Comparing Models of Strategic Thinking
in Van Huyck, Battalio, and Beil's Coordination Games

By Miguel A. Costa-Gomes, Vincent P. Crawford, and Nagore Iriberri[1]

Revised 23 August 2008

Abstract: This paper compares the leading models of strategic thinking with subjects' initial responses to Van Huyck et al.'s (1990, 1991; "VHBB") coordination games. The data favor models in which players treat their partners' decisions as correlated rather than independent. Among the equilibrium selection criteria we compare, payoff-dominant equilibrium fits better than risk-dominant equilibrium or a simple maximin model. Among the individualistic models we compare, level-*k* and cognitive hierarchy models perform best overall, usually fitting better than noisy introspection and logit quantal response equilibrium models. In VHBB's games, however, payoff-dominant equilibrium usually fits better than any individualistic model.

Keywords: noncooperative games, experimental economics, behavioral game theory, level-*k* models, cognitive hierarchy, quantal response equilibrium, noisy introspection, coordination games, equilibrium selection

JEL codes: C72, C92, C51

---

[1]Costa-Gomes: Department of Economics, University of Aberdeen Business School, University of Aberdeen AB24 3QY, United Kingdom, e-mail m.costagomes@abdn.ac.uk; Crawford: Department of Economics, University of California, San Diego, 9500 Gilman Drive, La Jolla, California, 92093-0508, e-mail vcrawfor@dss.ucsd.edu. Iriberri: Departament d´Economia i Empresa, Universitat Pompeu Fabra, Barcelona, Spain, e-mail: nagore.iriberri@upf.edu. We are grateful to Colin Camerer, Jacob Goeree, Charles Holt, and Thomas Palfrey for helpful discussions. Iriberri thanks Ministerio de Educación y Ciencia (SEJ2006-05455 and SEJ2007-64340) and Barcelona Economics-Xarxa CREA for research support.

1. Introduction

Many applications of game theory involve settings where players have had enough experience with analogous games to make equilibrium a reasonable assumption. If only long-run outcomes matter and convergence and equilibrium selection do not depend on the details of learning, such applications can rely entirely on equilibrium. Because the cognitive requirements for learning to converge to equilibrium in a stationary setting are mild—even reinforcement learning, in which players need not even know they are playing a game, usually suffices—there is then no need for a deeper understanding of strategic thinking.

Many other applications involve games played without clear precedents in which initial outcomes matter. Such applications, which include most questions involving comparative statics or mechanism design, depend on predicting initial responses to games even if eventual convergence to equilibrium is assured. In other applications, convergence to equilibrium is assured and only long-run outcomes matter, but the equilibrium is selected from multiple equilibria via history-dependent learning dynamics (Van Huyck et al. 1990, 1991 ("VHBB"), Crawford 1995). Such applications also depend on predicting initial responses, and may depend on the structure of players' learning rules as well.

The cognitive requirements for initial responses to be in equilibrium are far more stringent than the requirements for learning to converge to equilibrium: Players must have perfectly coordinated beliefs, which without precedents on which to base them requires players to have accurate models of each other's decisions (or at least their probability distributions). It is easy to imagine strategic thinking being this accurate in simple games such as those that are dominance-solvable in a very small number of rounds. But the thinking required for equilibrium initial responses in more complex games is often behaviorally far-fetched: Even players who are capable of such thinking may doubt that others are capable of it, or doubt that others believe others are capable of it. Moreover, there is a growing body of laboratory evidence that initial responses often deviate systematically from equilibrium, especially when it requires thinking that is not straightforward.

As Costa-Gomes and Crawford (2006; henceforth "CGC") note, modeling initial responses more accurately promises several benefits. It can establish the robustness of the conclusions of equilibrium analyses in games where boundedly rational rules mimic equilibrium, and challenge the conclusions of applications to games where equilibrium is implausible without learning. It can resolve empirical puzzles by explaining the systematic deviations from equilibrium some games evoke. More generally, it can yield insights into cognition that elucidate other aspects of strategic behavior, including the structure of learning rules, where assumptions about cognition determine which analogies between current and previous games players recognize and sharply distinguish reinforcement from beliefs-based and more sophisticated rules.

