Online Appendix for "Testing Game Theory in the Field: Swedish LUPI Lottery Games"

Robert Östling, Joseph Tao-yi Wang, Eileen Y. Chou and Colin F. Camerer

A. The Symmetric Fixed-N Nash Equilibrium

Let there be a finite number of N players that each pick an integer between 1 and K. If there are numbers that are only chosen by one player, then the player that picks the lowest such number wins a prize, which we normalize to 1, and all other players get zero. If there is no number that only one player chooses, everybody gets zero.

To get some intuition for the equilibrium in the game with many players, we first consider the cases with two and three players. If there are only two players and two numbers to choose from, the game reduces to the following bimatrix game.

	1	2
1	$0,0$	$1,0$
2	$0,1$	$0,0$

This game has three equilibria. There are two asymmetric equilibria in which one player picks 1 and the other player picks 2, and one symmetric equilibrium in which both players pick 1.

Now suppose that there are three players and three numbers to choose from (i.e., $N = K = 3$). In any pure strategy equilibrium it must be the case that at least one player plays the number 1, but not more than two players play the number 1 (if all three play 1, it is optimal to deviate for one player and pick 2). In pure strategy equilibria where only one player plays 1, the other players can play in any combination of the other two numbers. In pure strategy equilibria where two players play 1, the third player plays 2. In total there are 18 pure strategy equilibria. To find the symmetric mixed strategy equilibrium, let p_1 denote the probability with which 1 is played and p_2 the probability with which 2 is played. The expected payoff from playing the pure strategies if the other two players randomize is given by

$$\begin{aligned}
\pi(1) &= (1-p_1)^2, \\
\pi(2) &= \left[(1-p_1-p_2)^2 + p_1^2\right], \\
\pi(3) &= \left[p_1^2 + p_2^2\right].
\end{aligned}$$

Setting the payoff from the three pure strategies yields $p_1 = 2\sqrt{3} - 3 = 0.464$ and $p_2 = p_3 = 2 - \sqrt{3} = 0.268$.

In the game with N players, there are numerous asymmetric pure strategy equilibria as in the three-player case. For example, in one type of equilibrium exactly one player picks 1 and the other players pick the other numbers in arbitrary ways. In order to find symmetric mixed strategy equilibria, let p_k denote the probability put on number k.[1]

[1] We have not been able to show that there is a unique symmetric equilibrium, but when numerically solving for a symmetric equilibrium we have not found any other equilibria than the ones reported below. Existence of a symmetric equilibrium is guaranteed since players have finite strategy sets, see Lemma 6 in Partha Dasgupta and Eric Maskin (1986).

In a symmetric mixed strategy equilibrium, the distribution of guesses will follow the multinomial distribution. The probability of x_1 players guessing 1, x_2 players guessing 2 and so on is given by

$$f(x_1, ..., x_K; N) = \begin{cases} \frac{N!}{x_1! \cdots x_K!} p_1^{x_1} \cdots p_K^{x_K} & \text{if } \sum_{i=1}^{K} x_i = N, \\ 0 & \text{otherwise,} \end{cases}$$

where we use the convention that $0^0 = 1$ in case any of the numbers is picked with zero probability. The marginal density function for the k^{th} number is the binomial distribution

$$f_k(x_k; N) = \frac{N!}{x_k!(N - x_k)!} p_k^{x_k} (1 - p_k)^{N - x_k}.$$

Let $g_k(x_1, x_2, ..., x_k; N)$ denote the marginal distribution for the first k numbers. In other words, we define g_k for $k < K$ as

$$g_k(x_1, x_2, ..., x_k; N) = \sum_{x_{k+1} + x_{k+2} + \cdots + x_K = N - (x_1 + x_2 + \cdots + x_k)} \frac{N!}{x_1! x_2! \cdots x_K!} p_1^{x_1} p_2^{x_2} \cdots p_K^{x_K}.$$

Using the multinomial theorem we can simplify this to[2]

$$g_k(x_1, x_2, ..., x_k; N) = \frac{N!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} \frac{(p_{k+1} + p_{k+2} + \cdots + p_K)^{N - (x_1 + x_2 + \cdots + x_k)}}{(N - (x_1 + x_2 + \cdots + x_k))!}.$$

If $k = K$, then $g_k(x_1, x_2, ..., x_k; N) = f(x_1, x_2, ..., x_k; N)$. Finally, let $h_k(N)$ denote the probability that nobody guessed k and there is at least one number between 1 to $k - 1$ that only one player guessed. This probability is given by (again if $k < K$)

$$h_k(N) = \sum_{\substack{(x_1, ..., x_{k-1}):\text{ some } x_i = 1 \\ \& \ x_1 + \cdots + x_{k-1} \leq N}} g_k(x_1, x_2, ..., x_{k-1}, 0; N).$$

If $k = K$, then this probability is given by

$$h_K(N) = \sum_{\substack{(x_1, ..., x_{k-1}):\text{ some } x_i = 1 \\ \& \ x_1 + \cdots + x_{k-1} = N}} f(x_1, x_2, ..., x_{K-1}, 0; N).$$

The probability of winning when guessing 1 and all other players follow the symmetric mixed strategy is given by

$$\pi(1) = f_1(0; N - 1) = (1 - p_1)^{N-1}.$$

[2]The multinomial theorem states that the following holds

$$(p_1 + p_2 + \cdots + p_K)^N = \sum_{x_1 + x_2 + \cdots + x_K = N} \frac{N!}{x_1! x_2! \cdots x_K!} p_1^{x_1} p_2^{x_2} \cdots p_K^{x_K},$$

given that all $x_i \geq 0$.

The probability of winning when playing $1 < k < K$ is given by[3]

$$\pi(k) = f_k(0; N-1) - h_k(N-1),$$
$$= (1 - p_k)^{N-1} - h_k(N-1).$$

Similarly, the probability of winning when playing $k = K$ is given by

$$\pi(K) = f_K(0; N-1) - h_K(N-1).$$

In a symmetric mixed strategy equilibrium, the probability of winning from all pure strategies in the support of the equilibrium must be the same. In the special case when $N = K$ and all numbers are played with positive probability, we can simply solve the system of $K - 2$ equations where each equation is

$$(1 - p_k)^{N-1} - h_k(N-1) = (1 - p_1)^{N-1},$$

for all $2 < k < K$ and the Kth equation

$$(1 - p_K)^{N-1} - h_K(N-1) = (1 - p_1)^{N-1}.$$

In principle, it is straightforward to solve this system of equations. However, computing the h_k function is computationally explosive because it requires the summation over a large set of vectors of length $k - 1$. Assuming $N = K$, the table below show the equilibrium for up to eight players.[4]

	3x3	**4x4**	**5x5**	**6x6**	**7x7**	**8x8**
1	0.4641	0.4477	0.3582	0.3266	0.2946	0.2710
2	0.2679	0.4249	0.3156	0.2975	0.2705	0.2512
3	0.2679	0.1257	0.1918	0.2314	0.2248	0.2176
4		0.0017	0.0968	0.1225	0.1407	0.1571
5			0.0376	0.0216	0.0581	0.0822
6				0.0005	0.0110	0.0199
7					0.0004	0.0010
8						0.0000

[3]The easiest way to see this is to draw a Venn diagram. More formally, let $A = \{\text{No other player picks } k\}$ and let $B = \{\text{No number below } k \text{ is unique}\}$, so that $P(A) = f_k(0; N-1)$ and $P(B) = h_k(N-1)$. We want to determine $P(A \cap B)$, which is equal to

$$P(A \cap B) = P(A) + P(B) - P(A \cup B).$$

To determine $P(A \cup B)$, note that it can be written as the union between two independent events

$$P(A \cup B) = P(B \cup (B' \cap A)).$$

Since B and $B' \cap A$ are independent,

$$P(A \cup B) = P(B) + P(B' \cap A).$$

Combining this with the expression for $P(A \cap B)$ we get

$$P(A \cap B) = P(A) - P(A \cap B').$$

[4]See Online Appendix B for details about how these probabilities were computed.

These probabilities are close to the Poisson-Nash equilibrium probabilities. To see this, the table below shows the Poisson-Nash equilibrium probabilities when N is equal to K for 3 to 8 players. Note that all the fixed-N and Poisson-Nash probabilities for all strategies in the 5x5 game and larger are within 0.02.

	3x3	**4x4**	**5x5**	**6x6**	**7x7**	**8x8**
1	0.4773	0.4057	0.3589	0.3244	0.2971	0.2747
2	0.3378	0.3092	0.2881	0.2701	0.2541	0.2397
3	0.1849	0.1980	0.2046	0.2057	0.2030	0.1983
4		0.0870	0.1129	0.1315	0.1430	0.1492
5			0.0355	0.0575	0.0775	0.0931
6				0.0108	0.0234	0.0385
7					0.0020	0.0064
8						0.0002

In order to calculate the Nash equilibrium with a fixed number of players for higher N and K, we have to use a different method (explained further in Online Appendix B). As is shown in Figure A1, the Poisson and fixed-N equilibrium are practically indistinguishable for $n = 27$ and $K = 99$ (parameters similar to those used in the lab experiment).

B. Computational and Estimation Issues

This appendix provides details about the numerical computations and estimations that are reported in the paper. We have used STATA and MATLAB for all computations and estimations.

Poisson-Nash Equilibrium

The Poisson-Nash equilibrium was computed in MATLAB through iteration of the equilibrium condition (1). Unfortunately, MATLAB cannot handle the extremely small probabilities that are attached to high numbers in equilibrium, so the estimated probabilities are zero for high numbers (17 and above for the laboratory and 5519 and above for the field).

Fixed-N Nash Equilibrium

To compute the equilibrium when the number of players is fixed and commonly known for up to eight players, we programmed the functions f_k, f_K, h_k and h_K in MATLAB and then solved the system of equations characterizing equilibrium using MATLAB's solver fsolve. Note that the h_k function includes the summation of a large number of vectors. For high k and N the number of different vectors involved in the summation grows explosively.

To compute the equilibrium for larger N,[5] let p_k be the symmetric equilibrium probabilities and let q_k be the conditional probability that a player picks number k conditional on not having guessed a lower number (so q_k depends on $p_1, p_2, ..., p_k$). This implies that

[5] We thank Dylan Thurston for suggesting this algorithm to speed up the computation.

$q_1 = p_1$, $q_2 = p_2/(1-q_1)$, $q_3 = p_3/\left[(1-q_1)(1-q_2)\right]$, $q_4 = p_4/\left[(1-q_1)(1-q_2)(1-q_3)\right]$, and so on.

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

$$l_{k+1,m} = \sum_{m' \ge m, m' \ne m+1} l_{k,m'} \binom{m'}{m} q_k^{(m'-m)} \left(1-q_k\right)^m .$$

Based on these probabilities the probability of winning on each number is

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

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

$$\begin{aligned}
l_{2,2} &= \sum_{m' \ge m, m' \ne m+1} l_{k,m'} \binom{m'}{m} q_k^{(m'-m)} \left(1-q_k\right)^m = l_{1,2}\left(1-g_1\right)^2 = \left(1-q_1\right)^2 \\
l_{2,1} &= l_{1,1}\left(1-q_1\right) = 0 \\
l_{2,0} &= l_{1,0} + l_{1,2} q_1^2 \left(1-q_1\right)^0 = q_1^2 \\
l_{3,2} &= l_{2,2}\left(1-q_2\right)^2 = \left(1-q_1\right)^2\left(1-q_2\right)^2 \\
l_{3,1} &= l_{2,1}\left(1-q_2\right) = 0 \\
l_{3,0} &= l_{2,0} + l_{2,2} q_2^2 = q_1^2 + \left(1-q_1\right)^2 q_2^2 = \left(2\sqrt{3}-3\right)^2 + \left(1-2\sqrt{3}+3\right)^2 0.5^2
\end{aligned}$$

The winning probabilities are given by

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

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

Cognitive Hierarchy with Quantal Response

Calculating the cognitive hierarchy prediction for a given τ and λ is straightforward. However, the cognitive hierarchy prediction is non-monotonic in τ and λ, implying that the log-likelihood function is not generally smooth.

In order to calculate the log-likelihood, we assume that all players play according to the same aggregate cognitive hierarchy prediction, i.e., the log-likelihood function is calculated using the multinomial distribution as if all players played the same strategy. For the field data, we calculated the log-likelihood for the daily average frequency for each week, but the frequency was rounded to integers in order to be able to calculate the log-likelihood. For the lab data, we instead calculated the log-likelihood by summing the frequencies for each week since we did not want unnecessary estimation errors due to rounding off to integers.

Maximum likelihood estimation for the field data is computationally demanding so we used a relatively coarse two-dimensional grid search. We used a 50x50 grid and restricted τ to be between 0.05 and 12, and restricted λ to be between 0.0001 and 0.05. We tried different bounds on the parameters and different grid sizes, but that did not change the results. The log-likelihood function for the first week is shown in Figure A8. The log-likelihood appears relatively smooth, but since we have been forced to use a very coarse grid we might not have found the global maximum.

For the maximum likelihood estimation of the lab data, we used a one-dimensional 100,000 grid search and let λ vary between 0.001 and 30. The log-likelihood function for the first week is shown in Figure A9. It is clear that the log-likelihood function non-smooth, but given that we have used such a fine grid we are confident that the estimated parameters are global optima. Figure A10 shows the log-likelihood function from the first week when τ is fixed at 1.5. Figure A7 shows the cognitive hierarchy prediction week-by-week for the laboratory data when τ is 1.5.

