

Testing Cognitive Hierarchy Theories of Beauty Contest Games

July 2010

P. RICHARD HAHN

*Department of Statistical Science, Duke University
Durham, North Carolina 27708-0251, U.S.A.
hahn@stat.duke.edu*

KRISTIAN LUM

*Department of Statistical Science, Duke University
Durham, North Carolina 27708-0251, U.S.A.
kristian.lum@stat.duke.edu*

CARL MELA

*T. Austin Finch Foundation Professor of Business Administration
Fuqua School of Business, Duke University
Durham, North Carolina 27708-0251, U.S.A.
mela@duke.edu*

SUMMARY

Behavioral game theory experiments consistently reveal that individuals deviate from theoretically optimal (Nash equilibrium) strategies even in simple games. The α -beauty contest is among the simplest games that elicit such non-optimal behavior; accordingly, there is substantial interest in formally characterizing the observed play for this game.

Several authors, beginning with the works of Stahl and Wilson (1995, 1994) and Nagel (1995), have proposed an intuitively appealing and formally elegant cognitive hierarchy (CH) model of strategic reasoning for these instances where Nash equilibrium is patently unrealistic. In a CH model the player population is partitioned according to how many steps of iterated reasoning players perform when strategizing and each subpopulation plays conditionally optimally according to their level of strategic foresight. In this paper we specifically consider two common instantiations of the general CH model, the CH-Poisson model of Camerer et al. (2004) and the level- k model as described in Crawford and Iriberri (2007), evaluating how well each succeeds at describing empirical beauty contest data.

Our evaluation proceeds by first developing a highly flexible semiparametric CH model which includes these two commonly studied models as special cases. We then describe an experiment to collect data specifically tailored to test key assumptions of the CH framework. Finally, we describe an appropriate null model against which to evaluate the ability of CH models to characterize our experimental data.

The key finding is that while the CH-Poisson and the level- k models do not describe our new data well, the more general semiparametric CH model significantly outperforms a non-CH alternative model, lending support to the possibility that a cognitive hierarchy strategic structure underpins the observed bids.

Some key words: behavioral game theory, cognitive hierarchy models, model assessment, bounded rationality.

1 INTRODUCTION

Experiments consistently demonstrate that people do not always strategize the way that mathematical game theory says they ought to (Camerer, 2003). Cognitive hierarchy (CH) theories of strategic reasoning elegantly account for this fact by taking into consideration players' beliefs about how their opponents will play (Crawford, 2007). Contrary to the Nash equilibrium of a game, which one arrives at by assuming that all players are capable of reasoning their way to the equilibrium solution and that all players assume as much about one another (Bosch-Domènech et al., 2002; Stahl and Wilson, 1995), CH models posit that people do not reason all the way to equilibrium because doing so simply requires too much effort and/or ability¹ (Costa-Gomes and Crawford, 2006; Crawford, 2007; Stahl and Wilson, 1994, 1995; Nagel, 1995).

CH models propose instead that there exists a hierarchy of player types, corresponding to the different numbers of steps that players reason ahead in a game. Some people – call them level-0 players – simply play at random. Level-1 thinkers reason that people play randomly in this way, and they play the optimal strategy given this assumption. Level-2 players assume that some fraction of players are using a random strategy and that the remainder of players are level-1 players, and they play the optimal strategy given this assumption, and so on. CH models often generate better predictions of behavior than Nash equilibrium. While this marks CH models as *better* descriptions of empirical game play, here we attempt to determine whether such descriptions are in fact *accurate* ones. That is, we attempt to assess the statistical evidence for the hypothesis that people are playing according to a cognitive hierarchy.

Although CH models have been developed for a wide variety of games, we base our

¹Indeed, the sort of reasoning required to arrive at the equilibrium solution is formal mathematical induction, a process with which people are known to struggle (Newell and Simon, 1972; Johnson et al., 2002).

analysis on the α -beauty contest game (Moulin, 1986; Nagel, 1995), owing to its simplicity and high profile in the existing literature. On the one hand, this specificity is inherently limiting. On the other hand, this very simple game is the quintessential CH model, because “the sharpest evidence on iterated dominance comes from α -beauty contest games (Camerer, 2003)”. Our purpose is to critically assess that evidence.

We do this by comparing the CH model to an appropriate non-CH alternative, an approach that is relatively uncommon in the literature, a notable exception being Stahl and Wilson (1995). We share these author’s conviction that “[f]or the purposes of hypothesis testing of alternative theories, it is necessary to construct an encompassing econometric model.” The majority of our paper is devoted to developing just such an encompassing model for the α -beauty contest, which we call the semiparametric cognitive hierarchy (SPCH) model. This model nests most published CH models as special cases.

We collect new beauty contest data from an experiment designed to highlight the signature patterns of CH play. Analyzing this data, we find that our flexible SPCH model convincingly outperforms earlier variants in the literature (called the CH-Poisson model and the level- k model, to be defined) in terms predicting player behavior. Moreover, these earlier models do worse than a non-CH null model, in which players’ behavior is not constrained to reflect a cognitive hierarchy at all, at predicting the observed behavior, while the newly introduced SPCH model does better. More plainly, we find no evidence in support of the specific earlier CH variants, but do find compelling evidence of behavior consistent with some cognitive hierarchy.

The paper proceeds as follows. In this first section we briefly review the α -beauty contest and rehearse the previously proposed CH-Poisson and level- k models (Camerer et al., 2004; Crawford, 2007). In the second section we outline the statistical challenges presented by CH models generically and describe how we might overcome these challenges with new experimental data and a novel Bayesian analysis. Both of these new elements are introduced

specifically with the goal of measuring how well CH-style strategizing accounts for the observed game play. In section three we report the results of our analysis on new data collected as described. Section four concludes with a brief discussion.

1.1 The Beauty Contest

The goal of each player of the α -beauty contest² (Moulin, 1986; Nagel, 1995) is to report a number b – referred to here as a *bid*³ – that is as close as possible to α times the whole group’s average bid. Bids are restricted to lie within some fixed interval (L, U) . Formally we can say that for a beauty contest played among N players, player i has a payoff defined by

$$u_i(\mathbf{b}) = M - d(b_i - \alpha \bar{b}) \tag{1}$$

where $d()$ is some distance metric, \mathbf{b} denotes the realized N -vector of bids, $M = d(U - L)$, and $\bar{b} = N^{-1} \sum_{j=0}^N b_j$. For example, a beauty contest on an interval from 0 to 100 could have a payoff of $\$(100 - |b_i - \alpha \bar{b}|)$.

The Nash equilibrium for $\alpha \in (0, 1)$ is 0. Everyone in the group is trying to undercut everyone else’s bid by the fraction α , driving the equilibrium strategy to zero. Nonetheless, experiments consistently reveal that many, if not most, people do not play the zero strategy.

The beauty contest game has many desirable properties from an analyst’s point of view, two of which we note here. First, it is a symmetric game, meaning that all players have the

²Beauty contest games are so called after a quote by Keynes (1936), first cited in this context by Nagel (1995): “professional investment may be likened to those newspaper competitions in which the competitors have to pick out the six prettiest faces from a hundred photographs, the price being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole . . . 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 average opinion to be. And there are some, I believe, who practise the fourth, fifth, and higher degrees.”

³We introduce this term as an intuitive one for the non-specialist; it is not intended to evoke an auction setting. If preferred, the b could be read as “behavior”.

same payoff function. Second, for large N ,

$$d(b_i - \alpha \bar{b}) - d(b_i - \alpha \bar{b}_{-i}) \approx 0$$

so that payoff maximization effectively depends only on the average bid of the other $N - 1$ players. This follows because the boundedness of the bids entails that the contribution of any one bid to the overall mean grows like $1/N$. From this perspective, it becomes natural to ask if the observed non-Nash play in the α -beauty game is a result of rational agents acting on the conviction that their opponents are acting irrationally so that $\alpha \bar{b}_{-i} \neq 0$. If a given player does not trust that his opponents can reason their way to the Nash equilibrium strategy, then the Nash equilibrium solution is no longer optimal or rational for that player.