A variety of models have been proposed to describe experimental subjects' initial responses to games. These models normally allow players' responses to be in equilibrium, but do not assume it. They include simply adding noise to equilibrium predictions ("equilibrium plus noise"); McKelvey and Palfrey's (1995) notion of quantal response equilibrium ("QRE") and its leading special case, logit QRE ("LQRE"); the level-*k* models of Nagel (1995), Stahl and Wilson (1995), Ho et al. (1998), Costa-Gomes et al. (2001), and CGC; Camerer et al.'s (2004; "CHC") closely related cognitive hierarchy ("CH") model; and Goeree and Holt's (2004; "GH") model of noisy introspection ("NI").

Level-*k*/CH models have now been compared with LQRE in several experimental datasets (Chong et al. 2005, Crawford and Iriberri 2007ab, Camerer et al. 2007) and at least one field setting (Östling et al. 2008). In most cases level-*k*/CH models have better fits, but the results have been suggestive rather than conclusive. To our knowledge NI models have only been compared with other models in 2×2 or 3×3 games, and only with equilibrium, LQRE, and a single-type level-*k* model (GH, Costa-Gomes and Weizsäcker, 2008), or a non-logit strengthening of QRE called regular QRE (Goeree et al. 2005).

This paper brings additional evidence to bear on the comparison of equilibrium plus noise, LQRE, level-*k*/CH, and NI models, analyzing subjects' initial responses to the several different games in VHBB's (1990, 1991) famous coordination experiments. The variety and structural simplicity of VHBB's games and their larger strategy spaces (seven decisions rather than the usual two or three) allow more informative tests. VHBB's data also shed light on the important but seldom studied issue (but see Ho et al. 1998) of whether people playing *n*–person games take the independence of others' responses into account. Finally, they allow us to consider how the leading models of initial responses address the issue of equilibrium selection, and how they fare in comparison to coordination refinements such as risk- or payoff-dominance.

The rest of the paper is organized as follows. Section 2 reviews the leading models of strategic thinking and discusses their strengths and weaknesses. Section 3 introduces VHBB's games and uses their data to compare the models.

2. Alternative Models of Initial Responses to Games

Until recently, the choices for modeling non-equilibrium initial responses to games were limited. Any notion that is to be taken to data must allow for errors in some way. The most obvious choice, adding mean-zero noise with a specified distribution and an estimated precision parameter to equilibrium predictions ("equilibrium plus noise"), sometimes does well. However, even in games with unique equilibria, equilibrium plus noise often misses systematic patterns in subjects' deviations from equilibrium, which tend to be sensitive to out-of-equilibrium payoffs in patterns that it cannot account for. And in games with multiple equilibria, particularly VHBB's where every feasible decision is part of some symmetric pure-strategy equilibrium, equilibrium plus noise is incomplete in that it does not specify a unique (even though probabilistic) prediction conditional on the value of its behavioral parameters (in this case, the precision). Such multiplicity has previously been dealt with by estimating an unrestricted probability distribution over the equilibria (Bresnahan and Reiss 1991), but such a model very badly overfits VHBB's data. To put equilibrium plus noise on a more equal footing with the other models considered here, which are complete in the above sense, we consider two natural variants, risk-dominant equilibrium ("RDE") and payoff-dominant equilibrium ("PDE") plus noise, based on Harsanyi and Selten's (1988) refinements.[2] We also consider maximin (VHBB's "secure") decisions, which VHBB gave a prominent role, and which functions like an equilibrium refinement in these symmetric games.

To capture the payoff-sensitivity of deviations from equilibrium, McKelvey and Palfrey (1995) proposed the notion of QRE, in which players' decisions are noisy, with the probability density of each decision increasing in its expected payoff, evaluated taking the noisiness of others' decisions into account. A QRE is then a fixed point in the space of decision probability distributions, with each player's distribution a noisy best response to the others'. As the distributions' precision increases, QRE approaches equilibrium; and as it approaches zero, QRE approaches uniform randomization over players' feasible decisions. A QRE model is closed by specifying a response distribution, which is logit in almost all applications. The resulting logit QRE ("LQRE") implies error distributions that respond to out-of-equilibrium payoffs, often in plausible ways.[3] In applications LQRE's precision is estimated econometrically or calibrated from previous analyses. With estimated precision, LQRE often fits subjects' initial responses better than an equilibrium model (McKelvey and Palfrey 1995, Goeree et al. 2002, Weizsäcker 2003).

