Testing Game Theory in the Field:
Swedish LUPI Lottery Games[†]

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

Game theory is usually difficult to test in the field because predictions typically depend sensitively on features that are not controlled or observed. We conduct one such test using both laboratory and field data from the Swedish lowest unique positive integer (LUPI) game. In this game, players pick positive integers and whoever chooses the lowest unique number wins. Equilibrium predictions are derived assuming Poisson distributed population uncertainty. The field and lab data show similar patterns. Despite various deviations from equilibrium, there is a surprising degree of convergence toward equilibrium. Some deviations can be rationalized by a cognitive hierarchy model. (JEL C70, C93, D44, H27)

G ame theory predictions are challenging to test with field data because those predictions are usually sensitive to details about strategies, information and payoffs which are difficult to observe in the field. As Robert Aumann pointed out: "In applications, when you want to do something on the strategic level, you must have very precise rules [...] An auction is a beautiful example of this, but it is very special. It rarely happens that you have rules like that (Eric van Damme 1998)."

In this paper we exploit such a happening, using field data from a Swedish lottery game. In this lottery, players simultaneously choose positive integers from 1 to K. The winner is the player who chooses the lowest number that nobody else picked. We call this the LUPI game, because the *l*owest *u*nique *p*ositive *i*nteger wins.[1] Because strategies and payoffs are known, the field setting is unusually well-structured

1

compared to other strategic field data on contracting, pricing, entry, information disclosure, or auctions. The price one pays for clear structure is that the game does not very closely resemble any other familiar economic game. Gaining structure at the expense of generality is similar to the tradeoff faced in using data from game shows and sports to understand general strategic principles.

This paper analyzes LUPI theoretically and reports data from the Swedish field lottery and from parallel lab experiments. The paper has several theoretical and empirical parts. The parts have a coherent narrative flow because each part raises some new question which is answered by the next part. The overarching question, which is central to all empirical game theory, is this one: What strategic models best explain behavior in games?

The first specific question is *Q1: What does an equilibrium model of behavior predict in these games?* To answer this question, we first note that subjects do not know exactly how many other players are participating in the game and that the actual number of players varies from day to day. We therefore approximate the equilibrium by applying the theory of Poisson games.[2] In Poisson games, the number of players is Poisson-distributed (Roger B. Myerson 1998).[3] Remarkably, assuming a variable number of players rather than a fixed number makes computation of equilibrium *simpler* if the number of players is Poisson-distributed.

The number of players in the Swedish LUPI game actually varies too much from day-to-day to match the cross-day variance implicit in the Poisson assumption. However, the Poisson-Nash equilibrium is (probably) the only computable equilibrium benchmark. Field tests of theory always violate some of the assumptions of the theory, to some degree; it is an empirical question whether the equilibrium benchmark fits reasonably well despite resting on incorrect assumptions. (We revisit this important issue in the conclusion after all the data are presented.)

After deriving the Poisson equilibrium in order to answer Q1, we compare the Poisson equilibrium to the field data. In our view, the equilibrium is surprisingly close (given its complexity and counterintuitive properties). However, there are clearly large deviations from the equilibrium prediction and some behaviorally interesting fine-grained deviations. These empirical results raises question *Q2: Can non-equilibrium behavioral models explain the deviations when the game is first played?*

The simple LUPI structure allows us to provide tentative answers to Q2 by comparing Poisson-Nash equilibrium predictions with predictions of a particular parametric model of boundedly rational play: the level-*k* or cognitive hierarchy (CH) approach. CH predicts too many low-number choices (compared to the Poisson-Nash), capturing some deviation of the field data.

Because the LUPI game is simple, it is easy to go a step further and run a lab experiment that matches many of the key features of the game played in the field.

[2] As Milton Friedman (1953) famously noted, theories with false assumptions could often predict well (and, in economics, often do).

[3] This also distinguishes our paper from contemporaneous research on unique bid auctions by Jürgen Eichberger and Dmitri Vinogradov (2008); Andrea Gallice (2009); Yaron Raviv and Gabor Virag (2009); Amnon Rapoport et al. (2009), and Harold Houba, Dinard van der Laan, and Dirk Veldhuizen (forthcoming), which all assume that the number of players is fixed and commonly known.

The lab data enable us to address one more question: *Q3: How well does behavior in a lab experiment designed to closely match features of a field environment parallel behavior in the field?* Q3 is important because of an ongoing debate about lab-field parallelism in economics, rekindled with some skepticism by Steven D. Levitt and John A. List (2007) (see Armin Falk and James J. Heckman 2009 and Camerer forthcoming for replies). We conclude that the basic empirical features of the lab and field behavior are comparable. This close match adds to a small amount of evidence of how well experimental lab data can generalize to a particular field setting when the experiment was specifically intended to do so.

The ability to track decisions by each player in the lab also enables us to answer some minor questions that cannot be answered by field data. For example, it appears that players tend to play recent winning numbers more, sociodemographic variables do not correlate strongly with performance, and there are not strong identifiable differences in skill across players (measured by winning frequency).

Before proceeding, we must mention an important caveat. LUPI was not designed by the lottery creators to be an exact model of a particular economic game. However, it combines some strategic features of interesting naturally occurring games. For example, in games with congestion, a player's payoffs are lower if others choose the same strategy. Examples include choices of traffic routes and research topics, or buyers and sellers choosing among multiple markets. LUPI has the property of an extreme congestion game, in which having even one other player choose the same number reduces one's payoff to zero.[4] Indeed, LUPI is similar to a game in which being first matters (e.g., in a patent race), but if players are tied for first they do not win. One close market analogue to LUPI is the lowest unique bid auction (LUBA; see Eichberger and Vinogradov 2008; Gallice 2009; Raviv and Virag 2009; Rapoport et al. 2009; and Houba, van der Laan, and Veldhuizen forthcoming). In these auctions, an object is sold to the lowest bidder whose bid is unique (or in some versions, to the highest unique bidder). LUPI is simpler than LUBA because winners do not have to pay the amount they bid, and there are no private valuations and beliefs about valuations of others. However, LUPI contains the same essential strategic conflict: between wanting to choose low numbers and wanting to choose unique numbers.

Finally, the scientific value of LUPI games is like the scientific value of data from game shows and professional sports, such as Deal or No Deal (e.g., Steffen Andersen et al. 2008, and Thierry Post et al. 2008). Like the LUPI lottery, game shows and sports are not designed to be replicas of typical economic decisions. Nonetheless, game shows and sports are widely studied in economics because they provide very clear field data about actual economic choices (often for high stakes), and they have simple structures that can be analyzed theoretically. The same is true for LUPI.

The next section provides a theoretical analysis of a simple form of the LUPI game, the Poisson-Nash equilibrium. Section II reports the basic field data and

---

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

compare them to the Poisson-Nash approximate benchmark. It also introduces the cognitive hierarchy model and asks whether it can explain the field data. Section III describes the lab replication. Section IV concludes the paper. Supporting material is available in a separate online Appendix.

I. Theory

In the simplest form of LUPI, the number of players, N, has a known distribution, the players choose integers from 1 to K simultaneously, and the lowest unique number wins. The winner earns a payoff of 1, while all others earn 0.[5]

We first analyze the game when players are assumed to be fully rational, best responding, and have equilibrium beliefs. We assume that the number of players N is a random variable that has a Poisson distribution.[6] The Poisson assumption proves to be easier to work with than a fixed N, both theoretically and computationally. The actual variance of N in the field data is much larger than in the Poisson distribution so the Poisson-Nash equilibrium is only a computable approximation to the correct equilibrium. Whether it is a good approximation will partly be answered by looking at how well the theory fits the field data.[7] In addition, we implement the Poisson distribution of N exactly in lab experiments.

A. *Properties of Poisson Games*

In this section, we briefly summarize the theory of Poisson games developed by Myerson (1998, 2000). The theory is then used in Section IB to characterize the Poisson-Nash equilibrium in the LUPI game.

Games with population uncertainty relax the assumption that the exact number of players is common knowledge. In particular, in a *Poisson game* the number of players N is a random variable that follows a Poisson distribution with mean n. We have

$$N \sim \text{Poisson}(n): \quad N = k \text{ with probability } \frac{e^{-n}n^k}{k!}$$

and, in the case of a Bayesian game (or the cognitive hierarchy model developed below), players' types are independently determined according to the probability distribution $r = (r(t))_{t \in T}$ on some type space T. Let a type profile be a vector of non-negative integers listing the number of players of each type t in T, and let $Z(T)$ be the set of all such type profiles in the game. Combining N and r we can describe

---

[5] In this stylized case, we assume that if there is no lowest unique number then there is no winner. This simplifies the analysis because it means that only the probability of being unique must be computed. In the Swedish game, if there is no unique number then the players who picked the smallest and least-frequently-chosen number share the top prize.

[6] Players did not know the number of total bets in both the field and lab versions of the LUPI game. Although players in the field could get information about the current number of bets that had been made so far during the day, players had to place their bets before the game closed for the day and therefore could not know with certainty the total number of players that would participate in that day.

[7] For small N, we show in online Appendix A that the equilibrium probabilities for fixed-N Nash and Poisson-Nash equilibrium are practically indistinguishable (Figure A1).

the *population uncertainty* with the distribution $y \sim Q(y)$ where $y \in Z(T)$ and $y(t)$ is the number of players of type $t \in T$.

Players have a common finite action space C with at least two alternatives, which generates an *action profile* $Z(C)$ containing the number of players that choose each action. Utility is a bounded function $U: Z(C) \times C \times T \rightarrow \mathbb{R}$, where $U(x, b, t)$ is the payoff of a player with type t, choosing action b, and facing an opponent action profile of x. Let $x(c)$ denote the number of other players playing action $c \in C$.

Myerson (1998) shows that the Poisson distribution has two important properties that are relevant for Poisson games and simplify computations dramatically. The first is the *decomposition property*, which in the case of Poisson games imply that the distribution of type profiles for any $y \in Z(T)$ is given by

$$Q(y) = \prod_{t \in T} \frac{e^{-nr(t)}\big(nr(t)\big)^{y(t)}}{y(t)!}.$$

Hence, $\tilde{Y}(t)$, the random number of players of type $t \in T$, is Poisson with mean $nr(t)$, and is independent of $\tilde{Y}(t')$ for any other $t' \in T$. Moreover, suppose each player independently plays the mixed strategy σ, choosing action $c \in C$ with probability $\sigma(c \mid t)$ given his type t. Then, by the decomposition property, the number of players of type t that chooses action c, $\overline{Y}(c, t)$, is Poisson with mean $nr(t)\sigma(c \mid t)$ and is independent of $\overline{Y}(c', t')$ for any other c', t'.

The second property of Poisson distributions is the *aggregation property*, which states that any sum of independent Poisson random variables is Poisson distributed. This property implies that the number of players (across all types) who choose action c, $\tilde{X}(c)$, is Poisson with mean $\sum_{t \in T} nr(t)\sigma(c \mid t)$, independent of $\tilde{X}(c')$ for any other $c' \in C$. We refer to this property of Poisson games as the *independent actions* (IA) property.

Myerson (1998) also shows that the Poisson game has another useful property: *environmental equivalence* (EE). Environmental equivalence means that conditional on being in the game, a type t player would perceive the population uncertainty as an outsider would, i.e., $Q(y)$. If the strategy and type spaces are finite, Poisson games are the only games with population uncertainty that satisfy both IA and EE (Myerson 1998).

EE is a surprising property. Take a Poisson LUPI game with 27 players on average. In our lab implementation, a large number of players are recruited and are told that the number of players who will be active (i.e., play for real money) in each period varies. Consider a player who is told she is active. On the one hand, she might then act as if she is playing against the number of opponent players that is Poisson-distributed with a mean of 26 (since her active status has lowered the mean of the number of remaining players). On the other hand, the fact that she is active is a clue that the number of players in that period is large, not small. If N is Poisson-distributed the two effects *exactly* cancel out so all active players in all periods act as if they face a Poisson-distributed number of opponents. EE, combined with IA, makes the analysis rather simple.