This fact suggests that characterizing players' beliefs about the strategies others play may be one route to accurately characterizing actual bidding behavior. Such an approach poses two questions. First, can we come up with plausible restrictions on the belief distributions so as to constrain the possible behavior that would qualify as rational? Second, how might we test if those restrictions are actually obeyed in practice? Cognitive hierarchy models (Stahl and Wilson, 1995, 1994; Nagel, 1995; Camerer et al., 2004) are a natural candidate to address the first question and we describe these models in the next section. Then, the remainder of the paper takes up the second question.

1.2 Cognitive hierarchies

A CH model is built upon several *prima facie* reasonable premises:

1. Players are distributed among a discrete collection of *strategy classes* defined by the number of steps ahead in the game players will reckon when formulating their strategies.
2. Players have strategy-class-specific beliefs about the relative proportion of players in

strategy classes lower than themselves.

3. Players assume that they are thinking at least one step ahead of any other player.
4. Players will best respond in the sense of maximizing expected payoff conditional on their beliefs.

The first and second conditions mean that an agent's strategic beliefs cannot be wholly idiosyncratic. Condition three is a convenient and plausible restriction, which can be thought of as an "arrogance" assumption. The final condition is the usual payoff maximization assumption.

These assumptions alone leave too many degrees of freedom in that both the distribution of the players across the various strategy classes and also the strategy-class-specific belief distributions remain undetermined. Even if all these various distributions took simple parametric forms, the model would pose estimability difficulties, with N latent class memberships and up to N class-specific belief distributions free to vary.

1.3 The CH-Poisson Model

Camerer et al. (2004) handle the indeterminacy of CH models by fiat, adding three additional – and quite restrictive – assumptions to those above:

- P5. Players are distributed among strategy classes via a $\text{Poisson}(\tau)$ distribution.
- P6. Players have accurate beliefs about the relative proportions of players in strategy classes lower than themselves.
- P7. Players in the lowest strategy class issue bids uniformly at random over the allowed interval.

Taken together, these additional assumptions define game play for any strategy class: players will maximize their expected payoffs with respect to their class-specific belief distribution on strategy classes, given as $g_k(h) = f_\tau(h) / \sum_{l=1}^{k-1} f_\tau(l)$, where f is the probability mass function of the Poisson distribution. A step- k player's best response can then be found iteratively by computing the best response for all strategy classes below k . Model fitting is thereby reduced to the estimation of a single parameter τ . Model assessment or evaluation can then be conducted according to some criterion, conditional on this parameter estimate.

The first of these additional assumptions, the parametric assumption, is less restrictive than the other two. Condition P6 is easily the most restrictive because it implies that the bidding behavior of all the strategy classes is coordinated purely by the true underlying distribution. This condition would be equally constrictive even if the underlying distribution were not a single-parameter distribution like the Poisson. Condition P7 puts an upper bound on where any (non-level-0) strategy class can bid by fixing mean bid the level-0 players.

Thus, the CH-Poisson model consists of a discrete component, constituting a countably infinite collection of bids, the values of which are determined by a single parameter, τ , and a continuous component, which is the uniform distribution from which the bids of level-0 players are assumed to be drawn. On the face of it, the actual data⁴ (Ho et al., 1998) exhibit

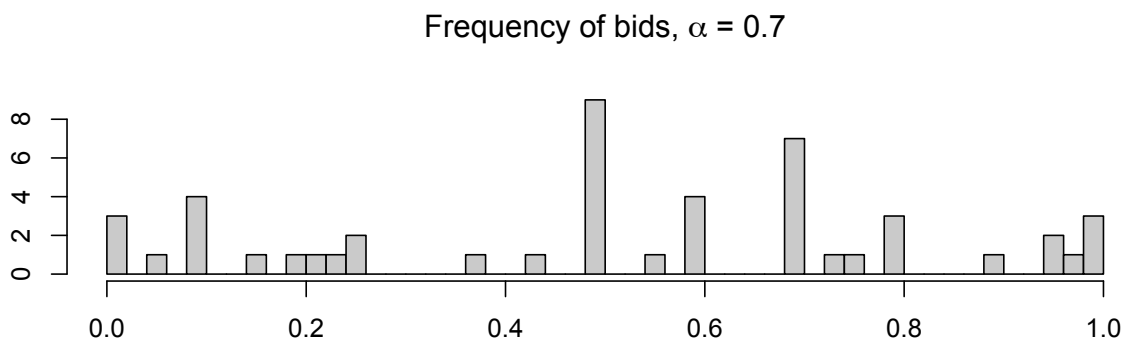


Figure 1: Strategic play is not overwhelmingly apparent from the raw data, which appears roughly uniform. We have rescaled here to the unit interval (as we will throughout).

many properties that would seem to rule out the CH-Poisson model, including a lack of many identical bids corresponding to the discrete component of the CH-Poisson and also approximately equal number of bids below $1/2$ as above (or equal to). This misfit may be formalized somewhat by comparing the sample mean, the estimator for τ used in Camerer et al. (2004), to $\hat{\tau} \equiv -\log(2N^{-1} \sum_{j=1}^N \mathbb{1}(b_j > 1/2))$, a consistent estimator based explicitly on counting the known level-0 players. For the data shown this degree of sophistication is unnecessary, as the sample mean is an unattainable (under CH-Poisson) 0.52 . That is, the population mean of a CH-Poisson model with uniform random level-0 players can never be greater than $1/2$ for any value of τ , so that a sample mean of 0.52 yields an undefined estimate.

1.4 Level- k model

The level- k model, as described in Crawford and Iriberri (2007), is a CH model in which every player assumes that all the other players are in the strategy class immediately below them. That is, it modifies P5-P7 as follows:

LK5. Players are distributed among K strategy classes via a multinomial distribution with probability weights $\boldsymbol{\pi}$.

LK6. Players believe that all of their opponents reason exactly one step less than they do.

LK7. Players in the lowest strategy class issue bids uniformly at random over the allowed interval.

Notice that LK7 and P7 are identical, that LK5 is less restrictive than P5, but most importantly that these assumptions, like P5-P7, uniquely define optimal play across all

⁴We thank Teck Ho for making these data available to us. The data we have shown here aggregate seven groups of seven players each, all playing with the same $\alpha = 0.7$.

strategy classes. One of the upshots of our analysis is the ability to determine which set of assumptions, if any, is a good match to observed bidding behavior.

2 A SEMIPARAMETRIC CH (SPCH) MODEL FOR BEAUTY CONTEST DATA

Our aim is to develop a model that affords great flexibility in the range of beliefs it permits a rational player to hold, while still admitting statistical analysis. Moreover, it should nest the more restrictive assumptions of the CH-Poisson and level- k formulations to facilitate model comparison. In the following subsections we develop these properties of our generalized CH model from the ground up. As a CH model, our model will retain conditions 1-4 from above. We will replace conditions P5-P7 of the CH-Poisson and LK5-LK7 of the level- k model with less rigid analogues.

2.1 Monotonically decreasing target bids

Rather than explicitly specifying each strategy class's belief distributions, we adopt a less strict characterization which only specifies how the various strategy classes bid relative to one another.

It will be valuable from here out to distinguish carefully between three related, but distinct quantities. First we denote by T_k the *target bid* of a strategy class k player – α times the value that such a player expects will be the mean play of his opponents. Secondly, we will denote by b_i the observed bid of the i th player. Lastly, we will denote by b_i^* the utility-maximizing bid for agent i .

Thus equipped, we can express our first restriction as

$$T_k < T_{k'} \text{ whenever } k > k'. \quad (2)$$

In words, higher step-ahead thinkers are required to have lower target bids. Understanding

this requirement is aided by some notation. Recall that for relatively large N (tens or hundreds), we can express T_k as

$$T_k = \alpha \sum_{j=1}^{k-1} g_k(j) T_k(j) \quad (3)$$

where $T_k(j)$ is a level- k player's belief as to the target value of a level $j < k$ player.