From the point of view of describing strategic thinking, LQRE's fit comes at a cost: Players must not only respond to a nondegenerate probability distribution of other players' responses but also find a generalized equilibrium that is a fixed point in a large space of response distributions. If equilibrium reasoning is cognitively taxing, LQRE reasoning is doubly taxing. Further, the mathematical complexity of LQRE means that it must almost always be solved for computationally and is not easily adapted to analysis. Finally, in some settings LQRE fits worse than equilibrium (Camerer et al. 2007, Chong et al. 2005, Crawford and Iriberri 2007a), sometimes even making systematic qualitative errors (Crawford and Iriberri 2007b, Östling et al. 2008).

Motivated by these considerations and experimental evidence, a different vein of work on strategic thinking considers models that treat deviations from equilibrium as an integral part of the structure rather than as errors or responses

---

[2] Haruvy and Stahl (2007) take an approach that is similar in spirit to 3×3 normal-form games.

[3] Haile et al. (2008) have shown that the distributional assumptions are crucial, in that with an unrestricted distribution QRE can "explain" any given dataset. The use of the logit distribution has been guided more by fit, custom, and choice axioms than independent evidence.

to errors. Although the number of possible non-equilibrium structures seems daunting, much of the experimental evidence supports a particular class of models called level-*k* or cognitive hierarchy ("CH") models, which alleviate the cognitive and computational complexity concerns mentioned above.

The flavor of this evidence is illustrated by Nagel's (1995) results for *n*-person guessing games. Her games are dominance-solvable in infinite numbers of rounds, so that equilibrium requires "only" iterated knowledge of rationality, with no further restrictions on beliefs. But her subjects never played their equilibrium strategies initially, and their response distributions resembled neither equilibrium plus noise nor LQRE. Instead there were spikes that suggest a discrete, heterogeneous distribution of strategic thinking "types."

The spikes' locations and how they vary across treatments are consistent with two plausible interpretations. In one, subjects follow finitely iterated dominance rules in which each does *k*-1 rounds of iterated dominance for some small number, k = 1, 2, or 3, and then best responds to a uniform prior over his partner's remaining strategies. In another, subjects follow "level-*k*" rules in which each starts with a uniform prior over others' possible guesses and then iterates the best response mapping *k* times, again with k = 1, 2, or 3. In Nagel's games these rules yield identical guesses, and theorists often interpret her results as evidence that her subjects performed iterated dominance. In some more recent experiments (Stahl and Wilson 1995, Ho et al. 1998) the rules are weakly separated, and in others they are separated mostly by information search implications (Costa-Gomes et al. 2001) or elicited beliefs (Costa-Gomes and Weizsäcker 2008) rather than by their implications for decisions. In CGC's experiments, however, the rules are strongly separated by decisions (as well as search), and the results clearly favor level-*k* over iterated dominance rules.

In a level-*k* model, as suggested by these results, players' types are allowed to be heterogeneous, but each player's type is drawn from a common distribution. Type *Lk* anchors its beliefs in a nonstrategic *L0* type, which represents players' models of others' instinctive reactions to the game and is usually taken as uniformly random over the feasible strategies, and adjusts them via thought-experiments with iterated best responses: *L1* best responds to *L0*, *L2* to *L1*, and so on. Like equilibrium players, *L1* and higher types are rational, with perfect models of the game. Their only departure from equilibrium is replacing its perfect model of others with a simplified model of others. *L1* and higher types make undominated decisions, and in many games *Lk* complies with *k* rounds of iterated dominance, so its decisions are *k*–rationalizable.

In applications the population type frequencies are inferred from data-fitting exercises or calibrated from previous analyses. The estimated frequency of *L0* is normally zero or small; and the type distribution is fairly stable across games, with most weight on *L1* and *L2* (see fn. 11). Unlike LQRE, a level-*k* model's point predictions do not depend on estimated precisions, only on the estimated type frequencies. In applications it is usually assumed that *L1* and higher types make errors, which are often taken to be logit as in LQRE. However, despite the noisiness of types' decisions, a level-*k* model requires neither that players respond to nondegenerate distributions of others' responses (except *L1*'s response to *L0*, whose uniform randomness is simple to respond to) nor that they find fixed points. This simple recursive structure avoids the common criticism of LQRE that finding a fixed point in the space of distributions is too taxing for a realistic model of strategic thinking.