An *equilibrium* for the Poisson game is defined as a strategy function σ such that every type assigns positive probability only to actions that maximize the expected utility for players of this type; that is, for every action $c \in C$ and every type $t \in T$,

$$\text{if } \sigma(c \mid t) > 0 \text{ then } \overline{U}(c \mid t, \sigma) = \max_{b \in C} \overline{U}(b \mid t, \sigma)$$

for the expected utility

$$\overline{U}(b \mid s, \sigma) = \sum_{x \in Z(C)} \prod_{c \in C} \left(\frac{e^{-n\tau(c)}(n\tau(c))^{x(c)}}{x(c)!} \right) U(x, b, s),$$

where

$$\tau(c) = \sum_{t \in T} r(t) \sigma(c \mid t)$$

is the marginal probability that a random sampled player will choose action c under σ. Note that this equilibrium is by definition symmetric; asymmetric equilibria where players of the same type could play differently are not defined in games with population uncertainty since ex ante we do not know the list of participating players.

Myerson (1998) proves existence of equilibrium under all games of population uncertainty with finite action and type spaces, which includes Poisson games.[8] This existence result provides the basis for the following characterization of the Poisson-Nash equilibrium.

B. *Poisson Equilibrium for the LUPI Game*

In the (symmetric) Poisson equilibrium, all players employ the same mixed strategy $\mathbf{p} = (p_1, p_2, \ldots, p_K)$ where $\sum_{i=1}^{K} p_i = 1$. Let the random variable $X(k)$ be the number of players who pick k in equilibrium. Then, $\Pr(X(k) = i)$ is the probability that the number of players who pick k in equilibrium is exactly i. By environmental equivalence (EE), $\Pr(X(k) = i)$ is also the probability that i opponents pick k. Hence, the expected payoffs for choosing different numbers are:

$$\pi(1) = \Pr(X(1) = 0) = e^{-np_1}$$

$$\pi(2) = \Pr(X(1) \neq 1) \cdot \Pr(X(2) = 0) = (1 - np_1 e^{-np_1}) \cdot e^{-np_2}$$

$$\pi(3) = \Pr(X(1) \neq 1) \cdot \Pr(X(2) \neq 1) \cdot \Pr(X(3) = 0)$$

$$\vdots$$

$$\pi(k) = \left(\prod_{i=1}^{k-1} \Pr(X(i) \neq 1) \right) \cdot \Pr(X(k) = 0)$$

$$= \left(\prod_{i=1}^{k-1} [1 - np_i e^{-np_i}] \right) \cdot e^{-np_k}$$

---

[8] For infinite types, Myerson (2000) proves existence of equilibrium for Poisson games alone.

for all $k > 1$. If both k and $k + 1$ are chosen with positive probability in equilibrium, then $\pi(k) = \pi(k + 1)$. Rearranging this equilibrium condition implies

$$(1) \qquad\qquad e^{np_{k+1}} = e^{np_k} - np_k.$$

Alternatively, this condition can be written as

$$(2) \qquad\qquad p_k - p_{k+1} = -\frac{1}{n} \ln(1 - np_k e^{-np_k}).$$

In addition to condition 1 (or 2), the probabilities must sum up to one and the expected payoff from playing numbers not in the support of the equilibrium strategy cannot be higher than the numbers played with positive probability.

   The three equilibrium conditions allow us to characterize the equilibrium and show that it is unique.

PROPOSITION 1: *There is a unique mixed equilibrium* $\mathbf{p} = (p_1, p_2, \cdots, p_K)$ *of the Poisson LUPI game that satisfies the following properties:*

   1) *Full support:* $p_k > 0$ *for all* k.

   2) *Decreasing probabilities:* $p_{k+1} < p_k$ *for all* k.

   3) *Convexity/concavity:* $(p_{k+1} - p_{k+2}) > (p_k - p_{k+1})$ *for* $p_{k+1} > 1/n$, *and* $(p_{k+1} - p_{k+2}) < (p_k - p_{k+1})$ *for* $1/n > p_k$.

   4) *Convergence to uniform play with many players: for any fixed* K, $n \to \infty$ *implies* $p_{k+1} \to p_k$.

   5) *Probability asymptotes to zero with more numbers to guess: for any fixed* n, $K \to \infty$ *implies* $p_K \to 0$.

   The proof is given in the Appendix. The intuition for the results in Proposition 1 are as follows. For the first property, first note that if k is chosen, so is $k + 1$, since deviating from k to $k + 1$ would otherwise be profitable. Nothing matters if there is a smaller number than k uniquely chosen by an opponent, but if not, picking k wins only if nobody else chooses k, while picking $k + 1$ wins if either nobody chooses k or if more than two opponents choose k. Together with the fact that 1 has to be chosen guarantees full support. Secondly, lower numbers should be chosen more often because the LUPI rule favors lower numbers. For example, if everyone is choosing uniformly, you should pick 1. However, as more people participate in the game, this advantage disappears which implies convergence to uniform (property 4).[9] Thirdly, condition 2 shows that the difference between p_k and p_{k+1} solely

---

[9] For example, when $K = 100$ and $n = 500$, the mixture probabilities start at $p_1 = 0.0124$ and end with $p_{97} = 0.0043$, $p_{98} = 0.0038$, $p_{99} = 0.0031$, $p_{100} = 0.0023$; so the ratio of highest to lowest probabilities is about
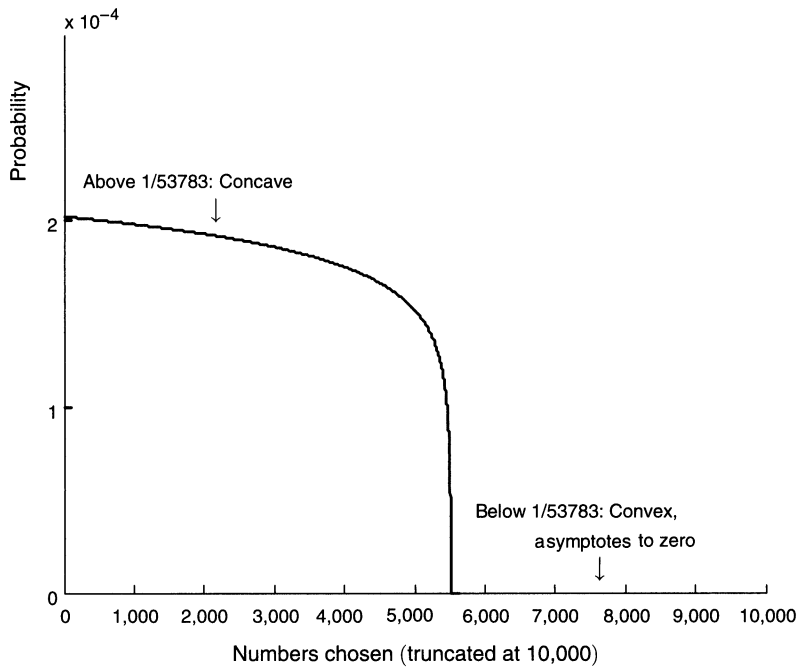
FIGURE 1. POISSON-NASH EQUILIBRIUM FOR THE FIELD LUPI GAME $(n = 53{,}783, K = 99{,}999)$

depends on the function $f(x) = xe^{-x}$ where $x = np_k$. Since $f'(x) > 0$ if $x < 1$, and $f'(x) < 0$ if $x > 1$, the critical cutoff for concavity/convexity happens at $p_k = 1/n$. Lastly, since p_K is the smallest among all probabilities, if p_K does not converge to zero as K becomes large, the probabilities will not sum up to one.

In the Swedish game the average number of players was $n = 53{,}783$ and number choices were positive integers up to $K = 99{,}999$. As Figure 1 shows, the equilibrium involves mixing with substantial probability between 1 and 5,500, starting from $p_1 = 0.0002025$. The predicted probabilities drop off very sharply at around 5,500. This is due to the concavity/convexity switch around $1/n$, which happens at $T = 5{,}513$ $(p_T = 0.00001857)$; the difference $(p_k - p_{k+1})$ increases as you move toward $1/n$ from either side. Figure 1 shows only the predicted probabilities for 1 to 10,000, since probabilities for numbers above 5,518 are positive but minuscule. Note that $n = 53{,}783 < K = 99{,}999$ implies that $K/n > 1$, and hence, the concavity/convexity switch (and the "sharp drop") has to occur at $T < K$.[10]

The central empirical question that will be answered later is how well actual behavior in the field matches the equilibrium prediction in Figure 1. Keep in mind that the simplified game analyzed in this section differs in some potentially important ways from the actual Swedish game. Computing the equilibrium is complicated and its

---

six-to-one. When $K = 100$ and $n = 5{,}000$, all mixture probabilities for numbers 1 to 100 are 0.01 (up to two-decimal precision).

[10] p_T is close to $1/n$ by the concavity/convexity switch. So, T is positively related to n. Since p_K converges to zero for large K due to Property 5 of Proposition 1, T does not depend on K as long as K is large (and "non-binding").

properties are not particularly intuitive. It might therefore be surprising if the actual data matched the equilibrium closely. Because there are 49 days of data, we can also see whether choices move in the direction of the Poisson-Nash benchmark over time.

II. The Field LUPI Game

The field version of LUPI, called Limbo, was introduced by the government-owned Swedish gambling monopoly Svenska Spel on January 29, 2007. This section describes its essential elements; additional description is in online Appendix C.

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

During the first three to four weeks, it was only possible to play the game at physical branches of Svenska Spel by filling out a form (Figure A11 in the online Appendix). The form allowed players to bet on up to six numbers,[11] to play the same numbers for up to 7 days in a row, or to let a computer choose random numbers for them (a "HuxFlux" option). During the following weeks it was also possible to play online, see online Appendix C for a description of the online interface.

Daily data were downloaded for the first seven weeks, ending on the 18th of March 2007. The game was stopped on March 24th, one day after a newspaper article claimed that some players had colluded in the game, but it is unclear whether collusion actually occurred.

Unfortunately, we have only gained access to aggregate daily frequencies, not to individual level data. We also do not know how many players used the randomization HuxFlux option. However, because the operators told us how HuxFlux worked, we can estimate that approximately 19 percent of players were randomizing in the first week.[12]

Note that the theoretical analysis of the LUPI game in the previous section differs from the field LUPI game in three ways. First, the theory used a tie-breaking rule in which nobody wins if there is no uniquely chosen number (to simplify expected payoff calculations enormously). In the field game, however, players who tie by choosing the smallest and least-frequently chosen number share the prize. This is a minor difference because the probability that there is no unique number is very small and it never happened during the 49 days for which we have data. A second, more important, difference is that we assume that each player can only pick one

---

[11] The rule that players could only pick up to six numbers a day was enforced by the requirement that players had to use a "gambler's card" linked to their personal identification number when they played. Colluding in LUPI can conceivably increase the probability of winning but would require a remarkable degree of coordination across a large syndicate, and is also risky if others collude in a similar way.

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

TABLE 1—DESCRIPTIVE STATISTICS AND POISSON-NASH EQUILIBRIUM PREDICTIONS FOR FIELD LUPI GAME

	All days				Week 1	Week 7	Eq
	Avg.	SD	Min.	Max.	Avg.	Avg.	Avg.
Number of bets	53,783	7,782	38,933	69,830	57,017	47,907	53,783
Average number played	2,835	813	2,168	6,752	4,512	2,484	2,595
Median number played	1,675	348	436	2,273	1,204	1,936	2,542
Winning number	2,095	1,266	162	4,465	1,159	1,982	2,595
Lowest number not played	3,103	929	480	4,597	1,745	3,462	4,938
Above $T = 5,513$ (percent)	6.65	6.24	2.56	30.32	20.11	2.80	0.00
First (25 percent) quartile	780	227	66	1,138	425	914	1,251
Third (75 percent) quartile	2,898	783	2,130	7,610	3,779	3,137	3,901
Below 100 (percent)	6.08	4.84	2.72	2.97	15.16	3.19	2.02
Below 1,000 (percent)	32.31	8.14	21.68	63.32	44.91	27.52	20.03
Below 5,000 (percent)	92.52	6.44	68.26	96.74	78.75	96.44	93.32
Below 10,000 (percent)	96.63	3.80	80.70	98.94	88.07	98.81	100.00
Even numbers (percent)	46.75	0.58	45.05	48.06	45.91	47.45	50.00
Divisible by 10 (percent)	8.54	0.466	7.61	9.81	8.43	9.01	9.99
Proportion 1900–2010 (percent)	71.61	4.28	67.33	87.01	79.39	68.79	49.78
11, 22, …, 99 (percent)	12.2	0.82	10.8	14.4	12.39	11.44	9.09
111, 222, …, 999 (percent)	3.48	0.65	2.48	4.70	4.27	2.78	0.90
1111, 2222, …, 9999 (1/1,000)	4.52	0.73	2.81	5.80	4.74	3.95	0.74
11111, 22222, …, 99999 (1/1,000)	0.76	0.84	0.15	5.45	2.26	0.21	0

Notes: Proportion of numbers between 1900 and 2010 refers to the proportion relative to numbers between 1844 and 2066. "11, 22, …, 99" refers to the proportion relative to numbers below 100, "111, 222, …, 999" refers to numbers below 1,000, and so on. The "Eq Avg" predictions refers to the predictions of the Poisson-Nash equilibrium with $n = 53,783$, and $K = 99,999$.

number. In the field game, players are allowed to bet on up to six numbers. This does play a role for the theoretical predictions, since it allows players to coordinate one's guesses to avoid choosing the same number more than once (as could be the case when each bid is placed by a different player). Finally, we do not take the second and third prizes present in the field version into account, but this is unlikely to make a big difference given the strategic nature of the game.

