A COGNITIVE HIERARCHY MODEL OF GAMES[1]

Colin F. Camerer

Teck-Hua Ho

Juin-Kuan Chong

Abstract

Players in a game are "in equilibrium" if they are rational, and accurately predict other players' strategies. In many experiments, however, players are not in equilibrium. An alternative is "cognitive hierarchy" (CH) theory, where each player assumes his strategy is the most sophisticated. The CH model has inductively-defined strategic categories: Step 0 players randomize; and step k thinkers best-respond, assuming other players are distributed over step 0 through step $k-1$. This model fits empirical data, and explains why equilibrium theory predicts behavior well in some games and poorly in others. An average of 1.5 steps fits data from many games.

[1]February 25, 2004. This research was supported by NSF grants SES-0078911 and SES-0078853. Thanks to C. Mónica Capra, Haitao Cui, Paul Glimcher, and Roger Myerson. Sara Robinson was a helpful wordsmith and James Montier provided the beauty-contest data on consulting clients. Former math nerd Matthew Rabin directed our attention to the golden ratio. Meghana Bhatt, Ming Hsu, Brian Rogers, and Lisa Wang provided excellent research assistance. Useful comments were received from seminars at university of California- Berkeley, California Institute of Technology, University of Chicago, Carnegie Mellon University, Columbia University, Harvard University, New York University, University of Pittsburgh, Stanford University, the Nobel Symposium in Sweden (December 2001), and from four referees and an unusually helpful editor (Edward Glaeser).

I. Introduction

Most theories of behavior in games assume that players think strategically, meaning that they form beliefs by analyzing what others might do, and choose rational responses given their beliefs. But these assumptions, by themselves, cannot precisely predict outcomes. This is because players can act rationally given their beliefs, but have mistaken beliefs about what others will do. Thus, game theorists add the assumption that players are mutually consistent; that is, each player's belief is consistent with what the other players actually do. Taken together, mutual rationality and mutual consistency define equilibrium.

In many real-world games, however, such as the stock market, some players believe, incorrectly and overconfidently, that the other participants are not doing as much thinking as they themselves are. In these situations, the players are not in equilibrium because some players' beliefs are mistaken.

In his book *General Theory of Employment, Interest, and Money*, Keynes likens the stock market to a newspaper contest in which people guess which faces others will judge to be the most beautiful. "It is not the case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree, where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practice the fourth, fifth, and higher degrees" [1936, p. 156].

The essence of Keynes' observation is captured in the "beauty contest" game, in which players are asked to pick numbers from 0 to 100 and the player whose number is closest to $\frac{2}{3}$ of the average wins a prize. Equilibrium theory predicts each contestant will reason as follows: "Even if all the other players guess 100, I should guess no more than $\frac{2}{3}$ times 100, or 67. Assuming the other contestants reason similarly, however, I should guess no more than 45 ..." and so on, finally concluding that the only rational and consistent choice for all the players is zero.

When the beauty contest game is played in experimental settings, the group average is typically between 20 and 35. Apparently, some players are not able to reason their way to the equilibrium value, or they assume that others are unlikely to do so. If the game is played multiple times with the same group, the average moves close to 0.

There are other games where players are surprisingly close to equilibrium, however. Consider a stylized business entry game, in which n players decide simultaneously whether to enter a market with known demand d, where $d < n$. The payoffs are such that players prefer to enter if a total of d or fewer players enter, and they prefer to stay out otherwise. Equilibrium theory predicts the total amount of entry will be very close to the demand d, and this is exactly what happens, even in one-shot experiments where the usual forces that might lead to equilibration (such as learning) haven't yet had a chance to operate.

In this paper we propose an alternative to equilibrium theory– a *cognitive hierarchy* model – that explains empirical behavior in both of these games. In cognitive hierarchy theories, each player believes he understands the game better than the other players. Specifically, CH models posit *decision rules* which reflect an iterated process of strategic thinking [see Binmore, 1988]. The iteration formalizes Selten's [1998, p. 421] intuition that

> the natural way of looking at game situations...is not based on circular concepts, but rather on a step-by-step reasoning procedure.

The CH model consists of iterative decision rules for players doing k steps of thinking, and the *frequency distribution* $f(k)$ (assumed to be Poisson) of step k players. The iterative process begins with "step 0" types who don't assume anything about their opponents and merely choose according to some probability distribution (for simplicity, we assume uniform). "Step k" thinkers assume their opponents are distributed, according to a normalized Poisson distribution, from step 0 to step $k-1$; that is, they accurately predict the relative frequencies of players doing fewer steps of thinking, but ignore the possibility that some players may be doing as much or more. Step 2 players of the beauty contest game, for example, assume the other players are a combination of step 0 players (whose average guess is 50), and step 1 players (who guess $\frac{2}{3}$ times 50.)

A Poisson distribution is described by a single parameter, τ, which is the mean and variance.[2] In 24 beauty contest data sets, the median estimate is $\hat{\tau} = 1.61$. This value explains why the convergence process stops at an average around 30 in the beauty contest game, rather than converging to the equilibrium of zero. Our model, with a similar τ value, also offers an explanation of the "instant equilibration" that occurs in business

[2]Future work could endogenize the distribution $f(k)$ based on tradeoffs between benefits and costs of thinking harder.

entry games. Indeed, values of τ between 1 and 2 explain empirical results for nearly 100 games, suggesting that assuming a τ value of 1.5 could give reliable predictions for many other games as well.

The paper is organized as follows. The next section describes the model. Section III collects some theoretical results. Section IV reports estimation of the τ parameter from six classes of games. Section V explores the "economic value" of CH and other theories. Section VI notes how the Poisson-CH model can help account for two patterns of broad economic interest–speculation and money illusion. Section VII concludes and sketches future research. More details are in our longer paper [Camerer, Ho, and Chong, 2002].

II. The Poisson Cognitive Hierarchy (CH) Model

II.A. Decision Rules

Denote player i's j-th strategy by s_i^j and assume i has finitely-many (m_i) strategies. The iterative rules for our model start with 0-step players, who choose according to a probability distribution which is not derived from strategic thinking. A convenient special case used in this paper, is uniform randomization (a placeholder assumption which could easily be relaxed in later research[3]). Assuming uniform randomization, the 0-step thinkers' choice probabilities are $P_0(s_i^j) = \frac{1}{m_i}\ \forall j$.

Denote a k-step player's belief about the proportion of h-step players by $g_k(h)$. We assume that players doing $k \geq 1$ steps do not realize others are using *more than* k steps

[3] In maximum-likelihood estimation, algorithms find a value of the model free parameter (in this case, τ) which implies predicted probabilities that maximize the product of the predicted likelihoods of all the strategies that are actually chosen by the subjects. If there is any chosen strategy that is predicted to have zero probability, then the product of *all* the likelihoods is zero. This is a problem because one parameter value might yield much more accurate predictions than another parameter value, but if the more accurate model includes a single zero probability the product is zero. Assuming that 0-step thinkers randomize across *all* possible strategies means that the predicted probability of each strategy is at least $f(0)/m_i$ (if there are m_i strategies). Therefore, no chosen strategies have zero predicted probability and the zero-likelihood problem does not arise. Readers more familiar with game theory will appreciate that having all strategies chosen with positive probability also solves two familiar theoretical problems– eliminating noncredible threats (since all threats are "tested" by randomizing 0-step thinkers) as subgame perfection does; and eliminating *ad hoc* rules for Bayesian updating after zero probability events (since all events have probability greater than zero.)

of thinking (that is, $g_k(h) = 0, \forall\ h \geq k+1$). This is plausible because the brain has limits (such as working memory in reasoning through complex games) and also does not always understand its own limits. We also assume that people are overconfident and do not realize there are others using exactly as many thinking steps as they are (i.e., $g_k(k) = 0$). This is consistent with psychological evidence of persistent overconfidence about relative skill in many domains [e.g., Camerer and Lovallo, 1999]. Both assumptions imply, for example, a 1-step player optimizes against perceived random response.[4]

We assume that k-step players have an accurate guess about the relative proportions of players who are doing less thinking than they are. They normalize these actual frequencies to form their beliefs about the competition, so that $g_k(h) = \frac{f(h)}{\sum_{l=0}^{k-1} f(l)}, \forall\ h < k$. This specification exhibits "increasingly rational expectations": As k increases, the total absolute deviation between the actual frequencies $f(h)$ and the beliefs $g_k(h)$ shrinks. This algebraic property implies that as k grows large, players doing k and $k+1$ steps of thinking will, in the limit, have the same beliefs, make the same choices, and have the same expected payoffs. Thus, for a large k, there is no marginal benefit for a k-step player to think harder. This could prove useful for establishing limited thinking through some kind of cost-benefit analysis [see for example, Gabaix and Laibson, 2000; Gabaix et al., 2003; Chen, Iyer and Pazgal, 2003].

Denote another player $-i$'s strategy by $s_{-i}^{j'}$, and player i's payoffs from choosing s_i^j when the other player chooses s_{-i}^j by $\pi_i(s_i^j, s_{-i}^{j'})$. Given the k-step thinker's beliefs, the expected payoff to a $k-$step thinker from choosing strategy s_i^j is $E_k(\pi_i(s_i^j)) = \sum_{j'=1}^{m_{-i}} \pi_i(s_i^j, s_{-i}^{j'}) \{\sum_{h=0}^{k-1} g_k(h) \cdot P_h(s_{-i}^{j'})\}$. For simplicity, we assume players best-respond ($P_k(s_i^*) = 1$ iff $s_i^* = argmax_{s_i^j} E_k(\pi_i(s_i^j))$, and randomize equally if two or more strategies have identical expected payoffs.[5] Given a specific distribution $f(k)$, the model can be solved recursively, starting with 0-step player behavior and iterating to compute $P_1(s_i^j), P_2(s_i^j), \cdots$. (In practice, we truncate the recursion at a k large enough that

[4]This is the familiar "principle of insufficient reason" of Laplace [see Banks, Camerer and Porter, 1994; Haruvy and Stahl, 1998].

[5]Allowing stochastic response gives rise to other models (and could easily be incorporated into CH). In quantal response equilibrium [Rosenthal, 1989; McKelvey and Palfrey, 1995, 1998], players' beliefs and choices of other players are consistent, but are stochastic. The Poisson-CH model retains best-response (except for 0-step thinkers) and weakens equilibrium (i.e., belief-choice consistency). QRE retains equilibrium but weakens best-response. Weiszacker [2003] allows the degree of stochastic response λ_i and beliefs about stochastic response of others $\hat{\lambda}_j$ to be different ($\lambda_i = \infty$ and $\hat{\lambda}_j = 0$ is 1-step thinking). Capra's "thinking tree" [1999] uses one parameter which simultaneously weakens best-response and equilibrium.

the remaining frequencies, $f(k')$ for $k' > k$, are tiny.)

Our Poisson-CH model has some distinct advantages over alternative CH models, such as making $g_k(k-1) = 1$– that is, k-step players think *all others* do only $k-1$ steps of thinking [see Nagel, 1995; Stahl and Wilson, 1995; Ho, Camerer and Weigelt, 1998; Costa-Gomes, Crawford and Broseta, 2001; Costa-Gomes and Crawford, 2004]. This alternative model fits data about as well as our specification but exhibits increasingly *ir*rational expectations– i.e., $g_k(h)$ gets further from the true $f(h)$ as k grows, rather than closer– and makes implausible predictions in some games.[6] Assuming players respond stochastically instead of best-responding is obviously a plausible alternative too, but requires an extra parameter and generally improves fit only a little in most games we have studied.