In CHC's closely related CH model, type *Lk* best responds not to *Lk-1* alone but to a mixture of lower-level types, and the type frequencies are treated as a parameterized Poisson distribution. Unlike in a level-*k* model, *L1* and higher types are usually assumed not to make errors; instead the uniformly random *L0*, which has positive frequency in the Poisson distribution, doubles as an error structure for the higher types. As in a level-*k* model, players need not respond to the noisiness of others' decisions (except *L0*'s) or find fixed points, but they do need to respond to a nondegenerate distribution of lower types' responses, in proportions determined by an estimated Poisson parameter. Like a level-*k* model, a CH model makes point predictions that do not depend on estimated precisions, only on the Poisson parameter. It also has a recursive structure, albeit somewhat more complex one than a level-*k* model's structure.

Like RDE, PDE, maximin, and LQRE, level-*k* and CH models are applicable to "any" game and have small numbers of behavioral parameters. Because in many games *Lk* complies with *k* rounds of iterated dominance, a distribution of level-*k* types that is realistically concentrated on low levels of *k* mimics equilibrium in games that are dominance-solvable in a few rounds, but deviates systematically in some more complex games, in predictable ways.[4] This allows level-*k* and CH models, like LQRE, to capture the sensitivity of deviations from equilibrium to out-of-equilibrium payoffs; and they often fit subjects' initial responses better than PDE or RDE. In some applications the Poisson constraint is not very restrictive (Chong et al. 2005), and the CH model fits as well as or better than a level-*k* model; but in others (CGC, Crawford and Iriberri 2007ab) that constraint is strongly binding.

Although LQRE has been the most popular model of initial responses, not all researchers consider it suitable for that purpose.[5] GH suggest using LQRE for limiting outcomes, instead proposing an NI model to describe initial responses. Their NI model relaxes LQRE's equilibrium assumption while maintaining its assumption that players respond to a nondegenerate probability

[4] Level-*k* and CH models thus provide a concrete, evidence-based way to think about the robustness of mechanisms. Because *L1* and all higher types respect simple dominance, mechanisms that implement desired outcomes in dominant strategies may have an advantage over more complex mechanisms that implement superior outcomes, but only in equilibrium.

[5] McKelvey and Palfrey (1995) suggest using LQRE for both initial responses and limiting outcomes, with increasing precision as a reduced-form model of learning. An appendix at http://dss.ucsd.edu/~vcrawfor/#VHBB discusses LQRE as a model of limiting outcomes.

distribution of other players' responses. Instead players form beliefs by iterating best responses as in a level-*k* model, but their higher-order beliefs reflect increasing amounts of noise, converging to uniform randomness. For a given noise distribution, the NI model makes probabilistic predictions that depend on how fast the noise grows. In the extreme case in which the noise does not grow with the number of iterations, NI mimics LQRE. Other extremes mimic level-*k* types: If the noise jumps immediately to ∞, NI beliefs are *L0*; if it is zero for one iteration and then jumps to ∞, NI beliefs are *L1*, and so on.[6]

In applications GH assume that the noisiness of higher-order beliefs grows geometrically with iterations, which yields beliefs similar but not identical to *Lk*'s, with slower noise growth like a higher *k*. The resulting NI model is more flexible than LQRE, and cognitively less taxing because it does not require fixed-point reasoning; but such an NI model is more taxing than a level-*k* or CH model because players' choices are indefinitely iterated best responses to noisy higher-order beliefs (although for computational purposes in applications GH truncate the iteration to ten rounds). NI's structure, like LQRE's, is not directly grounded in evidence; in fact the evidence from Nagel's (1995) and subsequent experiments suggests that the indefinite iteration of best responses and the assumed homogeneity of strategic thinking are not very realistic.

2. Van Huyck, Battalio, and Beil's (1990, 1991) coordination games

This section compares RDE and PDE, maximin, level-*k*, CH, LQRE, and NI models in VHBB's (1990, 1991) coordination games.[7] VHBB's subjects played symmetric coordination games in which they chose among seven effort levels, with payoffs determined by their own efforts and an order statistic, the minimum or median, of their own and others' efforts. We consider five of their treatments, in all of which subjects chose among efforts {1, …,7}: their 1990 "minimum" treatment A, in which groups of 14-16 subjects played games in which, denoting subject *i*'s effort x_i and the group minimum N, subject *i*'s payoff in (1987) dollars was $0.2N - 0.1x_i + 0.6$; their 1990 minimum treatment B, in which the same groups played the same games but with the cost of effort lowered to 0, making effort 7 a weakly dominant strategy; their 1990 minimum treatment C_d, in which subjects subsequently played a two-person game with the same payoff function as in treatment A, with a new, randomly selected partner each period; their 1991 "median" treatment Γ, in which groups of 9 subjects played games in which, denoting the group median M, subject *i*'s payoff was $0.1M - 0.05(M - x_i)^2 + 0.6$; and their 1991 median treatment Ω, in which subject *i*'s payoff was $0.1M + 0.6$ when $x_i = M$ but was 0 when $x_i \neq M$.[8]