Nevertheless, these three differences between the payoff structures of the game analyzed theoretically, and the field game as it was played, are a motivation for running laboratory experiments with single bets, no opportunity for direct collusion, and only a first prize, which match the game analyzed theoretically more closely.

A. Descriptive Statistics

Table 1 reports summary statistics for the first 49 days of the game. Two additional columns display the corresponding daily averages for the first and last weeks to see how much learning takes place. The last column displays the corresponding statistics that would result from play according to the Poisson equilibrium.

Overall, the average number of bets N was 53,783, but there was considerable variation over time. There is no apparent time trend in the number of participating players, but there is less participation on Sundays and Mondays (see Figure A2 in the online Appendix).[13] The variation of the number of bets across days is therefore

---

[13] The Sunday–Monday average N (standard deviation) is 44,886 (4,001) and the Tuesday–Saturday average is 57,341 (5,810). Dividing the sample in this way does reduce the variance in N by almost half. However, the sum-

much higher than what the Poisson distribution predicts (its standard deviation is 232). However, note that larger variance in N means sometimes there are many fewer players (so chosen numbers should be smaller) and sometimes there are many more players (so chosen numbers should be larger). Fixing the mean of N and increasing the variance might therefore have little overall impact on the equilibrium number distribution (and has little effect in the lab data reported later).

Despite some differences between the simplified theory and the way the field lottery game was implemented, the average number chosen overall was 2,835, which is close to the equilibrium prediction of 2,595. The mean number in the last week is 2,484, compared to the prediction of 2,595.[14] The median converged toward the equilibrium prediction of 2,542, from 1,204 in the first week to 1,936 in the last week. Winning numbers, and the lowest numbers not chosen by anyone, also varied a lot over time.

In equilibrium, the first and foremost prediction is that essentially nobody (fewer than 0.01 percent) should choose a number above $T = 5,513$. In the field lottery game, 20 percent chose these high numbers in the first week, but in the last week only 2.8 percent did. For numbers above 10,000, 12 percent chose these extremely high numbers in the first week, but in the last week only 1 percent did. This indicates both compelling convergence, as well as initial deviation. In fact, the third quartile (75 percent) was 7,610 in day 1, but quickly dropped below 3,200, resulting in an average of 3,779 for week 1 and 2,443 for week 2. Then, the third quartile converged back toward the equilibrium prediction (3,901), ending up at 3,137 in week 7. *All* other aggregate statistics in Table 1 are closer to the equilibrium predictions in the last week than in the first week. Many of the statistics converge rather swiftly and closely. For example, although 20 percent chose numbers above $T = 5,513$ in week 1, less than 5 percent did each day from week 3 to 7.[15]

An interesting feature of the data is a tendency to avoid round or focal numbers and choose quirky numbers that are perceived as "anti-focal" (as in hide-and-seek games, see Vincent P. Crawford and Nagore Iriberri 2007a). Even numbers were chosen less often than odd ones (46.75 percent versus 53.25 percent). Numbers divisible by 10 are chosen a little less often than predicted. Strings of repeating digits (e.g., 1111) are chosen too often. Players also overchoose numbers that represent years in modern time (perhaps their birth years). If players had played according to equilibrium, the fraction of numbers between 1900 and 2010 divided by all numbers between 1844 and 2066 should be 49.78 percent, but the actual fraction was 70 percent.[16]

---

mary statistics in the two groups are very close together (the means are 2,792 and 2,941).

[14] To judge the significance of the difference between theory and data we simulated 1000 weekly average numbers from the Poisson-Nash equilibrium. That is, $7 \times 53,783$ i.i.d. draws were drawn from the equilibrium distribution and the average number was computed. This yields one simulated average. The procedure was then repeated a total of 10,000 times to create 10,000 simulated averages. The low and high range of 9,500 of these simulated averages—a 95 percent confidence interval—is 2,590 to 2,600. Since the weekly averages in the data lie outside this extremely tight interval, we can conclude that the data are significantly different than those predicted by theory. But note that this is an extremely demanding test because the very large sample sizes mean that the data must lie very close to the theory to *not* reject the theory.

[15] Figure A3 (in online Appendix) provides weekly boxplots of the data and Figure A4 plots average daily frequencies for week 1, 3, 5, and 7 for those who are interested in weekly changes in the distribution and percentiles.

[16] We compare the number of choices between 1900 and 2010 to the number of choices between 1844 and 2066 since there are twice as many strategies to choose from in the latter range compared to the first. If all players
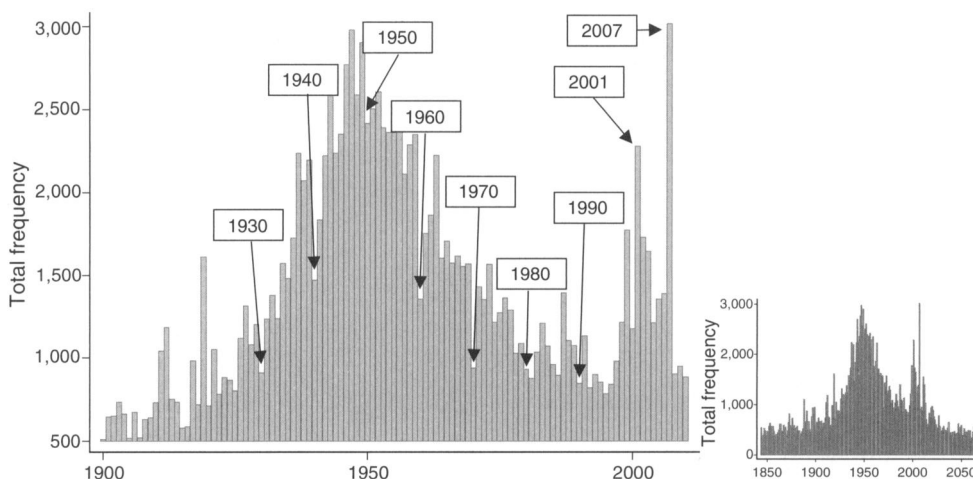
FIGURE 2. NUMBERS CHOSEN BETWEEN 1900 AND 2010, AND BETWEEN 1844 AND 2066,
DURING ALL DAYS IN THE FIELD

Figure 2 shows this focality in a histogram of numbers between 1900 and 2010 (aggregating all 49 days). Note that although the numbers around 1950 are most popular, there are noticeable dips at focal years that are divisible by ten.[17] Figure 2 also shows the aggregate distribution of numbers between 1844 and 2066, which clearly shows the popularity of numbers around 1950 and 2007. There are also spikes in the data for special numbers like 2121, 2222, and 2345. Explaining these focal numbers with the cognitive hierarchy model presented below is not easy (unless the 0-step player distribution is defined to include focality), so we will not comment on them further (see Crawford and Iriberri 2007a for a successful application in simpler hide-and-seek games).

B. *Results*

Do subjects in the field LUPI game play according to the Poisson-Nash equilibrium benchmark? In order to investigate this, we assume that the number of players is Poisson distributed with mean equal to the empirical daily average number of numbers chosen (53,783). As noted previously, this assumption is wrong because the variation in number of bets across days is much higher than what the Poisson distribution predicts.

Figure 3 shows the average daily frequencies from the first week together with the equilibrium prediction (the dashed line), for all numbers up to 99,999 and for the restricted interval up to 10,000. Recall that in the Poisson-Nash equilibrium, probabilities of choosing higher numbers first decrease slowly, drop quite sharply

---

randomized uniformly (an approximation to the equilibrium strategy with large n and K), the proportion of numbers between 1900 and 2010 would be 50 percent.

[17] Note that it would be unlikely to observe these dips reliably with typical experimental sample sizes. It is only with the large amount of data available from the field (2.5 million observations) that these dips are visually obvious and different in frequency than neighboring unround numbers.
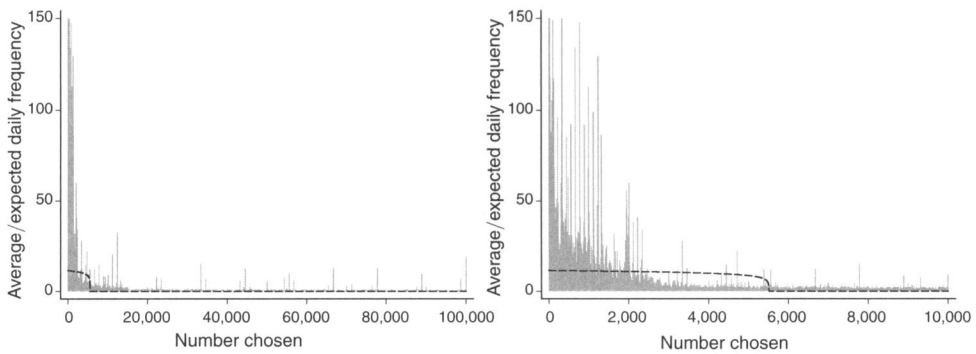
FIGURE 3. AVERAGE DAILY FREQUENCIES AND POISSON-NASH EQUILIBRIUM PREDICTION FOR THE FIRST WEEK
IN THE FIELD ($n = 53{,}783$, $K = 99{,}999$)

at around 5,500, and asymptotes to zero after $p_{5513} \approx 1/n$ (recall Proposition 1 and Figure 1). Compared to equilibrium, there is overshooting at numbers below 1,000, undershooting at numbers between 2,000 and 5,500, and again overshooting beyond 5,500.[18] It is also noteworthy how spiky the data is compared to the equilibrium prediction, which is a reflection of clustering on special numbers, as described above.

Nonetheless, the ability of the very complicated Poisson-Nash equilibrium approximation to capture some of the basic features of the data is surprisingly good. For example, most of the guesses (79.89 percent in the first week) are concentrated at or below $T = 5{,}513$. As a referee nicely expressed this central result: "To me, the truncation of the distribution (i.e., the set $[T, K]$ has negligible mass) is the first-order effect of equilibrium reasoning. Furthermore, the relationship between k, K, and T is not obvious so the finding that, by the seventh week, almost all of the mass of the empirical distribution is concentrated in $[0, T]$ is quite striking."

Figure 4 shows average daily frequencies of choices from the second through the last (7th) week. Behavior in this period is closer to equilibrium than in the first week. In particular, the overshooting of high numbers almost vanishes—only 4.41 percent of the choices are above 5,513. However, when only numbers below 10,000 are plotted, the overshooting of low numbers and undershooting of intermediate numbers is still clear (although the undershooting region shrinks to numbers between 4,000 and 5,500) and there are still many spikes of clustered choices.

The first three columns of Table 2 provide the frequencies for the last week of field data (Figure A4 in the online Appendix) with a bin size of 500 up to number 5,500 and compare it with the Poisson-Nash equilibrium prediction. The χ^2 test statistic is 53,864.6, strongly rejecting the equilibrium model. This suggests that although there is only substantial undershooting in the last three bins, the data from the final week is still far from the Poisson-Nash equilibrium. Moreover, the drop from around 43,000 (in the first four bins) to around 30,000 (in the next four bins) is also difficult to account for.

