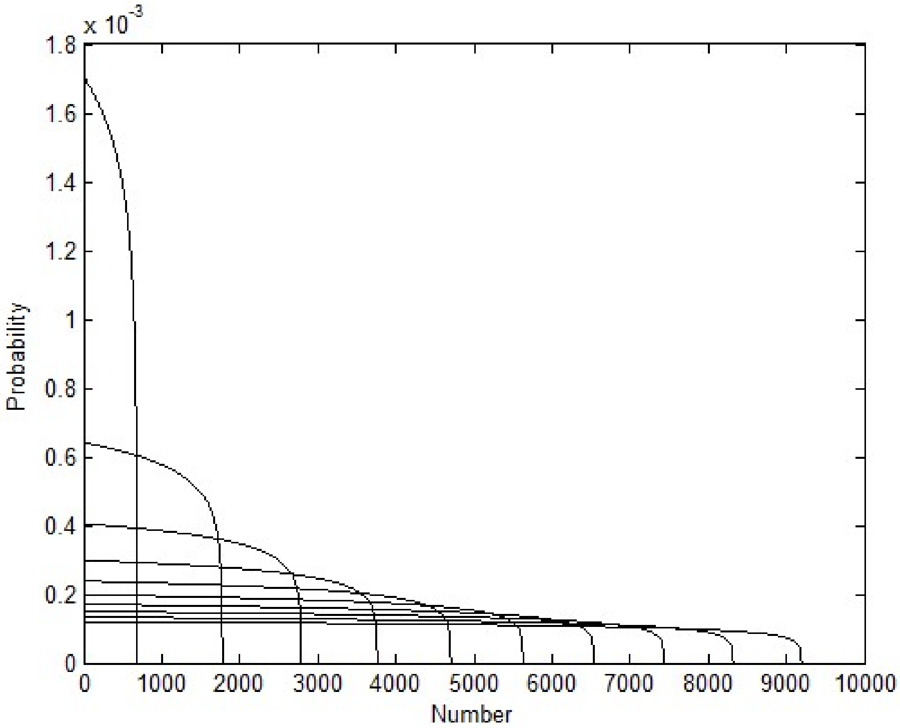


**Figure F1**

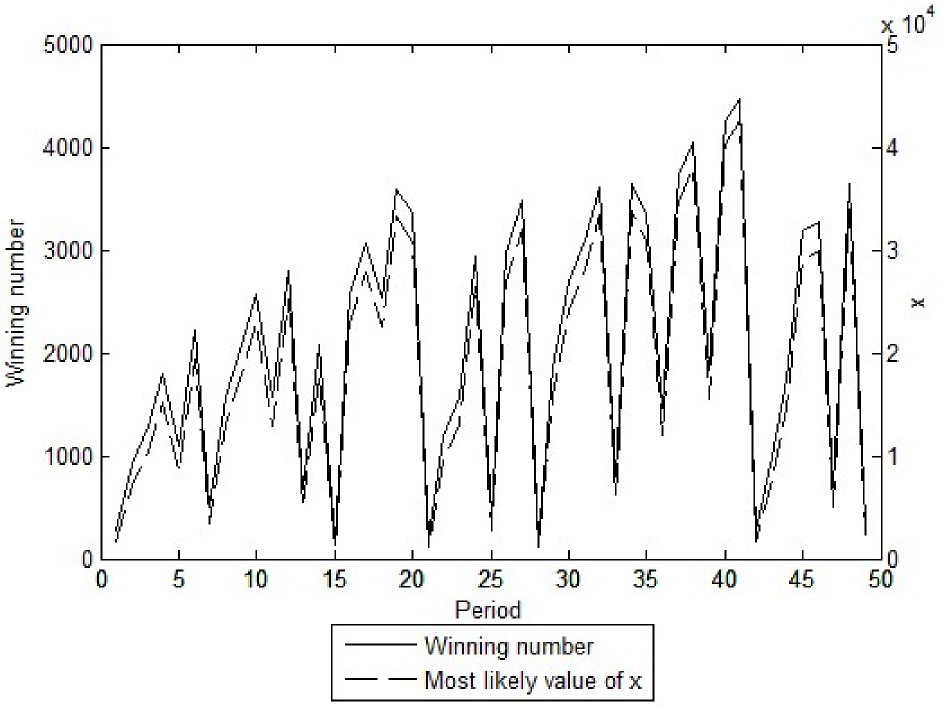


**Figure F2**

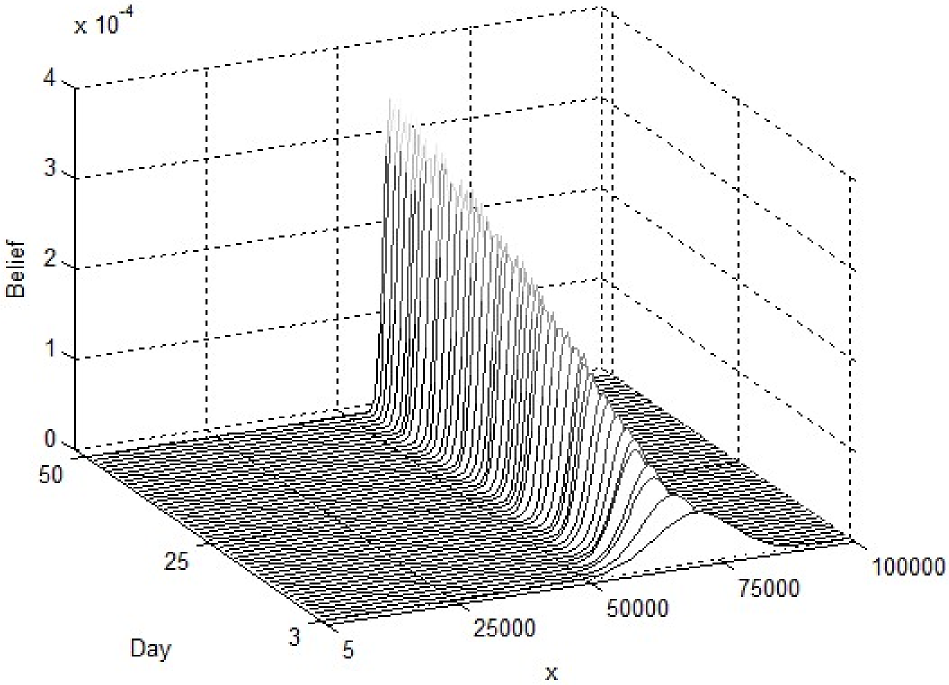