Another possibility is to assume that k-step players realize there are other $k-$step thinkers ($g_k(k) > 0$). Self-awareness of this sort is a step away from the goal of creating a precise disequilibrium theory, because it imposes consistency of beliefs and choices *within* each group of same-step thinkers. Such a model is also more difficult to solve since finding a solution requires finding a fixed point at each step of thinking. Moreover, when evaluated relative to the $g_k(k) = 0$ specification for five data sets (in our working paper) we show that assuming $g_k(k) = 0$ achieves a better fit.

II.B. The Distribution $f(k)$

One way of getting a rough idea of a natural distribution of thinking steps, $f(k)$, is to let $f(0), f(1), \cdots f(k)$ be free parameters up to some reasonable k, then use data to estimate each $f(k)$ separately using maximum likelihood [cf. Stahl and Wilson, 1995; Ho, Camerer and Weigelt, 1998; Bosch-Domenech et al., 2002]. Estimation of this sort reveals substantial frequencies of levels 0-2 (see our working paper), and medians of 1-2.[7]

[6] In the entry game described in more detail in section III.C below, the $g_k(k-1) = 1$ specification leads to cycles in which $e(i, d)$ entry functions predict entry for $d < .5$ and staying out for $d > .5$ for i odd, and the opposite pattern for i even. Averaging across these entry functions gives a step function $E(k, d)$ with a single step at $d = .5$. This is too simple because entry frequencies are smoothly monotonic across d rather than jumping at $d = .5$.

[7] This fact is consistent with the last part of Keynes' passage on newspaper beauty contests and the stock market: "there are some, I believe, who practise the fourth, fifth and higher degrees [of reasoning about reasoning]." Keynes's wording suggests he believed that few investors do more than three steps, an intuition corroborated by the experimental estimates fifty years later.

To determine a precise parametric distribution $f(k)$, we first outline a list of properties such a distribution should have: Since the thinking steps are integers, a discrete distribution of $f(k)$ is natural [see also Stahl, 1998]. The decision rules described above also require more and more steps of computation as k rises, because a $k-$step thinker does all the computations the lower-step thinkers do, then combines the results to calculate her own expected payoffs. If this process is sharply constrained by working memory, it is plausible that as k rises, fewer and fewer players do the next step of thinking beyond k.[8] A reduced-form way to express this constraint is that the relative proportion $f(k)/f(k-1)$ declines with k. If this decline is captured by assuming $f(k)/f(k-1)$ is proportional to $1/k$, then $f(k)$ is the Poisson distribution, $f(k) = e^{-\tau}\tau^k/k!$, which is characterized by one parameter τ (both its mean and variance). Values of τ can be further pinned down by restrictions on particular $f(k)$ values (e.g., if $k = 1$ is the mode then $\tau \in (1, 2)$).[9]

We focus on the one-parameter Poisson distribution for $f(k)$ because the simpler one-parameter Poisson form fits almost as well as a 7-parameter model (with frequencies $f(k)$ up to $k = 7$)– allowing each $f(k)$ to be independent results in less than a 1 percent decrease in log likelihood– in four of the five data sets we examined. The Poisson model is also easier to compute and estimate, and easier to work with theoretically (see section III).

[8]Devetag and Warglien [2003] report evidence that working memory plays a role in strategic thinking. They measure the amount of working memory using a classic “digit span” task (i.e., how many digits a person can remember from a long string). Working memory is modestly correlated with the tendency to eliminate iteratedly-dominated strategies. Other evidence of limited thinking is reported by Hedden and Zhang [2002].

[9]If $f(1)$ is maximized compared to the neighboring frequencies $f(0)$ and $f(2)$, or if 0- and 2-step thinking are equally common, then $\tau = \sqrt{2}$. If $f(0) + f(1) = 2f(2)$ then $\tau = \frac{\sqrt{5}+1}{2} \approx 1.618$, a remarkable constant known as the “golden ratio” [see Livio, 2002]. The golden ratio is the limit of the ratios of adjacent numbers in the Fibbonaci sequence. It is often used in architecture because rectangles with golden ratio proportions are aesthetically pleasing. It also occurs in spiral patterns of seashells and hawks circling their prey. None of these different assumptions (and resulting τ) are more compelling than the others. They just show how an exact τ value can be derived by adding a simple restriction to a one-parameter distribution.

III. Some Theoretical Properties of the Poisson-CH Model

The combination of optimizing decision rules and the one-parameter Poisson structure makes the Poisson-CH model relatively easy to work with theoretically. This section illustrates some of its properties.

III.A. Dominance-Solvable Games

When $f(k)$ is Poisson-distributed, the relative proportions of types one step below and two steps below a k-step thinker, $f(k-1)/f(k-2) = \tau/(k-1)$, puts overwhelming weight on the $k-1$ types if τ is very large (i.e., $k \ll \tau$). In that case, a k-step thinker acts as if almost all others are using $k-1$ steps. This property of the Poisson distribution provides a simple way to link thinking steps to iterated deletion of dominated strategies.[10] First note that 1-step thinkers will never choose weakly-dominated strategies, because those strategies are never best responses to the random strategies of 0-step types.[11] Now assume τ is very large. Then 2-step thinkers act as if they are playing a mixture of (almost) all 1-step thinkers who have deleted weakly-dominated strategies, and a small percentage of 0-step thinkers who are random. These 2-step thinkers will not play strategies which are strictly-dominated *or* strategies which are weakly-dominated after deleting weakly-dominated strategy play by others. This logic can be extended to iteratively eliminate as many dominated strategies as one likes, because k-step thinkers will act as if almost all other thinkers are one step below them (i.e., $g_k(k-1) = 1 - \epsilon$) when k is much smaller than τ.

Another important property of our CH model is if a k-step thinker plays a (pure) equilibrium strategy, then all higher-step thinkers will play that strategy too.[12] This

[10]This means first eliminating dominated strategies, then eliminating any remaining strategies which are dominated (assuming the strategies eliminated in the first round will not be played), and so on. Note that different outcomes may result depending on whether the eliminated strategies are strictly dominated (i.e., always yields a lower payoff than another strategy) or weakly dominated (i.e., never yields a better payoff than another strategy, and sometimes yields a lower payoff).

[11]This property holds even if 0-step thinkers do not randomize uniformly, as long as all their strategy choices have strictly positive probability.

[12]Simple proof: The k-step thinker plays the equilibrium strategy, call it s_e, against a perceived

means that once a type k reaches a pure equilibrium strategy all higher types will play it too.

Thus, as $\tau \to \infty$ the prediction of the Poisson-CH model will converge to any Nash equilibrium which is reached by finitely-many iterated deletions of weakly-dominated strategies. It is *not* generally true, however, that CH converges to Nash as $\tau \to \infty$.

The Poisson-CH model makes an interesting prediction about the beauty contest games that Nash equilibrium does not. In beauty contest games with two or more players, the game is dominance-solvable and has an unique Nash equilibrium. However, the two-person beauty contest game is special because it can be solved by one step of weak dominance. In the two-person game, one player will always be high and one low, and $\frac{2}{3}$ times the average will be closer to the lower player's number. Therefore, rational players want to choose the lowest number possible- 0. In the Poisson-CH model (with any distribution $f(k)$) applied to the two-person game, all players using one or more thinking steps will choose zero (i.e., the data should consist of approximately $1 - f(0)$ players choosing exactly zero). This is not true in the three-player game; a smart player wants to choose a number between the other two numbers if they are sufficiently far apart. In experiments by Grosskopf and Nagel [2001] and new results we report below, there *are* more choices of 0 in two-player games than in three-player games, although not nearly as many as the Poisson-CH model predicts.

III.B. Coordination Games

Many interesting games, and models of the macroeconomy, have multiple equilibria [e.g., Cooper, 1999]. This raises an important question of how players, or an entire economy, can coordinate on an equilibrium, and *which* types of equilibria are most likely to arise. A large game theory literature on "refinements" has struggled with the problem of how to add further mathematical restrictions in order to refine or limit the number of plausible equilibria. The holy grail being sought in this scientific process is a definition that would guarantee existence of a unique refined type of equilibrium. No such definition has been

mixture of types 0 to $k-1$. The $k+1$-step thinker faces a perceived mixture of types 0 to $k-1$ (with relative weight $\sum_{h=0}^{k-1} f(h)/\sum_{h=0}^{k} f(h)$ and type k (with relative weight $f(k)/\sum_{h=0}^{k} f(h)$). But by definition s_e is a best-response to the mixture of types 0 to $k-1$, and a best response to k's play of s_e (since it is a pure equilibrium strategy). By linearity of the expected payoffs, s_e is therefore a best-response to the mixture of types from 0 to $k-1$ and type k.

discovered. Our CH model goes in an opposite direction. Since our CH model allows players to have incorrect beliefs about each other, it can be seen as a *behavioral* refinement which makes a precise prediction about what will happen in coordination games.[13]

Multiple equilibria typically arise *because* of the mutual consistency assumption. For instance, there are many economic situations where there is one equilibrium that is Pareto- or payoff-dominant (i.e., better for everybody) but seems intuitively riskier than another, Pareto-inferior equilibrium [e.g., Cooper, 1999; Camerer, 2003, chapter 7]. Players will enter a Pareto-inferior equilibrium if each correctly believes the others will play the less risky strategy, even though it is not optimal.

A model of this type of situation is the "stag hunt" (or "assurance") games. Table I shows a stag hunt game in which each of n players simultaneously choose either L or H. The group choice is H if everyone chooses H, and L otherwise (i.e., if at least one person picks L). The row player earns 1 if everyone chooses H, earns 0 if she chooses H and the group outcome is L, and earns x ($0 < x < 1$) if she chooses L (and the group choice is therefore L). Everyone choosing H is a Pareto-dominant equilibrium, but reaching it depends on everyone thinking everyone else will choose H. Choosing L is also an equilibrium, but pays less than if players could somehow coordinate on everyone choosing H.

Game theorists have developed concepts to refine intuitions about when the (L,L) or (H,H) equilibria are likely to arise. Our CH model can replicate some of these intuitions and predicts an important effect of group size which has been observed in experiments– namely, that as the group size n increases, the group is more likely to get drawn into the inefficient (L,L) equilibrium [Van Huyck, Battalio and Beil, 1990; Camerer, 2003, chapter 7].

In the two-player stag hunt game, 1-step thinkers choose H if $x < 1/2$ and choose L if $x > 1/2$ because they optimize against 0-step thinkers who randomize. (If $x = 1/2$ then all players randomize equally.) Higher-step thinkers do exactly what the 1-step thinkers do. In the three-player game, however, a 1-step player thinks she is facing two 0-step players who randomize independently; so the chance of at least one L is .75. As a

[13] In an extensive-form game, each equilibrium may require a different level of cognitive effort from the players. Ho and Weigelt [1996] show that players use simplicity as the refinement criterion and choose the equilibrium that needs the least cognitive effort. The multiple equilibria in this paper require the same level of cognitive effort so the simplicity criterion does not help.

result, the 1-step player (and higher-level players) choose H iff $x \leq .25$. Thus, for values $.25 \leq x \leq .5$, the Poisson-CH model predicts mostly H play in 2-player games and mostly L-play in 3-player games (the frequencies of H and L play, respectively, are $1-(f(0))/2$, or 89 percent for $\tau = 1.5$). This is a simple way of expressing the idea that there is more strategic uncertainty in games with more players, and fits the experimental fact that the inefficient (L,L) outcome occurs more often in larger groups.

There is an interesting connection between the "selection principle" of risk-dominance in 2x2 coordination games [Harsanyi and Selten, 1988] and our CH model. In 2-player symmetric coordination games, CH predicts that all players (except half the 0-step thinkers) will choose the risk-dominant equilibrium (see our working paper for details), which matches a wide range of experimental data showing that players tend to choose risk-dominant strategies rather than payoff-dominant ones in these games [Camerer, 2003, chapter 7].