Model Selection

Since the Poisson-Nash equilibrium probabilities are zero for high numbers, the likelihood of the equilibrium prediction is always zero. However, to be able to compare the equilibrium prediction with the cognitive hierarchy model, we calculate the log-likelihoods using only data on numbers up to 5518 (field) and 16 (laboratory). These log-likelihoods cannot be directly compared with the log-likelihoods in Table 3 and 5, however, since those are calculated using data on all numbers. For comparison, we therefore compute the log-likelihoods for the cognitive hierarchy model in the same way as for the equilibrium prediction. In order for these probabilities to sum up to one, we divide the probabilities by the total probability attach to numbers up to the threshold (5518 or 16). Using the estimated parameters reported in Table 3, Table A1 shows the log-likelihoods only based on numbers up to 5518.

Week	1	2	3	4	5	6	7
Log-likelihood Eq. (<5519)	-43364	-32072	-28452	-27759	-28087	-21452	-19719
Log-likelihood CH (< 5519)	-23072	-20035	-17141	-15527	-15475	-15534	-14591

Table A1: Log-likelihoods for Poisson-Nash equilibrium and cognitive hierarchy for field data up to 5518

The log-likelihoods are higher for the cognitive hierarchy model in all weeks. The cognitive hierarchy model is estimated with two parameters, while the equilibrium prediction has no free parameters. One way to compare the models is to use Gideon Schwarz (1978) information criterion which penalizes a model depending on the number of estimated parameters by subtracting a factor $\log(n) \times m/2$ from the log-likelihood value, where n is the number of observations and m the number of estimated parameters. The log-likelihoods in Table A1 are calculated based on daily averages, so the penalty for the cognitive hierarchy model is approximately $\log(53783) = 10.9$, indicating that the cognitive hierarchy model is the better model in all weeks. Schwarz information criterion penalizes the number of estimated parameters more harshly than for example Aikake's information criterion. However, it should be kept in mind that the two parameters in cognitive hierarchy model are estimated using the data, whereas the equilibrium prediction is not estimated at all, so any comparison based on information criteria is likely to be unfair.

Week	1	2	3	4	5	6	7
Log-likelihood eq. (<17)	-192.9	-95.3	-91.3	-81.4	-93.0	-59.7	-145.2
Log-likelihood CH (<17)	-76.6	-61.2	-54.13	-45.5	-49.2	-43.9	-54.7
Log-likelihood CH $\tau = 1.5$ (<17)	-79.0	-52.9	-63.4	-56.5	-56.9	-65.3	-70.3
BIC eq. (<17)	-192.9	-95.3	-91.3	-81.4	-93.0	-59.7	-145.2
BIC CH (<17)	-79.8	-64.5	-57.4	-48.8	-52.5	-47.2	-58.0
BIC CH $\tau = 1.5$ (<17)	-82.3	-56.3	-66.7	-59.8	-60.2	-68.6	-73.6

Table A2: Log-likelihood and Schwarz information criterion (BIC) for the cognitive hierarchy and equilibrium models in the laboratory (up to 16)

Table A2 reports the restricted log-likelihoods and the corresponding values of the Schwarz information criterion for the laboratory data. Based on Schwarz information criterion, the cognitive hierarchy model outperforms equilibrium in all weeks, but the equilibrium prediction does better than the cognitive hierarchy model with $\tau = 1.5$ in the sixth week.

Equivalent Number of Observations

To calculate ENO for the Poisson LUPI game, suppose the model predicts probability p_k for choosing number k. By the independent actions property of Poisson games, each number $1, \ldots, K$ can be viewed as an independent condition, each predicting on average $G_k = n \cdot p_k$ observations where players pick k.[6]

In the data, we observe exactly N observations, and hence, set $n = N$ (fixed) for the Poisson model. Let x_{ki} be observation i for number k where $k = 1, 2, \ldots, K$ and $i = 1, 2, \ldots, N$. Then, $x_{ki} = 1$ if player i chooses k and 0 otherwise. The mean of the other $N - 1$ observations (except i) for each of the K conditions jointly form the vector $\bar{X}'_{oi} = (\bar{X}_{o1i}, \cdots, \bar{X}_{oki}, \cdots, \bar{X}_{oKi})$, in which

$$\bar{X}_{oki} = \sum_{j \neq i} x_{kj} = \frac{N \cdot \bar{X}_k - x_{ki}}{N - 1}$$

where $\bar{X}_k = \sum_{j=1}^{N} x_{kj}$ is the mean of all N subjects in condition k.

[6] This result depends crucially on the Poisson assumption. Also, since ENO is a property of the model for a specific dataset, one cannot directly compare different ENOs calculated using different datasets. In particular, when the variation in the data is large, so is ENO since a few noisy observations is insufficient to outweigh the model.

Let $\bar{X}_o = (\bar{X}_{o1}, \cdots, \bar{X}_{oi}, \cdots, \bar{X}_{oN})$ be the matrix of "mean of other $N-1$ observations," and $G' = (G_1, \cdots, G_k, \cdots, G_K)$ be vector of model predictions. Then, the mean squared errors (MSE) for the data and the model are

$$MSE(\bar{X}_o) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\sum_{i=1}^{N}\left(x_{ki} - \bar{X}_{oki}\right)^2$$

$$MSE(G) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\sum_{i=1}^{N}\left(x_{ki} - G_k\right)^2$$

Consider the pooled error variance (across all conditions):

$$S^2 = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N-1}\sum_{i=1}^{N}\left(x_{ki} - \bar{X}_k\right)^2$$

Note that (by the definition of $\bar{X}_{oki}$)

$$\begin{aligned} MSE(\bar{X}_o) &= \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\sum_{i=1}^{N}\left(x_{ki} - \frac{N\cdot\bar{X}_k - x_{ki}}{N-1}\right)^2 \\ &= \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\sum_{i=1}^{N}\left(\frac{N\cdot x_{ki} - N\cdot\bar{X}_k}{N-1}\right)^2 = \frac{N}{N-1}S^2 \end{aligned}$$

Let $\hat{N}$ be the corresponding N such that $MSE(\bar{X}_o) = MSE(G)$, indicating that the data and the model are equally good in prediction. This requires

$$MSE(G) = MSE(\bar{X}_o) = \frac{\hat{N}}{\hat{N}-1}S^2 = \frac{1}{\hat{N}-1}S^2 + S^2$$

Then, we define

$$ENO = \hat{N} - 1 = \frac{S^2}{MSE(G) - S^2}$$

To see how ENO can be interpreted, consider the following "restricted regression with recycling" that predicts each observation x_{ki} with two forecasts, the model prediction $G_{ki}(= G_k$ independent of $i)$ and the mean of the other $N-1$ observations $\bar{X}_{oki}$, with the restriction $\beta_1 + \beta_2 = 1$:

$$x_{ki} = \beta_1(G_{ki}) + \beta_2(\bar{X}_{oki}) + \epsilon_{ki}$$

Clive W. J. Granger and Ramu Ramanathan (1984) show that the optimal weight $\hat{\beta}_1$ that minimizes sum of squared residuals $\sum_{k=1}^{K}\sum_{i=1}^{N}\hat{\epsilon}_{ki}^2$ is the same as that of the ordinary least square regression

$$x_{ki} - \bar{X}_{oki} = \beta_1\left(G_{ki} - \bar{X}_{oki}\right) + \tilde{\epsilon}_{ki}$$

Therefore, the optimal weighted average predictor for the hold-out observation x_{ki} has

$$\hat{\beta}_1 = \frac{\sum_{k=1}^{K}\sum_{i=1}^{N}\left(x_{ki}-\bar{X}_{oki}\right)\cdot\left(G_{ki}-\bar{X}_{oki}\right)}{\sum_{k=1}^{K}\sum_{i=1}^{N}\left(G_{ki}-\bar{X}_{oki}\right)^2}$$

$$= \frac{\frac{1}{KN}\sum_{k=1}^{K}\sum_{i=1}^{N}\left[\left(x_{ki}-\bar{X}_{oki}\right)^2+\left(x_{ki}-\bar{X}_{oki}\right)\cdot\left(G_{ki}-x_{ki}\right)\right]}{\frac{1}{KN}\sum_{k=1}^{K}\sum_{i=1}^{N}\left[\left(G_{ki}-x_{ki}\right)^2+\left(x_{ki}-\bar{X}_{oki}\right)^2+2\left(G_{ki}-x_{ki}\right)\cdot\left(x_{ki}-\bar{X}_{oki}\right)\right]}$$

$$= \frac{MSE(\bar{X}_o)-CD\left(\bar{X}_o,G\right)}{MSE(\bar{X}_o)+MSE(G)-2CD\left(\bar{X}_o,G\right)}$$

where the "common deviation" is

$$CD(\bar{X}_o,G) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\sum_{i=1}^{N}\left(x_{ki}-\bar{X}_{oki}\right)\left(x_{ki}-G_{ki}\right)$$

Since $\bar{X}_k - \bar{X}_{oki} = \bar{X}_k - \frac{N\cdot\bar{X}_k - x_{ki}}{N-1} = \frac{1}{N-1}\left(x_{ki}-\bar{X}_k\right)$ and $\sum_{i=1}^{N}\left(x_{ki}-\bar{X}_k\right)=0$,

$$\begin{aligned}
CD(\bar{X}_o,G) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\sum_{i=1}^{N}\Big[&\left(x_{ki}-\bar{X}_k\right)^2+\left(\bar{X}_k-\bar{X}_{oki}\right)\cdot\left(x_{ki}-\bar{X}_k\right)\\
&+\left(x_{ki}-\bar{X}_k\right)\cdot\left(\bar{X}_k-G_{ki}\right)+\left(\bar{X}_k-\bar{X}_{oki}\right)\cdot\left(\bar{X}_k-G_{ki}\right)\Big]\\
= \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\sum_{i=1}^{N}\frac{N}{N-1}&\left(x_{ki}-\bar{X}_k\right)^2+\frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\sum_{i=1}^{N}\left(\bar{X}_k-\bar{X}_{oki}\right)\cdot\left(\bar{X}_k-G_{ki}\right)\\
= S^2+\frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}&\sum_{i=1}^{N}\left(\bar{X}_{oki}-\bar{X}_k\right)\cdot\left(G_{ki}-\bar{X}_k\right)
\end{aligned}$$

Ido Erev, Alvin Roth, Robert Slonim and Greg Barron (2007) show that under the assumption that $CD(\bar{X}_o,G)=S^2$, or when the errors of the two predictors are not correlated, we have (since $MSE(\bar{X}_o)=\frac{N}{N-1}S^2$)

$$\hat{\beta}_1 = \frac{\frac{N}{N-1}S^2-S^2}{\frac{N}{N-1}S^2+MSE(G)-2S^2} = \frac{\frac{S^2}{N-1}}{\frac{S^2}{N-1}+(MSE(G)-S^2)} = \frac{ENO}{ENO+(N-1)}$$

Hence, ENO is the weight given to the model prediction G_{ki}.[7]

Estimating the Poisson-Nash Equilibrium Week-by-Week

Using week-by-week data, we estimate the best-fitting n for the Poisson-Nash equilibrium by minimizing mean squared error:

$$MSE(G) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{N}\sum_{i=1}^{N}\left(x_{ki}-G_k\right)^2$$

[7] In private communication, Ido Erev gave the following illustration: Suppose you have a model of NBA star Shaquille O'Neal's free-throw shooting predicting an average of 50 percent. You see him make 8 out of 10 free-throws today. What is your updated belief that he will make the next one? Here, you have m=10 observations that predict 80 percent and a model that predicts 50 percent. Regardless of your prediction, the weight you put on the model is your ENO.

Note that we cannot employ the method of maximizing likelihood, as used to estimate the cognitive hierarchy model, because large number guesses occur in the data, while the last property of Proposition 1 states the predicted probabilities for large numbers should be practically zero.

The estimation process starts with $n_l = 1$ and $n_h = 100,000$, and consists of iterative grid searches (with a grid size of 30 for each iteration). We then picked the best-fitting integer values around these estimates. The top panel of Table A3 displays the results. In particular, the estimated n is 22319 for week 1, 28593 for week 2, and gradually increases to 33862 for week 5, before jumping up to 40331 and 41549 for week 6 and 7, respectively. Since T (the maximum number guessed with positive probability) is positively correlated with n, this reflects the week-by-week transition "filling up the gap between 3,000 and 5,500" illustrated in Figure A4.

Week	1	2	3	4	5	6	7
Estimated Poisson-Nash equilibrium							
Estimated n	22319	28593	31947	33184	33862	40331	41549
Actual n	57017	54955	52552	50471	57997	55583	47907
χ^2(for average frequency)	180.66	50.50	28.66	37.89	25.13	7.81	12.47
(Degree of freedom)	***(4)	***(6)	***(6)	***(7)	***(7)	(8)	(8)
Proportion below (%)	54.36	68.57	74.08	77.58	79.23	80.49	80.49
ENO	2512.9	6805.7	8793.1	11570.1	15861.4	19913.8	19421.2
Poisson-Nash equilibrium ($n = 53,783$)							
χ^2(for average frequency)	372.64	160.01	109.63	131.09	113.98	43.70	41.85
(Degree of freedom)	***(4)	***(6)	***(6)	***(7)	***(7)	***(8)	***(8)
Proportion below (%)	48.95	61.29	67.14	67.44	69.93	76.25	76.23
ENO	2176.4	4964.4	6178.4	7032.4	8995.0	14056.8	13879.3
Cognitive hierarchy model							
Log-likelihood	-53740	-31881	-22085	-19672	-19496	-19266	-17594
τ	1.80	3.17	4.17	4.64	5.02	6.76	6.12
λ	0.0034	0.0042	0.0058	0.0068	0.0069	0.0070	0.0064
χ^2(for average frequency)	41.83	35.23	5.44	3.88	4.49	4.59	5.48
(Degree of freedom)	***(4)	***(6)	(6)	(7)	(7)	(8)	(8)
Proportion below (%)	62.58	72.57	78.65	80.17	82.09	82.43	82.24
ENO	3188.8	7502.5	9956.0	12916.1	17873.0	21469.6	21303.0

*=10 percent, **=5 percent and ***=1 percent significance level.
The degree of freedom for a χ^2 test is the number of bins minus one.
The proportion below the theoretical prediction refers to the fraction of the empirical density that lies below the theoretical prediction, or one minus the fraction of overshooting.