This expression makes clear the impossibility a formally distinguishing between a given player type's belief distribution, $g_k(j)$, and his beliefs about the other player's beliefs, from which $T_k(j)$ is derived. But if we make an additional assumption that

$$\text{players of strategy level } k \text{ know } g_{k'} \text{ for all } k' < k \quad (4)$$

we can further extend our analysis. At an intuitive level, the appeal of this assumption is that it jibes with a strong conception of a cognitive hierarchy – not only do some players reason more steps ahead than others, those players are also assumed to have the capacity to project themselves into the strategic viewpoint of lower strategy classes, though the reverse is not possible⁵.

This new assumption, along with the assumption that all players are conditionally rational actors, entails that

$$T_k(j) = T_j \quad (5)$$

⁵We might add that under some interpretations of the level- k model (Costa-Gomes et al., 2001), a level- k player does not have any beliefs about the strategies of players lower than $k - 1$. Observed bids alone cannot distinguish between this case and the case where players have beliefs about all players, but simply believe that there are no players lower than $k - 1$ in the population ($g_k(k - 1) = 1$). This is a special case of the unidentifiability of $g(\cdot)$ and we do not address this point further. We stress, however, that if the bidding data alone does not support either model, the need to distinguish between the two is moot.

and we have the following recursive identity:

$$T_k = \alpha \sum_{j=1}^{k-1} g_k(j) T_j, \quad (6)$$

from which we can investigate what sorts of restrictions on the belief distributions g are implied by the order restriction given in (2). A straightforward calculation shows

$$\frac{T_{k+1}}{T_k} = \frac{\alpha \sum_{j=0}^k g_{k+1}(j) T_j}{\alpha \sum_{h=0}^{k-1} g_k(h) T_h} \quad (7)$$

$$= \frac{\sum_{j=0}^{k-1} g_{k+1}(j) T_j + g_{k+1}(k) T_k}{\sum_{h=0}^{k-1} g_k(h) T_h} \quad (8)$$

$$= \frac{(1 - g_{k+1}(k)) \sum_{j=0}^{k-1} \frac{g_{k+1}(j)}{(1 - g_{k+1}(k))} T_j + g_{k+1}(k) T_k}{\mathbf{E}_k(T)} \quad (9)$$

$$= (1 - g_{k+1}(k)) \frac{\mathbf{E}_{k+1}(T_j \mid j < k)}{\mathbf{E}_k(T)} + \alpha g_{k+1}(k) \quad (10)$$

$$\leq 1. \quad (11)$$

That is, dealing with (2) is equivalent to working with distributions g which satisfy (10); this is restriction SP6. This expression is relatively easy to interpret, the left-hand side being a convex combination of α and the ratio of the expectations of the $(k + 1)$ -level and k -level player regarding what the opposition will bid. As might be expected, the requirements on the belief distributions g thus take the form of a moment constraint only, meaning that we obtain considerable variety in the shapes of distributions permitted under (2).

Indeed, the CH-Poisson model and the level- k model both use distributions satisfying (2). In the first case this follows because $\frac{\mathbf{E}_{k+1}(T_j \mid j < k)}{\mathbf{E}_k(T_j)} = 1$ under the CH-Poisson model so that we have

$$\alpha g_{k+1}(k) \leq g_{k+1}(k)$$

which is true whenever $\alpha \leq 1$. In the second case, $g_{k+1}(k) = 1$ so that we find the condition

is again satisfied whenever $\alpha \leq 1$.

2.2 Incorporating Error

As noted, empirical α -beauty data exhibit a greater variety of observed bids than the CH-Poisson or level- k models would suggest. In particular, the degree of observed heterogeneity points to additional sources besides the level-0 players. Because any CH-model, as described so far, permits just one optimal bid per strategy class, we find that, (excepting level-0 players), the number of unique plays we observe must correspond to the number of strategy classes appearing in our sample. Lying back of this mathematical observation is the simple fact that any realistic CH model should allow players to deviate to various degrees from their optimal target bid. The observation that individuals will often bid distinct amounts in separate instances of the α -beauty contest played some duration apart is strong motivating evidence for building “jitter” into our CH model. Others that have taken this approach are De Giorgi and Reimann (2008); Stahl and Wilson (1995); Haruvy et al. (2001) and Bosch-Domènech et al. (2010).

2.3 Conditional Rationality

Fortunately, incorporating optimization error can be done without violating conditional rationality, subject to a few natural assumptions. First, we assume that each agent’s bid comes from a class-specific distribution with class-specific mean given by the target bid for that class, T_k , as previously defined. Secondly, we assume that the payoff function uses squared distance so that in (1) we have

$$d(\cdot) \equiv (\cdot)^2. \tag{12}$$

We now demonstrate how these two assumptions preserve conditional rationality for each strategy class.

Let

$$\begin{aligned}
B^0 &\sim F_0 & \mathbb{E}(B^0) &= T_0 \\
B^1 &\sim F_1 & \mathbb{E}(B^1) &= \alpha T_0 = T_1 \\
B^2 &\sim F_2 & \mathbb{E}(B^2) &= \alpha[g_2(0)T_0 + g_2(1)T_1] \\
&\vdots \\
B^k &\sim F_k & \mathbb{E}(B^k) &= \alpha \sum_{j=0}^{k-1} g_k(j)T_j
\end{aligned}$$

be the random variables describing the bids of players from the various strategy classes and let γ_i be the indicator variable denoting strategy class membership of the i th individual so that $b_i(\gamma_i)$ denotes the observed bid of individual i subject to being a level- γ_i player. Under the squared-error loss function we then have that the optimal bid for player i if he knew *exactly* the bids of the other players is given as

$$\begin{aligned}
b_i(\gamma_i) \mid \mathbf{b} &= \arg \max_b - (b - \alpha \bar{b})^2 \\
&\approx \arg \max_b - (b - \alpha \bar{b}_{-i})^2.
\end{aligned}$$

Then, integrating over player i 's strategy-class-specific beliefs (given as g_{γ_i}) we find the optimal play can be written

$$b_i(\gamma_i) = \arg \max_b - \mathbb{E}_{g_{\gamma_i}} \left[\left(b - \alpha \sum_{m \neq i} \frac{B_m}{N-1} \right)^2 \right]$$

where one can assume equality for N large. More suggestively we can note that $\sum_{m \neq i} \frac{B_m}{N-1}$ can be written as a sum of $(N-1)$ independent and identically distributed draws from a distribution $G^{\gamma_i} = \sum_{j=0}^{k-1} g_{\gamma_i}(j)F_j$ which has mean $\alpha^{-1}T_{\gamma_i}$ (as defined above). Applying a well-known result from decision theory which states that the optimal solution under expected

squared loss is the mean, we see that T_{γ_i} is indeed conditionally optimal⁶. That is,

$$b_i^*(\gamma_i) = T_{\gamma_i}$$

so that the cognitive hierarchy with optimization error still coheres as long as players assume that everyone will play optimally in the mean. Notationally, it may be helpful to think of the observed plays as

$$b_i(\gamma_i) = T_{\gamma_i} + \epsilon_i$$

where $E(\epsilon_i) = 0$ for all γ_i .

To summarize, T_k is what a level- k player “intends” to play, which is his optimal play subject to his beliefs about the other players given that they too intend to play optimally; what he actually plays is $b_i(k)$, which can be thought of as an observation of a random variable $B^k \sim F_k$ with $E(B^k) = T_k$.

Even though individuals play with random errors about their class-specific mean, the mean structure itself is rational even with respect to this randomness in the bids.

2.4 Error distribution

So far we have defined a nonparametric analytical model for the cognitive hierarchy. Agents playing according to this model are conditionally rational, organized hierarchically, free to hold flexible class-specific belief distributions, and free to make mistakes in their utility maximization, so long as they get things right in expectation. However, for testing purposes we are free to employ a flexible parametric model.