III.C. Market Entry Games

In section IV below we report experimental results from a simple business entry game. In this game, N entrants simultaneously decide whether to enter a market or stay out (denoted 1 and 0, respectively). Denote the market demand by $d < N$ (expressed as a fraction of the number of potential entrants N, so $0 < d < 1$). If d or fewer players enter (i.e., supply is equal to or less than demand), the entrants all earn a payoff of 1. If more than d enter (i.e., supply is greater than demand), the entrants earn zero, while staying out yields a certain payoff of 0.5. For theoretical simplicity, assume there are infinitely many atomistic entrants. (In our empirical estimation we do not make this assumption.) If entrants are atomistic and risk-neutral, they only care about whether the fraction of others entering is above d or not: if the fraction of others entering is below d, they should enter; if it is above, they stay out. Denote the entry function of step k players for demand d by $e(k,d) : d \to [0,1]$; this function maps the demand into a decision to enter (1) or stay out (0). Denote the interim total entry function for all steps up to and including k by $E(k,d) : d \to [0,1]$. The function $E(k,d)$ adds up the entry functions of the types up to and including k, and normalizes (by dividing by $\sum_{h=0}^{k} f(h)$). The prediction of the model about how many players will enter for each value of d is the limiting case $E(\infty, d)$ (which will depend on τ).

Appendix 1 shows that in the Poisson-CH model, a particular thinking-step type k

has a series of cutpoints which prescribe values of d at which the k-step thinker will enter or not (which, naturally, depend on τ). For example, a 1-step thinker will stay out for $d > .5$ and enter for $d < .5$ (and is indifferent when $d = .5$); the entry-function therefore has one cutpoint at .5. The Appendix proves that the cumulative entry function $E(\infty, d)$ will be weakly monotonic in d– markets with bigger demand attract more entrants– iff $1 + 2\tau < e^{\tau}$, or $\tau < 1.25$.[14]

Figure I shows the predicted entry functions for CH players using 0, 1, and 2 levels of reasoning $(e(0, d), e(1, d), e(2, d))$ and the interim cumulative entry function $(E(2, d))$, for $\tau = 1.5$. Note how the entry function $e(2, d)$ of the 2-step type "smoothes" the cumulative entry function $E(2, d)$. The 2-step thinkers only enter when they think they can exploit the fact that too few lower-0- and 1-step thinkers entered (for $\frac{.5}{1+\tau} < d < .5$ and $\frac{.5+\tau}{1+\tau} < d < 1$), and stay out when they think too many 0- and 1-step types entered (for $0 < d < \frac{.5}{1+\tau}$ and $.5 < d < \frac{.5+\tau}{1+\tau}$).

Note that if the entry game were *actually* played sequentially, it is easy to see how perfect equilibration could occur: Exactly d of N players would enter, because all the later entrants would know how many earlier players had entered. (For example, in a subgame perfect equilibrium the first d entrants would enter and all the subsequent entrants would stay out). Because the Poisson-CH model is recursive, the game becomes effectively "pseudo-sequential": Higher-level players act as if they are moving "after" they have observed what other players do, even though they are actually playing simultaneously. The pseudo-sequentiality created by the recursive structure of the CH model approximates the equilibration which would occur if the game were actually played sequentially.

A wide variety of experimental data show that in entry games like these, the entry rate *is* usually remarkably monotonic in demand d even though players do not communicate and have no way to organize their choices to enter with the correct frequency [e.g., Rapoport and Seale, in press; Camerer, 2003, chapter 7]. Remarking on the surprising

[14] When τ is a little greater than 1.25, as it seems to be empirically in many experimental datasets, there is a small non-monotonicity in which there is more entry at values just below $d = .5$ than the rate of entry at higher demands just above $d = .5$. This is because 1-step thinkers never enter when $d < .5$ and always enter when $d > .5$. Their anticipated entry function $e(1, d)$ invites all higher-step thinkers to enter for values just below .5, and to stay out for values just above .5. The cumulative effect of the higher-step thinkers' entry is never overturned as k grows large. The downward-blip in overall entry just below and above $d = .5$ is probably a small effect empirically but also illustrates a sharp, counterintuitive prediction that could be tested in experiments with a large entry pool $N \approx 100$ and values of d bracketed closely around .5, like $d = .49$ and $d = .51$.

similarity between predicted entry rates across values of d and actual entry in pilot experiments he conducted, Daniel Kahneman [1988] wrote that “to a psychologist, it looks like magic”. The Appendix proof shows how the Poisson-CH model can produce entry which is monotonic in d and approximates equilibrium– the ‘magic” which surprised Kahneman. However, players also collectively overenter at low values of d and underenter at high values of d so their behavior is not entirely in equilibrium. The Poisson-CH model also accounts for overentry at low d and underentry at high d, due to the lingering effect of 0-step thinkers who enter half the time regardless of d. The Poisson-CH can therefore explain the magic of approximate equilibration– monotonicity of entry with the demand d– as well as systematic departures from equilibrium observed in the data.

IV. Estimation and Model Comparison

This section estimates values of τ in the Poisson-CH model and compares its fit to Nash equilibrium. Exploring a wide range of games and models is useful in the early stage of a research program. Models which sound appealing (perhaps because they are conventional) may fit surprisingly badly, thus redirects attention to novel ideas. Fitting a wide range of games turns up clues about where models fail and how to improve them.

Since our model is designed to be general, it is particularly important to check its robustness across different types of games and see how regular the best-fitting values of τ are. Once τ is specified, the model's predictions about the distribution of choices can be easily derived by iterating steps of thinking (and bounding the procedure at a high value of k).

IV.A. The Beauty Contest Games

An empirical warm-up example is the beauty contest game described above, in which the player whose number (from 0 to 100) is closest to $\frac{2}{3}$ times the average wins a fixed prize. Table II shows estimates of τ in 24 p-beauty contest games (p is the multiplier, which, so far, has been $\frac{2}{3}$), which were chosen to minimize the (absolute) difference between the predicted and actual mean of chosen numbers (see our working paper). The table is ordered from top to bottom by the mean number chosen. The first seven lines show games in which the equilibrium is not zero; in all the others the equilibrium is zero.

The first four columns describe the game or subject pool, the source, group size, and total sample size. The fifth and sixth columns show the Nash equilibrium and the difference between the equilibrium and the average choice. The middle three columns show the mean, standard deviation, and mode in the data. The mean choices are generally far off from the equilibrium; they choose numbers which are too low when the equilibrium is high (first six rows) and numbers which are too high when the equilibrium is low (lower rows). The rightmost six columns show the estimate of τ from the Poisson-CH model, and the mean, prediction error, standard deviation, and mode predicted by the best-fitting estimate of τ, and the 90 percent confidence interval for τ estimated from a randomized resampling (bootstrap) procedure.

There are several interesting patterns in Table II. The prediction errors of the mean (column 13, "error") are extremely small, less than .6 in all but two cases. This is no surprise since τ is estimated (separately in each row) to minimize this prediction error. The pleasant surprise is that the predicted standard deviations and modes which result from the error-minimizing estimate of τ are also fairly close (across rows, the correlation of the predicted and actual standard deviation is .72) even though τ's were not chosen to match these moments.

The values of τ have a median and mean across rows of 1.30 and 1.61, close to the golden ratio (1.618...) and $\sqrt{2}$ ($\approx$ 1.41) values derived from simple axioms mentioned above. The confidence intervals have a range of about one in samples of reasonable size (50 subjects or more).

Note that nothing in the Poisson-CH model, per se, requires τ to be fixed across games or subject pools, or across details of how games are presented or choices are elicited.[15] Outlying low and high values of τ are instructive about how widely τ might vary, and why. Estimates of τ are quite low (0-.1) for the p-beauty contest game when $p > 1$ and, consequently, the equilibrium is at the upper end of the range of possible choices (rows 1-2). In these games, subjects seem to have trouble realizing they should choose very large numbers when $p > 1$ (though they equilibrate rapidly by learning; see Ho, Camerer and Weigelt [1998]). Low τ's are also estimated among the PCC subjects playing two- and three-player games (rows 8 and 10). High values of τ ($\approx$ 3-5) appear in games where the equilibrium is in the interior, 72, (rows 7-10)– small incremental steps toward the

[15]Our working paper notes that in games where beliefs about the choices of others are elicited (along with one's own choices), the number of thinking steps sometimes shifts upward. Forcing somebody to articulate what they think others will do apparently gets them to think harder.

equilibrium in these games produce high values of τ. High τ values are also estimated in games with an equilibrium of zero when subjects are professional stock market portfolio managers (row 19), Caltech students (row 20), game theorists (row 24), and subjects self-selecting to enter newspaper contests (row 25). The latter subject pools show that in highly analytical and educated subject pools (especially with self-selection) τ can be much higher than in other subject pools.

A sensible intuition is that when stakes are higher, subjects will use more steps of reasoning (and may think others will think harder too). Rows 3 and 6 compare low stakes ($1 per person per period) and high stakes ($4) in games with an interior equilibrium of 72. When stakes are higher τ is estimated to be twice as large (5.01 versus 2.51), which is a clue that some sort of cost-benefit analysis may underlie steps of reasoning.

Notwithstanding these interesting outliers, there is also substantial regularity across very diverse subject pools and payoff levels. About half the samples have confidence intervals which include $\tau = 1.5$. Subsamples of corporate CEOs (row 13), high-functioning 80-year old spouses of memory-impaired patients [Kovalchik et al., in press; row 15], and high school students (row 16) all have τ values from 1.1-1.7.

Since CH-type models are ideally suited to capture limited equilibration in dominance-solvable games like the beauty contest game, it is important to see how well the same model and τ values fit games with different structures. So we fit five other data sets using maximum likelihood estimation (MLE) procedure (see Appendix 2 for a list): Three sets of matrix games with 2-4 strategies (33 games in total); the binary entry game described above with 12 players and demands $d \in \{2, 4, 6, 8, 10\}$, and 22 games with mixed equilibria.

The estimation aims to answer two questions: Is the estimated value of τ reasonably regular across games with very different structures? And how accurate is the Poisson-CH specification compared to Nash equilibrium?

IV.B. How Regular Is τ?

Table III shows game-by-game MLE estimates of τ in the Poisson CH model, and estimates when τ is constrained to be common across games within each data set. The interquartile range across the 60 estimates is (.98,2.21) and the median is 1.55. Five of 60 game-specific τ estimates are high (four or more) and a few are zero.

The Appendix Table A.I shows bootstrapped 95 percent confidence intervals for the τ estimates. Most of the intervals have a range of about one. The common τ estimates are roughly 1-2; a τ of around 1.5 is enclosed in the 90 percent interval in three data sets, and τ seems to be about one in the Cooper-Van Huyck and entry data. These reasonably regular τ's suggests that the Poisson-CH model with $\tau = 1.5$ can be used to reliably predict behaviors in new games.

In future work, variation in estimates of τ could be useful in sharpening a theory of how steps of thinking are chosen endogeneously. While endogenizing thinking steps or τ is beyond the scope of this paper, it is likely that some kind of model comparing perceived benefits of thinking further, with thinking costs (constrained by working memory, and permitting individual differences) will do better. Three pieces of evidence point to the promise of a cost-benefit endogenization: (1) τ is estimated to be quite large in p-beauty contest games in subject pools with unusual analytical skill (e.g., Caltech undergraduates) or special training (game theorists and computer scientists who study multi-agent machine learning), which is a clue that lowering thinking costs due to skill or training leads to higher τ. (2) Unpublished data we have collected also show larger estimates of τ (about .5 steps more) in Caltech undergraduates than in comparable students from a nearby community college (more evidence of skill or lower cognitive cost as an important variable). (3) Unpublished data show that in incomplete-information signaling games, τ is estimated to be lower (less than 1). Bayesian updating on what another player's signal choice reveals about her likely thinking-step type presumably consumes more working memory than simply computing expected payoffs (raising thinking costs), so lower τ's in these games is also consistent with cost-benefit calculus.