Table A3: Goodness-of-fit for the estimated Poisson-Nash equilibrium and cognitive hierarchy for field data

Nevertheless, these estimated n are far less than the actual number of guesses. As shown in Table A3, the average number of guesses per day is $57,017$ for week 1 and fluctuates quite a bit but eventually drops to $47,907$ for week 7. Table 1 reports the average number of guesses for all 49 days as $53,783$ and the minimum number of actual guesses is $38,933$. (See Figure A2 for the entire distribution.) Hence, the *maximum* estimated n (week 6 and 7) are only comparable to the *minimum* of the actual number of guesses, bounding the distribution of estimated n away from the majority of n that actually occurred.

Though it is not clear how these incorrect beliefs sustained throughout the 49 days, we can still compare the goodness-of-fit results of the estimated Poisson-Nash equilibrium with to the cognitive hierarchy model. Table A3 provides this comparison, reporting three goodness-of-fit measurements for the estimated Poisson-Nash equilibrium in the top panel, and replicate the parameter-free Poisson-Nash equilibrium and cognitive hierarchy results of Table 3 in the middle and bottom panel. To ensure predicted observations to be at least 5 for the estimated Poisson-Nash equilibrium, we set the χ^2 test bin numbers to 5, 7, 7, 8, 8, 9, and 9, for week 1 to 7. This is done for all three models to obtain comparable test results (and differs from Table 3).

As shown in the top panel of Table A3, the χ^2 test rejects the estimated Poisson-Nash equilibrium for week 1-5 at the 1% level, even when we bin 500 numbers into one category and consider the rounded average of each bin. However, the estimated equilibrium performs better than the original Poisson-Nash equilibrium (which is parameter free), and cannot be rejected for week 6 and 7. But according to the χ^2 test statistics, the cognitive hierarchy model fits better in all weeks.

The proportion below theoretical prediction and ENO also yield similar results: Though better than the original parameter-free Poisson-Nash equilibrium, the weekly estimated Poisson-Nash equilibrium fits worse than the cognitive hierarchy in the first week (54.4% vs. 62.6%; ENO: 2513 vs. 3189), but improves substantially and and is very similar to the cognitive hierarchy model in the last week 6 (80.5% vs. 82.2%; ENO: 19914 vs. 21303).

C. Additional Details About the Field LUPI Game

This part of the Appendix provides some additional details about the field game that was not discussed in the main text.[8]

The prize guarantee for the winner of 100,000 SEK was first extended until the 11th of March and then to the 18th of March, so the prize guarantee covered all days for which we have data. The thresholds for the second and third prizes were determined so that the second prizes constituted 11 percent of all bets and the third prizes 17.5 percent. The winner of the first prize also won the possibility to participate in a "final game".[9] The final game ran weekly and had four to seven participants. The "final game" consisted of three rounds where the participants chose two numbers in each round. The rules of this game were very similar to the original game, but what happened in this game did not depend on what number you chose in the main game, so we leave out the details about this game.

The Hux Flux randomization option involved a uniform distribution where the support of the distribution was determined by the play during the 7 previous days.[10] It became

[8]Stefan Molin at Svenska Spel told us that he invented the game in 2001 after taking a game theory course from the Swedish theorist and experimenter Martin Dufwenberg.

[9]3.5 percent of all daily bets were reserved for this "final game".

[10]In the first week HuxFlux randomized numbers uniformly between 1 and 15000. After seven days of play, the computer randomized uniformly between 1 and the average 90th percentile from the previous

possible to play the game on the Internet sometime between the 21st and 26th of February 2007. The web interface for online play is shown in Figure A12. This interface also included the option HuxFlux, but in this case players could see the number that was generated by the computer before deciding whether to place the bet.

We use daily data from the first seven weeks. The reason is that the game was withdrawn from the market on the 24th of March 2007 and we were only able to access data up to the 18th of March 2007.

The game was heavily advertised around the days when it was launched. The main message was that this was a new game in which you should strive to be the only person choosing the lowest unique number. The winning numbers (for the first, second, and third prizes) were reported on TV, text-TV and the Internet every day. In the TV programs they reported not only the winning numbers, but also commented briefly about how people had played previously.

The richest information about the history of play was given on the home page of Svenska Spel. People could display and download the frequencies of all numbers played for all previous days. However, this data was presented in a raw format and therefore not very accessible. The homepage also displayed a histogram of yesterday's guesses which made the data easier to digest. An example of how this histogram looked is shown in Figure A13. The homepage also showed the total number of bets that had been made so far during the day.

The web interface for online play also contained some easily accessible information. Besides links to the data discussed above as well as information about the rules of the game, there were some pieces of statistics that could easily be displayed from the main screen. The default information shown was the first name and home town of yesterday's first prize winner and the number that that person guessed. By clicking on the pull-down menu in the middle, you could also see the seven most popular guesses from yesterday. This information was shown in the way shown in Figure A14. By moving the mouse over the bars you can see how many people guessed that number. In this example, the most popular number was 1234 with 85 guesses! Note that this information was not easily available before online play was possible. From the same pull-down menu, you could also see the total number of distinct numbers people guessed on during the last seven days. Finally, you could display the numbers of the second- and third prize winners of yesterday.

In addition to this information, Svenska Spel also published posters with summary statistics for previous rounds of the game (see Figure A15). The information given on these posters varied slightly, but the one in Figure A15 shows the winning numbers, the number of bets, the size of the first prize and if there was any numbers below the winning number that no other player chose. It also shows the average, lowest and highest winning number, as well as the most frequently played numbers.

D. Additional Details About the Lab Experiment

Screenshots from the input and results screens of the laboratory experiment are shown in Figure A16 and A17. Figure A18 shows screenshots from the post-experimental questionnaire and Figure A19 a screenshot from the CRT.

seven days. However, the only information given to players about HuxFlux was that a computer would choose a number for them.

Figure A20 displays the aggregate data from non-selected and selected subjects' choices. Subjects are slightly more likely to play high numbers above 20 when they are not selected to participate, but overall the pattern looks very similar. This implies that subjects' behavior in a particular round is almost unaffected depending on whether they had marginal monetary incentives or not.

Experimental Instructions

Instructions for the laboratory experiment are as follows (translated directly by one of the authors from the Swedish field instructions, but modified in order to fit the laboratory game):

Instruction for Limbo[11]

Limbo is a game in which you choose to play a number, between 1 and 99, that you think nobody else will play in that round. The lowest number that has been played only once wins.

The total number of rounds will not be announced. At the beginning of each round, the computer will indicate whether you have been selected to participate in that round. The computer selects participating players randomly so that the average number of participating players in each round is 26.9. Please choose a number even if you are not selected to participate in that round.

[**Instructions where the Poisson distribution is explicitly described:**

The game is played in 49 rounds. In the beginning of each round, the computer will indicate whether you have been selected to participate in this round. The computer selects participating players randomly so that the average number of participating players in each round is 26.9.

Specifically, the number of players in each round is pre-drawn from a so called Poisson distribution. The diagram below shows the Poisson distribution with mean 26.9. The horizontal axis shows different possible numbers of participants, and the vertical axis shows the probability of having that many participants. Notice that in some rounds there are more than 27 players and in other rounds there are fewer than 27 players. You will not know how many players are participating in each round. All you know is the probabilities of what the number of players might be, given by the distribution shown in the diagram.

On the second screen you can indicate which number between 1 and 99 that you want to play in that round.

Note: We also attached Figure A5 at the end of the instructions.]

After all participating players have selected a number, the round is closed and all bets are checked. The lowest unique number that has been received is identified and the person that picked that number is awarded a prize of 7$.

The winning number is reported on the screen and shown to everybody after each round.

Prizes are paid out to you at the end of the experiment.

If you have any questions, raise your hand to get the experimenter's attention. Please be quiet during the experiment and do not talk to anybody except the experimenter.

[11] In order to mirror the field game as closely as possible, we referred to the LUPI game as "Limbo" in the lab.

Individual Lab Results

The regression results in Table 6 mask a considerably degree of heterogeneity between individual subjects. Based on the responses in the post-experimental questionnaire, we coded four variables depending on whether they mentioned each aspect as a motivation for their strategy.

Random All subjects who claimed that they played numbers randomly were coded in this category.[12]

Stick All subjects who stated that they stuck to one number throughout parts of the experiment were included in this category. Many of these subjects explained their choices by arguing that if they stuck with the same number, they would increase the probability of winning.

Lucky This category includes all subjects who claimed that they played a favorite or lucky number.

Strategic This category includes all players who explicitly motivated their strategy by referring to what the other players would do.[13]

Several subjects were coded into more than one category.[14]

How well does the classification based on the self-reported strategies explain behavior? Table A4 reports regressions where the dependent variables are four summary statistics of subjects' behavior—the number of distinct choices, the mean number, the standard deviation of number, and the total payoff. In the first column for each measure of individual play only the four categories above are included as dummy variables. There are few statistically significant relationships. Subjects coded into the "Stick" category did tend to choose fewer and less dispersed numbers, and subjects coded as "Lucky" tend to pick higher and more dispersed numbers. Table A4 also report regressions for the same dependent variables and some demographic variables. The only statistically significant relationship is that subjects familiar with game theory tend to pick less dispersed numbers (though their payoffs are not higher). Note that the explanatory power is very low and that there are no significant coefficients in the regressions on the total payoff from the experiment. This suggests that it is hard to affect the payoff by using a particular strategy, which is consistent with the fully mixed equilibrium (where payoffs are the same for all strategies).

[12]For example, one subject motivated this strategy choice in a particular sophisticated way: "First I tried logic, one number up or down, how likely was it that someone else would pick that, etc. That wasn't doing any good, as someone else was probably doing the exact same thing. So I started mentally singing scales, and whatever number I was on in my head I typed in. This made it rather random. A couple of times I just threw curveballs from nowhere for the hell of it. I didn't pay any attention to whether or not I was selected to play that round after the first 3 or so."

[13]For example, one subject stated the following: "I tried to pick numbers that I thought other people wouldn't think of—whatever my first intuition was, I went against. Then I went against my second intuition, then picked my number. After awhile, I just used the same # for the entire thing."

[14]For example, the following subject was classified into all but the "Lucky" category: "At first I picked 4 for almost all rounds (stick) because it isn't considered to be a popular number like 3 and 5 (strategic). Afterwards, I realized that it wasn't helping so I picked random numbers (random)."

	# Distinct		Mean		Std. dev.		Payoff	
Random	0.77		-0.37		-0.54		-0.26	
	(1.44)		(-0.42)		(-0.61)		(-0.21)	
Stick	-1.48^{***}		-1.12		-1.50^{*}		-0.36	
	(-2.80)		(-1.30)		(-1.70)		(-0.29)	
Lucky	1.24		4.43^{***}		3.73^{***}		-0.39	
	(1.60)		(3.52)		(2.88)		(-0.22)	
Strategic	0.35		-0.65		-0.54		1.42	
	(0.68)		(-0.78)		(-0.63)		(1.21)	
Age		-0.02		-0.00		0.02		0.26
		(-0.18)		(-0.02)		(0.11)		(1.32)
Female		-0.23		-0.92		-1.03		-1.10
		(-0.46)		(-1.12)		(-1.23)		(-0.99)
Income (1-4)		-0.13		-0.35		-0.50		0.38
		(-0.48)		(-0.81)		(-1.17)		(0.67)
Lottery player		0.17		0.59		0.39		-0.13
		(0.34)		(0.70)		(0.47)		(-0.12)
Game theory		0.25		0.23		-1.48^{*}		-0.55
		(0.63)		(-0.28)		(-1.74)		(-0.49)
R^2	0.08	0.01	0.10	0.02	0.08	0.04	0.01	0.02
Obs.	152	152	152	152	152	152	152	152

Only selected choices are included in the calculation of the dependent variables. $t-$statistics within parentheses. Constant included in all regressions. *=10 percent, **=5 percent and ***=1 percent significance level.

Table A4: Linear regressions explaining individual behavior

The questionnaire in two of the sessions also contained the three-question Cognitive Reflection Test (CRT) developed by Shane Frederick (2005).[15] The purpose with collecting subjects' responses to the CRT is to get some measure of cognitive ability. In line with the results reported in Frederick (2005), a majority of the UCLA subjects answered only zero or one questions correctly. Interestingly, there does not appear to any relation between player's behavior or payoff in the LUPI game and the number of correctly answered questions, but the sample size is small ($n = 76$). The number of correctly answered CRT questions is not significant when the four measures in Table A4 are regressed on the CRT score.

Figure A21 shows a histogram of the number of distinct numbers that subjects played during the experiments. Based only on choices when players were selected to participate, subjects played on average 9.65 different numbers, compared to 10.9 expected in Poisson-Nash equilibrium. Figure A21 also shows a simulated distribution of how many distinct numbers players would pick if they played according to the equilibrium distribution.

References

Dasgupta, Partha, and Eric Maskin. 1986. "The Existence of Equilibrium in Discontinuous Economic Games, I: Theory." *Review of Economic Studies*, 53(1): 1–26.

Erev, Ido, Alvin Roth, Robert Slonim, and Greg Barron. 2007. "Learning and Equilibrium as Useful Approximations: Accuracy of Prediction on Randomly Selected Constant Sum Games." *Economic Theory*, 33(1): 29–51.

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives*, 19(4): 25–42.

Granger, Clive W. J., and Ramu Ramanathan. 1984. "Improved Methods of Combining Forecasts." *Journal of Forecasting*, 3(2): 197–204.

Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *Annals of Statistics*, 6(2): 461–464.

[15] The CRT consists of three questions, all of which would have an instinctive answer, and a counterintuitive, but correct, answer. See Frederick (2005) or the screenshot in Figure A19 for the questions that we used.