---

[18]This overshooting-undershooting-overshooting pattern could explain why in Table 1, the average number played is close to the equilibrium prediction, while the first quartile and median are always too low (though converging up) and the percentage above 5,513 is always too high.
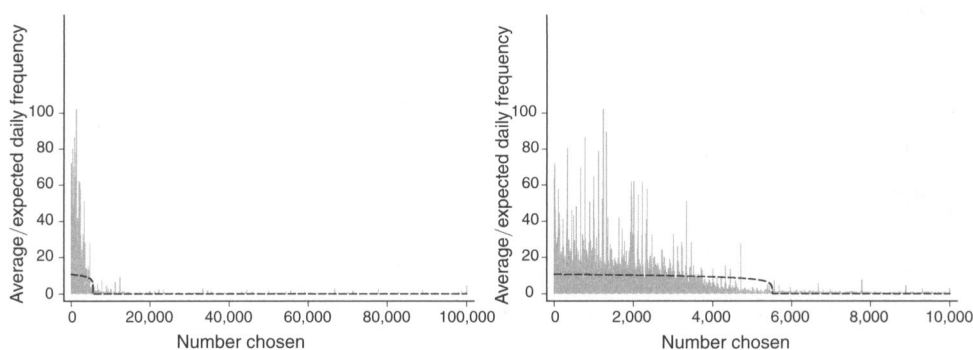
FIGURE 4. AVERAGE DAILY FREQUENCIES AND POISSON-NASH EQUILIBRIUM PREDICTION
FOR WEEK 2–7 IN THE FIELD ($n = 53,783$, $K = 99,999$)

TABLE 2—FREQUENCY TABLE FOR THE LAST WEEK OF FIELD DATA, THE POISSON-NASH EQUILIBRIUM AND
COGNITIVE HIERARCHY MODEL

| | Week 7 versus equilibrium | | | | Week 7 versus cognitive hierarchy | | | |
| | Total frequency (for all numbers) | | Average frequency (for each number) | | Total frequency (for all numbers) | | Average frequency (for each number) | |
Bin range	Week 7	Eq	Week 7	Eq	Week 7	CH	Week 7	CH
1–500	47,047	33,796.5	94	67.6	47,047	43,538.8	94	87.1
501–1,000	45,052	33,448.3	90	66.9	45,052	42,641.2	90	85.3
1,001–1,500	41,489	33,060.8	83	66.1	41,489	42,343.9	83	84.7
1,501–2,000	43,815	32,624.0	88	65.2	43,815	41,257.8	88	82.5
2,001–2,500	33,827	32,123.5	68	64.2	33,827	39,631.4	68	79.3
2,501–3,000	29,850	31,537.8	60	63.1	29,850	36,794.6	60	73.6
3,001–3,500	33,115	30,832.0	66	61.7	33,115	32,437.4	66	64.9
3,501–4,000	25,765	29,943.8	52	59.9	25,765	25,532.5	52	51.1
4,001–4,500	16,810	28,745.2	34	57.5	16,810	16,006.9	34	32.0
4,501–5,000	6,614	26,891.5	13	53.8	6,614	6,401.8	13	12.8
5,001–5,500	2,463	22,130.7	5	44.3	2,463	2,075.7	5	4.2
χ^2	53,864.6***		107.6***		2,891.4***		5.66	

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Nevertheless, given there are so much data, it is not surprising that the equilibrium model is rejected. One possible remedy is to use the *average* number of guesses for each number (in the bin range), instead of the *total* number of guesses. This is done by simply dividing the bin totals by 500. We then round these numbers into integers so we can perform a χ^2 test (like for the total number of guesses). The results are shown in column 4 and 5 of Table 2. The χ^2 test statistic is 107.6, still rejecting the equilibrium model, but much smaller than that of the total.

The top panel of Table 3 provides additional weekly goodness-of-fit measures for the Poisson-Nash equilibrium. Weekly χ^2 test results are shown in the first two rows. In particular, the χ^2 test statistics drop sharply from more than 640 in the first week to less than 110 in the last week. Nevertheless, the equilibrium prediction is rejected at the 0.1 percent level for all weeks.

TABLE 3—GOODNESS-OF-FIT FOR POISSON-NASH EQUILIBRIUM AND COGNITIVE HIERARCHY FOR FIELD DATA

Week	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Poisson-Nash equilibrium							
χ^2 (for average frequency)	640.45	323.76	257.42	259.27	261.19	121.29	107.60
(Degree of freedom)	(10)***	(10)***	(10)***	(10)***	(10)***	(10)***	(10)***
Proportion below (percent)	48.95	61.29	67.14	67.44	69.93	76.25	76.23
ENO	2,176.4	4,964.4	6,178.4	7,032.4	8,995.0	14,056.8	13,879.3
Cognitive hierarchy model							
Log-likelihood	−53,740	−31,881	−22,085	−19,672	−19,496	−19,266	−17,594
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)	77.92	52.21	7.64	3.90	4.60	4.64	5.48
(Degree of freedom)	(9)***	(9)***	(8)	(8)	(8)	(9)	(9)
Proportion below (percent)	62.58	72.57	78.65	80.17	82.09	82.43	82.24
ENO	3,188.8	7,502.5	9,956.0	12,916.1	17,873.0	21,469.6	21,303.0

Notes: 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.
*** Significant at the 1 percent level.
** Significant at the 5 percent level.
* Significant at the 10 percent level.

Another possibility is to calculate the proportion of the empirical density that lies below the predicted density. This measure is one minus the summed "miss rates", the differences between actual and predicted frequencies, for numbers which are chosen more often than predicted. In Table 3, the percentage of empirical data that lies below equilibrium density is reported in the third row, increasing from just below 50 percent in the first week to more than 75 percent in the last week.

Finally, when all models are not literally true, one can compare models using the "equivalent number of observations" (ENO) of the relevant model computed from raw mean squared errors. ENO was first proposed by Ido Erev et al. (2007) to compare relative performance of different learning models (that were all rejected) predicting subject behavior in games with mixed strategy equilibria, a similar situation to what we have here. Roughly speaking, ENO represents the weight one should put on the model relative to the existing data when predicting new observations. As stated in Erev et al. (2007): "The ENO of the model is the value of [N] (the size of the experiment) that is expected to lead to a prediction that is as accurate as the model's prediction." As shown in the fourth row of Table 3, the ENO of the Poisson-Nash equilibrium increases from about 2,200 in week 1 to almost 14,000 in week 7, demonstrating the improvement of equilibrium from week 1 to 7.[19]

The next question is whether an alternative theory can explain both the degree to which the equilibrium prediction is surprisingly accurate *and* the degree to which there is systematic deviation.

---

[19] However, since there are on average 47,907($\approx$ 335,347/7) guesses every day in week 7, even an ENO of 14,000 can still be easily outweighed by one day of data.
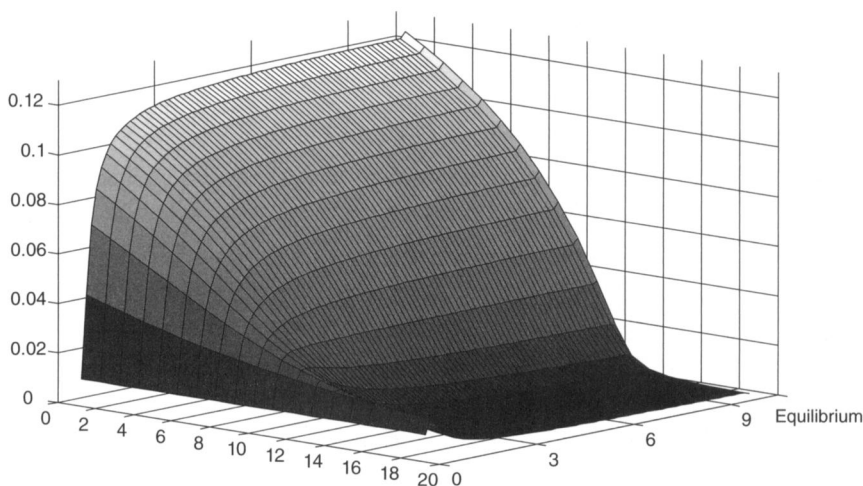
FIGURE 5. PROBABILITY OF CHOOSING NUMBERS 1–20 IN SYMMETRIC QRE ($n = 26.9$, $K = 99$, $\lambda = 0.001, \ldots, 10$) AND IN THE POISSON-NASH EQUILIBRIUM ($n = 26.9$, $K = 99$)

C. Rationalizing Non-Equilibrium Play

A natural way to model limits on strategic thinking is by assuming that different players carry out different numbers of steps of iterated strategic thinking in a cognitive hierarchy (CH). This idea has been developed in behavioral game theory by several authors (e.g., Rosemarie Nagel 1995; Dale O. Stahl and Paul W. Wilson 1995; Miguel Costa-Gomes, Crawford, and Bruno Broseta 2001; Camerer, Teck-Hua Ho, and Juin-Kuan Chong 2004; and Costa-Gomes and Crawford 2006) and applied to many games of different structures (e.g., Crawford 2003; Camerer, Ho, and Chong 2004; Crawford and Iriberri 2007b; and Tore Ellingsen and Östling 2010).[20]

One alternative candidate to cognitive hierarchy would be the quantal response equilibrium (QRE). QRE and CH have been compared to Nash predictions in many experimental studies, and they often explain deviations from Nash equilibrium in similar ways (e.g., Brian W. Rogers, Thomas R. Palfrey, and Camerer 2009). However, QRE and CH can be clearly distinguished in LUPI games since QRE seems to predict too few low-number choices. For example, for $n = 26.9$ players and number choices from 1 to $K = 99$ (as implemented in our lab experiment), Figure 5 shows a 3-dimensional plot of the QRE probability distributions for many values of λ, along with the Poisson-Nash equilibrium.[21] When λ is low, the distribution is approximately uniform. As λ increases more probability is placed on lower numbers 1–12. When λ is high enough the QRE closely approximates the Poisson-Nash equilibrium, which puts roughly linear declining weight on numbers 1 to 15 and

---

[20] A precursor to these models was the insight, developed much earlier in the 1980's by researchers studying negotiation, that people often "ignore the cognitions of others" in asymmetric-information bidding and negotiation games (Max H. Bazerman et al. 2000).

[21] The plot shows the QRE based on a power function, but the picture looks identical with a logit function.

infinitesimal weight on higher numbers. We therefore focus on a cognitive hierarchy model in this paper.

Level-k and cognitive hierarchy models require a specification of how k-step players behave and the proportions of players for various k. We follow Camerer, Ho, and Chong (2004) and assume that the proportion of players that do k thinking steps is Poisson distributed with mean τ, i.e., the proportion of players that think in k steps is given by

$$f(k) = e^{-\tau} \tau^k / k!.$$

We assume that k-step thinkers incorrectly believe that all other players can only think fewer steps than themselves, but correctly guess the proportions of players doing 0 to $k - 1$ steps (as a truncated Poisson distribution). In other words, level-1 thinkers believe all other players are level- 0 types. Level-2 thinkers believe there are level-0 types and level-1 types. Level-3 thinkers believe there are level-0, level-1 and level-2 types, and so on.[22] Then the conditional density function for the belief of a k-step thinker about the proportion of $l < k$ step thinkers is

$$g_k(l) = \frac{f(l)}{\sum_{h=0}^{k-1} f(h)}.$$

The IA and EE properties of Poisson games (together with the general type specification described earlier) imply that the number of players that a k-step thinker believes will play strategy i is Poisson distributed with mean

$$nq_i^k = n\sum_{j=0}^{k-1} g_k(j)p_i^j.$$

Hence, the expected payoff for a k-step thinker of choosing number i is

$$\pi^k(i) = \prod_{j=1}^{i-1} \left[1 - nq_j^k\, e^{-nq_j^k}\right] \cdot e^{-nq_i^k}.$$

To fit the data well, it is necessary to assume that players respond stochastically (as in QRE) rather than always choose best responses (see also Rogers, Palfrey, and Camerer 2009).[23] We assume that level-0 players randomize uniformly across all

---

[22] An alternative approach which often has advantages is that level-k types think *all* other players are exactly level $(k - 1)$. However, this does not work in LUPI games: If we start out with L0 types playing random, L1 types should all play 1. If L2 types believe there are only L1 types, they should never play 1. If L3 types best respond to only L2 types, then they should all play 1 (since they believe nobody is playing 1), and this logic will continue to cycle.