IV.C. Which Models Fit Best?

Table IV shows log likelihoods (LL) and mean-squared deviations for several model estimated game-by-game or with common parameters across games in a dataset.[16] This

[16]When the Stahl-Wilson games 2, 6, 8 are included the common τ is zero because these games swamp the other 10, so we excluded those games in doing the common-τ estimation. Poisson-CH fits badly in those games because the Nash strategy is not reached by any number of thinking steps, but is frequently chosen. The best the model can do is to pick $\hat{\tau} = 0$ so that 1/3 of the players are predicted to pick it (since there are three strategies). These games show boundary conditions under which the model fails badly. Modifying the model so that a fraction ϕ of the 0-step players are actually choosing Nash (which will then lead 1-step types to choose Nash if ϕ is large enough) would patch this problem. Including

table answers several questions. Focusing first on the Poisson-CH model, game-specific estimates of τ fit almost as well as common within-column estimates in most data sets (except for the Stahl-Wilson data). The Poisson-CH model also fits substantially better than Nash in every case. This shows that relaxing mutual consistency can be a fruitful approach to building a descriptive theory of disequilibrium behavior in games.

A graphical comparison of how much the theories' predictions deviate from the data gives a quick snapshot of how accurate they are. Each point in Figures II-III represents a distinct strategy in each of the 33 matrix games (Figure II) and 22 mixed games (Figure III). Each point represents the absolute deviation between the Nash prediction and the data (on the x-axis) and the Poisson-CH prediction (using a common τ within each dataset) and the data (on the y-axis). Points in the lower right of the graph represent strategies in which CH is more accurate than Nash; points in the upper left represent strategies in which Nash is more accurate than CH.

The graphs enable us to answer an important question visually: When the Nash predictions are good approximations, is Poisson-CH almost as accurate? The answer appears to be yes, because there are few points with low Nash deviations and high Poisson-CH deviations (i.e., few points in the upper left of the graphs). And when the Nash predictions are poor approximations, are CH predictions usually more accurate? The answer is also Yes. Figure II shows that particularly in the matrix games (where Nash often makes 0-1 pure strategy predictions), there are many strategies in which the Nash prediction is off by more than .50. For these strategies, the Poisson-CH prediction is usually off by less than .20. So Poisson-CH is able to correct the largest mistakes made by the equilibrium prediction. Figure III shows that both models are generally more accurate in mixed games than in the Figure II matrix games, and that Poisson-CH improves only a little on equilibrium.

IV.D. Predicting Across Games

Good theories should predict behavior accurately in new situations. A simple way to see how well Poisson-CH and equilibrium models can do this is to estimate the value of τ on $n-1$ datasets and forecast behavior in each holdout dataset separately. The result of this kind of cross-game estimation is reported in Table V. Across games, the Poisson-CH

some self-awareness would also explain these anomalies.

model fits only a little less accurately than when estimates are common within games. This suggests the Poisson-CH model has some promise for predicting behavior in one set of games, based on observations from other games.

IV.E. Two Examples

Two specific games help illustrate concretely how the Poisson-CH model can explain where Nash predictions succeed and fail. These games were chosen because they have the median likelihood ratio of Poisson-CH relative to Nash within their respective data sets, so they are statistically representative of the overall result and are not biased either for or against the CH and Nash models.

Table VI shows game 8 from Costa-Gomes et al. [2001]. The Nash prediction is pure play of (T,L). Most row players do choose T, but a third of the column players choose R instead of the equilibrium response L. The Poisson-CH model (using the common-$\hat{\tau}$ within the Costa-Gomes et al dataset) predicts 82 percent play T because it is a dominant strategy and so all players using more than 0 steps are predicted to choose it. It also predicts 45 percent of players will choose R because half the 0-step players and all the 1-step players choose it (though players using two or more steps pick L). So CH is able to approximate the accurate Nash prediction about dominant strategy play of T, but corrects Nash theory's mistaken prediction of how often R is played. The key point is that 1-step thinkers play R. Only column players doing two or more steps of thinking figure out that row players will pick T (and respond optimally with L), so the relative infrequency of high-step thinkers explain why R is chosen so often.

Table VII shows game 4 from Binmore et al. [2001]. This game has a mixed equilibrium in which row players are predicted to choose B, and column players are predicted to choose R, both at 67 percent of the time.

The column player prediction is plausible because R pays either 1 or 0, and players actually chose it 83 percent of the time. The Poisson-CH model reproduces this finding very closely because *all* players doing one or more steps of thinking are predicted to choose R, an aggregate frequency of 84 percent. The Nash prediction that row players choose B most often is less plausible because B has no positive payoffs; and in fact, it is the strategy chosen least often. In equilibrium, of course, players are predicted to choose B because they guess correctly that column players often choose R. In the Poisson-CH model, however, 1-step thinkers do not anticipate the play of R and mix between T and M. So the Poisson-CH model predicts 25 percent choice of each T and M, which is closer to what actually happens than the Nash prediction.

These examples illustrate how the Poisson-CH model can mix limited thinking with strategic thinking (through the behavior of players doing two or more steps of thinking), and as a result, generally fit data from one-shot games better than equilibrium models do.[17] Remember that these are statistically *typical* examples; they were not chosen to highlight where Poisson-CH does particularly well or poorly.

V. Economic Value of Theories

In 1960, Schelling [1960, p. 98] wrote, "A normative theory must produce strategies that are at least as good as what people can do without them". Schelling's definition suggests a simple measure of the value of an economic theory when applied to a particular game: How much greater a payoff do players earn when they best-respond to a theory's forecast rather than responding naively?

A reasonable way to measure the "economic value" of a theory applied to a particular game, is to take a set of experimental data and compute the difference between the expected payoff from using the best response given by the theory, and the average payoff subjects actually earned [see Camerer and Ho, 2001].

If a player's beliefs and the choices of others players are mutually consistent, then equilibrium theory predicts the game exactly. Thus, for a game where the players are in equilibrium, the economic value of equilibrium theory will be 0. On the other hand, if players are in equilibrium then models which assume they are *not* in equilibrium (such as the Poisson-CH model) will have negative economic value. Thus, the economic value of equilibrium for a game is a way of measuring the degree of equilibration.

Furthermore, if the Poisson-CH model is correct for a given game, then the best response the theory dictates corresponds to what the highest-step thinkers do. Thus, the economic value of Poisson-CH can be interpreted as the marginal payoff to using many steps of thinking, compared to average payoffs. If the economic value is low– i.e., the

[17] A web-based calculator is available to provide Poisson-CH prediction and τ estimate to fit data from one-shot games. The calculator is located at http://groups.haas.berkeley.edu/simulations/CH/.

marginal payoff to thinking very hard is low– this fact could be used as a justification for an evolutionary or "cognitive economics" explanation of why more players do not think harder, which could potentially endogenize the limits of thinking.

Table VIII reports the economic value of the Poisson-CH and Nash models across several data sets. The economic value of Poisson-CH is derived using parameters estimated on $n-1$ data sets to forecast the remaining data set.[18] The payoffs from predicting "clairvoyantly" (i.e., using the actual distribution of strategies chosen by all other subjects), are also reported because these represent an upper bound on economic value.

The Poisson-CH approach adds value in all data sets, from 20 to 70 percent of the maximum possible economic value. Nash equilibrium typically adds economic value, although only about half as much as Poisson-CH, and subtracts value in one data set. Recall that if players were in equilibrium, the Nash predictions would have zero economic value and disequilibrium models like CH would have negative economic value. The fact that this pattern is not observed is another way of saying players are not in equilibrium, and economic value measures the "degree" of disequilibrium.

VI. Economic Implications of Limited Strategic Thinking

Models of iterated thinking can be applied to several interesting problems in economics, including asset pricing, speculation, competition neglect in business entry, incentive contracts, and macroeconomics (see our longer paper for some ideas along these lines). An example is Crawford [2003]'s model of optimal lying. He shows that if "some of the people can be fooled some of the time", the presence of such foolable nonstrategic types

[18]Economic value is only slightly higher when parameters are estimated within each data set (see our working paper). One could argue that economic value measured using within-game τ is upward biased, because the model effectively has access to data about the particular game which the typical subject does not. The cross-game estimation below does not have this upward bias. It is true that the cross-game Poisson-CH forecast uses other data the subject did not see– namely, the behavior in the other four data sets– but this is typically the case in forecasting (after all, distillation of data is part of what people pay for when they buy forecasts). Furthermore, the subjects have "data" (or insight) which the model doesn't have– namely, how people like themselves and their fellow subjects might react to a particular game, and how they may have behaved in dozens of other experiments they participated in.

influences rational players to misrepresent their intentions (much as the presence of 0-step players influence the behavior of higher-step thinkers). Similarly, Cai and Wang [2003] explain the experimental tendency for players to overcommunicate private information with a model of limited thinking.

We illustrate further with two economic applications that have been studied experimentally: Speculation, and money illusion. The idea is to see whether the Poisson-CH model can help us understand something fundamental about economics.

VI.A. Speculation

In 1982, Milgrom and Stokey proved a remarkable "Groucho Marx theorem": If rationality is common knowledge, risk-averse players should not make speculative bets with one another (unless they have hedging motives). Of course, speculation goes on constantly, in the form of sports betting and a large fraction of trading in financial markets and other forums. It is difficult to know from field data which assumption of the Groucho Marx Theorem is violated: Is widespread speculation due to hedging (undoubtedly an important part of the operation of foreign exchange and futures markets)? Or to the extra fun from watching a sports event after betting on it? Or is speculation due to limits on knowledge of rationality? Since the Poisson-CH model does not impose common knowledge of rationality, it contradicts the Groucho Marx theorem and predicts that speculation will occur, even when hedging and spectator fun don't matter. In CH, some degree of betting comes immediately from the fact that 0-step and 1-step players are not thinking strategically about how the betting propensities of others depends on what those other players know [cf. Eyster and Rabin, 2002].

Table IX illustrates a betting game originally studied experimentally by Sonsino, Erev and Gilat [2002] and replicated by Sovik [2000]. There are four equally-likely states, {A,B,C,D}. After the state is determined, players are privately informed about a set of possible states including the true one. Player I either learns the state is A or B (denoted (A,B)) or C or D (i.e., (C,D)). Player II learns the state with certainty if it is A or D, or learns (B,C). Players then choose whether to bet, with payoffs given in Table IX. If both players bet, they win or lose the amounts in the table column corresponding to the true state.

In equilibrium, there should be no mutual betting if rationality is common knowledge and players think strategically. The proof begins with the fact that rational player II's will never bet when they know the state is A, and will always bet when they know the state is D. If player I is rational and believes player II's are too, she will figure out that she can never win by betting when her information is (A,B). Iterating further, player II should never bet in (B,C), which leads player I to not bet in information set (C,D). Therefore, players should never mutually bet.

This counterintuitive no-betting result is also the prediction of the Poisson-CH model with $\tau \to \infty$. With the typical empirical value of $\tau = 1.5$, however, a different picture emerges. One-step player I's will bet when they know (A,B) because they think they are equally likely to win 32 and lose 28 (they have not figured out that they will never win 32 because rational player I's never bet in state A). However, two-step thinkers know that one-step player II's won't bet in A, so they will not bet in (A,B). The eventual result of this iterated limited rationality across all information states is high betting rates in the information states (A,B) and (C,D) for player I, and in (B,C) for player II.