Specifically, we introduce a Beta error model and propose to learn about the latent

⁶Note that once players are asked to accommodate uncertainty (in this case, from two sources – the uncertainty over strategy class membership and that due to bidding error) the exact form of distance used in the payoff function becomes important in calculating the optimal CH bid; in particular, distinct nonlinear distance metrics will result in a distinct set of target bids.

strategy classes using a conjugate Bayesian Dirichlet-Multinomial model. Notice that this entails that the underlying distribution of strategy classes will be a discrete distribution of K strategy classes, equivalent to LK5 (but different than P5).

The assumption of Beta errors is convenient and innocuous from a purely game-theoretic perspective – none of the previous development hinged on particular features of the belief distributions beyond the first moments. Statistically, this approach entails that our inferences about model parameters are had conditionally on our parametric assumptions, but we take this as a virtue rather than a vice. The Beta is computationally tractable and contains the uniform distribution as a special case. See also Bosch-Domènech et al. (2010) for the use of Beta distributed errors in the context of beauty contest games.

2.5 Exploiting the Exogeneity of α to Infer Strategy-Class Membership

The SPCH model has so far been described for a fixed value of α . Consistent with prior literature, we take α to be functionally independent of the belief distributions g , the vector of strategy class indicators γ , and also the bidding errors ε . This exogeneity has two interesting consequences.

First, g not being a function of α implies immediately that $T_k(\alpha)$ is a decreasing function of α for $k > 0$. This fact suggests relaxing the assumption that the level-0 players have a constant bidding distribution across values of α (as in Camerer et al. (2004) where these players draw their bids from a fixed uniform distribution). A more flexible alternative is to let the level-0 players have target bids that follow a nondecreasing function of α , which will become condition SP7. Others that have investigated relaxing the uniform level-0 assumption include Haruvy and Stahl (2008) and Ho et al. (1998). In this case,

$$T_k(\alpha) \leq T_k(\alpha') \quad \text{for all } k \text{ and } \alpha < \alpha' \tag{13}$$

with no additional modifications necessary.

Second, γ not being a function of α implies that strategy-class membership is a fixed attribute of a player that does not change from game to game. Accordingly, having subjects bid *without feedback* for various values of α gives us multiple observations from which to infer class membership. Intuitively, observing bids over multiple values of α permits us to discern if any observed clustering of bids is a result of CH behavior by checking that those clusters evolve suitably with changing α (see figures 2 and 3). Formally, it yields the following factorization of the likelihood:

$$\sum_{k=1}^K \pi_k \left(\prod_{j=1}^J \text{Beta}(b(\alpha_j) \mid a_{k,j}, \beta_{k,j}) \right). \quad (14)$$

Keep in mind that this factorization is *in addition* to the previously described order conditions on the target bids, which enter the likelihood via a_j and β_j .

This factorization of the likelihood has important consequences for parameter estimation and model evaluation – two densities (for two α levels) that appear to match the data when looked at individually could not, in some instances, have plausibly come from a CH model if evaluated jointly across the two levels.

2.6 SPCH Summary

To recap, we have in place of A5-A7 or LK5-LK7 the following restrictions:

SP5. Players are distributed among K strategy classes via a multinomial distribution with probability weights $\boldsymbol{\pi}$.

SP6. Players' strategy-class-specific belief distributions g_k must satisfy (for all k and α)

$$(1 - g_{k+1}(k)) \frac{\mathbb{E}_{k+1}(T_j \mid j < k)}{\mathbb{E}_k(T)} + \alpha g_{k+1}(k) \leq 1$$

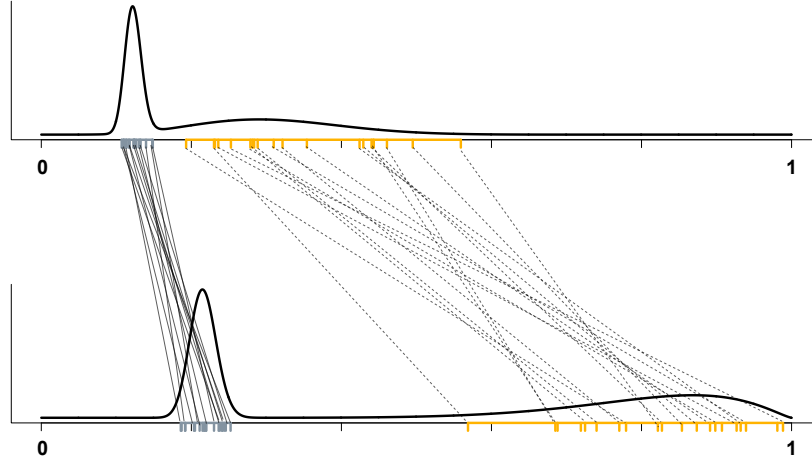


Figure 2: Lines connect players' bids across games with differing levels of α . This plot illustrates valid CH play wherein individuals do not switch mixture component/strategy class across games.

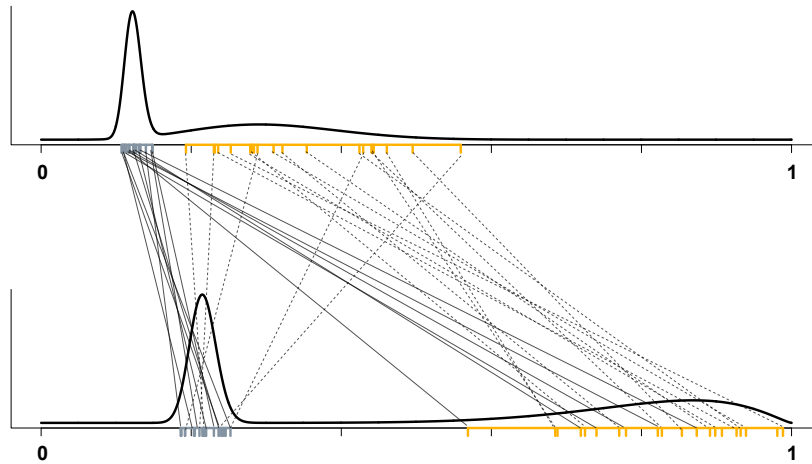


Figure 3: Switching class across α , as shown here, is not permitted under a CH model.

SP7. Target bids for level-0 players follow a nondecreasing function of α .

Working with these three assumptions, we avoid having to specify or estimate the belief distributions g . As a result, we are able to estimate target bids and strategy-class membership probabilities that are consistent with a wide range of possible CH models. While this set up cannot by itself distinguish finely between specific cases, it represents a benchmark CH

model for testing the assumptions of the CH framework generically.

3 DATA AND ANALYSIS

Finally, we describe the new α -beauty contest data we have collected, all the formal details of our model, as well as our analysis and conclusions. We begin by describing the data collection method. We then perform a Bayesian test for CH behavior based on a posterior hold out log likelihood measure, similar to a Bayes factor.

3.1 The α -beauty survey

Our α -beauty contest was played among over 300 internet respondents recruited by a third-party survey provider. Each participant was asked to play the beauty contest for six values of $\alpha \in \{0.05, 0.1, 0.25, 0.5, 0.75, 0.95\}$ (presented in a random order and with no feedback). Data from this experiment (described in greater detail in Appendix A) is represented in the following figures.

Several features of the data stand out immediately from these plots. First, that players exhibit substantially randomness in their bidding and/or that the vast majority of players are level-0 players. Second, we see, perhaps, hints of monotonically increasing mean bidding for some subset of the population, as indicated in the cluster of bids fanning out from near zero at $\alpha = 0.05$. Quantifying these impressions is one goal of our analysis.