[23] The CH model with best-response piles up most predicted responses at a very small range of the lowest integers (1-step thinkers choose 1, 2-step thinkers choose 2, and k-step thinkers will never pick a number higher than k). Assuming quantal response smoothes out the predicted choices over a wider number range.
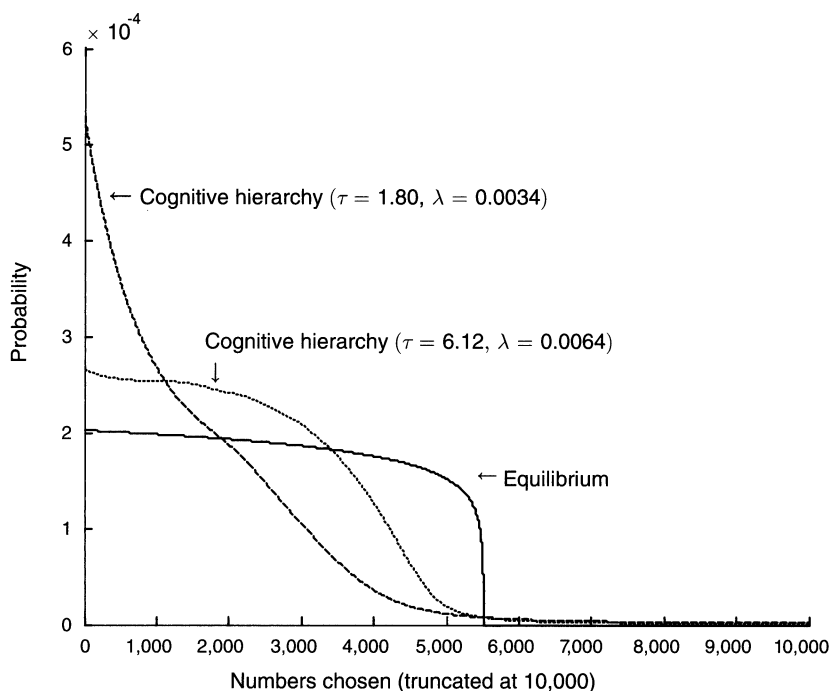
FIGURE 6. PROBABILITY OF CHOOSING NUMBERS 1–10,000 IN THE POISSON-NASH EQUILIBRIUM
AND THE COGNITIVE HIERARCHY MODEL ($n = 53,783$, $K = 99,999$)

numbers 1 to K, and higher-step players best respond with probabilities determined by a normalized power function of expected payoffs.[24]

The probability that a k step player plays number i is given by

$$
p_i^k = \frac{\left(\Pi_{j=1}^{i-1} \left[1 - nq_j^k \, e^{-nq_j^k} \right] e^{-nq_i^k} \right)^{\lambda}}{\sum_{l=1}^{K} \left(\Pi_{j=1}^{i-1} \left[1 - nq_j^k \, e^{-nq_j^k} \right] e^{-nq_l} \right)^{\lambda}},
$$

for $\lambda > 0$. Since q_j^k is defined recursively—it only depends of what lower step thinkers do—it is straightforward to compute the predicted choice probabilities numerically for each type of k-step thinker (for given values of τ and λ) using a loop, then aggregating the estimated p_k^k across steps k. Apart from the number of players and the number of strategies, there are two parameters: the average number of thinking steps, τ, and the precision parameter, λ.

Figure 6 shows the prediction of the cognitive hierarchy model for the parameters of the field LUPI game, i.e., when $n = 53,783$ and $K = 99,999$. The dashed line

---

[24] In many previous studies logit choice functions are typically used and they fit comparably to power functions (e.g., Camerer and Ho 1998 for learning models). Some QRE applications have used power functions and found better fits (e.g., in auctions, Jacob K. Goeree, Charles A. Holt, and Palfrey 2002). However, in this case a logit choice function fits substantially worse for the field data (with 99,999 numbers to choose from). The reason is that logit choice probabilities are convex in expected payoff. This implies numerically that probabilities are either substantial for only a small number of the 99,999 possible numbers, or are close to uniform across numbers. The logit CH model simply cannot fit the intermediate case in which thousands of numbers are chosen with high probability and many other numbers have very low probability (as in the data).
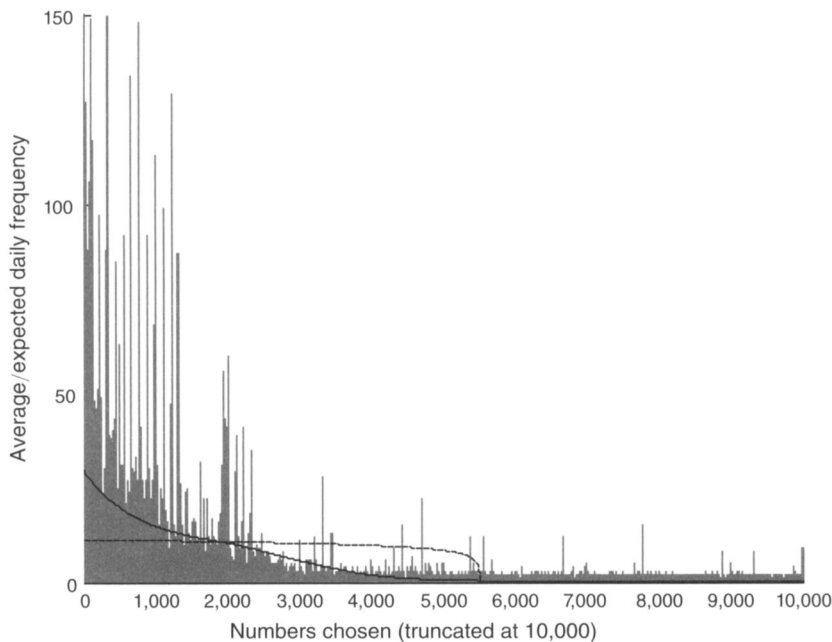
FIGURE 7. AVERAGE DAILY FREQUENCIES, COGNITIVE HIERARCHY (SOLID LINE) AND
POISSON-NASH EQUILIBRIUM PREDICTION (DASHED LINE) FOR THE FIRST WEEK IN THE FIELD
($n = 53{,}783$, $K = 99{,}999$, $\tau = 1.80$, $\lambda = 0.0034$)

corresponds to the case when players do relatively few steps of reasoning and their responses are very noisy ($\tau = 1.80$ and $\lambda = 0.0034$). The dotted line corresponds to the case when players do more steps of reasoning and respond more precisely ($\tau = 6.12$ and $\lambda = 0.0064$). Increasing τ and λ creates a closer approximation to the Poisson-Nash equilibrium, although even with a high τ there are too many choices of low numbers.

Can the cognitive hierarchy model account for the main deviations from equilibrium described in the previous section? The bottom panel of Table 3 reports the results from the maximum likelihood estimation of the data using the cognitive hierarchy model.[25] The best-fitting estimates week-by-week suggest that both parameters increase over time. The average number of thinking steps that people carry out, τ, increases from about 1.8 in the first week—an estimate in line with estimates from 1.0 to 2.5 that typical fit experimental datasets well (Camerer, Ho, and Chong 2004)—to 6 in the last week.

Figure 7 shows the average daily frequencies from the first week together with the CH estimation and the equilibrium prediction. The CH model does a reasonable job of accounting for the over- and undershooting tendencies at low and intermediate numbers (with the estimated $\hat{\tau} = 1.80$). In later weeks, the week-by-week

---

[25] It is difficult to guarantee that these estimates are global maxima since the likelihood function is not smooth and concave. We also used a relatively coarse grid search, so there may be other parameter values that yield slightly higher likelihoods and different parameter values.
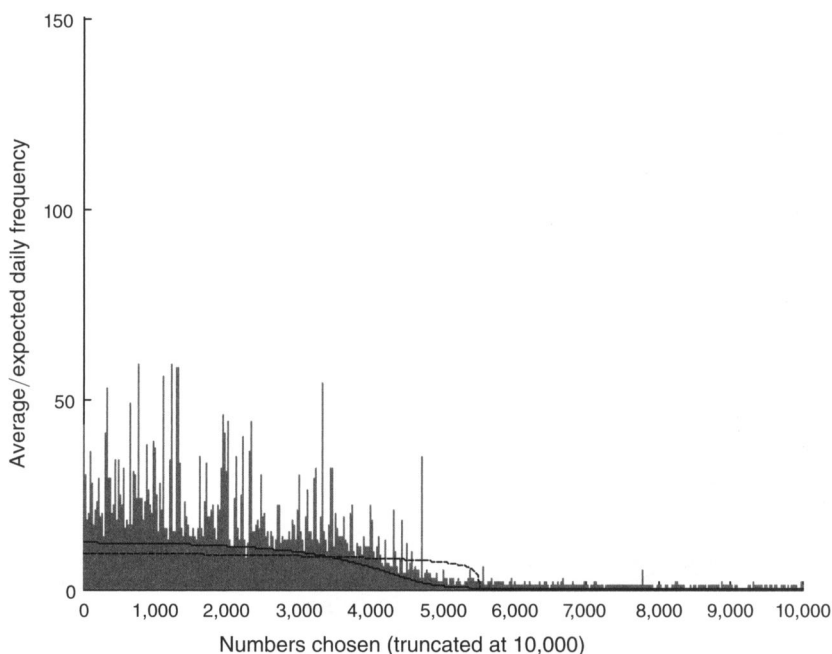
FIGURE 8. AVERAGE DAILY FREQUENCIES, COGNITIVE HIERARCHY (SOLID LINE) AND
POISSON-NASH EQUILIBRIUM PREDICTION (DASHED LINE) FOR THE LAST WEEK IN THE FIELD
$(n = 53{,}783, K = 99{,}999, \tau = 6.12, \lambda = 0.0064)$

estimates of τ drift upward a little (and λ increases slightly), which is a reduced-form model of learning as an increase in the mean number of thinking steps. In the last week the cognitive hierarchy prediction is much closer to equilibrium (because τ is around 6) but is still consistent with the smaller amounts of over- and under-shooting of low and intermediate numbers (see Figure 8).

To get some notion of how close to the data the fitted cognitive hierarchy model is, the bottom panel of Table 3 displays several goodness-of-fit statistics. First, the log-likelihoods reveal that the cognitive hierarchy model does better in explaining the data toward the last week and is always much better than Poisson-Nash.[26] However, as shown in the right panel of Table 2, though predictions of the cognitive hierarchy model are much closer to the data than that of equilibrium, the large number of observations simply forces the χ^2 test to reject the model, even when we bin 500 numbers into one category. In particular, the last week χ^2 test statistic for the cognitive hierarchy model is 2,891.4, much smaller than that of equilibrium (53,864.6), but still at an extremely high level of significance.

As discussed above, one remedy is to consider rounded averages for each bin. This is reported in the last two columns of Table 2. The last week χ^2 test statistic

[26] Since the computed Poisson-Nash equilibrium probabilities are ε for $k > 5{,}518$, the likelihood is always essentially zero for the equilibrium prediction. In online Appendix B, however, we compute the log-likelihood for the low numbers only. Based on Gideon Schwarz' (1978) information criterion, the cognitive hierarchy model still performs better in all weeks.

is 5.66, implying that the prediction is not statistically different from the cognitive hierarchy prediction. Since there is one cell that has a predicted value less than five, it might not be appropriate to conduct a χ^2 test including that cell. Hence, the bottom panel of Table 3 reports weekly χ^2 test results that focus on only the bins that have a predicted value greater than 5. The fifth row of the bottom panel provides the number of bins used, and the fourth row reports χ^2 test statistics. As can be seen, the cognitive hierarchy model is still strongly rejected in the first two weeks, though not as strongly as equilibrium, but not in the following weeks.