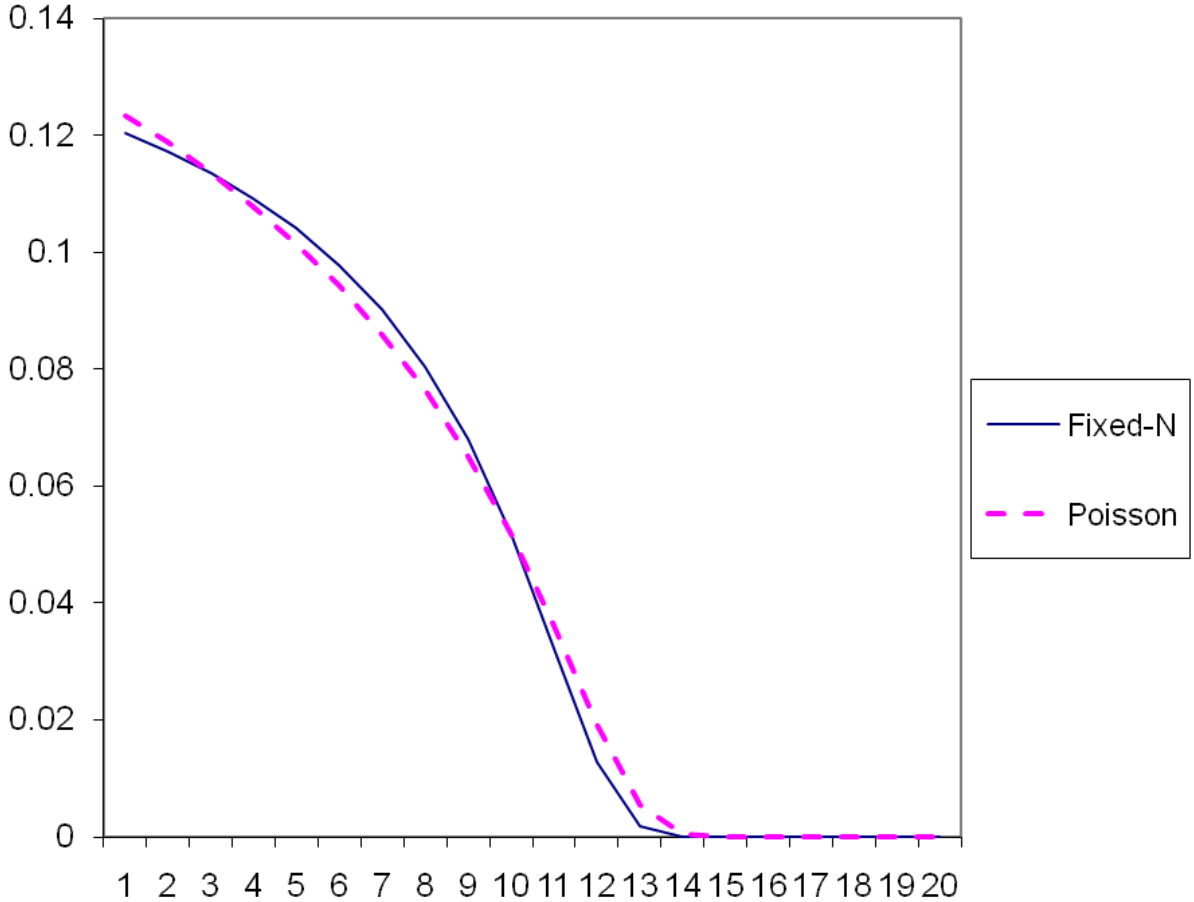


Figure A1. Probability of choosing numbers 1 to 20 in the Fixed-N Nash equilibrium and the Poisson-Nash equilibrium (*n*=27, *K*=99).

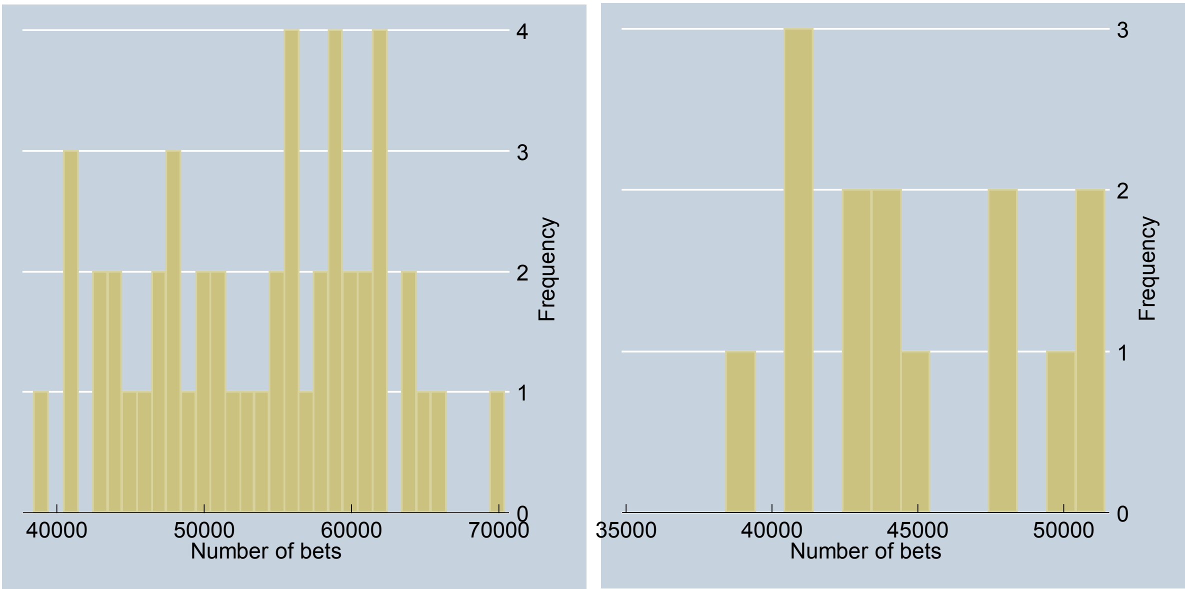


Figure A2. Total number of daily bets on all days (left) and Sundays and Mondays (right).

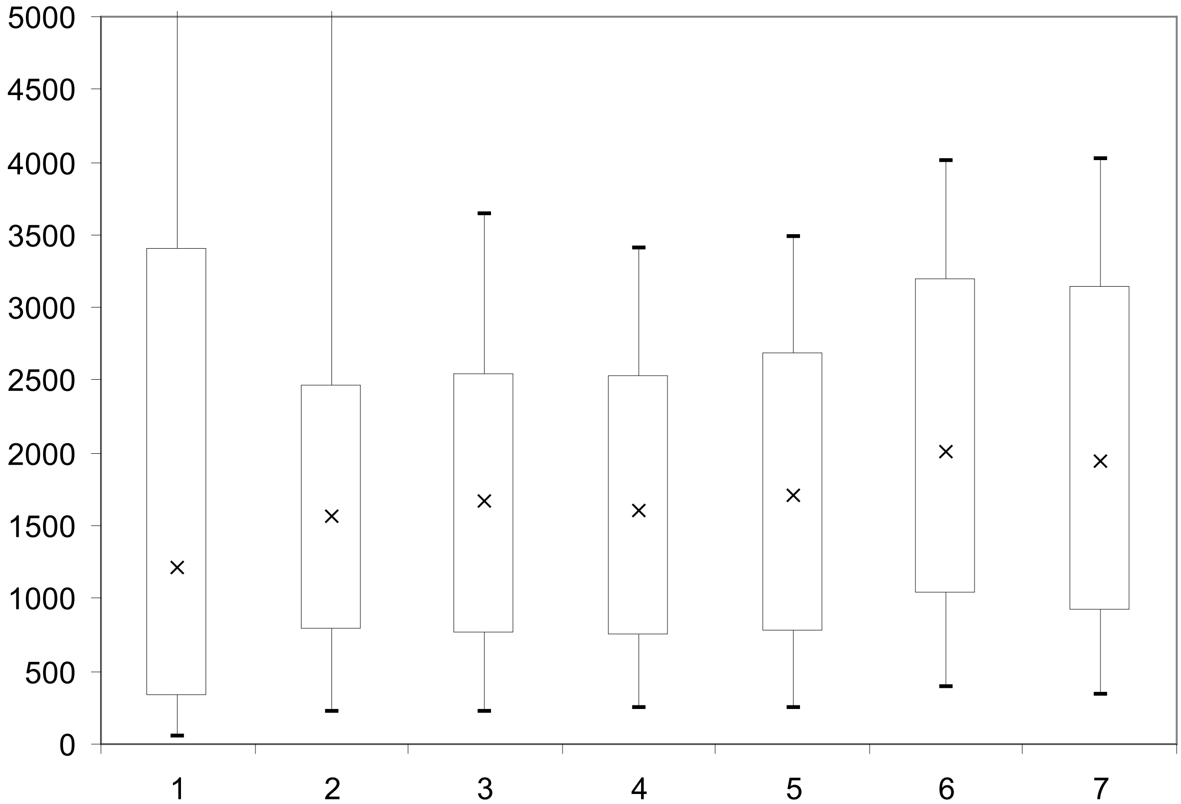
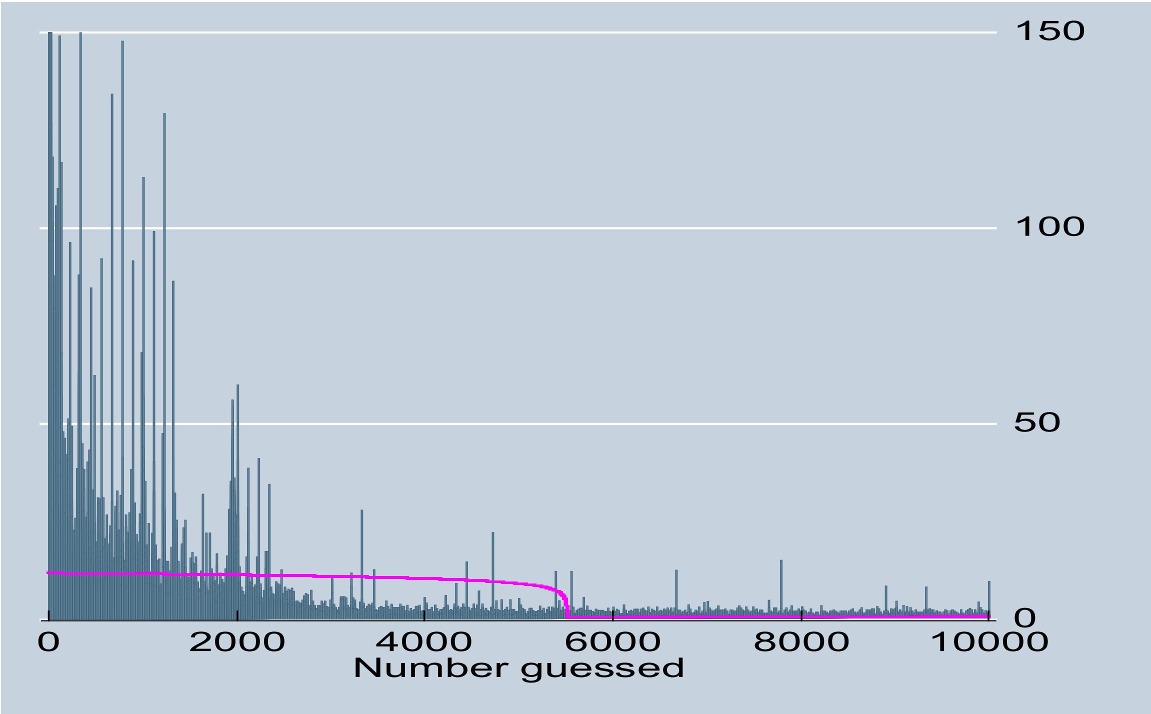
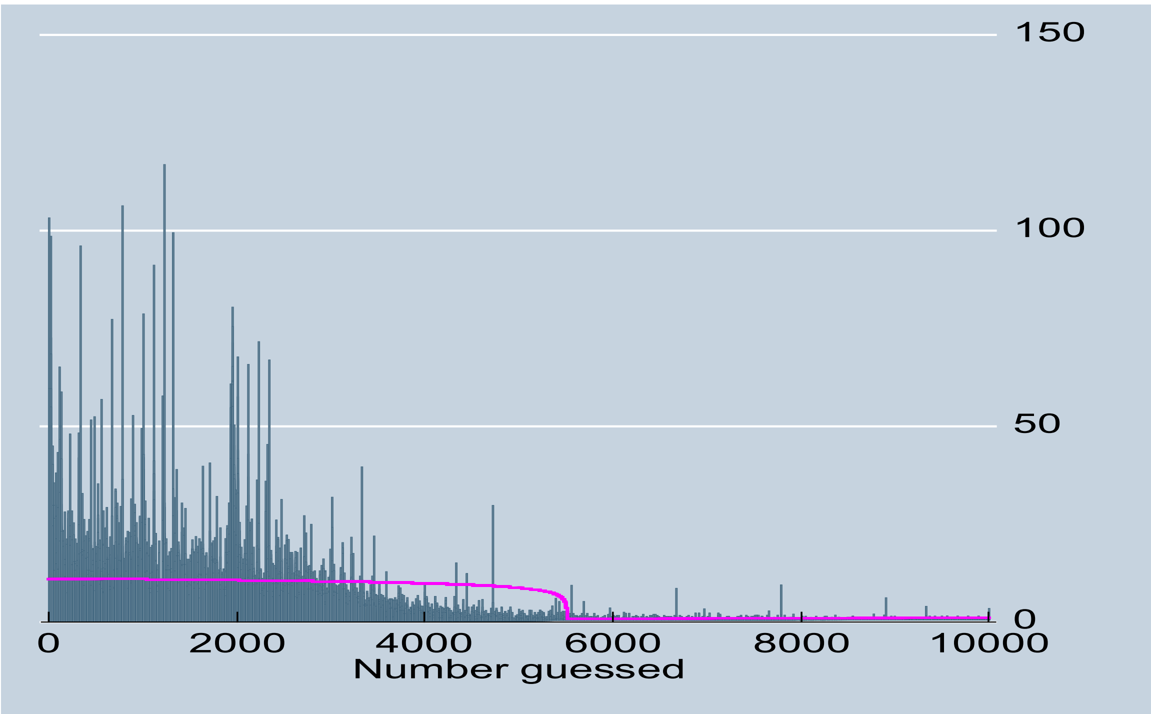


Figure A3. Weekly box plots of data (10-25-50-75-90 percentile box plots).