Table IX shows predicted betting rates (for $\tau = 1.5$) and Sovik [2000]'s first-round data.[19] The model does not track differences in betting rates across the three ambiguous information states particularly well, but it is an obvious improvement on the Nash prediction of no mutual betting in any states. Furthermore, the model makes testable comparative static predictions: For example, if the payoff in the C state is changed from 20 to 32, betting rates in (B,C) and (C,D) should fall dramatically (from 72 percent and 89 percent to 12 percent and 46 percent). This example shows how a small change in parameters can turn the predicted CH result from a gross violation of the Groucho Marx Theorem to a reasonable approximation of it. More generally, the example shows how CH captures partial awareness of adverse selection, which may be useful in understanding consumer product markets, the winner's curse in common-value auctions, and so forth.

VI.B. Money Illusion

A long-running debate in macroeconomics concerns the extent of "money illusion", the failure to adjust incomes and prices for inflation. Fehr and Tyran [2002] investigate

[19] The results of Sonsino, Erev and Gilat [2002] are even more consistent with CH, except that they report substantial betting and nonbetting rates by player II's in states A and D (about 20 percent) which suggests $\tau \approx 1$.

money illusion in two parallel pricing games. In their games, groups of four players choose integer prices from 1 to 30. In one game, prices are strategic substitutes– players earn more by pricing high when others price low, and vice versa. In the other game, prices are strategic complements– players earn more by matching prices of other players. Each player's nominal payoffs depend on his or her own price and on the (rounded) average price of the other three players in his or her group. These nominal payoffs are displayed in a 30 x 30 table with the player's own prices in 30 rows and the average price of others in the 30 columns. To compute their real payoffs, players had to divide the nominal numbers in the table they see in front of them by the average price of the other players in the group (which is clearly shown at the top of each column).

The research question is whether players use the nominal payoff, or act as if they calculated real payoffs. If players use real payoffs and are in equilibrium, they will either choose prices of 11 and 14 (depending on which of two cost structures, denoted x or y, they have) in both the substitutes and complements conditions. (This is also the prediction of CH with infinite τ.)

Fehr and Tyran found a striking regularity: In the substitutes condition, players choose prices very close to the Nash predictions of 11 and 14. But in the complements condition, prices were far from equilibrium, a median of 22-23. This pattern leaves unresolved whether players have money illusion or not. It appears that their money illusion depends on the strategic structure of the game, which is an unsatisfying conclusion.

The Poisson-CH model with the typical value $\tau = 1.5$ can account for this pattern (with one important modification[20]) reasonably well. The key is that in the complements case, 1-step thinkers have best responses which are prices in the 20's, well above the Nash equilibria, and 2- and higher-step thinkers also choose prices which are too high.

Table X shows summary statistics of predicted and actual prices and the Poisson-CH predictions for $\tau = 1.5$ (which is not the best-fitting value[21]). Median experimental prices

[20]The Poisson-CH model fits best if we assume the 1-step thinkers believe the 0's choices are perfectly correlated– i.e., the distribution of the *average* price is uniform. A similar improvement in fit from assuming that lower level types' choices are correlated occurs in p-beauty contests [see Ho, Camerer and Weigelt, 1998] and weak link games [Camerer, 2003, chapter 7]. The tendency for players to think that a single 0-step player's price distribution is a representative or exemplar of the average price is an example of the "representativeness" heuristic, which is well-documented in research on the psychology of judgment [e.g., Kahneman 2003].

[21]The best-fitting values are different in the two conditions, .6 in complements and 2.6 in substitutes.

are predicted exactly in three of four samples. Repeating the main empirical theme of this paper, the model can explain when Nash equilibrium is reached surprisingly quickly (in the substitutes treatment), and can also explain when behavior is far from equilibrium (in the complements treatment). More importantly, the model provides a clear answer to the central research question of whether money illusion occurs. The answer is that players *do* appear to have money illusion in both treatments, but differences in equilibration in both cases may result from a common process of limited strategic thinking.[22]

VII. Conclusion

This paper introduced a simple cognitive hierarchy (CH) model of games. The model is designed to be as general and precise as Nash equilibrium. In fact, it predicts players are unlikely to play Nash strategies which are refined away by subgame or trembling-hand perfection, and always selects one statistical distribution when there are multiple Nash equilibria, so it is even more precise than simple Nash equilibrium.

This paper uses both axioms and estimation to restrict the frequencies of players who stop thinking at various levels. Most players do *some* strategic thinking, but the amount of strategic thinking is sharply constrained by working memory. This is consistent with a Poisson distribution of thinking steps that can be characterized by one parameter τ (the mean number of thinking steps, and the variance). Plausible restrictions and estimates from many experimental data sets suggest that the mean amount of thinking τ is between one and two. The value $\tau = 1.5$ is a good omnibus guess which makes the Poisson-CH theory parameter-free and is very likely to predict as accurately as Nash equilibrium, or more accurately, in one-shot games.

The main contribution is showing that the same model can explain limited equilibration in dominance-solvable games (like p-beauty contests) *and* the surprising accuracy of Nash equilibrium in some one-shot games, such as simultaneous binary entry games in which players choose whether to enter a market with a fixed demand. In one-shot games

Fehr and Tyran note that this difference is consistent with the fact that in the substitutes case, misestimating what others do is a much more costly error than in the complements case. This is another clue for how the number of thinking steps may respond endogeneously to incentives.

[22]When the Poisson-CH model is applied to players with no money illusion, it fits the data poorly (since it predicts Nash-like play in the complements case). It also fits poorly if 2- and higher-step players do not have money illusion but lower-step players do.

with no communication, the rate of entry in these entry games is 'magically' monotonic in the demand d, but there is reliable over-entry at low values of d and under-entry at high values of d. The Poisson-CH approach predicts monotonicity (it is guaranteed when $\tau \leq 1.25$) and also explains over- and under-entry. Furthermore, the Poisson-CH approach creates a kind of endogenous purification that explain how a population mixture of players who use pure strategies (and perhaps regard mixing as nonsensical) can approximate a mixed equilibrium.

Because players do not appear to be mutually consistent in one-shot games where there is no opportunity to learn, a theory of how others are likely to play could have economic value– i.e, players could earn more if they used the model to recommend choices, compared to how much they actually earn. In fact, economic value is always positive for the Poisson-CH model, whether τ is estimated within a data set or across data sets. Economic value is 10-50 percent of the maximum possible economic value that could be achieved by knowing in advance the sample frequencies of how others actually play. The Nash approach adds less economic value, and sometimes subtracts economic value (e.g., in p-beauty contests with $p < 1$ players are better choosing on their own than picking the Nash recommendation of 0). Economic value provides a precise measure of Schelling's [1960] definition of how "normative" a theory is, and also measures the degree of disequilibrium in economic terms.

There are many challenges in future research. An obvious one is to endogenize the mean number of thinking steps τ, presumably from some kind of cost-benefit analysis in which players weigh the marginal benefits of thinking further against cognitive constraint. In a cost-benefit approach, the fact that beliefs (and hence, choices) converge as the number of steps rises creates diminishing marginal benefits which leads to a natural truncation which limits the amount of thinking.

Since the Poisson-CH model makes a prediction about the kinds of algorithms that players use in thinking about games, cognitive data other than choices– like prompting players to state beliefs (which might shift 0-step thinkers to one or more steps), information lookups, or brain imaging [e.g., Camerer et al., in press]– can be used to test the model. For example, Rubinstein [2003] reports response times in large web-based experiments that are consistent with slower reponses by players who use more thinking steps.

The model is easily adapted to incomplete information games because the 0-step players make choices that reach every information set. This eliminates the need to impose delicate refinements to make predictions. Explaining behavior in signaling games and other extensive-form games with incomplete information is therefore workable and a high priority. Extending the model to extensive-form games is easy by assuming that 0-step thinkers randomize independently at each information set, and higher-level types choose best responses at information sets using backward induction. Other models which link limited thinking about other players to limited look-ahead in extensive-form games could prove more interesting.

Finally, the ultimate goal of the laboratory honing of simple models is to explain behavior in the economy. Models of iterated thinking could prove useful in thinking about asset markets, speculation and betting, contract structure, and other phenomena [cf. Eyster and Rabin, 2002].

Appendix 1: Entry Game Analysis for Poisson-CH

We have:

$$\begin{aligned} e(0,d) &= \frac{1}{2}, \qquad \forall d \\ E(k,d) &= \frac{\sum_{j=0}^{k} f(j)\cdot e(j,d)}{\sum_{j=0}^{k} f(j)} = \frac{\sum_{j=0}^{k} f(j)\cdot e(j,d)}{F(j)}, \text{where } F(j) \equiv \sum_{j=0}^{k} f(j) \end{aligned}$$

In general, for $k \geq 1$

$$e(k,d) = \begin{cases} 0 & \text{if } E(k-1,d) > d \\ 1 & \text{if } E(k-1,d) < d \end{cases}$$

In general, $E(k,d)$ is a step function with the following cutpoint values (at which steps begin or end) with increasing d for $d < 1/2$
$\frac{\frac{1}{d}f(0)}{F(k)}, \frac{\frac{1}{2}f(0)+f(k)}{F(k)}, \frac{\frac{1}{2}f(0)+f(k-1)}{F(k)}, \frac{\frac{1}{2}f(0)+f(k-1)+f(k)}{F(k)}, \ldots \frac{\frac{1}{2}f(0)+f(2)+\cdots+f(k)}{F(k)}$

The cutpoint values for $d > 1/2$ are $\frac{\frac{1}{2}f(0)+f(1)}{F(k)}, \frac{\frac{1}{2}f(0)+f(1)+f(2)}{F(k)}, \ldots \frac{\frac{1}{2}f(0)+f(1)+f(2)+\cdots+f(k)}{F(k)}$
(For $d = 1/2$ atomistic entrants are all indifferent and randomize so $E(k,.5) = .5\ \forall k$.)

These cutpoints imply two properties: The cutpoints are always (weakly) monotonically increasing in d for the $d < 1/2$ segment as long as $f(k-1) > f(k)$, $\forall k \geq 2$.

For a Poisson $f(k)$, this is equivalent to $\tau \leq 2$. Furthermore, the last cutpoint for the $d < 1/2$ segment is smaller than the first cutpoint of the $d > 1/2$ segment iff $\frac{1}{2}f(0) + f(2) + f(3) + \cdots + f(k-1) + f(k) \leq \frac{1}{2}f(0) + f(1)$. This is equivalent to $f(1) \geq f(2) + f(3) + \cdots + f(k)$, which implies $f(1) \geq 1 - f(0) - f(1)$. For Poisson this implies $(1 + 2\tau) \geq e^{\tau}$ or $\tau \leq 1.25$. Thus, $\tau \leq 1.25$ implies weak monotonicity throughout both the left ($d < 1/2$) and right ($d > 1/2$) segments of the entry function $E(k, d)$ (since $\tau < 1.25$ satisfies the $\tau < 2$ condition *and* ensures monotonicity across the crossover from the left to right halves of $E(k, d)$).

Appendix 2: Details of Games and Experimental Methods

The matrix games are 12 games from Stahl and Wilson [1995], 8 games from Cooper and Van Huyck [in press] (used to compare normal- and extensive-form play), and 13 games from Costa-Gomes, Crawford and Broseta [2001]. All these games were played only once with feedback, with sample sizes large enough to permit reliable estimation.