Two other measures of goodness-of-fit previously discussed are also computed in order to compare the CH model with the equilibrium prediction. In particular, the proportion of the empirical density that lies below the predicted density is reported in sixth row of the bottom panel (in Table 3). The cognitive hierarchy model does better than the equilibrium prediction in all seven weeks based on this statistic. For example, in the first week, 63 percent of players' choices were consistent with the cognitive hierarchy model, whereas 49 percent were consistent with equilibrium. However, both models improve substantially across the weeks. On the other hand, weekly ENO is calculated and reported in the fourth row of the bottom panel (in Table 3). Again, the cognitive hierarchy model does better in all seven weeks, starting from around 3,200 in the first week and end up at 21,000 in the final two.[27]

In conclusion, the cognitive hierarchy model performs better than the Poisson-Nash equilibrium in all seven weeks of data regardless of what measure is used, explaining the systematic deviation from equilibrium. In particular, the cognitive hierarchy model can rationalize the tendencies that some numbers are played more, as well as the undershooting below the equilibrium cutoff. The value-added of the cognitive hierarchy model is not primarily that it gives a slightly better fit, but that it provides a plausible story for how players manage to play so close to equilibrium.[28]

III. The Laboratory LUPI Game

We conducted a parallel lab experiment for two reasons.

First, the rules of the field LUPI game do not exactly match the theoretical assumptions used to generate the Poisson-Nash equilibrium prediction. In the field data some choices were made by a random number generator, some players might have chosen multiple numbers or colluded, there were multiple prizes, and the variance in N is larger than the Poisson distribution variance.

In the lab, we can more closely implement the assumptions of the theory. If the theory fits poorly in the field and closely in the lab, then that suggests the theory is on the right track when its underlying assumptions are most carefully controlled. If

---

[27] Again, since the total number of guesses in week 7 is 335,347, even though the CH model has a much higher ENO than Poisson-Nash equilibrium, it also can be easily outweighed by merely one day of data.

[28] Nonetheless, one might wonder whether the parameter-free Poisson-Nash equilibrium does worse than cognitive hierarchy merely because the latter has two parameters. We address this issue in Online Appendix B by estimating the Poisson-Nash equilibrium model week-by-week to obtain the best n (mean of the Poisson distribution) that minimizes mean squared error instead of maximizing empirical likelihood. As shown in Table A1, the estimated Poisson-Nash model still performs worse than the cognitive hierarchy model in week 1–5, but catches up and is comparable to cognitive hierarchy in week 6 and 7. Nevertheless, to make this prediction, the estimated n have to be much lower than the actual n and it is not clear how such incorrect beliefs could be sustained.

the theory fits closely in both cases, this suggests that the additional factors in the field that are excluded from the theory do not matter. If the theory fits both well, but slightly better in the lab, this is also reassuring since it indicates the additional factors in the field contributed merely noise.

Second, because the field game is rather simple, it is possible to design a lab experiment which closely matches the field in its key features. How closely the lab and field data match provides some evidence in ongoing debate about how well lab results generalize to comparable field settings (e.g., Levitt and List 2007, and Falk and Heckman 2009).

In designing the laboratory game, we compromise between two goals: to create a simple environment in which theory should apply (theoretical validity), and to recreate the features of the field LUPI game in the lab (specialized external validity). Because we use this opportunity to create an experimental protocol that is closely matched to a particular field setting, we often sacrificed theoretical validity in favor of close field replication.

The first choice is the scale of the game: The number of players (N), possible number choices (K), and stakes. We choose to scale down the number of players and the largest payoff by a factor of 2,000. This implies that there were on average 26.9 players and the prize to the winner in each round was \$7. We scaled down K by a factor of 1,000 since $K = 99$ allows for focal numbers such as 66, 77, 88, and 99 to be chosen, and the shape of the equilibrium distribution has some of the basic features of the equilibrium distribution for the field data parameters (e.g., most numbers should be below 10 percent of K). Since the field data span 49 days, the experiment also has 49 rounds in each session. (We typically refer to experimental rounds as "days" and seven-"day" intervals as "weeks" for semantic comparability between the lab and field descriptions.)

The number of players in each round was drawn from a distribution with mean 26.9. In three of the four sessions, subjects were told the mean number of players, and that the number varied from round to round, but did not know the distribution (in order to match the field situation in which players were very unlikely to know the total number playing each day). Due to a technical error, in these three sessions, the variance was lower than the Poisson variance (7.2 to 8.6 rather than 26.9). However, this mistake is likely to have little effect on behavior because subjects only know the winning number in each round and can draw little inference about the underlying distribution of the number of players. In the last session, the number of players in each round was drawn from a Poisson distribution with mean 26.9 and the subjects were informed about this (Figure A5 in the online Appendix). Furthermore, the data from the true Poisson session and the lower-variance sessions look statistically similar so we pool them for all analysis (see below).

Some design choices made the lab setting different from the field setting but closer to the assumptions of the theory. In contrast to the field game, in the lab each player was allowed to choose only one number, they could not use a random number generator, there was only one prize per round, \$7, and if there was no unique number nobody won.

In the field data, we do not know how much Swedish players learned each day about the full distribution of numbers that were chosen. The numbers were available

online and partially reported on a TV show. To maintain parallelism with the field, only the winning number was announced in the lab.

Four laboratory sessions were conducted at the California Social Science Experimental Laboratory (CASSEL) at the University of California Los Angeles on March 22nd and March 25th of 2007, and on March 3, 2009. The experiments were conducted using the Zürich Toolbox for Ready-made Economic Experiments (zTree) developed by Urs Fischbacher, as described in Fischbacher (2007). Within each session, 38 graduate and undergraduate students were recruited, through CASSEL's web-based recruiting system. All subjects knew that their payoff will be determined by their performance. We made no attempt to replicate the demographics of the field data, which we unfortunately know very little about. However, the players in the laboratory are likely to differ in terms of gender, age and ethnicity compared to the Swedish players. In the four sessions, we had slightly more male than female subjects, with the great majority clustered in the age bracket of 18 to 22, and the majority spoke a second language. Half of the subjects had never participated in any form of lottery before. Subjects had various levels of exposure to game theory, but very few had seen or heard of a similar game prior to this experiment.

A. *Experimental Procedure*

At the beginning of each session, the experimenter first explained the rules of the LUPI game. The instructions were based on a version of the lottery form for the field game translated from Swedish to English (see online Appendix D). Subjects were then given the option of leaving the experiment, in order to see how much self-selection influences experimental generalizability. None of the recruited subjects chose to leave, which indicates a limited role for self-selection (after recruitment and instruction).

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

In the beginning of each round, subjects were informed whether they would actively participate in the current round (i.e., if they had a chance to win). They were required to submit a number in each round, even if they were not selected to participate. The difference between behavior of selected and non-selected players gives us some information about the effect of marginal incentives on performance (see Camerer and Robin M. Hogarth 1999).

When all subjects had submitted their chosen numbers, the lowest unique positive integer was determined. If there was a lowest unique positive integer, the winner earned $7; if no number was unique, no subject won. Each subject was privately informed, immediately after each round, what the winning number was, whether they had won that particular round, and their payoff so far during the experiment.
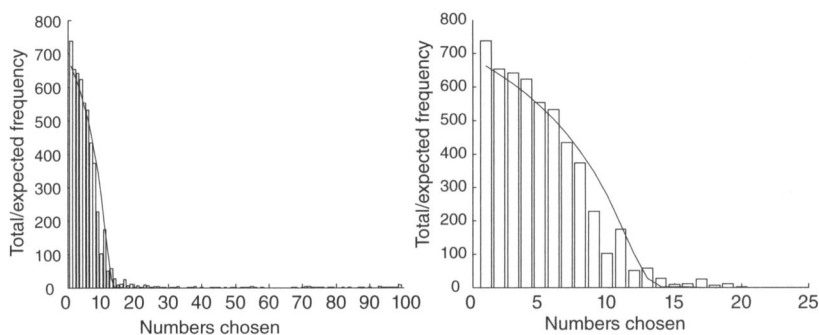
FIGURE 9. LABORATORY TOTAL FREQUENCIES AND POISSON-NASH EQUILIBRIUM PREDICTION
(all sessions, participating players only, $n = 26.9$, $K = 99$)

This procedure was repeated 49 times, with no practice rounds (as is the case of the field). After the last round, subjects were asked to complete a short questionnaire which allowed us to build the demographics of our subjects and a classification of strategies used. In two of the sessions, we included the cognitive reflection test as a way to measure cognitive ability (to be described below). All sessions lasted for less than an hour, and subjects received a show-up fee of \$8 or \$13 in addition to earnings from the experiment (which averaged \$8.60). Screenshots from the experiment are shown in online Appendix D.

B. *Lab Descriptive Statistics*

In the remainder of the paper, we focus only on the choices from incentivized subjects that were selected to actively participate. It is noteworthy, however, that the choices of participating and non-participating subjects did not significantly differ (p-value 0.66, Mann-Whitney). The choices of participating subjects from the session with the announced Poisson distribution and the pooled other three sessions do not significantly differ at the five percent level ($p = 0.058$, Mann-Whitney, $p = 0.59$ based on t-test with clustered standard errors). In the remainder of the paper we therefore pool all four sessions.

Figure 9 shows the data for the choices of participating players (together with the Poisson-Nash equilibrium prediction). There are very few numbers above 20 so the numbers 1 to 20 are the focus in subsequent graphs. In line with the field data, players have a slight predilection for certain numbers, while others are avoided. Judging from Figure 9, subjects avoid some even numbers, especially 10, while they endorse the odd (and prime) numbers 11, 13, and 17. Interestingly, only one subject played 20, while 19 was played ten times and 21 was played seven times.

Table 4 shows some descriptive statistics for the participating subjects in the lab experiment. As in the field, some players in the first week tend to pick very high numbers (above 20) but the percentage shrinks by the seventh week. The average number chosen in the last week corresponds closely to the equilibrium prediction (5.8 versus 5.2) and the medians are identical (5.0). Both the average winning numbers and the lowest unchosen numbers are relatively close to the equilibrium

TABLE 4—DESCRIPTIVE STATISTICS FOR LABORATORY DATA

	All rounds				R. 1–7	R. 43–49	Eq
	Avg.	SD	Min.	Max.	Avg.	Avg.	Avg.
Average number played	5.96	1.43	4.32	12.55	8.56	5.84	5.22
Median number played	4.65	1.03	3	10	6.14	5.00	5
First (25 percent) quartile	2.50	0.54	2	4	3.00	2.43	3
Third (75 percent) quartile	7.24	1.72	5	17	10.21	7.14	8
Below 20 (percent)	98.02	2.77	81.98	100.00	93.94	98.42	100.00
Above $T = 11$ (percent)	5.60	6.52	0	42.34	16.52	4.64	2.44
Even numbers (percent)	45.19	4.47	35.16	53.47	42.11	49.15	46.86
Session 1							
Winning number	6.02	9.38	1	67	13.00	2.50	5.22
Lowest number not played	8.08	2.57	1	12	4.86	8.14	8.44
Session 2							
Winning number	5.07	2.59	1	10	5.83	5.14	5.22
Lowest number not played	7.47	2.96	1	12	6.29	8.43	8.44
Session 3							
Winning number	5.61	3.26	1	14	6.14	5.67	5.22
Lowest number not played	7.53	2.68	2	13	7.43	10.00	8.44
Session 4 (Poisson)							
Winning number	5.81	3.62	1	17	6.71	3.14	5.22
Lowest number not played	7.61	3.30	1	13	5.14	8.14	8.44

Notes: Summary statistics are based only on choices of subjects who are selected to participate. The equilibrium column refers to what would result if all players played according to equilibrium ($n = 26.9$ and $K = 99$).

prediction. The tendency to pick odd numbers decreases over time—42 percent of all numbers are even in the first week, whereas 49 percent are even in the last week. As in the field data, the overwhelming impression from Table 4 is that convergence to equilibrium is quite rapid over the 49 periods (despite receiving feedback only about the winning number).

C. Aggregate Results

In the Poisson equilibrium with 26.9 average number of players, strictly positive probability is put on numbers 1 to 16, while other numbers have probabilities numerically indistinguishable from zero. Figure 10 shows the average frequencies played in week 1 to 7 together with the equilibrium prediction (dashed line) and the estimated week-by-week results using the cognitive hierarchy model (solid line). These graphs clearly indicate that learning is quicker in the laboratory than in the field. Despite that the only feedback given to players in each round is the winning number, behavior is remarkably close to equilibrium already in the second week. However, we can also observe the same discrepancies between the equilibrium prediction and observed behavior as in the field. The distribution of numbers is too spiky and there is overshooting of low numbers and undershooting at numbers just below the equilibrium cutoff (at number 16).