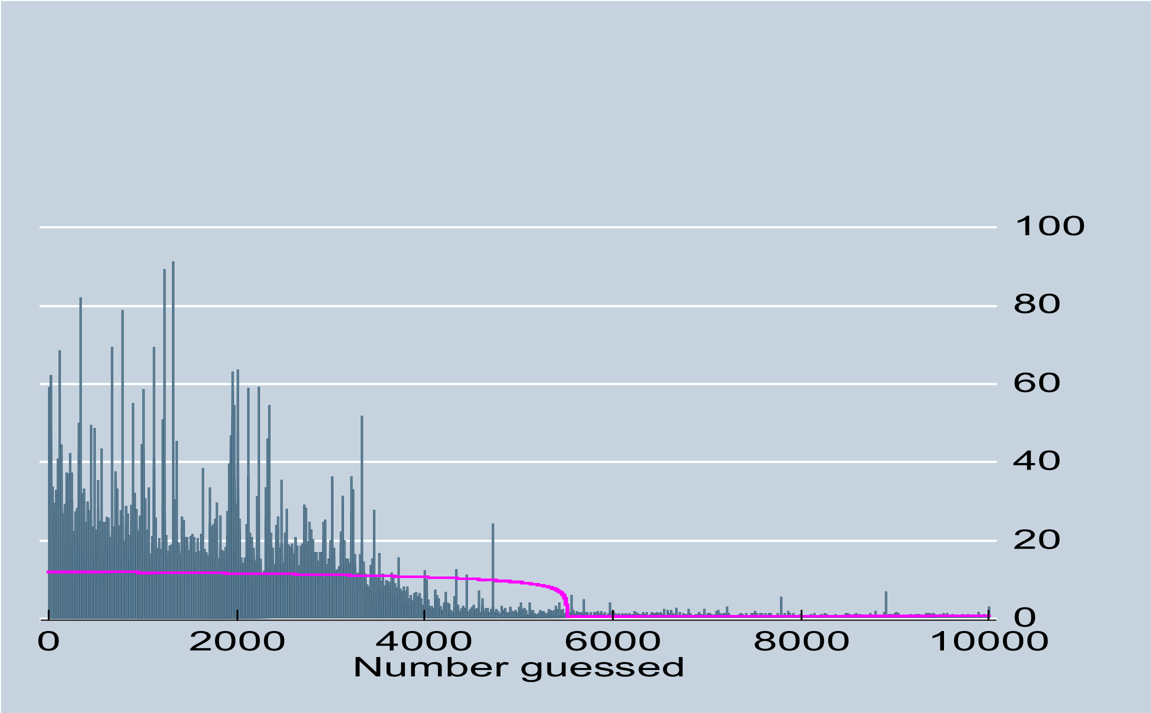
Week 1


150
100
50
0
0
2000
4000
6000
8000
10000
Number guessed

Week 3


150
100
50
0
0
2000
4000
6000
8000
10000
Number guessed

Week 5



Week 7

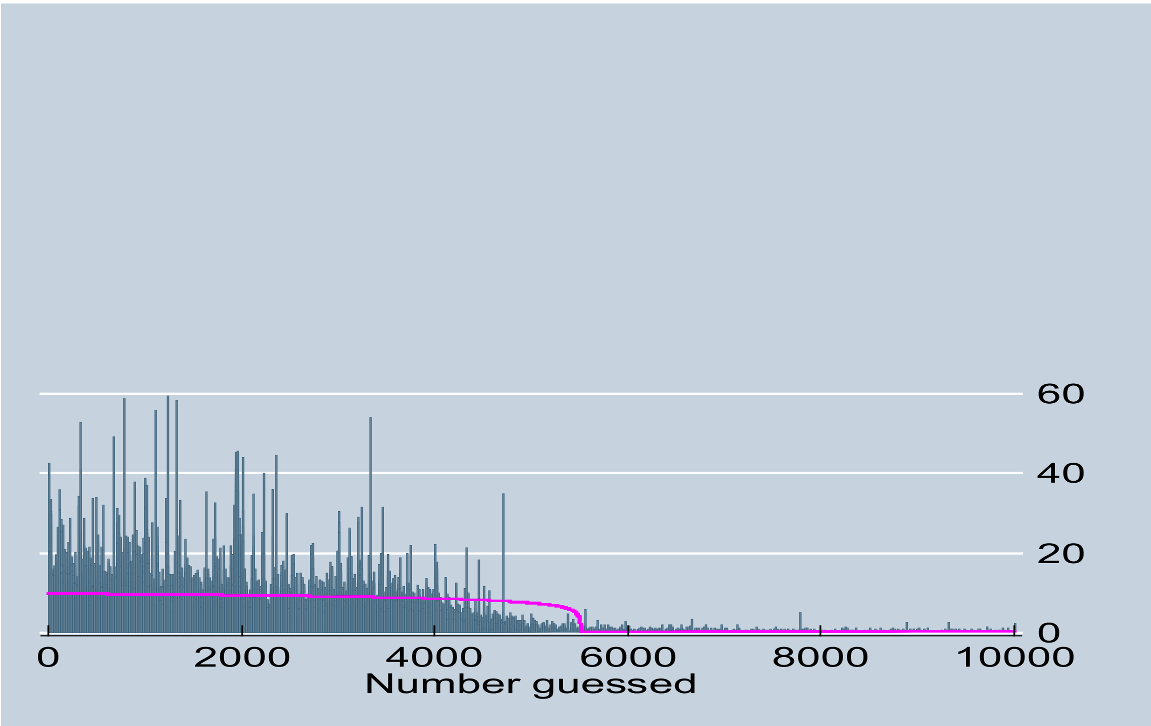


Figure A4. Average daily frequencies and equilibrium prediction for week 1, 3, 5, 7 in the field.

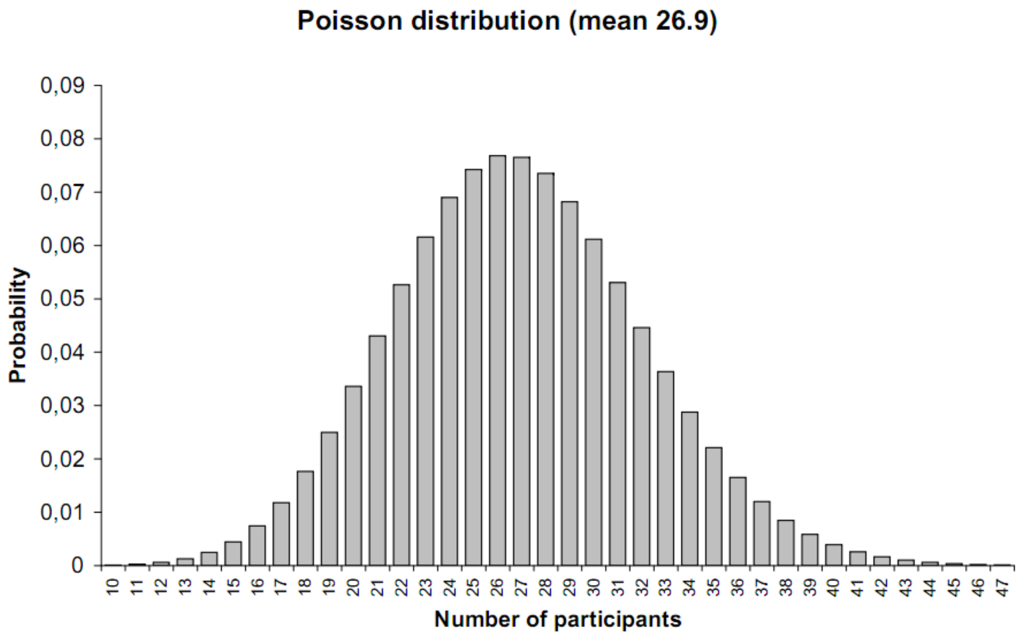


Figure A5. Poisson distribution pdf shown in the instructions of the fourth (Poisson) session.

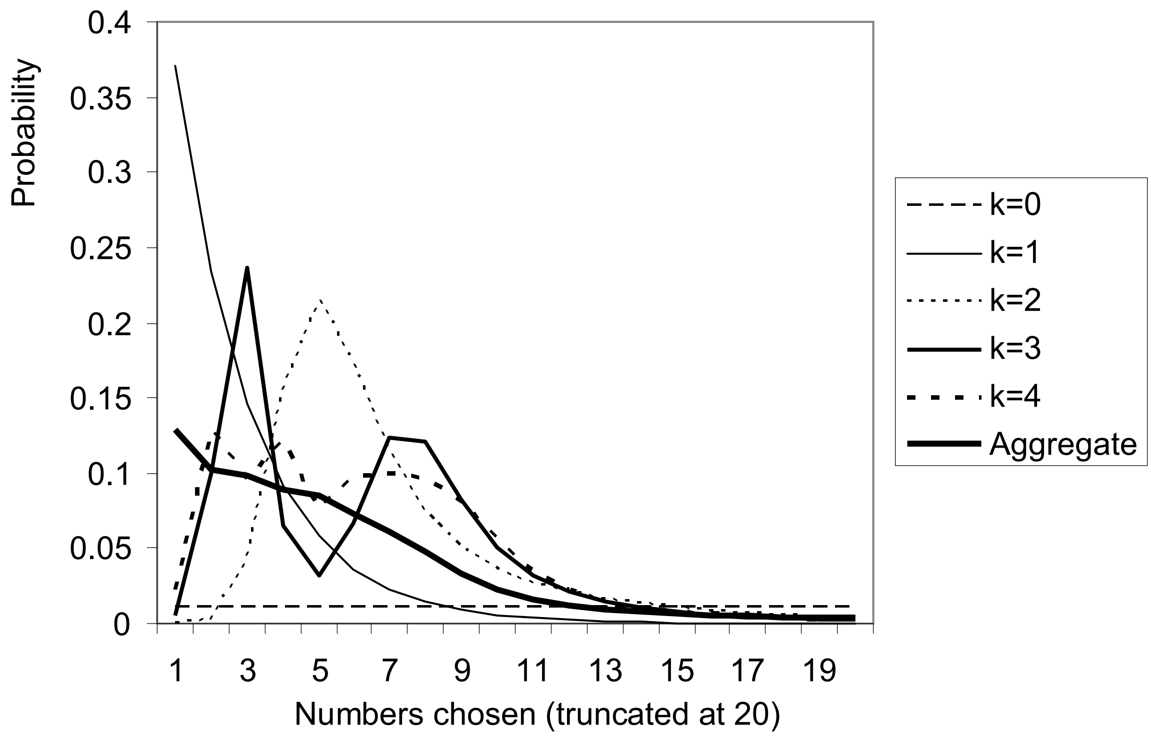


Figure A6. Probability of choosing numbers 1 to 20 in cognitive hierarchy model (*n*=26.9, *K*=99, *τ*=1.5, *λ*=2).

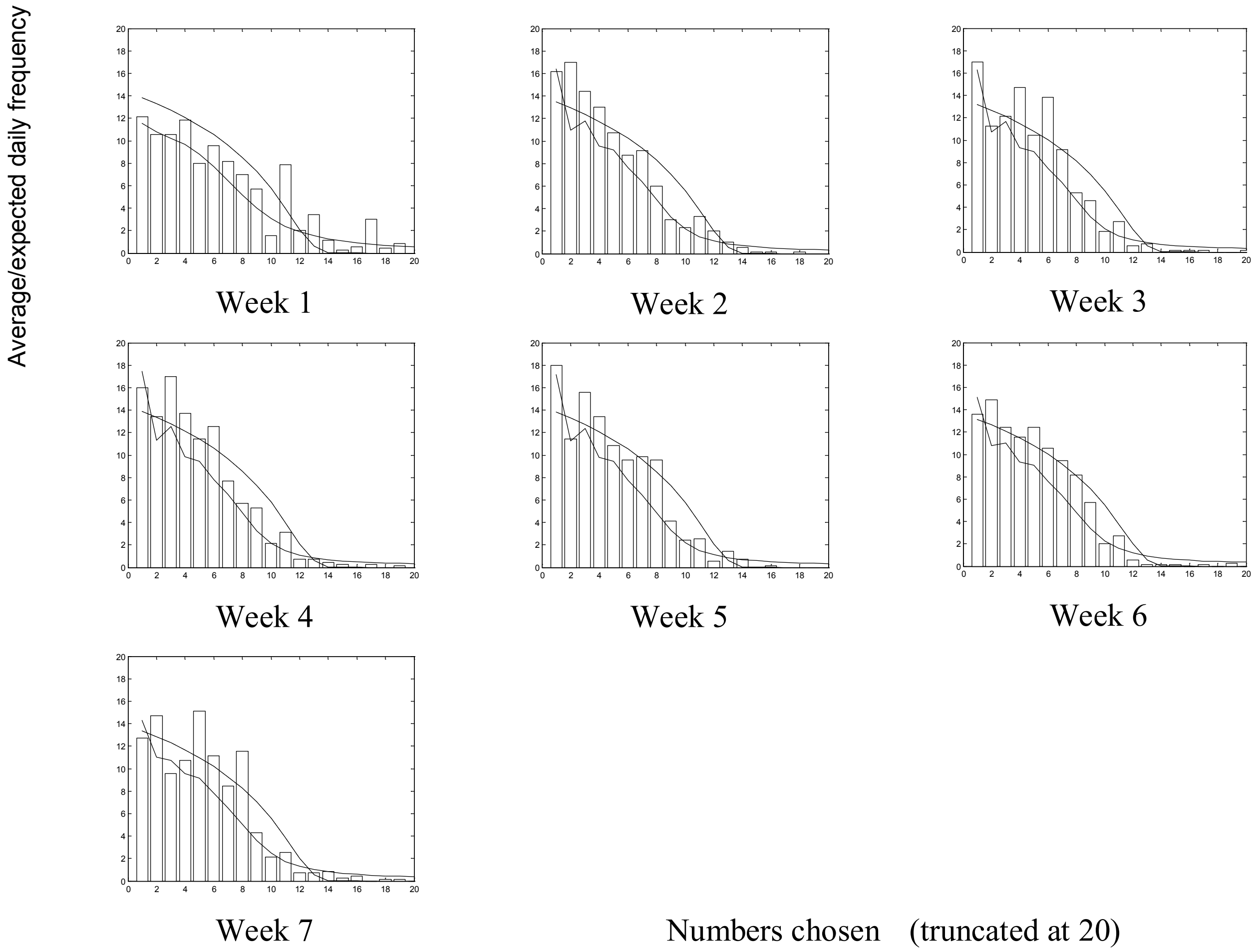


Figure A7. Average daily frequencies in the laboratory, Poisson-Nash equilibrium prediction (dashed lines) and estimated cognitive hierarchy (solid lines) when tau = 1.5 (line), week 1 to 7.