In our experiments, subjects played five entry games with no feedback. If the number of entrants was above (less than or equal to) d the entrants earned 0 ($1); nonentrants earned $.50. Subjects also played 22 matrix games with mixed equilibria. The 22 games with mixed-equilibria are taken from those reviewed by Camerer [2003, chapter 3], with payoffs rescaled so subjects win or lose about $1 in each game. The 22 mixed games are (in order of presentation to the subjects): Ochs [1995], (matching pennies plus games 1-3); Bloomfield [1994]; Binmore et al. [2001] Game 4; Rapoport and Amaldoss [2000]; Binmore et al. [2001], games 1-3; Tang [2001], games 1-3; Goeree, Holt, and Palfrey [2000], games 2-3; Mookerjee and Sopher [1997], games 1-2; Rapoport and Boebel [1992]; Messick [1967]; Lieberman [1962]; O'Neill [1987]; Goeree, Holt, and Palfrey [2000], game 1. Four games were perturbed from the original payoffs: The row upper left payoff in Ochs's original game 1 was changed to 2; the Rapoport and Amaldoss [2000] game was computed for r=15; the middle row payoff in Binmore et al [2001] game 2 was 30 rather than -30; and the lower left row payoff in Goeree, Holt and Palfrey's [2000] game 3 was 16 rather than 37. Original payoffs in games were multiplied by the following conversion factors: 10, 10, 10, 10, 0.5, 10, 5, 10, 10, 10,1,1,1,0.25,0.1,30,30,30,5,3,10,0.25. Currency units were then equal to $.10.

The entry and mixed-equilibrium games were run in four experimental sessions of 12 subjects each. Each game was played (with no feedback) against a random opponent in the same session and earnings accumulated. Two sessions used undergraduates from Caltech and two used undergraduates from Pasadena City College (PCC), which is near Caltech. (Individual-level estimation in progress suggests the PCC subjects do about .5 steps of thinking fewer than Caltech students, men do about .2 steps more thinking than women, and the average number of thinking steps drifts up by about .7 and the first and second halves of 22 games.) Mixed equilibrium games were run on the "playing in the dark" software developed by McKelvey and Palfrey. The entry games and some beauty contest games were run on software Taizan Chan wrote, which is available from us.

California Institute of Technology
Haas School of Business, University of California at Berkeley
NUS Business School, National University of Singapore

References

Banks, Jeffrey, Colin F. Camerer, and David Porter, "Experimental Tests of Nash Refinements in Signaling Games," *Games and Economic Behavior*, VI (1994), 1-31.

Binmore, Kenneth, "Modeling rational players: Part II," *Economics and Philosophy*, IV (1988), 9-55.

—–, Joe Swierzbinski and Chris Proulx, "Does Maximin Work? An Experimental Study," *Economic Journal*, CXI (2001), 445-464.

Bloomfield, Robert, "Learning a Mixed Strategy Equilibrium in the Laboratory," *Journal of Economic Behavior and Organization*, XXV (1994), 411-436.

Bosch-Domenech, Antoni, Jose G Montalvo, Rosemarie Nagel and Albert Satorra, "One, Two, (Three), Infinity,...: Newspaper and Lab Beauty-contest Experiments," *American Economic Review*, XCII(5) (2002), 1687-1701.

Cai, Hongbin and Joseph Wang, “Over-communication and bounded rationality in strategic information transmission games: An experimental investigation,” UCLA Department of Economics, 2003. http://econweb.sscnet.ucla.edu/

Camerer, Colin F., *Behavioral Game Theory: Experiments on Strategic Interaction.* (Princeton, NJ: Princeton University Press, 2003).

—— and Teck-Hua Ho, “Strategic Learning and Teaching,” in S.J. Hoch and H.C. Kunreuther (Eds.), *Wharton on Making Decisions* (New York: John Wiley and Sons, 2001), 159-176.

——, —— and Juin-Kuan Chong, “A Cognitive Hierarchy Theory of One-shot Games: Some Preliminary Results,” CalTech working paper, 2002. http://hss.caltech.edu/ camerer

——, and Daniel Lovallo, “Overconfidence and Excess Entry: An Experimental Approach,” *American Economic Review*, LXXXIX (1999), 306-18.

——, Drazen Prelec and George Loewenstein, “Neuroeconomics,” *Scandinavian Journal of Economics*, in press. http://www.hss.caltech.edu/ camerer/webmaterial/ scanjecon61.pdf

Capra, C. Mónica, *Noisy Expectation Formation in One-shot Games*, Ph.D. thesis, 1999.

Chen, Yuxin, Ganesh Iyer and Amit Pazgal, “Limited Memory and Market Competition,” Working Paper, University of California, Berkeley, 2003.

Cooper, David and J. Van Huyck, “Evidence on the Equivalence of the Strategic and Extensive Form Representation of Games,” *Games and Economic Behavior*, in press.

Cooper, Russell W., *Coordination Games: Complementarities and Macroeconomics.* (Cambridge, U.K.: Cambridge University Press, 1999).

Costa-Gomes, M. and V. Crawford, ”Cognition and Behavior in Two-Person

Guessing Games: An Experimental Study," UCSD Working paper, 2004. http://weber.ucsd.edu/ vcrawfor/Guess .

—–, —– and B. Broseta, "Cognition and Behavior in Normal-form Games: An Experimental Study," *Econometrica*, LXIX (2001), 1193-1235.

Crawford, Vince, "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions", *American Economic Review*, XCIII, (2003), 133-149.

Devetag, G. and M. Warglien, "Games and phone numbers: Do short-term memory bounds affect strategic behavior?" *Journal of Economic Psychology*, XXIV(2), (2003), 189.

Eyster, E. and M. Rabin, "Cursed equilibrium," UC-Berkeley working paper, 2002.

Fehr, E. and J.-R. Tyran, "Limited rationality and strategic interaction: The impact of the strategic environment on nominal intertia," Institute for Empirical Research in Economics No. 130, University of Zurich, 2002. http://www.iew.unizh.ch/wp/iewwp130.pdf

Gabaix, Xavier and David Laibson, "A boundedly rational decision algorithm," *American Economic Review*, XC (May 2000), 433-438.

—–, —–, Guillermo Moloche, and Stephen Weinberg. "The allocation of attention: Theory and evidence," MIT working paper, August 2003.

Goeree, J., C. Holt and T. Palfrey, "Risk Aversion in Games with Mixed Strategies," University of Virginia Department of Economics, May 2000.

Grosskopf, Brit and Rosemarie Nagel, "Rational Reasoning or Adaptive Behavior? Evidence from Two-Person Beauty Contest Games," Harvard NOM Research Paper No. 01-09, 2001.

Harsanyi, John C., and Reinhard Selten, *A General Theory of Equilibrium*

Selection in Games. (Cambridge, MA: MIT Press, 1988).

Haruvy, E. and D. Stahl, “An Empirical Model of Equilibrium Selection in Symmetric Normal-form Games,” University of Texas Department of Economics Working Paper, 1998.

Hedden, T. and J. Zhang, “What do you think I think you think?: Strategic reasoning in matrix games,” *Cognition*, LXXXV (2002), 1-36.

Ho, Teck-Hua, Colin Camerer, and Keith Weigelt, “Iterated Dominance and Iterated Best-response in p-Beauty Contests,” *American Economic Review*, LXXXVIII (1998), 947-969.

— and Keith Weigelt, “Task Complexity, Equilibrium Selection, and Learning: An Experimental Study,” *Management Science*, XVII (1996), 659-679.

Kahneman, D., “Experimental Economics: A Psychological Perspective,” in R. Tietz, W. Albers, and R. Selten (Eds.) *Bounded Rational Behavior in Experimental Games and Markets*, (New York: Springer-Verlag, 1988), 11-18.

—, ”Maps of bounded rationality: Psychology for behavioral economics,” *American Economic Review*, LXXXXIII (2003), 1449-1475.

Keynes, John Maynard, *The general theory of interest, employment and money.* (London: Macmillian, 1936).

Kovalchik, Stephanie, Colin F. Camerer, David M. Grether, Charles R. Plott and John M. Allman, ”Aging and Decision Making: A broad comparative study of decision behavior in neurologically healthy elderly and young individuals,” *Journal of Economic Behavior and Organization*, in press.

Lieberman, Bernhardt, “Experimental Studies of Conflict in Some Two-person and Three-person Games,” in J. H. Criswell, H. Solomon, and P. Suppes (Eds.), *Mathematical Models in Small Group Processes.* (Stanford, CA: Stanford University Press, 1962), 203-220.

Livio, Mario, *The Golden Ratio*, (Broadway Books, 2002).

McKelvey, Richard D. and Thomas R. Palfrey, “Quantal Response Equilibria for Normal-form Games,” *Games and Economic Behavior*, VII (1995), 6-38.

—— and ——, “Quantal Response Equilibria for Extensive Form Games,” *Experimental Economics*, I (1998), 9-41.

Messick, David M., “Interdependent Decision Strategies in Zero-sum Games: A Computer-controlled Study,” *Behavioral Science*, XII (1967), 33-48.

Milgrom, Paul, and Nancy Stokey, “Information, Trade and Common Knowledge,” *Journal of Economic Theory*, XXVI (1982), 17-27.

Mookerjee, Dilip and Barry Sopher, “Learning and Decision Costs in Experimental Constant-sum Games,” *Games and Economic Behavior*, XIX (1997), 97-132.

Nagel, R., “Unraveling in Guessing Games: An Experimental Study,” *The American Economic Review*, LXXXV (1995), 1313-1326.

Ochs, Jack, “Games with Unique, Mixed Strategy Equilibria: An Experimental Study,” *Games and Economic Behavior*, X (1995), 202-217.

O’Neill, Barry, “Nonmetric Test of the Minimax Theory of Two-person Zero-sum Games,” *Proceedings of the National Academy of Sciences*, LXXXIV (1987), 2106-2109.

Rapoport, Amnon and Amaldoss, Wilfred, “Mixed Strategies and Iterative Elimination of Strongly Dominated Strategies: An Experimental Investigation of States of Knowledge,” *Journal of Economic Behavior and Organization*, XLII (2000), 483-521.

Rapoport, Amnon and Boebel, Richard B., “Mixed Strategies in Strictly Competitive Games: A Further Test of the Minimax Hypothesis,” *Games and*

Economic Behavior, IV (1992), 261-283.

Rapoport, Amnon and Darryl A. Seale, "Coordination success in noncooperative large group market entry games," In C. R. Plott and V. L. Smith (Eds.), *Handbook of Experimental Economic Results*, in press.

Rosenthal, Robert W., "A bounded rationality approach to the study of noncooperative games," *International Journal of Game Theory*, XVIII(3) (1989), 273-92.

Rubinstein, A. "Instinctive and cognitive reasoning: Response times study," Tel-Aviv and NYU working paper 2003, http://arielrubinstein.tau.ac.il/papers /Response.pdf

Schelling, Thomas, *The Strategy of Conflict.* (Cambridge, MA: Harvard University Press, 1960). Selten, Reinhard, "Features of Experimentally Observed Bounded Rationality," *European Economic Review*, XLII (1998), 413-436.

Sonsino, D., I. Erev and S. Gilat, "On the Likelihood of Repeated Zero-sum Betting by Adaptive (Human) Agents," Technion, Israel Institute of Technology, 2002.

Sovik, Y., "Impossible Bets: An Experimental Study," University of Oslo Department of Economics, 2000.

Stahl, D. O., "Is Step-j Thinking an Arbitrary Modeling Restriction or a Fact of Human Nature?" *Journal of Economic Behavior and Organization*, XXXVII (1998), 33-51.

Stahl, D. O. and P. Wilson, "On Players Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, X (1995), 213-54.

Tang, F.-F., "Anticipatory Learning in Two-person Games: Some Experimental Results," *Journal of Economic Behavior and Organization*, XLIV (2001), 221-32.

Van Huyck, John; Ray Battalio, and Richard Beil, "Tacit Cooperation Games, Strategic Uncertainty, and Coordination Failure," *The American Economic Review*, LXXX (1990), 234-248.

Weizsacker, G. "Ignoring the rationality of others: Evidence from experimental normal-form games," *Games and Economic Behavior*, XLIV (2003), 145-171.