Figure 10 also displays the estimates from a maximum likelihood estimation of the cognitive hierarchy model presented in the previous section (solid line).[29] In this

---

[29] To illustrate how the CH model behaves, consider $N = 26.9$ and $K = 99$, with $\tau = 1.5$ and $\lambda = 2$. Figure A6 (in the online Appendix) shows how 0 to 5 step thinkers play LUPI and the predicted aggregate frequency, summing
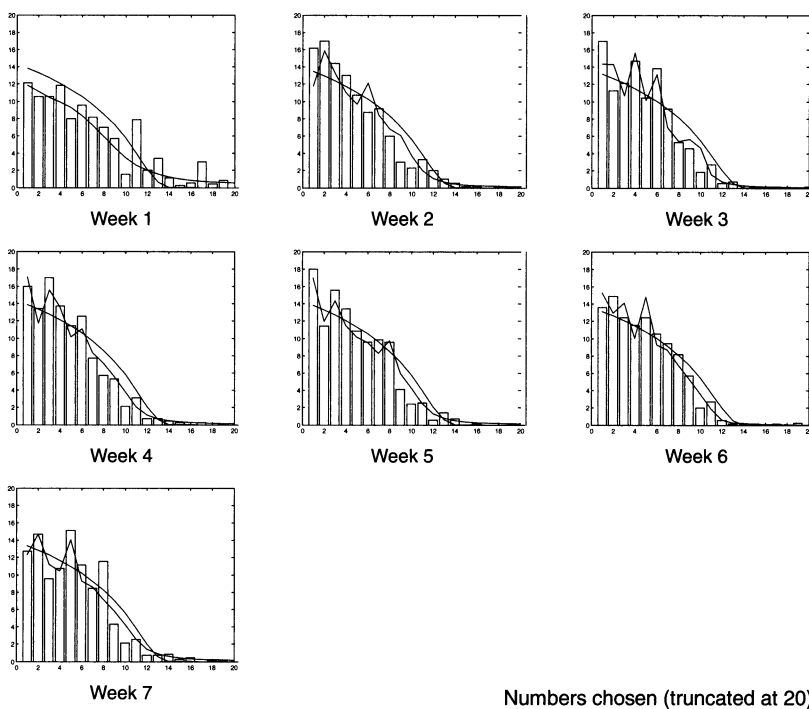
FIGURE 10. AVERAGE DAILY FREQUENCIES IN THE LABORATORY, POISSON-NASH EQUILIBRIUM PREDICTION
(DASHED LINES) AND ESTIMATED COGNITIVE HIERARCHY (SOLID LINES), WEEK 1–7 ($n = 26.9$, $K = 99$)

estimation, we use the estimated weekly τ from the field data and estimate λ only.[30] We do not estimate both parameters since we are most likely over-fitting by allowing two free parameters to estimate relatively few choice probabilities.[31] The cognitive hierarchy model can account both for the spikes and the over- and undershooting. Table 5 shows the estimated λ parameter. There is no clear time trend in this parameter and the λ parameter moves around quite a lot over the weeks. We also estimate the precision parameter λ while keeping the average number of thinking steps fixed at 1.5, which has been shown to be a value of τ that predicts experimental data well in a large number of games (Camerer, Ho, and Chong 2004). The estimated precision parameter is in this case considerably lower in the first week, but is then relatively constant.[32]

---

across all thinking steps. In this example, 1-step thinkers put most probability on number 1, 2-step thinkers put most probability on number 5, and 3-step thinkers put most probability on numbers 3 and 7.

[30] The alternative would be to fix λ and estimate τ, but there is no way to tell what a "reasonable" value of λ is. The precision parameter λ depends on scaling of payoffs, the number of strategies, etc., and can not be interpreted across games.

[31] When trying to estimate both parameters simultaneously, we found different estimates for different grid sizes and initial values. Most estimates of τ were between 6 and 12 and λ were most often between 10 and 20 (apart from the first week which always resulted in a λ of 1.32). The log-likelihood is neither smooth nor concave with fixed τ either, but with only one parameter to estimate we could use a very fine grid to search for the best-fitting parameter.

[32] Figure A7 (in the online Appendix) shows the fitted cognitive hierarchy model when τ is restricted to 1.5. It is clear that the model with $\tau = 1.5$ can account for the undershooting also when the number of thinking steps is fixed, but it has difficulties in explaining the overshooting of low numbers. The main problem is that with $\tau = 1.5$,

TABLE 5—GOODNESS-OF-FIT FOR POISSON-NASH EQUILIBRIUM AND COGNITIVE HIERARCHY MODEL FOR LABORATORY DATA

Week	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Poisson-Nash equilibrium							
χ^2 (for average frequency)	24.69	24.14	18.86	21.81	20.17	11.58	21.39
(Degree of freedom)	(5)***	(5)***	(5)***	(5)***	(5)***	(5)**	(5)***
Proportion below (percent)	82.25	88.55	87.61	88.64	88.64	92.86	87.06
ENO	158.5	202.5	173.2	239.5	244.1	844.4	200.3
Cognitive hierarchy model							
Log-likelihood	−210.4	−104.3	−88.6	−88.7	−87.5	−80.2	−99.4
τ (from field)	1.80	3.17	4.17	4.64	5.02	6.76	6.12
λ	1.26	5.97	16.89	5.59	5.28	22.69	4.52
χ^2 (for average frequency)	24.31	18.80	8.49	4.57	6.83	2.74	10.06
(Degree of freedom)	(5)***	(5)***	(5)	(5)	(5)	(5)	(5)*
Proportion below (percent)	84.62	87.44	90.52	92.54	92.42	91.11	91.07
ENO	296.0	263.3	466.8	4,909.9	3,475.8	894.7	502.9
Restricted cognitive hierarchy model							
λ ($\tau = 1.5$)	1.09	2.52	2.57	2.63	2.60	2.31	2.08

Notes: 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.
 ***Significant at the 1 percent level.
 **Significant at the 5 percent level.
 *Significant at the 10 percent level.

Table 5 provides some goodness-of-fit statistics for the cognitive hierarchy model and the Poisson-Nash equilibrium prediction. Consistent with results in the field, the cognitive hierarchy model fits data slightly better than the (parameter free) Poisson-Nash equilibrium in most weeks. In particular, the χ^2 test rejects equilibrium for all 7 weeks, but cannot reject the cognitive hierarchy model starting from week 3, even when we only bin 2 numbers into one category.[33] Similarly, based the proportion below the predicted density, the equilibrium prediction does remarkably well, while the cognitive hierarchy model does even better in all but the second and sixth weeks.[34]

The ENO results also confirm that equilibrium does pretty well starting from the second week, while cognitive hierarchy always does better than equilibrium.

D. *Individual Results*

An advantage of the lab over the field, in this case, is that the behavior of individual subjects can be tracked over time and we can gather more information about them to link to choices. Online Appendix D discusses some details of these analyses but we summarize them here only briefly.

---

there are too many zero-step thinkers that play all numbers between 1 and 99 with uniform probability. The log-likelihoods for the CH model with $\tau = 1.5$ range from −241 in week 1 to −212 in week 2, which are much worse than when using the field values of τ.

[33] We only use 6 bins (up to number 12) here to prevent the predicted number of observations to drop below 5. Even if we do not bin numbers at all, the χ^2 test (up to number 12) yields similar results, rejecting the equilibrium prediction for all weeks, and rejecting the cognitive hierarchy model for week 1, 2, 3, 6, and 7 (at the 5 percent level) and marginally for week 4 and 5 (at the 10 percent level).

[34] In online Appendix B we calculate the log-likelihoods using data from numbers 1 to 16, which allows us to compare the equilibrium prediction with cognitive hierarchy. Based on the Schwarz (1978) information criterion, the cognitive hierarchy model outperforms equilibrium in all weeks.

TABLE 6—PANEL DATA REGRESSIONS EXPLAINING INDIVIDUAL NUMBER CHOICES IN THE
LABORATORY

	All periods	Week 1	Week 2	Week 3–7
Round (1–49)	0.001	–0.109	–0.065	0.023
	(0.13)	(–0.42)	(–0.62)	(1.58)
$t - 1$ winner	0.178***	0.148**	0.304***	0.059*
	(4.89)	(2.38)	(2.98)	(1.89)
$t - 2$ winner	0.133***	0.096	0.242**	0.038*
	(2.98)	(1.18)	(2.40)	(1.68)
$t - 3$ winner	0.083*	0.052	–0.050	0.030
	(1.94)	(0.65)	(–0.63)	(1.18)
Fixed effects	Yes	Yes	Yes	Yes
Observations	4,360	421	585	3,354
R^2	0.03	0.09	0.01	0.00

Notes: The table report results from a linear subject fixed effects panel regression. Only actively participating subjects are included. t-statistics based on clustered standard errors are within parentheses.

  ***Significant at the 1 percent level.
   **Significant at the 5 percent level.
     *Significant at the 10 percent level.

In a post-experimental questionnaire, we asked people to state why they played as they did. We coded their responses into four categories (sometimes with multiple categories): "Random," "stick" (with one number), "lucky," and "strategic" (explicitly mentioning response to strategies of others). The four categories were coded 35, 30, 11, and 38 percent of the time. These categories had some relation to actual choices because "stick" players chose fewer distinct numbers and "lucky" players had number choices with a higher mean and higher variance. The only demographic variable with a significant effect on choices and payoffs was "exposure to game theory"; those subjects chose numbers with less variation across rounds. A measure of "cognitive reflection" (Shane Frederick 2005), a short-form IQ test, did not correlate with choice measures or with payoffs.

As is often seen in games with mixed equilibria, there is some mild evidence of "purification" since subjects chose only 9.65 different numbers on average (see online Appendix D), compared to 10.9 expected in Poisson-Nash equilibrium.

In the post-experimental questionnaire, several subjects said that they responded to previous winning numbers. To measure the strength of this learning effect we regressed players' choices on the winning number in the three previous periods. Table 6 shows that the winning numbers in previous rounds do affect players' choices early on, but this tendency to respond to previous winning numbers is considerably weaker in later weeks (3 to 7). The small round-specific coefficients in Table 6 also show that there does not appear to be any general trend in players' choices over the 49 rounds.

IV. Conclusion

It is often difficult to test game theory using field data because equilibrium predictions depend so sensitively on strategies, information and payoffs, which are usually not observable in the field. This paper exploits an empirical opportunity to test game

theory in a field setting which is simple enough that clear predictions apply (when some simplifying assumptions are made). The game is a LUPI lottery, in which the lowest unique positive integer wins a fixed prize. LUPI is a close relative of auctions in which the lowest unique bid wins.

One contribution of our paper is to characterize the Poisson-Nash equilibrium of the LUPI game and analyze behavior in this game using both a field dataset, including more than two million choices, and parallel laboratory experiments which are designed to first permit a clear test of the theory while also matching the field setting. In both the field and lab, players quickly learn to play close to equilibrium, but there are some diagnostic discrepancies between players' behavior and equilibrium predictions.

As noted earlier, the variance in the number of players in the field data is much larger than the variance assumed in the Poisson-Nash equilibrium. So the field data is not an ideal test of this theory, strictly speaking. Therefore, the key issues are how much the theory's predictions vary with changes in $\text{var}(N)$, and how much behavior changes in response to $\text{var}(N)$. If either theory or behavior is insensitive to $\text{var}(N)$, then the Poisson-Nash equilibrium could be a useful approximation to the field data.

As for theory: For the simple examples in which fixed-N and Poisson equilibria can be computed, zero variance (fixed-N) and Poisson variance equilibria are almost exactly the same (see online Appendix A). In fact, as shown in Figure A1 (in the online Appendix), the equilibrium probabilities for the fixed-N and Poisson-Nash equilibrium for the LUPI game with $n = 27$ and $K = 99$ (similar to the values used in the lab) are practically indistinguishable. Keep in mind that increasing $\text{var}(N)$ (holding n constant) implies that sometimes there are a lot of extra players so number choices should be higher, and sometimes there are fewer players so number choices should be lower. These two opposing effects could minimize the effect of variance on mean choices (as the low-K cases in online Appendix A suggest they do).