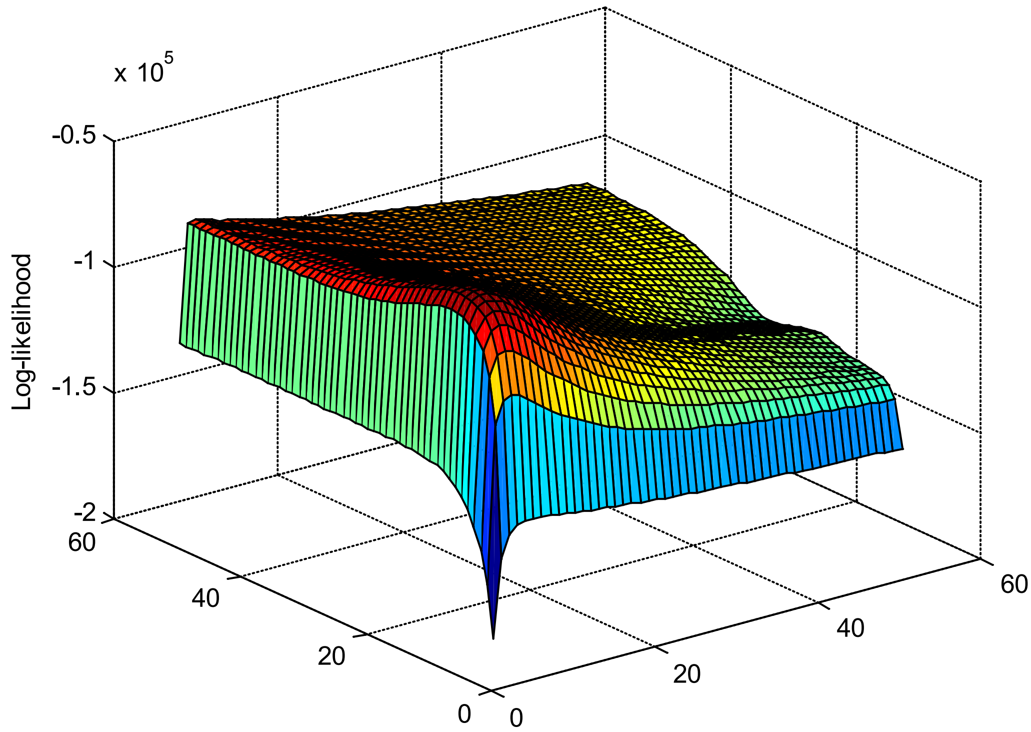


Figure A8. Log-likelihood for cognitive hierarchy in the field (first week).

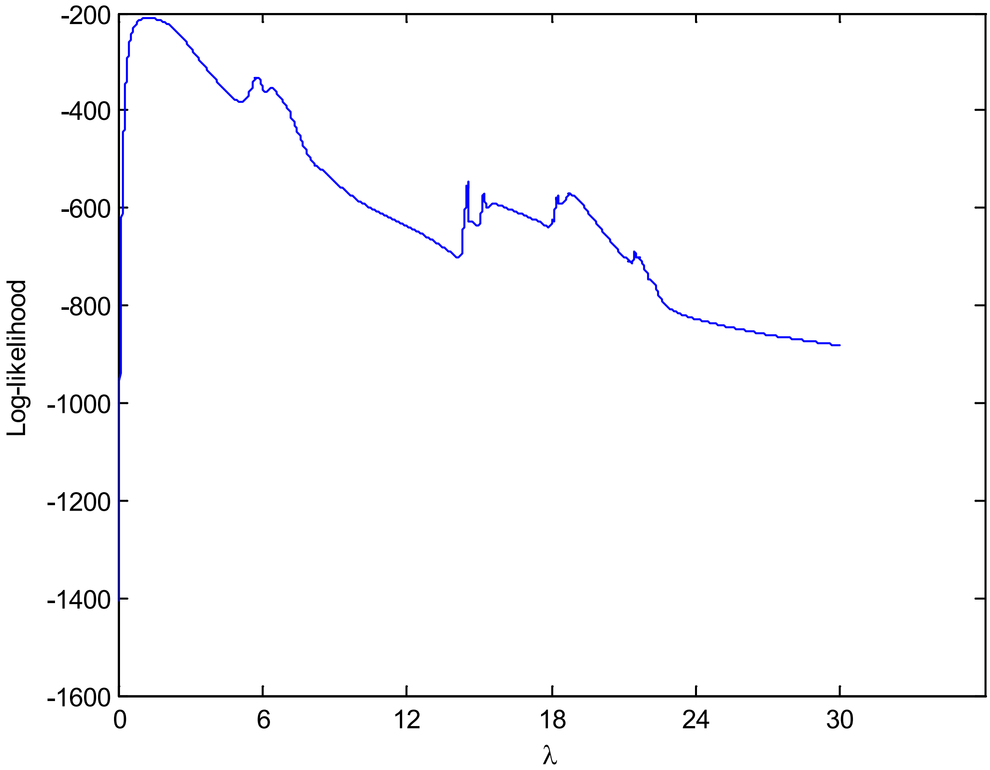


Figure A9. Log-likelihood for cognitive hierarchy in the laboratory (first week).

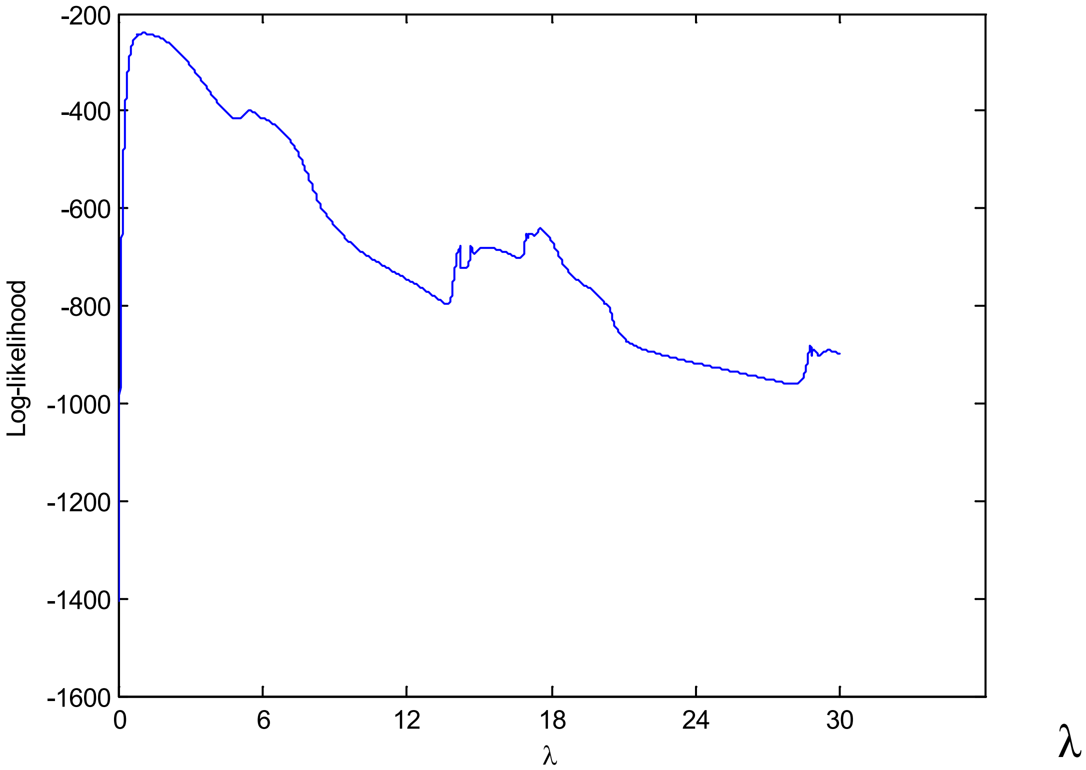
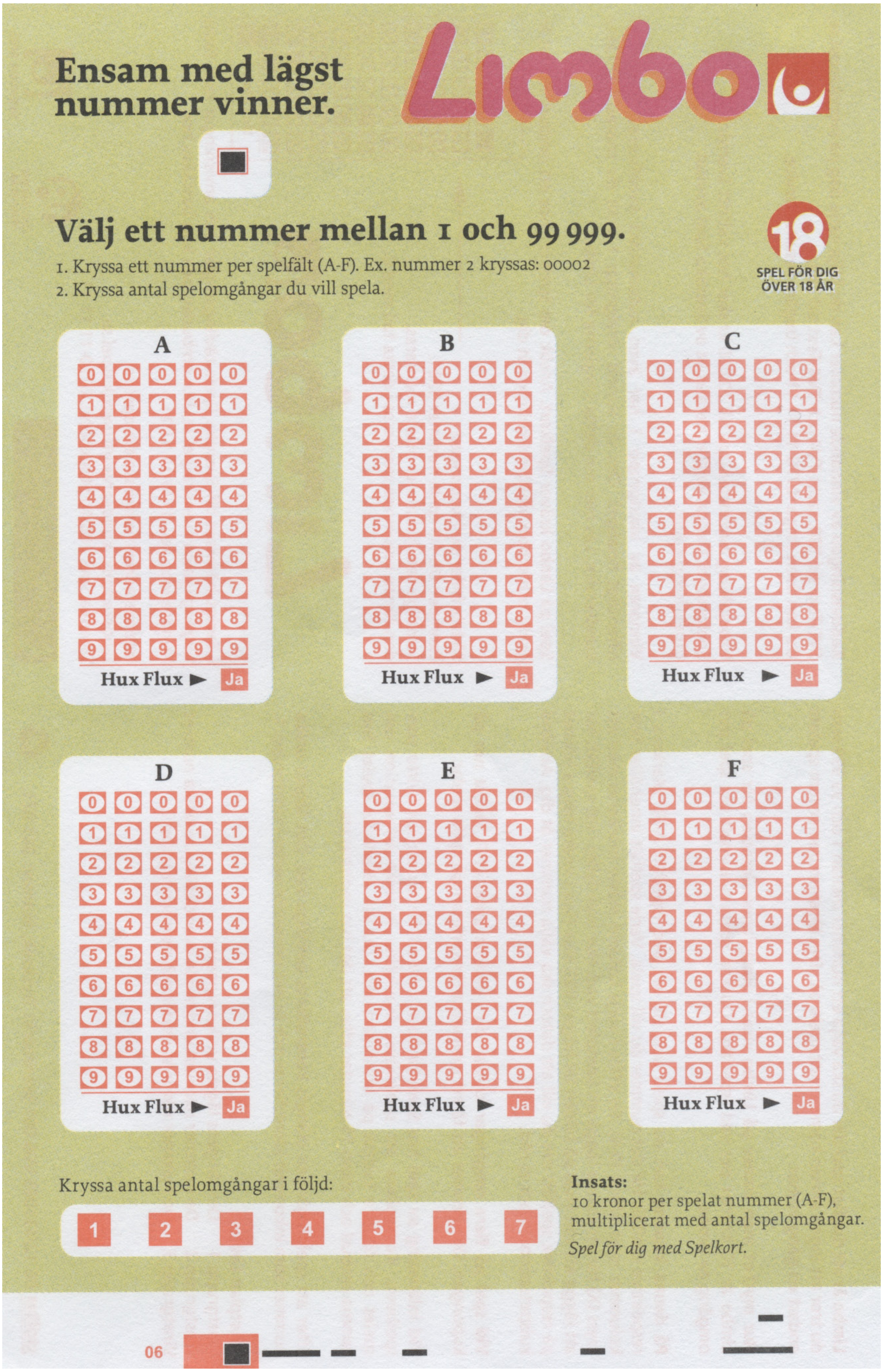


Figure A10. Log-likelihood function for cognitive hierarchy in the laboratory (first week, τ=1.5).


Ensam med lägst nummer vinner.

Limbo

Välj ett nummer mellan 1 och 99 999.

1. Kryssa ett nummer per spelfält (A-F). Ex. nummer 2 kryssas: 00002
2. Kryssa antal spelomgångar du vill spela.

SPEL FÖR DIG ÖVER 18 ÅR

A — Hux Flux ► Ja

B — Hux Flux ► Ja

C — Hux Flux ► Ja

D — Hux Flux ► Ja

E — Hux Flux ► Ja

F — Hux Flux ► Ja

Kryssa antal spelomgångar i följd:

1 2 3 4 5 6 7

Insats:
10 kronor per spelat nummer (A-F), multiplicerat med antal spelomgångar.