3.2 Posterior Inference

Central to our computational algorithm from estimating the SPCH model is a $K \times J$ matrix of target bids, which we denote \mathbf{T} . Each column represents a game, with increasing values of α from left to right. Each row represents a strategy class, with increasing strategy levels going down the rows. Therefore, if the matrix \mathbf{T} has the property that its entries increase

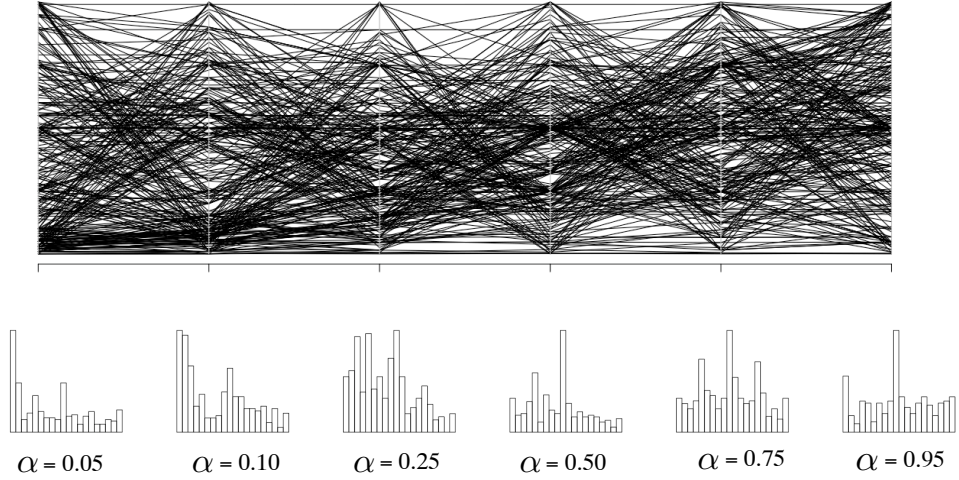


Figure 4: Six vertical lines mark the bidding distribution at the α level of the corresponding histogram. Line segments link players' bids across the various values. The bidding behavior across rounds appears largely haphazard.

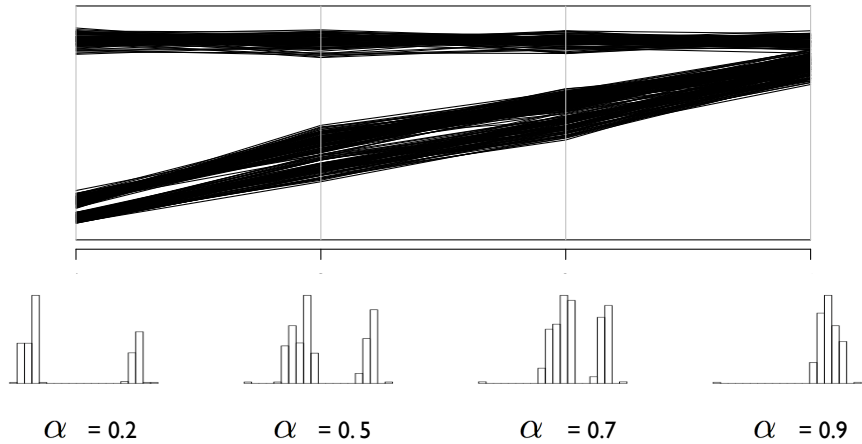


Figure 5: By contrast, simulated data drawn from a CH-Poisson model (with $\tau = 1$, Beta errors and a level-0 mean play of 0.85), exhibits clear structure, with clustering of bids that are consistent across α levels and a general upward trend of those clusters as α increases.

from left to right across each row and decrease going top to bottom down each column we may associate each entry T_{kj} with the target bid of a level- k player at the j th smallest value of α . Additionally, the entries of \mathbf{T} must lie within the unit interval to correspond to the (normalized) range of allowed bids in a beauty contest game.

We construct \mathbf{T} by first building a matrix \mathbf{C} with the relevant order restrictions, but which has elements on the real line.

1. Set element $C_{1J} = c$. This will be the largest value of \mathbf{C} .
2. Generate a decreasing sequence of numbers, beginning with c , by cumulatively subtracting arbitrary positive numbers, which we can denote by s_1, \dots, s_{J+K-2} . This sequence represents the first row and first column of \mathbf{C} , filling in from right to left along the first row and then down along the first column.
3. To create the remaining elements of \mathbf{C} , beginning with C_{22} , apply the following definition

$$C_{k+1,j+1} \equiv \phi_{k,j} C_{k,j+1} + (1 - \phi_{k,j}) C_{k+1,j} \quad (15)$$

where $\phi_{k,j} \in [0, 1]$ so that the remaining entries of \mathbf{C} are all convex combinations of the elements immediately above and immediately to the right.

A simple inductive argument shows that this construction maintains the required orderings. To arrive at \mathbf{T} we just set $\Phi(\mathbf{C}) \equiv \mathbf{T}$, where Φ is the Gaussian CDF⁷.

The following toy example helps illustrate how \mathbf{T} is built from the elements of $\boldsymbol{\theta} = (c, \mathbf{s}, \boldsymbol{\phi})$. Let $J = 3$ and $K = 2$, and set $c = 2$, $s_1 = 0.2$, $s_2 = 0.1$, $s_3 = 0.7$, $\phi_{1,1} = 0.3$ and

⁷The Gaussian CDF appears here simply as a mapping from the real line to the unit interval. There is no statistical motivation behind this choice; other mappings would have been comparably suitable.

$\phi_{1,2} = 0.9$. These values yield

$$\begin{aligned}
C_{1,3} &= 2 \\
C_{1,2} &= C_{1,3} - S_1 = 1.8 \\
C_{1,1} &= C_{1,3} - S_1 - S_2 = 1.7 \\
C_{2,1} &= C_{1,3} - S_1 - S_2 - S_3 = 1 \\
C_{2,2} &= \phi_{1,1}C_{1,2} + (1 - \phi_{1,1})C_{2,1} = 1.24 \\
C_{2,3} &= \phi_{1,2}C_{2,2} + (1 - \phi_{1,2})C_{1,3} = 1.316.
\end{aligned}$$

For this example, then, we have

$$\mathbf{T} = \Phi \left(\begin{bmatrix} 1.7 & 1.8 & 2 \\ 1 & 1.24 & 1.316 \end{bmatrix} \right) = \begin{bmatrix} 0.9554 & 0.9641 & 0.9772 \\ 0.8413 & 0.8925 & 0.9059 \end{bmatrix}.$$

With \mathbf{T} in hand, we have the strategy-level-specific Beta distributions' means for each value of α , so what remains is to specify the variance. We parametrize this feature of the model with a strategy-level-specific parameter $\nu_k \in [0, 1]$ which is the fraction of the maximum possible variance of a Beta distribution with a given mean. Throughout, we will parametrize the Beta distribution this way, in terms of mean T and variance v . The usual shape and scale parameters can be recovered by a straightforward calculation. If $y \sim \text{Beta}(a, \beta)$ it follows that $E(y) = \frac{a}{a+\beta} \equiv \mu$ and $\text{Var}(y) = \frac{a\beta}{(a+\beta)^2(a+\beta+1)} \equiv v$ for $a > 0$ and $\beta > 0$. From these equations we may deduce that

$$\begin{aligned}
a &= \frac{\mu^2(1-\mu)}{v} \\
\beta &= \frac{a(1-\mu)}{\mu} = \frac{\mu^2(1-\mu)^2}{v\mu}.
\end{aligned}$$

As per condition SP5 we assume the indicator variable γ_i is drawn independently (for each player i) from a multinomial distribution with unknown probabilities $\boldsymbol{\pi}$, which are given a Dirichlet prior distribution. Finally, we may write our likelihood conditional on γ_i (Tanner and Wong, 1987) as

$$f(\mathbf{b}_i \mid \gamma_i = k, \mathbf{T}) = \prod_{j=1}^J \text{Beta}(b_i(\alpha_j) \mid T_{k,j}, v_{k,j}) \quad (16)$$

$$\gamma_i \mid \boldsymbol{\pi} \sim \text{MN}(\boldsymbol{\pi}) \quad (17)$$

so that integrating over $\boldsymbol{\gamma}$ yields the likelihood in terms of $\boldsymbol{\pi}$ as in (14):

$$\sum_{k=1}^K \pi_k \left(\prod_{j=1}^J \text{Beta}(b_i(\alpha_j) \mid T_{k,j}, v_{k,j}) \right). \quad (18)$$

Our prior on \mathbf{T} is somewhat less straightforward, using an induced prior on the so-called “working” parameters $\boldsymbol{\theta} = (c, \mathbf{s}, \boldsymbol{\phi})$ (Ghosh and Dunson, 2009; Gelman, 2006; Meng and Van Dyk, 1999). The utility of this formulation is that elements of this parameter can be independent of one another and still satisfy the necessary order restrictions on \mathbf{T} . Specifically, it permits us to write our prior on \mathbf{T} as

$$\Pr(\mathbf{T} \in \Omega_T) = \int_{\Omega_T(\boldsymbol{\theta})} p(c) \prod_{h=1}^{(J-1)(K-1)} p(\phi_h) \prod_{q=1}^{J+K-2} p(s_q) d\boldsymbol{\phi} d\mathbf{s} dc. \quad (19)$$

where $\Omega_T(\boldsymbol{\theta})$ is understood to be the region of $\boldsymbol{\theta}$'s support such that $\mathbf{T}(\boldsymbol{\theta}) \in \Omega_T$. As a practical matter, (19) may be difficult to compute. For inferential purposes, however, our sampling chain can be defined in terms of $\boldsymbol{\theta}$ directly. Though the individual elements of $\boldsymbol{\theta}$ are unidentified, our posterior samples of the elements of \mathbf{T} will be identified.