**Figure F3**

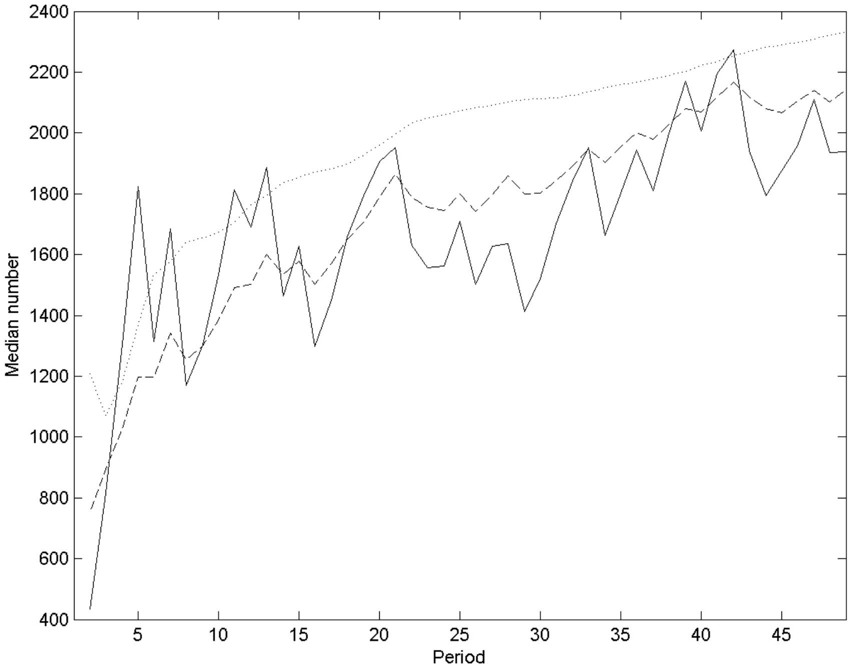


**Figure F4**

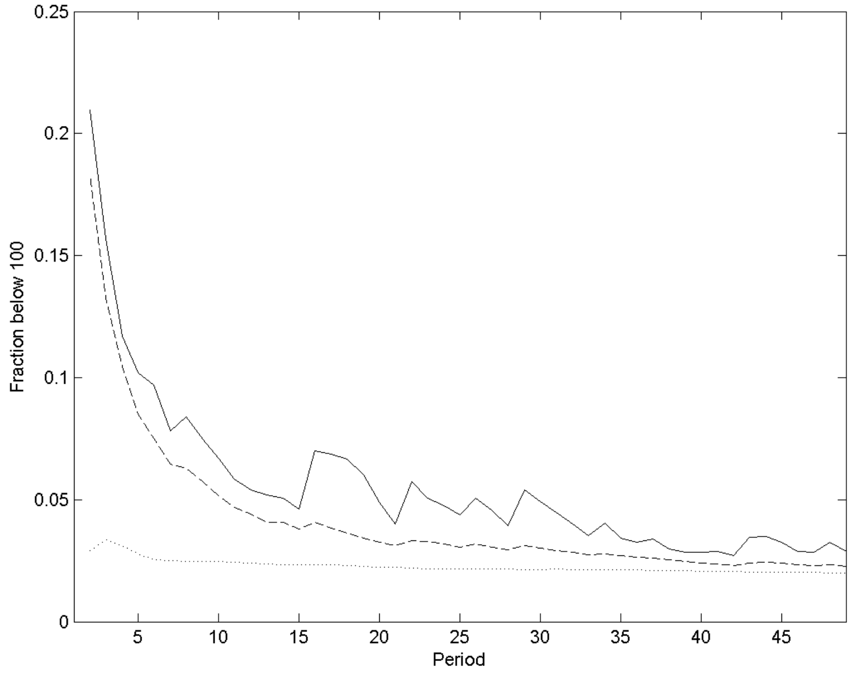


**Figure F5**