Spel för dig med Spelkort.

06

Figure A11. The paper entry form for the Swedish LUPI (Limbo) game.

Figure A12. Online entry interface for the Swedish LUPI (Limbo) game.

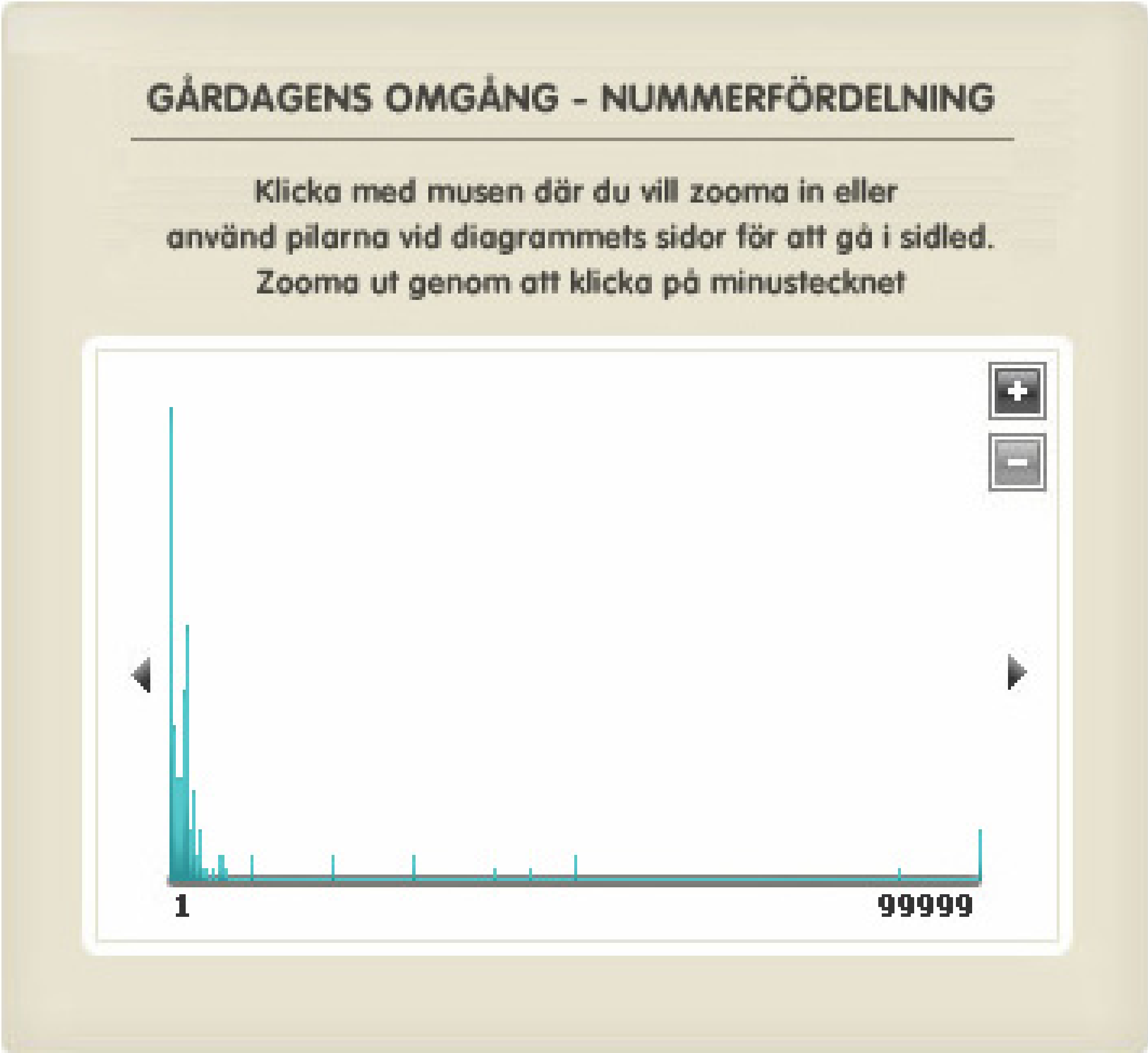


Figure A13. Histogram of yesterday's bets as shown online.

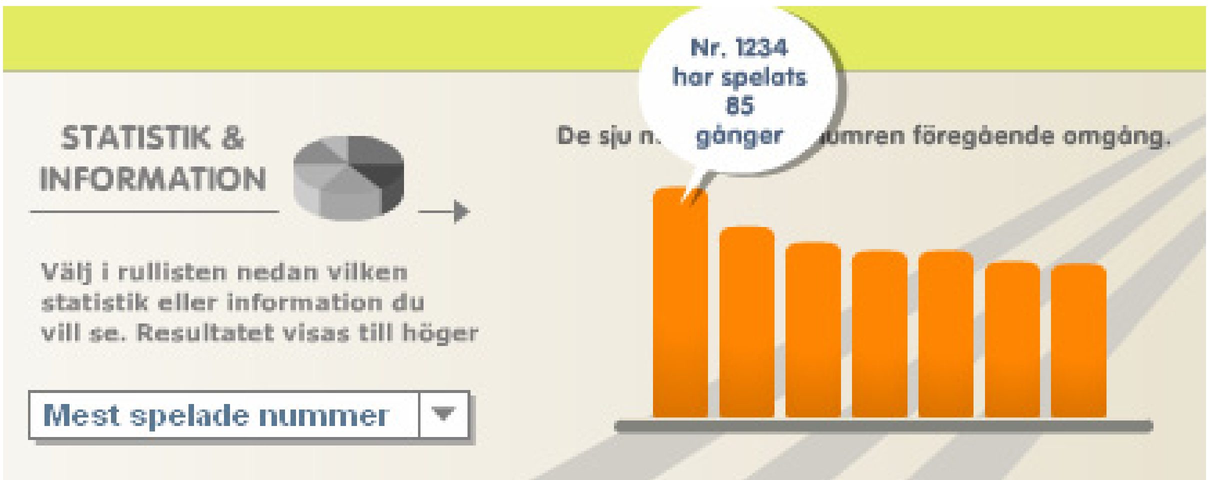


Figure A14. Most popular numbers yesterday as shown online.

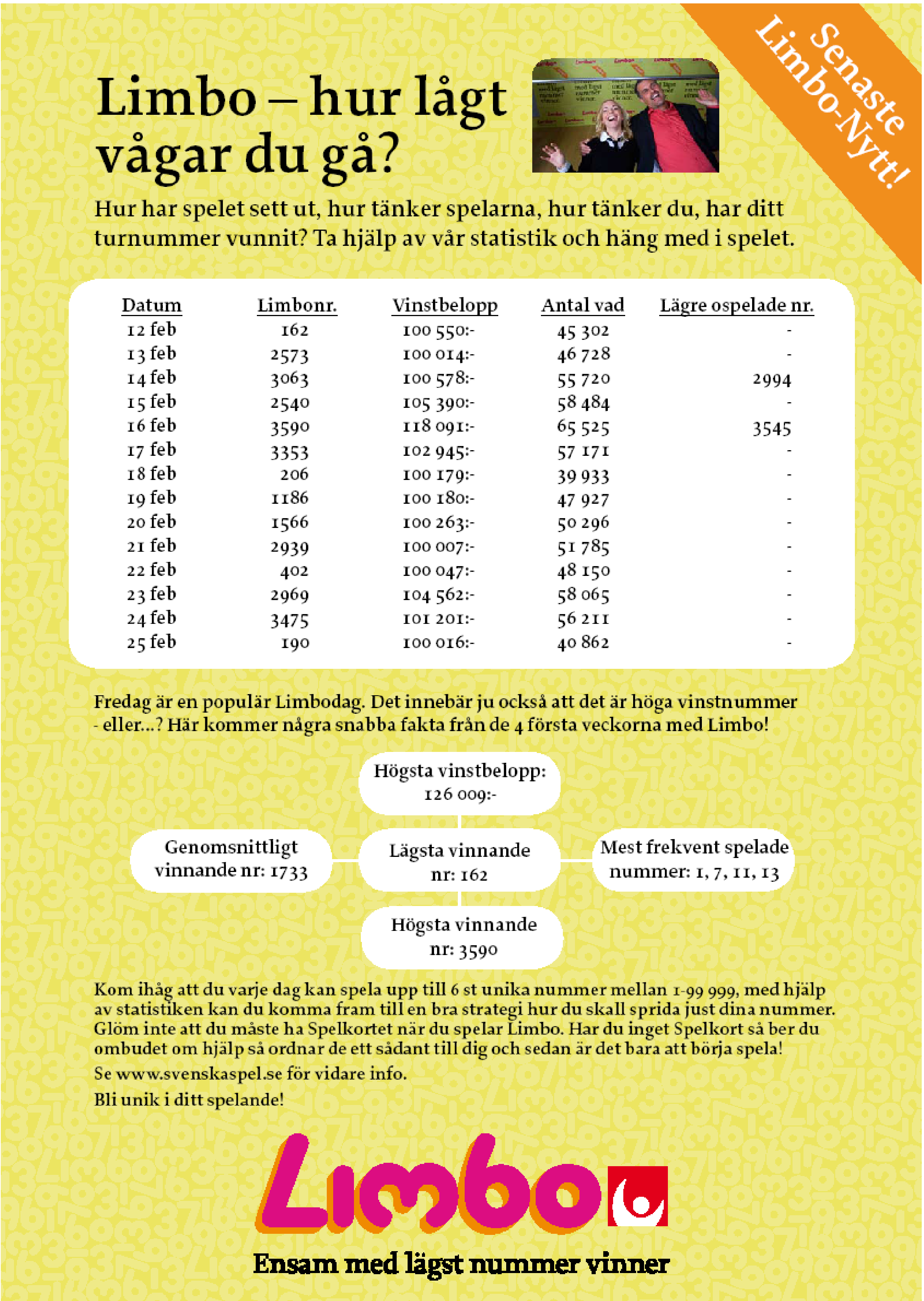


Figure A15. Example of Limbo poster.

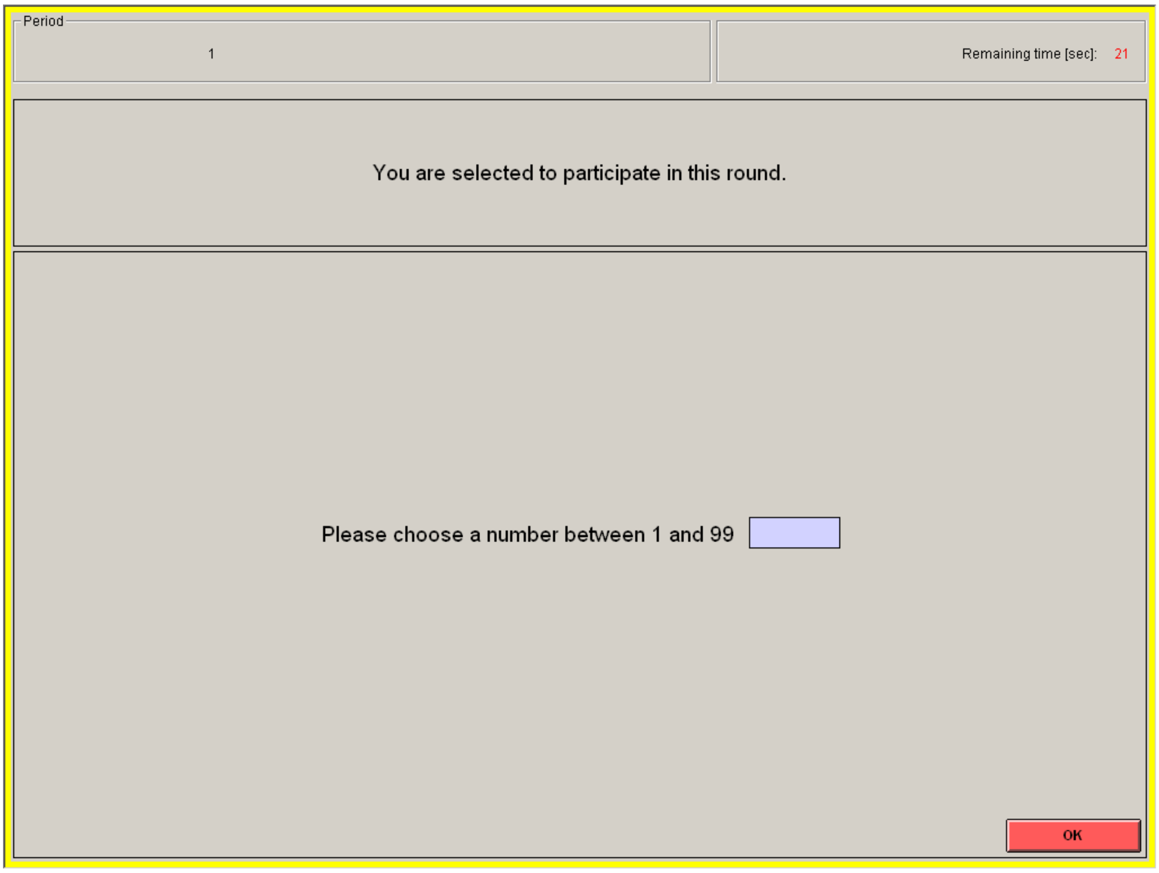


Figure A16. Screenshot of input screen in the laboratory experiment.