Choosing priors for the working parameters was done by first picking the distributional forms of these parameters on the basis of convenience and then selecting hyperparameter

values so as to produce draws from the prior predictive distribution that looked, to the eye, like what we would expect from a cognitive hierarchy model. Example draws are shown below. For completeness, the priors used on the remaining elements of the SPCH model are as follows:

$$\begin{aligned}
\xi_k &\sim \text{N}(-5/4, 2/3) \\
\nu_k &= \Phi(\xi_k) \\
c &\sim \text{N}(1, 1/5) \\
s_h &\sim \text{N}((J + K - 1)^{-1}, 1/5) \\
\phi_q &\sim \text{U}(0, 1).
\end{aligned}$$

We draw our posterior samples of \mathbf{T} using a Gibbs sampler, where each full conditional is drawn using a random walk Metropolis-Hastings algorithm. In sketch, this algorithm can be described by the following steps:

1. One element at a time, draw a candidate replacement θ_j^* from the random walk proposal distribution and form $\boldsymbol{\theta}^*$.
2. From this single-element change, generate $\mathbf{T}^* = \mathbf{T}(\boldsymbol{\theta}^*)$.
3. Accept this draw with probability proportional to

$$\frac{\prod_{i=1}^N \left(\prod_{j=1}^J \text{Beta}(b_i(j) \mid T_{\gamma_i,j}, v_{\gamma_i,j}) \text{SPCH}(T_{\gamma_i,j}, v_{\gamma_i,j}) \right)}{\prod_{i=1}^N \left(\prod_{j=1}^J \text{Beta}(b_i(j) \mid T_{\gamma_i,j}^*, v_{\gamma_i,j}) \text{SPCH}(T_{\gamma_i,j}^*, v_{\gamma_i,j}) \right)}. \quad (20)$$

4. If accepted, set $\mathbf{T} = \mathbf{T}^*$.

We sample $\boldsymbol{\nu}$ by a similar procedure; conjugate Gibbs updates are available for $\boldsymbol{\gamma}$ and $\boldsymbol{\pi}$.

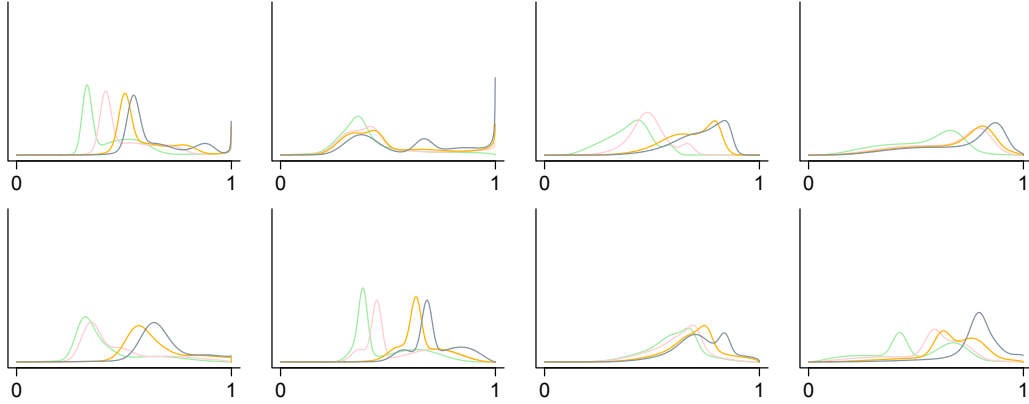


Figure 6: These draws from the SPCH prior demonstrate the key feature of evolving together to maintain the relevant order restrictions on the target bids across four levels of α . Each panel shows a single four-component ($K = 4$) mixture density over four values of α ascending from green to pink to orange to gray.

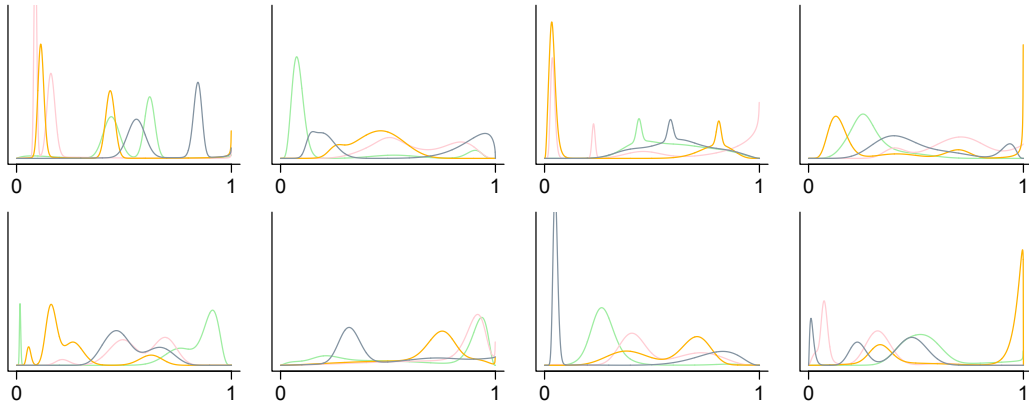


Figure 7: By contrast, these draws from the null latent class distribution clearly display non-order-restricted cluster means.

3.3 Results

3.3.1 Model Comparison

The main objective of our analysis is to contrast our flexible SPCH model to an appropriate null model to ascertain whether there is any evidence for CH play. For this task we use an unrestricted latent class mixture model, identical to the SPCH model, only less the order restrictions on of the target bids T (appearing in the model via the means of the class-specific

Beta distributions over bids). Such a model allows dependence between bids across α , but this dependence does not have to be consistent with the provisions of a CH model.

We evaluate the competing models by considering log likelihood scores of hold out data. We hold out all six bids (one for each level of α) of 30 randomly selected individuals for each model. We repeat this process for 10 such randomly selected subsets. The use of the log likelihood permits us to evaluate the shape of the density. Approaches relying only on distance from the winning bid are too coarse-grained, in that they are unable to distinguish between two models with the same mean, no matter how dissimilar they otherwise are.

The hold out data approach inherently enforces a complexity penalty; intuitively, a too-complex model will tend to overfit the in-sample training data and so do relatively poorly on out-of-sample test data. In our case, if the data-generating mechanism were in fact a CH strategy, then the model that assumes this during the estimation phase should outperform the more flexible unrestricted model, which has more propensity to be led astray by noise artifacts. This measure is conceptually similar to a Bayes factor, the main difference being that we use some portion of the data to fit the model first and then integrate over the resulting posterior; Bayes factors integrate directly over the prior. In cases, such as ours, where the specifics of the prior distribution are uncertain or unmotivated, this step ensures that conclusions are less sensitive to initial prior specification (Berger and Pericchi, 1996). Similarly, a sensitivity analysis can be performed, duplicating the analysis under slightly different priors. While we conducted no systematic study in this regard, we did confirm that our basic conclusions were insensitive to various specifications generating similar prior predictive distributions.