TABLE I

Row player's payoffs in an n-person stag hunt game

Row choice		group outcome	
		L	H
	L	x	—
	H	0	1

TABLE II

Data and CH estimates of τ in various p-beauty contest games

subject pool or game	source[1]	group size	sample size	Nash equil'm	pred'n error	data mean	data std dev	data mode	fit of CH model τ	fit of CH model mean	fit of CH model error	fit of CH model std dev	fit of CH model mode	bootstrapped 90% c.i.
p=1.1	HCW (98)	7	69	200	47.9	152.1	23.7	150	**0.10**	151.6	-0.5	28.0	165	(0.0,0.5)
p=1.3	HCW (98)	7	71	200	50.0	150.0	25.9	150	**0.00**	150.4	0.5	29.4	195	(0.0,0.1)
high $	CHW	7	14	72	11.0	61.0	8.4	55	**4.90**	59.4	-1.6	3.8	61	(3.4,4.9)
male	CHW	7	17	72	14.4	57.6	9.7	54	**3.70**	57.6	0.1	5.5	58	(1.0,4.3)
female	CHW	7	46	72	16.3	55.7	12.1	56	**2.40**	55.7	0.0	9.3	58	(1.6,4.9)
low $	CHW	7	49	72	17.2	54.8	11.9	54	**2.00**	54.7	-0.1	11.1	56	(0.7,3.8)
.7(Mean+18)	Nagel (98)	7	34	42	-7.5	49.5	7.7	48	**0.20**	49.4	-0.1	26.4	48	(0.0,1.0)
PCC	CHC (new)	2	24	0	-54.2	54.2	29.2	50	**0.00**	49.5	-4.7	29.5	0	(0.0,0.1)
p=0.9	HCW (98)	7	67	0	-49.4	49.4	24.3	50	**0.10**	49.5	0.0	27.7	45	(0.1,1.5)
PCC	CHC (new)	3	24	0	-47.5	47.5	29.0	50	**0.10**	47.5	0.0	28.6	26	(0.1,0.8)
Caltech board	Camerer	73	73	0	-42.6	42.6	23.4	33	**0.50**	43.1	0.4	24.3	34	(0.1,0.9)
p=0.7	HCW (98)	7	69	0	-38.9	38.9	24.7	35	**1.00**	38.8	-0.2	19.8	35	(0.5,1.6)
CEOs	Camerer	20	20	0	-37.9	37.9	18.8	33	**1.00**	37.7	-0.1	20.2	34	(0.3,1.8)
German students	Nagel (95)	14-16	66	0	-37.2	37.2	20.0	25	**1.10**	36.9	-0.2	19.4	34	(0.7,1.5)
70 yr olds	Kovalchik	33	33	0	-37.0	37.0	17.5	27	**1.10**	36.9	-0.1	19.4	34	(0.6,1.7)
US high school	Camerer	20-32	52	0	-32.5	32.5	18.6	33	**1.60**	32.7	0.2	16.3	34	(1.1,2.2)
econ PhDs	Camerer	16	16	0	-27.4	27.4	18.7	N/A	**2.30**	27.5	0.0	13.1	21	(1.4,3.5)
1/2 mean	Nagel (98)	15-17	48	0	-26.7	26.7	19.9	25	**1.50**	26.5	-0.2	19.1	25	(1.1,1.9)
portfolio mgrs	Camerer	26	26	0	-24.3	24.3	16.1	22	**2.80**	24.4	0.1	11.4	26	(2.0,3.7)
Caltech students	Camerer	17-25	42	0	-23.0	23.0	11.1	35	**3.00**	23.0	0.1	10.6	24	(2.7,3.8)
newspaper	Nagel (98)	3696, 1460, 2728	7884	0	-23.0	23.0	20.2	1	**3.00**	23.0	0.0	10.6	24	(3.0,3.1)
Caltech	CHC (new)	2	24	0	-21.7	21.7	29.9	0	**0.80**	22.2	0.6	31.6	0	(4.0,1.4)
Caltech	CHC (new)	3	24	0	-21.5	21.5	25.7	0	**1.80**	21.5	0.1	18.6	26	(1.1,3.1)
game theorists	Nagel (98)	27-54	136	0	-19.1	19.1	21.8	0	**3.70**	19.1	0.0	9.2	16	(2.8,4.7)
								mean	**1.30**					
								median	**1.61**					

Note 1: HCW (98) is Ho, Camerer, Weigelt AER 98; CHC are new data from Camerer, Ho, and Chong;
CHW is Camerer, Ho, Weigelt (unpublished); Kovalchik is unpublished data collected by Stephanie Kovalchik

TABLE III

Parameter Estimate τ for Cognitive Hierarchy Models

Data set	Stahl & Wilson	Cooper & Van Huyck	Costa-Gomes et al.	Mixed	Entry
Game-specific τ					
Game 1	2.93	15.90	2.28	0.98	0.70
Game 2	0.00	1.07	2.27	1.71	0.85
Game 3	1.40	0.18	2.29	0.86	-
Game 4	2.34	1.28	1.26	3.85	0.73
Game 5	2.01	0.52	1.80	1.08	0.70
Game 6	0.00	0.82	1.67	1.13	
Game 7	5.37	0.96	0.88	3.29	
Game 8	0.00	1.54	2.18	1.84	
Game 9	1.35		1.89	1.06	
Game 10	11.33		2.26	2.26	
Game 11	6.48		1.23	0.87	
Game 12	1.71		1.03	2.06	
Game 13			2.28	1.88	
Game 14				9.07	
Game 15				3.49	
Game 16				2.07	
Game 17				1.14	
Game 18				1.14	
Game 19				1.55	
Game 20				1.95	
Game 21				1.68	
Game 22				3.06	
Median τ	1.86	1.01	1.89	1.77	0.71
Common τ	1.54	0.82	1.73	1.48	0.73

TABLE IV

Model Fit (Log-likelihood LL and Mean Squared Deviation MSD)

Data set	Stahl & Wilson	Cooper & Van Huyck	Costa-Gomes et al.	Mixed	Entry
<u>Log-likelihood</u>					
Cognitive Hierarchy (Game-specific τ)	-360	-838	-264	-824	-150
Cognitive Hierarchy (Common τ)	-458	-868	-274	-872	-150
Nash Equilibrium [1]	-1823	-5422	-1819	-1270	-154
<u>Mean Squared Deviation</u>					
Cognitive Hierarchy (Game-specific τ)	0.0074	0.0090	0.0035	0.0097	0.0004
Cognitive Hierarchy (Common τ)	0.0327	0.0145	0.0097	0.0179	0.0005
Nash Equilibrium	0.0882	0.2038	0.1367	0.0387	0.0049

Note 1: The Nash Equilibrium result is derived by allowing a non-zero mass of 0.0001 on non-equilibrium strategies.

TABLE V

Cross-game Fit (Log-likelihood LL and Mean Squared Deviation MSD)

Data set	Stahl & Wilson	Cooper & Van Huyck	Costa-Gomes et al.	Mixed	Entry
Log-likelihood					
Cognitive Hierarchy (Common τ)	-469	-956	-293	-884	-154
Mean Squared Deviation					
Cognitive Hierarchy (Common τ)	0.0416	0.0335	0.0237	0.0215	0.0046

TABLE VI

Game 8 from Costa-Gomes et al. [2001]

	L	R	data	Nash	CH fit
T	45, 66	82, 31	.92	1	.82
TM	22, 14	57, 55	0	0	.06
BM	30, 42	28, 37	0	0	.06
B	15, 60	61, 88	.08	0	.06
data	.64	.36			
Nash	1	0			
CH fit	.55	.45			

TABLE VII

Game 4 from Binmore et al. [2001]

	L	C	R	data	Nash	CH fit
T	0,0	2,-2	-1,1	.33	.17	.25
M	2,-2	0,0	-1,1	.29	.17	.25
B	-1,1	-1,1	0,0	.28	.67	.50
data	.13	.04	.83			
Nash	.17	.17	.67			
CH fit	.08	.08	.84			

TABLE VIII
Economic Value of Various Theories

Data set	Stahl & Wilson	Cooper & Van Huyck	Costa-Gomes et al.	Mixed	Entry
Observed Payoff	195	586	264	328	118
Clairvoyance Payoff	243	664	306	708	176
Economic Value					
Clairvoyance	48	78	42	380	58
Cognitive Hierarchy (Common τ)	13	55	22	132	10
Nash Equilibrium	5	30	15	-17	2
% Maximum Economic Value Achieved					
Cognitive Hierarchy (Common τ)	26%	71%	52%	35%	17%
Nash Equilibrium	10%	39%	35%	-4%	3%

Note 1: The economic value is the total value (in USD) of all rounds that a "hypothetical" subject will earn using the respective model to predict other's behavior and best responds with the strategy that yields the highest expected payoff in each round.

TABLE IX

Poisson CH Prediction and Empirical Frequency for Betting Game [Sovik, 2000]

State of the World	A	B	C	D
Player I				
Payoff for Betting	32	-28	20	-16
Information Set	(A or B)		(C or D)	
Betting Rate				
Data	77%		53%	
Poisson CH	46%		89%	
Player II				
Payoff for Betting	-32	28	-20	16
Information Set	A	(B or C)		D
Betting Rate				
Data	0%	83%		100%
Poisson CH	12%	72%		89%

TABLE X
Poisson CH and Nash Prediction for Pricing Game on Money Illusion
[Fehr and Tyran, 2002]

Nash Prediction:		Median		Mean		Standard Deviation	
		Actual	τ = 1.5	Actual	τ = 1.5	Actual	τ = 1.5
x plays 11	x-comps	21	21	17.65	19.75	4.23	4.73
	y-comps	22.5	26	20.55	23.95	6.35	6.18
y plays 14	x-subs	11	11	10.17	9.12	2.93	7.30
	y-subs	14	14	12.5	10.16	4.10	7.79

APPENDIX TABLE A.I

95% Confidence Interval for the Parameter Estimate τ of Cognitive Hierarchy Models

Data set	Stahl & Wilson		Cooper & Van Huyck		Costa-Gomes et al.		Mixed		Entry	
	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
Game-specific τ										
Game 1	2.40	3.65	15.40	16.71	1.58	3.04	0.67	1.22	0.21	1.43
Game 2	0.00	0.00	0.83	1.27	1.44	2.80	0.98	2.37	0.73	0.88
Game 3	0.75	1.73	0.11	0.30	1.66	3.18	0.57	1.37	-	-
Game 4	2.34	2.45	1.01	1.48	0.91	1.84	2.65	4.26	0.56	1.09
Game 5	1.61	2.45	0.36	0.67	1.22	2.30	0.70	1.62	0.26	1.58
Game 6	0.00	0.00	0.64	0.94	0.89	2.26	0.87	1.77		
Game 7	5.20	5.62	0.75	1.23	0.40	1.41	2.45	3.85		
Game 8	0.00	0.00	1.16	1.72	1.48	2.67	1.21	2.09		
Game 9	1.06	1.69			1.28	2.68	0.62	1.64		
Game 10	11.29	11.37			1.67	3.06	1.34	3.58		
Game 11	5.81	7.56			0.75	1.85	0.64	1.23		
Game 12	1.49	2.02			0.55	1.46	1.40	2.35		
Game 13					1.75	3.16	1.64	2.15		
Game 14							6.61	10.84		
Game 15							2.46	5.25		
Game 16							1.45	2.64		
Game 17							0.82	1.52		
Game 18							0.78	1.60		
Game 19							1.00	2.15		
Game 20							1.28	2.59		
Game 21							0.95	2.21		
Game 22							1.70	3.63		
Common τ	1.39	1.67	0.74	0.87	1.53	2.13	1.30	1.78	0.42	1.07