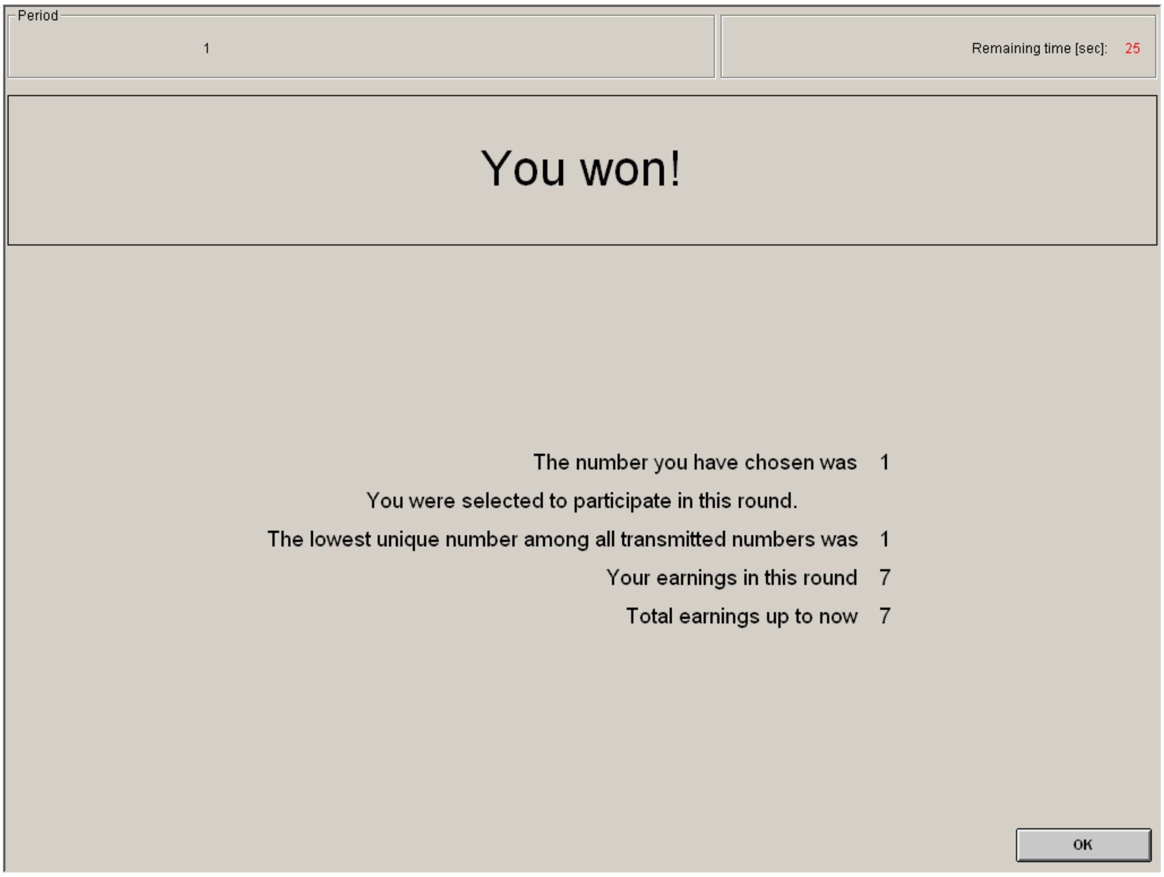


Figure A17. Screenshot of result screen in the laboratory experiment.

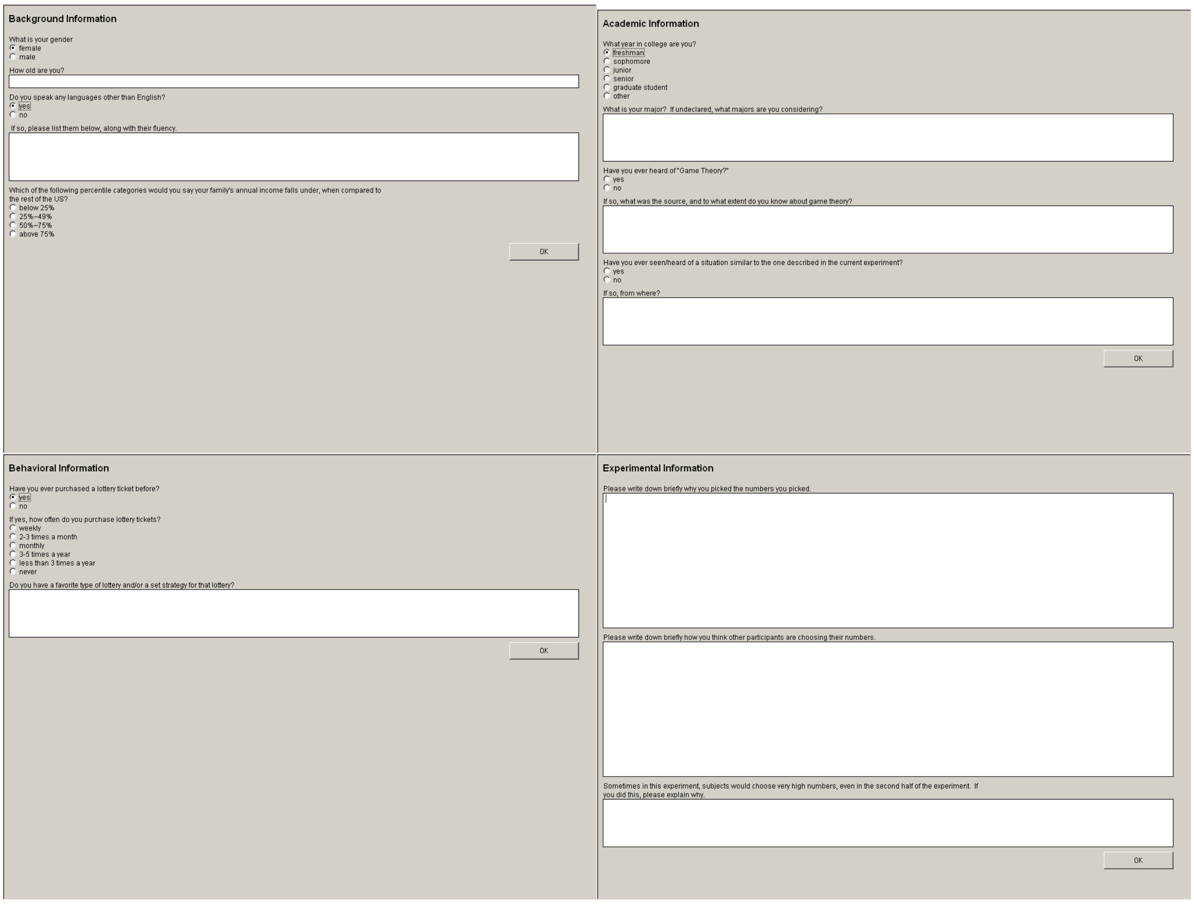


Figure A18. Screenshots of questionnaire in the laboratory experiment.

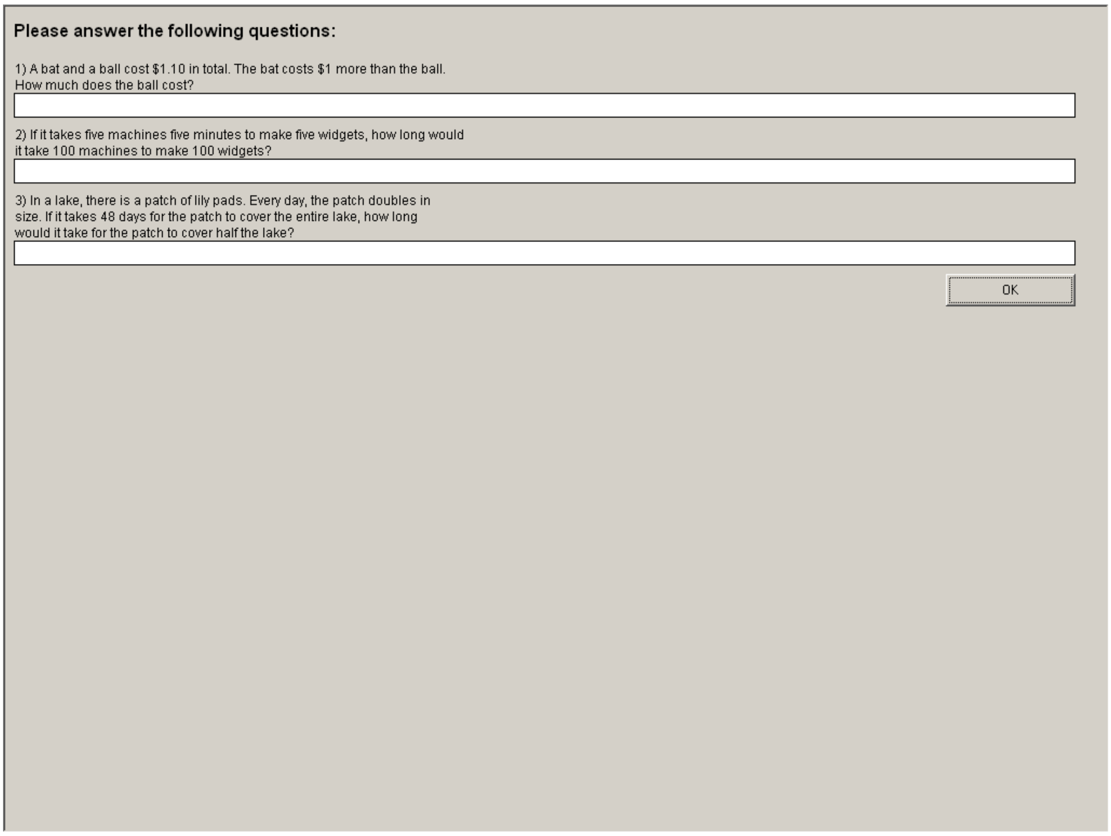


Figure A19. Screenshot of CRT in the laboratory experiment.

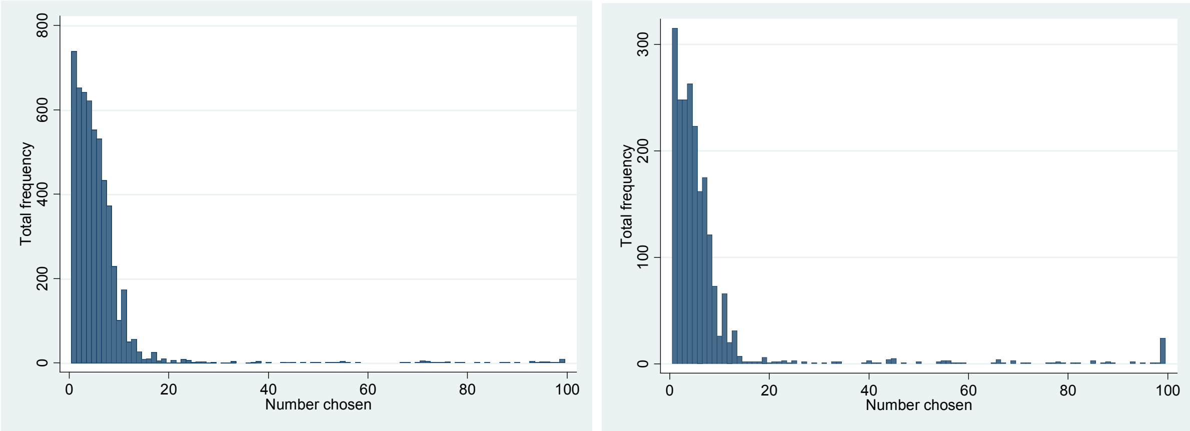


Figure A20. Laboratory total frequencies, selected (left) vs non-selected (right) subjects.

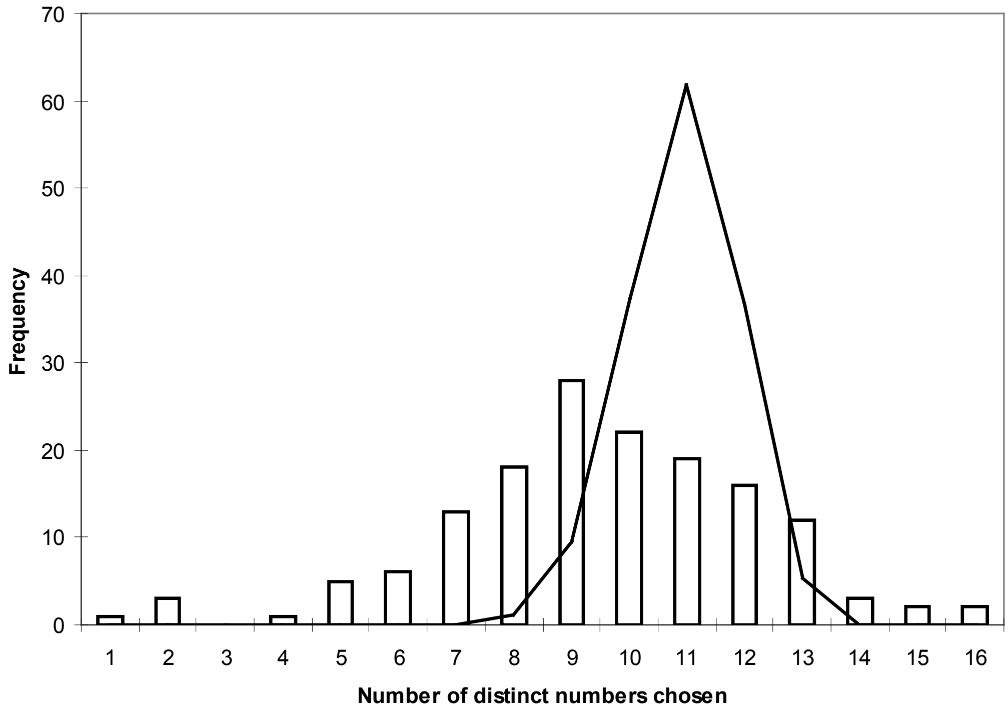


Figure A21. Histogram of the number of distinct numbers chosen by subjects (selected subjects' choices from all sessions, one subject choosing 27 distinct numbers excluded) and the corresponding simulated number of distinct numbers if subjects were playing the Poisson-Nash equilibrium.