Our results, reported in table 1, are unambiguous: the non-CH model performs better than the CH-P or the level- k model, while the SPCH model outperforms those models and also the non-CH null model⁸. Thus we see strong evidence for CH style play, but not play

⁸All models used Beta error distributions with the lowest strategy class mean free to vary; in other words

Table 1: Model Comparisons

Model	Log-Marginal (hold out) Likelihood
Uniform	0
level- k	49.6
CH-Poisson	55.8
SPCH	76
Null Latent-class mixture	63.5

specifically consistent with the popular simpler models, which do no better than the non-CH model.

3.3.2 Posterior Summaries

An additional benefit of the MCMC approach is the ability to examine interesting posterior quantities, such as the modal class membership. This provides us with a peek into how the player population may be partitioned. By looking at the observed bidding patterns isolated by these estimated class memberships we can hope to see CH-style reasoning in action. The story that emerges here is that while the CH assumption buys some predictive accuracy, the “crispness” of the model – how near to their optimal target bids people play – is weak. On the whole we observe CH trends, but the noise level about this trend is substantial; there appears to be a general upward trend with increasing α , as the CH model demands, but this tendency is clearly violated by many individual sets of bids.

Similarly we obtain a posterior mean for $\boldsymbol{\pi}$ of [0.0206 0.2973 0.3237 0.3584], suggesting that while a four-class model was fit, a three class model would likely suffice. This question could be taken up explicitly with slight modifications, by moving from a finite Dirichlet based model to a Dirichlet process based model (Escobar and West, 1995).

A most interesting finding is that the lowest strategy class exhibits a bimodal strategy,

SP7 was used for all models.

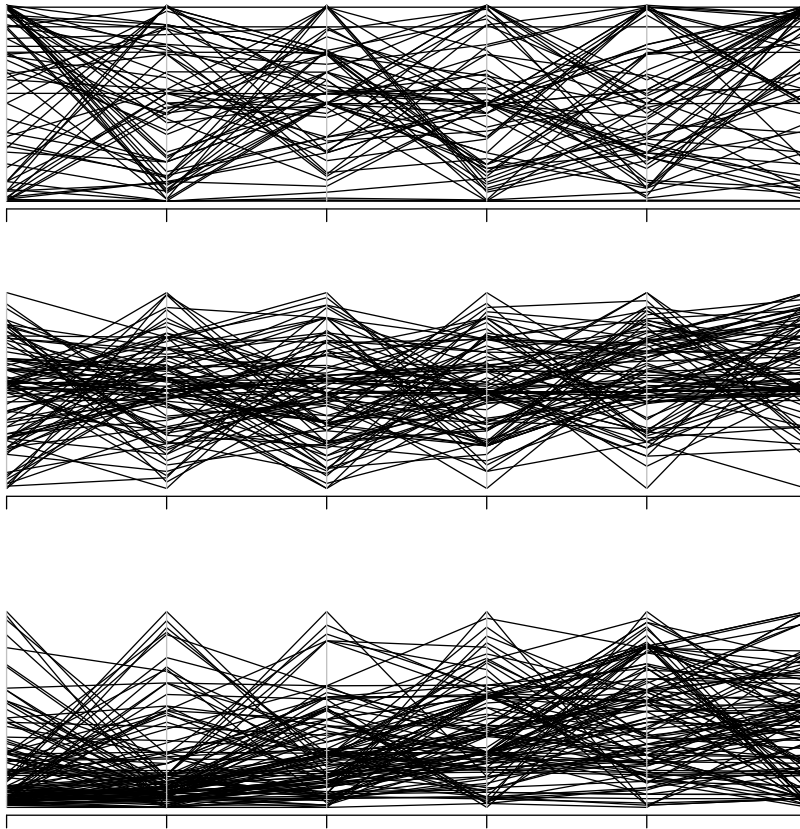


Figure 8: After fitting a four-class SPCH model, we can partition the player population by estimated modal class membership. This results in three populated strategy classes. Qualitatively this corresponds to a random class, and one and two step-ahead thinkers.

playing with high probability near the boundary of the interval, up near one or down near zero. Observations such as this could conceivably motivate new theories, CH or otherwise. In this case, the patterned play of the “random” class may be well described by an anchoring effect, where the endpoints have irrational psychological pull.

4 DISCUSSION

To be clear, our objective here has not been to develop a new theory of player behavior in beauty contest games, nor to conduct a “horse race” between the CH-Poisson and level- k

models. Rather our purpose is to explore, in a data-first fashion, the plausibility of existing theories as generically as possible.

Generally, we expect that any model allowing heterogeneity in strategic behavior will fit the data better than a model which does not. This is why CH models are trivially a descriptive improvement over Nash equilibrium models. To use this observation as evidence in favor of a given model without further scrutiny is to invite dramatic misevaluation of that model's descriptive power. Within the class of CH models, particular variations may be more or less accurate, such that comparing them pairwise is literally an exhausting task. The approach taken here allows testing the common CH assumptions used by all of these models simultaneously. This approach permits us to build confidence in a model of strategic behavior by judging its descriptive power relative to a null model of *greater* expressivity. It is the increased predictive accuracy of a more constrained model relative to a less constrained one that builds faith in the validity of those constraints. In the case of the SPCH model, this means comparing it to a less constrained null model, which is a latent class mixture model without the hallmark ordering restrictions of a CH model.

Our comparison has shown that a model which permits only bidding behavior *consistent with* a CH model outperforms a model without such a restriction. However, because the beauty contest does not require a player-level model to generate a bid, this evidence alone is insufficient to rule out non-CH theories – auxiliary information would be needed, as in Crawford and Iriberri (2007). However, had the non-CH model outperformed the very general SPCH model, auxiliary information would have been unnecessary; this is a key advantage of building highly generalized strategic models for testing purposes.

Our other main finding is that the CH-P and level- k models do not perform better than a non-CH latent class model for our new beauty contest data. So, while the SPCH model's good performance on the beauty contest game does not alone endorse it as a suitable model for more general games, the fact that the CH-Poisson and level- k models do not perform well

in this narrow context does rule out their candidacy as default models. Put another way, a necessary condition to be the standard bearer of cognitive hierarchy models is to accurately describe game play in this quintessential example.

The success of the SPCH model does encourage us to explore new CH variants, however. A natural next step would be to investigate alternative theory-motivated CH submodels that do better than the more general SPCH model on holdout evaluation tasks. The posterior summaries from our analysis can serve as an ideal launching point for generating such alternative theories, as they effectively quantify first impressions of the data or previous intuitions from the literature.

For example, our results suggest that strategy clustering may be a result of a simple priming effect. In our study we randomized the order of α so as to avoid an *order* effect, where the observed data patterns are driven by the relative order of the α 's. We may well still be seeing an *anchoring* effect (Tversky and Kahneman, 1974), however, meaning that a player's strategy may be dictated by which value of α they are first presented with. One notices that the bids, when grouped by modal class membership, tend to fall into low, high and medium clusters, which represent plausible anchor values at the high, low and middle regions of the allowable range.

Second, one might try to employ covariates to isolate membership in a given strategy class. This would be a formalization of the sort of *post-hoc* correlation analysis that has already been conducted on attributes such as education or profession (Chong et al., 2005). By incorporating these aspects directly into the model, we may avoid the fallacious over-interpretation that often accompanies latent variable models in general and mixture models in particular (Bauer and Curran, 2003). Our model has attempted to remedy this tendency by enforcing the implications of a CH model across values of α . Covariates would further strengthen the analysis. Experimental side information about the steps ahead of thinking is of course the gold standard in this regard (Crawford, 2007), though comparatively hard to

come by.

Finally, it would be intriguing to see how much predictive advantage follows from abandoning a player-level conception of player reasoning. Because the winning bid in the beauty contest game is a function of aggregate play only, it seems plausible that reasoning may not proceed from the “micro” level at all. In this spirit, we can cast the problem as a random effects regression model. Specifically, combining this approach with the anchoring hypothesis may be fruitful. On such a model, each player’s strategy would consist of first selecting an anchor value from a random effects distribution, then choosing the remaining bids for different α in an autoregressive fashion so as to maintain (approximate) self-consistency. Together with a contamination model or a screening process as in Stahl and Wilson (1995) for capturing those players that do not understand the rudiments of the game (chiefly, the role of varying α), this approach could yield a highly descriptive model of actual game play.