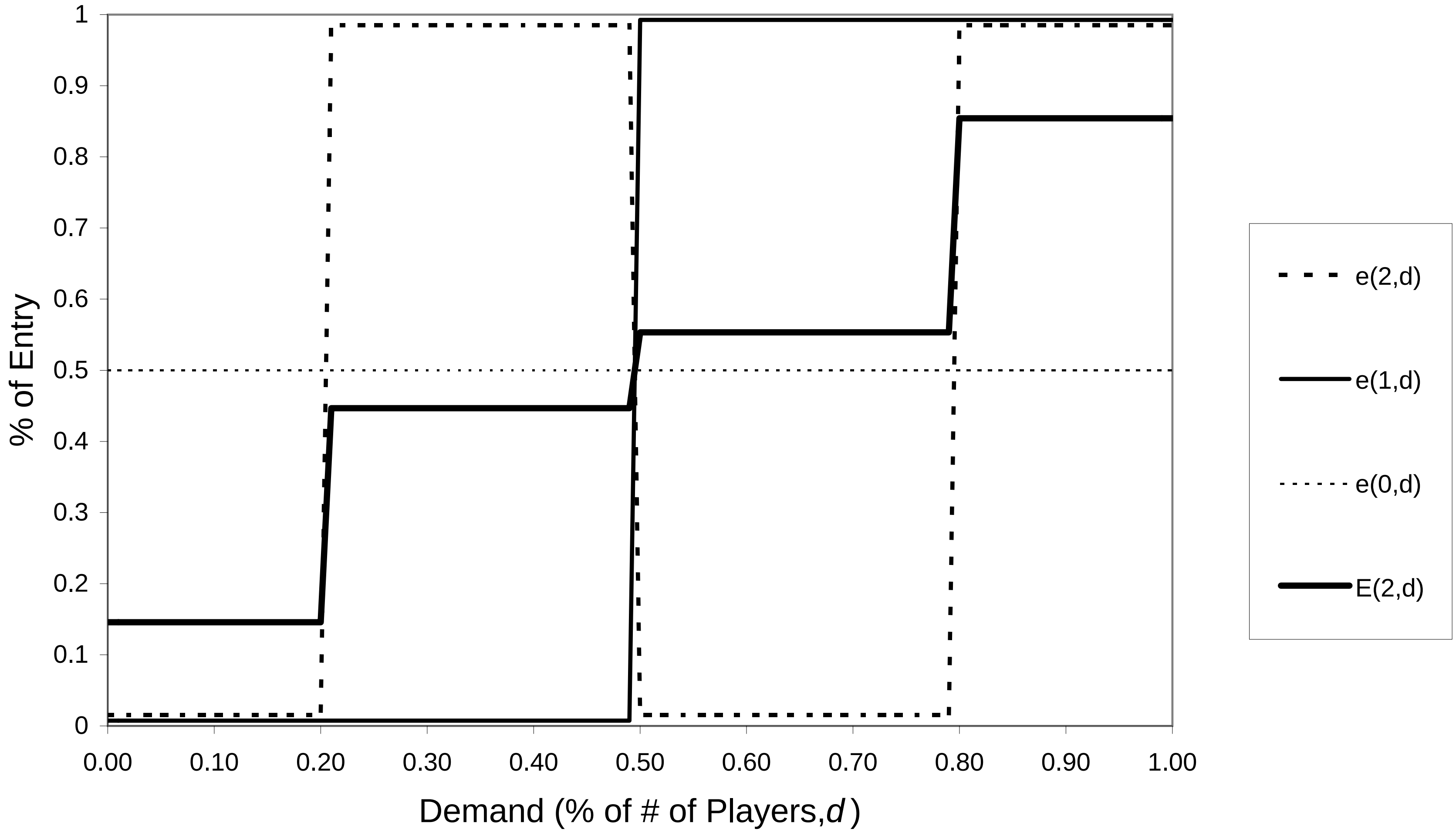

FIGURE I
Behaviors of Level 0, 1 and 2 Players (τ = 1.5)

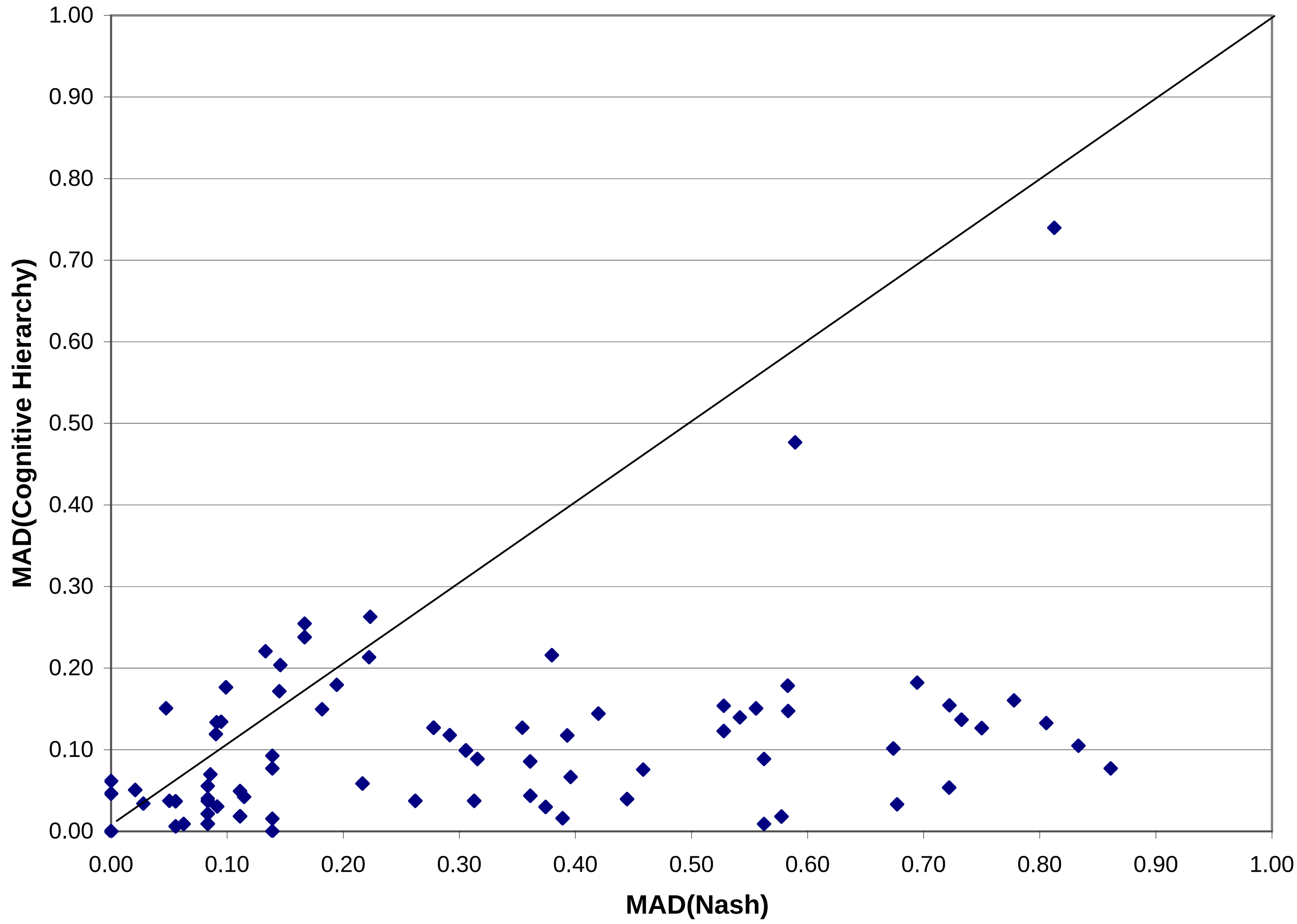

FIGURE II

Mean Adsolute Deviation for Matrix Games: Nash vs Cognitive Hierarchy (Common τ)

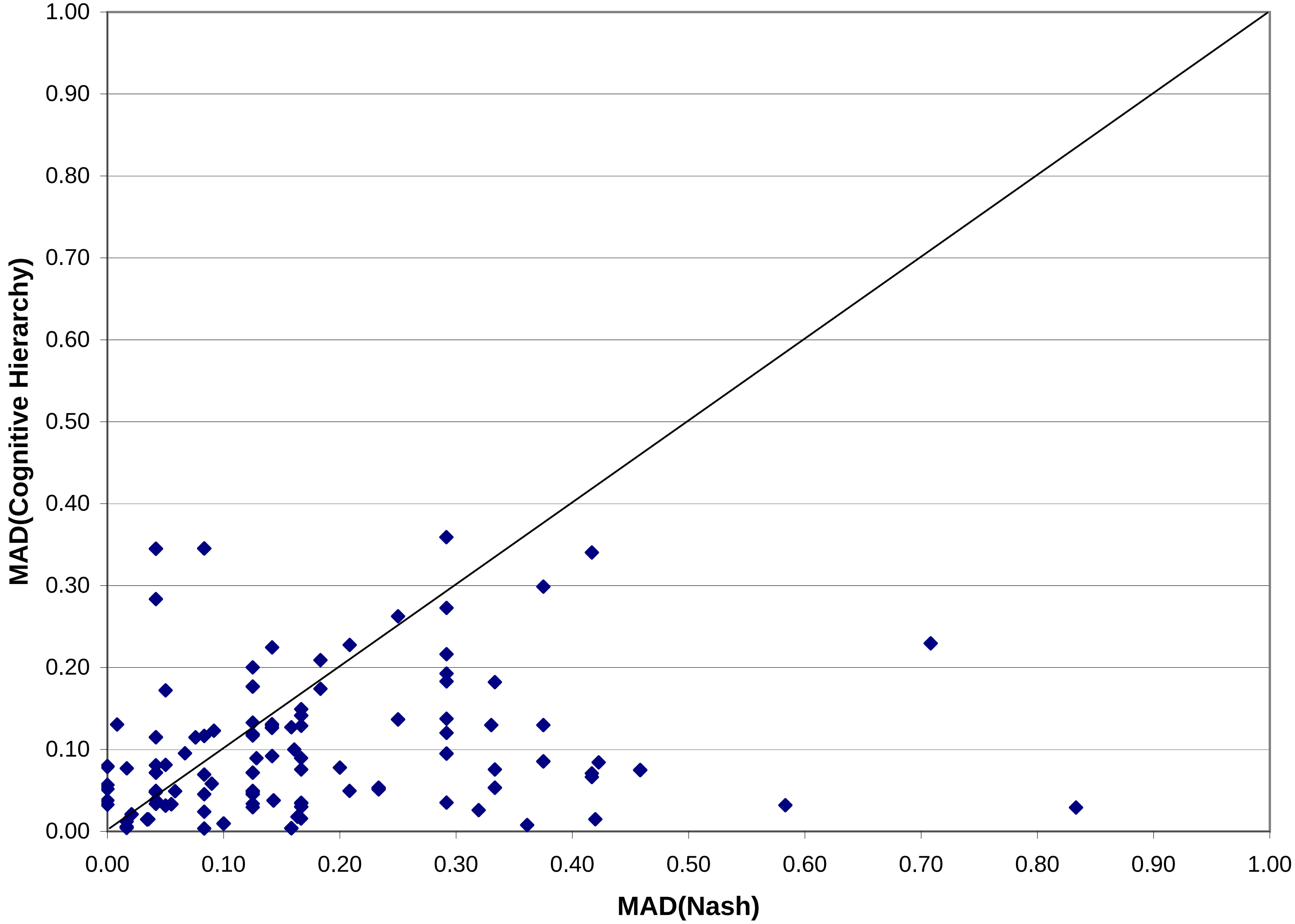

FIGURE III

Mean Adsolute Deviation for Mixed and Entry Games: Nash vs Cognitive Hierarchy (Common τ)