As for behavior: There are two sources of evidence that actual behavior is not too sensitive to $\text{var}(N)$. First, in the field data the Sunday and Monday sessions have lower n and lower standard deviation than all days, but choices are very comparable to data from all days (in which $\text{var}(N)$ about twice as large). Second, in the lab data different sessions with $\text{var}(N) \approx 8$ and $\text{var}(N) \approx 27$ lead to indistinguishable behavior.

These theoretical and behavioral considerations suggest why the "wrong" theory (Poisson-Nash) might approximate actual behavior surprisingly well in the field (despite the field $\text{var}(N)$ being empirically far from what the theory assumes).

A different way to describe our contribution is this: A LUPI game was actually played in the field, with specific rules. Can we produce *any* kind of theory which fits the data from this game? In this view, it does not matter whether the field setting matches the predictions of a theory exactly. Instead, all that matters is whether the theory fits well, even if its assumptions are wrong.

Here the answer is rather clear: The empirical distribution of choices clearly is moving in the direction of the Poisson-Nash equilibrium over the 49 days (as judged by every number choice statistic) and is numerically close. As a bonus, the CH model improves a little on the Poisson-Nash equilibrium, when optimally parameterized, in the sense that it can explain the key ways in which behavior departs from Poisson-Nash (too many low and very high numbers) in the short run. The estimated

number of thinking steps is in the first week 1.80, which is in line with estimates from many lab experiments.

Note that the point of the cognitive hierarchy model is not simply to fit the data better than Poisson-Nash, but also to show how people with limited computational power might start near, and converge to, such a complex equilibrium. However, the cognitive hierarchy model provides merely suggestive evidence regarding this convergence, and hence, should be viewed as a potential stepping stone (instead of the final word) to an investigation using a formal learning model.

Finally, because the LUPI field game is simple, it is possible to do a lab experiment that closely replicates the essential features of the field setting (which most experiments are not designed to do). This close lab-field parallelism in design adds evidence to the ongoing debate about when lab findings generalize to parallel field settings (e.g., Levitt and List 2007, and Falk and Heckman 2009). The lab game was described very much like the Swedish lottery (controlling context), experimental subjects were allowed to select out of the experiment after it was described (allowing self-selection), and lab stakes were made equal to the field stakes (in expected terms). Basic lab and field findings are fairly close: In both settings, choices are close to equilibrium, but there are too many large numbers and too few agents choose intermediate numbers at the high end of the equilibrium range. We interpret this as a good example of close lab-field generalization, when the lab environment is designed to be close to a particular field environment.[35]

APPENDIX: PROOF OF PROPOSITION 1

We first prove the five properties and then prove that the equilibrium is unique.

1) We prove this property by induction. For $k = 1$, we must have $p_1 > 0$. Otherwise, deviating from the proposed equilibrium by choosing 1 would guarantee winning for sure. Now suppose that there is some number $k + 1$ that is not played in equilibrium, but that k is played with positive probability. We show that $\pi(k + 1) > \pi(k)$, implying that this cannot be an equilibrium. To see this, note that the expressions for the expected payoffs allows us to write the ratio $\pi(k + 1)/\pi(k)$ as

$$\frac{\pi(k + 1)}{\pi(k)} = \frac{\Pr(X(k + 1) = 0) \cdot \prod_{i=1}^{k} \Pr(X(i) \neq 1)}{\Pr(X(k) = 0) \cdot \prod_{i=1}^{k-1} \Pr(X(i) \neq 1)}$$

$$= \frac{\Pr(X(k + 1) = 0) \cdot \Pr(X(k) \neq 1)}{\Pr(X(k) = 0)}.$$

If $k + 1$ is not used in equilibrium, $\Pr(X(k + 1) = 0) = 1$, implying that the ratio is above one. This shows that all integers between 1 and K are played with positive probability in equilibrium.

---

[35] Of course, it is also conceivable that there is a genuine lab-field behavioral difference but it is approximately canceled by differences in the design details which have opposite effects.

2) Rewrite condition (1) as

$$e^{np_{k+1}} - e^{np_k} = -np_k.$$

By the first property, both p_k and p_{k+1} are positive, so that the right hand side is negative. Since the exponential is an increasing function, we conclude that $p_k > p_{k+1}$.

3) Condition (2) can be re-written as

$$p_k - p_{k+1} = -\frac{1}{n} \ln(1 - f(np_k))$$

where $f(x) = xe^{-x}$, $f'(x) = (1 - x)e^{-x}$ and $x = np_k$. Hence, $f'(x) > 0$ if $x < 1$, and $f'(x) < 0$ if $x > 1$. If $p_{k+1} > 1/n$, by the second property, $np_k > np_{k+1} > 1$. So, $f(np_{k+1}) > f(np_k)$. It follows that

$$(p_{k+1} - p_{k+2}) = -\frac{1}{n} \ln(1 - f(np_{k+1}))$$

$$> -\frac{1}{n} \ln(1 - f(np_k)) = (p_k - p_{k+1}).$$

If $p_k < 1/n$, by the second property, $np_{k+1} < np_k < 1$. So, $f(np_{k+1}) < f(np_k)$. Thus,

$$(p_{k+1} - p_{k+2}) = -\frac{1}{n} \ln(1 - f(np_{k+1}))$$

$$< -\frac{1}{n} \ln(1 - f(np_k)) = (p_k - p_{k+1}).$$

4) Taking the limit of (2) as $n \to \infty$ implies that $p_{k+1} = p_k$.

5) Since $1 = \sum_{k=1}^{K} p_k > K \cdot p_K$ by the second property, we have $p_K < 1/K \to 0$ as $K \to \infty$.

In order to show that the equilibrium $\mathbf{p} = (p_1, p_2, \ldots, p_K)$ is unique, suppose by contradiction that there is another equilibrium $\mathbf{p}' = (p_1', p_2', \ldots, p_K')$. By the equilibrium condition (1), p_1 uniquely determines all probabilities $p_2, \ldots, p_K$, while p_1' uniquely determines $p_2', \ldots, p_K'$. Without loss of generality, we assume $p_1' > p_1$. Since in any equilibrium, p_{k+1} is strictly increasing in p_k by condition (1), it must be the case that all positive probabilities in $\mathbf{p}'$ are higher than in $\mathbf{p}$. However, since $\mathbf{p}$ is an equilibrium, $\sum_{k=1}^{K} p_k = 1$. This means that $\sum_{k=1}^{K} p_k' > 1$, contradicting the assumption that $\mathbf{p}'$ is an equilibrium.

REFERENCES

Andersen, Steffen, Glenn W. Harrison, Morten I. Lau, and E. Elisabet Rutström. 2008. "Risk Aversion in Game Shows." In *Risk Aversion in Experiments*, ed. James C. Cox and Glenn W. Harrison, 359–404. Bingley: Emerald.

Bazerman, Max H., Jared R. Curhan, Don A. Moore, and Kathleen L. Valley. 2000. "Negotiation." *Annual Review of Psychology*, 51: 279–314.

Camerer, Colin F. Forthcoming. "The Promise of and Success of Lab Generalizability in Experimental Economics: A Reply to Levitt and List (2007)." In *Methods in Experimental Economics*, ed. Andrew Caplin and Andrew Schotter, Oxford University Press..

Camerer, Colin, and Teck-Hua Ho. 1998. "Experience-Weighted Attraction Learning in Coordination Games: Probability Rules, Heterogeneity, and Time-Variation." *Journal of Mathematical Psychology*, 42(2–3): 305–26.

Camerer, Colin F., and Robin M. Hogarth. 1999. "The Effects of Financial Incentives in Experiments: A Review and Capital-Labor Production Framework." *Journal of Risk and Uncertainty*, 19(1–3): 7–42.

Camerer, Colin F., Teck-Hua Ho, and Juin-Kuan Chong. 2004. "A Cognitive Hierarchy Model of Games." *Quarterly Journal of Economics*, 119(3): 861–98.

Costa-Gomes, Miguel A., and Vincent P. Crawford. 2006. "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study." *American Economic Review*, 96(5): 1737–68.

Costa-Gomes, Miguel, Vincent P. Crawford, and Bruno Broseta. 2001. "Cognition and Behavior in Normal-Form Games: An Experimental Study." *Econometrica*, 69(5): 1193–1235.

Crawford, Vincent P. 2003. "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions." *American Economic Review*, 93(1): 133–49.

Crawford, Vincent P., and Nagore Iriberri. 2007*a*. "Fatal Attraction: Salience, Naïveté, and Sophistication in Experimental 'Hide-and-Seek' Games." *American Economic Review*, 97(5): 1731–50.

Crawford, Vincent P., and Nagore Iriberri. 2007*b*. "Level-k Auctions: Can a Nonequilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?" *Econometrica*, 75(6): 1721–70.

Eichberger, Jürgen, and Dmitri Vinogradov. 2008. "Least Unmatched Price Auctions: A First Approach." University of Heidelberg Department of Economics Discussion Paper 471.

Ellingsen, Tore, and Robert Östling. 2010. "When Does Communication Improve Coordination?" *American Economic Review*, 100(4): 1695–1724.

Erev, Ido, Alvin E. Roth, Robert L. 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.

Falk, Armin, and James J. Heckman. 2009. "Lab Experiments Are a Major Source of Knowledge in the Social Sciences." *Science*, 326(5952): 535–38.

Fischbacher, Urs. 2007. "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, 10(2): 171–78.

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

Friedman, Milton. 1953. *Essays in Positive Economics*. Chicago: University of Chicago Press.

Gallice, Andrea. 2009. "Lowest Unique Bid Auctions with Signals." Collegio Carlo Alberto Working Paper 112.

Goeree, Jacob K., Charles A. Holt, and Thomas R. Palfrey. 2002. "Quantal Response Equilibrium and Overbidding in Private-Value Auctions." *Journal of Economic Theory*, 104(1): 247–72.

Houba, Harold, Dinard van der Laan, and Dirk Veldhuizen. Forthcoming. "Endogenous Entry in Lowest-Unique Sealed-Bid Auctions." *Theory and Decision.*

Levitt, Steven D., and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *Journal of Economic Perspectives*, 21(2): 153–74.

Myerson, Roger B. 1998. "Population Uncertainty and Poisson Games." *International Journal of Game Theory*, 27(3): 375–92.

Myerson, Roger B. 2000. "Large Poisson Games." *Journal of Economic Theory*, 94(1): 7–45.

Nagel, Rosemarie. 1995. "Unraveling in Guessing Games: An Experimental Study." *American Economic Review*, 85(5): 1313–26.

Östling, Robert, Joseph Tao-yi Wang, Eileen Y. Chou, and Colin F. Camerer. 2011. "Testing Game Theory in the Field: Swedish LUPI Lottery Games: Dataset." *American Economic Journal: Microeconomics*. http://www.aeaweb.org/articles.php?doi=10.1257/mic.3.3.1.

Post, Thierry, Martijn J. van den Assem, Guido Baltussen, and Richard H. Thaler. 2008. "Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show." *American Economic Review*, 98(1): 38–71.

Rapoport, Amnon, Hironori Otsubo, Bora Kim, and William E. Stein. 2009. "Unique Bid Auction Games." Jena Economic Research Paper 2009–005.

Raviv, Yaron, and Gabor Virag. 2009. "Gambling by Auctions." *International Journal of Industrial Organization*, 27(3): 369–78.

Rogers, Brian W., Thomas R. Palfrey, and Colin F. Camerer. 2009. "Heterogeneous Quantal Response Equilibrium and Cognitive Hierarchies." *Journal of Economic Theory*, 144(4): 1440–67.

Rosenthal, Robert W. 1973. "A Class of Games Possessing Pure-Strategy Nash Equilibria." *International Journal of Game Theory*, 2(1): 65–67.

Schwarz, Gideon. 1978. "Estimating the Dimension of a Model." *Annals of Statistics*, 6(2): 461–64.

Stahl, Dale O., and Paul W. Wilson. 1995. "On Players' Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior*, 10(1): 218–54.

van Damme, Eric. 1998. "On the State of the Art in Game Theory: An Interview with Robert Aumann." *Games and Economic Behavior*, 24(1–2): 181–210.