Our analysis suggests we would do well to resist the appeal of analytically tidy CH variants like the CH-P and level- k unless they describe the data adequately. Any single beauty contest (for a specific value of α) can lead to the appearance that a CH-P or level- k model is a suitable fit to the data. By looking simultaneously at multiple α values we find that neither model is a convincing description of the data.

APPENDIX A: DATA COLLECTION DETAILS

Web participants were presented with the following text:

Should you decide to participate and complete the survey, in addition to your compensation from the panel company you will have the opportunity to win up to \$300.00. Specifically, each respondent will play a “move” in each of six games (to be described). In each game, one award of \$50 will be given to a player who makes a winning response. In the event of a tie, each of the players who submit a winning move will be entered in a raffle to win the \$50 prize for that round. The winning move depends on the play of all respondents.

You will be asked to play six games in this study.

In each of the six games, every player (yourself and all others responding to the survey at any point during the study) will choose a (real) number between 0 and 100. You are likely to be playing against a large number of players.

We will then compute the average number chosen by all respondents in each game.

The aim of the game is to pick a number that is closest among all respondents to a pre-specified percentage of the average response. This pre-specified percentage will be given in the questions below and will vary from question to question.

All in all, you will play this game six times (each time using a different percentage), hence the chance to win \$300.

Participants who agreed to participate were then presented with the following, for each of the six values of α (shown here for $\alpha = 0.95$):

The objective of this game is to select a number which is closer than anyone else’s to 95% of the average number chosen by all persons responding to this question. If

the average response is some number “X” and you select 0.95 times that number, then you win.

It should be noted that the payoff function used here is *not* the squared distance from the true target as described in the previous section. For practical reasons we were unable to offer payment to all players and were forced to resort to a raffle system. It may have been more elegant to enter players in a lottery with a chance to win proportional to their squared distance from the underlying target, which would have preserved the mean as their expected-payoff maximizing play, but we had to weigh this against the added layer of difficulty associated with having individuals reason explicitly about their odds of winning. That said, we conjecture that players’ bidding would be little affected by such a modification.

APPENDIX B: A BRIEF NOTE ON LEARNING

An important aspect of behavioral game theory that we have intentionally omitted here is a theory of learning across repeated games. Our main point – that a minimal condition for responsibly interpreting the parameters of a statistical model is reasonable fidelity to the data – stands separately from the repeated learning scenario, applying with full force to the one-shot game setting because, as Stahl and Wilson (1995) put it, initial, as opposed to learned, responses are “crucial to whatever learning follows.” Incorporating a learning component to our study of the α -beauty contest data would demand a dramatically more complex model, first because knowledge of the winning bid is by itself insufficient information to update one’s belief distribution and second because knowing that the other players are also going to update their beliefs means that players must additionally have a theory about how this updating occurs. Though well beyond the scope of our work, developing flexible models like the SPCH to test theories of strategic *learning* would surely be an interesting extension.

REFERENCES

- D. J. Bauer and P. J. Curran. Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, 8(3):338–363, 2003.
- J. O. Berger and L. R. Pericchi. Intrinsic Bayes factor for model selection and prediction. *Journal of the American Statistical Association*, 91, 1996.
- A. Bosch-Domènech, R. Nagel, A. Satorra, and J. G. Montalvo. One, two, (three), infinity: Newspaper and lab beauty-contest experiments. *American Economic Review*, 92(5):1687–1701, 2002.
- A. Bosch-Domènech, J. G. Montalvo, R. Nagel, and A. Satorra. Finite mixture analysis of beauty-contest data using generalised beta distributions. *Experimental Economics*, Forthcoming, 2010.
- C. Camerer. *Behavioral Game Theory*. Princeton University Press, 2003.
- C. F. Camerer, T.-H. Ho, and J.-K. Chong. A cognitive hierarchy model of games. *Quarterly Journal of Economics*, 119(3):861–898, 2004. doi: 10.1162/0033553041502225. URL <http://www.mitpressjournals.org/doi/abs/10.1162/0033553041502225>.
- J. K. Chong, C. Camerer, and T.-H. Ho. Cognitive hierarchy: A limited thinking theory in games. In *Experimental Business Research, Volume III: Marketing, Accounting and Cognitive Perspectives*, chapter 9. Kluwer Academic Press, 2005.
- M. A. Costa-Gomes and V. P. Crawford. Cognition and behavior in two-person guessing games: An experimental study. *American Economic Review*, 96(5):1737–1768, December 2006. URL <http://ideas.repec.org/a/aea/aecrev/v96y2006i5p1737-1768.html>.

- M. A. Costa-Gomes, V. P. Crawford, and B. Broseta. Cognition and behavior in normal-form games: An experimental study. *Econometrica*, 69(5):1193–1235, September 2001.
- V. P. Crawford. Let’s talk it over: Coordination via preplay communication with level- k thinking. Levine’s Bibliography 12224700000001449, UCLA Department of Economics, Sept. 2007. URL <http://ideas.repec.org/p/cla/levrem/12224700000001449.html>.
- V. P. Crawford and N. Iriberri. Level- k auctions: Can a non-equilibrium model of strategic thinking explain the winner’s curse and overbidding in private-value auctions? *econometrica*, 75:1721–1770, 2007.
- E. De Giorgi and S. Reimann. The $[\alpha]$ -beauty contest: Choosing numbers, thinking intervals. *Games and Economic Behavior*, 64(2):470–486, November 2008. URL <http://ideas.repec.org/a/eee/gamebe/v64y2008i2p470-486.html>.
- M. D. Escobar and M. West. Bayesian density estimation and inference using mixtures. *Journal of American Statistical Association*, 90(430), 1995.
- A. Gelman. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–533, 2006.
- J. Ghosh and D. B. Dunson. Default prior distributions and efficient posterior computation in Bayesian factor analysis. *Journal of Computational and Graphical Statistics*, 18(2):306–320, 2009.
- E. Haruvy and D. O. Stahl. Level- n bounded rationality and dominated strategies in normal-form games. *Journal of Economic Behavior & Organization*, 66(2):226–232, 2008.
- E. Haruvy, D. O. Stahl, and P. W. Wilson. Modeling and testing for heterogeneity in observed strategic behavior. *Review of Economics and Statistics*, 83(1):146–157, 2001.

- T.-H. Ho, C. Camerer, and K. Weigelt. Iterated dominance and iterated best response in experimental “*p*-beauty contests”. *American Economic Review*, 88(4):947–969, September 1998.
- E. Johnson, C. F. Camerer, S. Sen, and T. Rymon. Detecting failures of backward induction: Monitoring information search in sequential bargaining. *Journal of Economic Theory*, 104(1):16–47, 2002.
- J. M. Keynes. *The general theory of interest, employment and money*. Macmillan, London, 1936.
- X.-L. Meng and D. A. Van Dyk. Seeking efficient data augmentation schemes via conditional and marginal augmentation. *Biometrika*, 86(2):301–320, June 1999.
- H. Moulin. *Game Theory for Social Sciences*. New York University Press, 1986.
- R. Nagel. Unraveling in guessing games: An experimental study. *American Economic Review*, 85(5):1313–26, December 1995. URL <http://ideas.repec.org/a/aea/aecrev/v85y1995i5p1313-26.html>.
- A. Newell and H. Simon. *Human Problem Solving*. Prentice Hall, 1972.
- D. I. Stahl and P. W. Wilson. Experimental evidence on players’ models of other players. *Journal of Economic Behavior & Organization*, 25(3):309–327, December 1994. URL <http://ideas.repec.org/a/eee/jeborg/v25y1994i3p309-327.html>.
- D. O. Stahl and P. W. Wilson. On players’ models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10:218–254, 1995.
- M. A. Tanner and W. H. Wong. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540, June 1987.

A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131, September 1974.