---

[6] Compare Camerer et al. (2007), who also nest generalized variants of LQRE and CH models.
[7] Anderson et al. analyze limiting LQRE (as precision approaches infinity) in VHBB's (1990) minimum games, and Yi (2003) analyses limiting LQRE in VHBB's (1991) median games.
[8] Treatment C_d is best thought of as a game played by all 14-16 players in the group, evaluating expected payoffs before the uncertainty of pairing is resolved. Crawford (1995, fn. 10, p. 110)

In each case a subject's payoff was highest, other things equal, when his effort equaled the relevant order statistic, the group minimum in treatment A or B, the pair minimum in treatment C_d, or the median in treatment Γ or Ω. Any combination in which all players choose the same effort is an equilibrium; in these equilibria players' payoffs are higher, the higher the effort; and these Pareto-ranked equilibria are the only pure-strategy equilibria. Thus, all-7 is the payoff-dominant equilibrium in all the games we consider. The games are nonetheless non-trivial because there is a tension between the higher payoff of the all-7 equilibrium and its fragility, which is more extreme for minimum than median games; and for minimum games, the more players there are. As a result, the risk-dominant equilibrium is all-7 in treatments Γ, Ω, and B; all-4 in treatment C_d; and all-1 in treatment A (using Harsanyi and Selten's 1988 definition; see Crawford 1991, p. 56, fn. 27). The maximin decisions (and equilibria) are all-1 in treatments A and C_d, all-3 in Γ, and anything in B and Ω.

We focus on subjects' initial responses to each of the games they played (VHBB 1990, Tables 2 and 5; VHBB 1991, Table II; or see Crawford 1991, Table I).[9] We define Maximin, RDE and PDE plus noise, LQRE, level-*k* types, and NI with logit errors, each with estimated precision.[10]

In specifying the models for these *n*-person games, one important issue is whether players take the independence of others' efforts into account in forecasting the group minimum or median. Although independence is standard in game theory, and is normally built into all of the models compared here; there is experimental evidence that people often adopt a single model of others' choices, implicitly assuming that they are perfectly correlated (for example, Ho et al. 1998). This effectively reduces the game to a two-person game, and reduces the cognitive load. Because of the nonlinearity of the payoff functions, and the variation between two- and *n*-person versions of the "same" game, VHBB's games are ideally suited to testing for such mental simplifications. Accordingly, we consider two alternative versions of LQRE, level-*k*, NI, and CH, one in which a player views others' choices as independent, and one in which he views them as perfectly correlated. For the level-*k* and CH models, however, we take this to refer to *L0*, which is the channel by which the

---

shows that players' best responses are then given by an order statistic of the population effort distribution, which happens to be the group median for VHBB's payoffs. We omit median treatment Φ because it seemed to evoke framing effects, which none of the models considered here take adequately into account (but see Crawford and Iriberri (2007b). We omit the fixed-pairing minimum treatment C_f because it clearly elicited repeated-game effects.

9 Although each subject played a series of different games in fixed groups, the groups were large enough for subjects to treat their own influences on future choices as negligible, so that to a first approximation, their initial responses to each game can be viewed as responses to the game played in isolation. There was some evidence of order effects in later treatments, particularly in C_d, which was run last in a sequence; but these are beyond the scope of this paper's analysis.

10 Because Maximin does not (and cannot) have rational beliefs, we evaluate its expected deviation costs using the beliefs of the associated equilibrium.

correlation influences players' choices, through the higher-level types, in these models. Correlated Maximin and PDE are the same as the independent ones. Correlated RDE remains all-7 in treatments Γ, Ω, and B and all-4 in treatment C_d; and becomes all-4 in treatment A (because it makes A equivalent to C_d).

Table 1 summarizes the results of the comparisons. The left-most columns give the likelihoods of the empirical frequencies and of random frequencies, which provide upper and lower bounds on the attainable likelihoods for any model. (The upper bound is not 0, as it usually is for a perfect fit, because the estimated models all predict nondegenerate random distributions of outcomes.) In VHBB's symmetric games, for both the level-*k* and CH models, *L2* and higher types coincide with *L1*, so these models share the homogeneity of PDE, RDE, LQRE, and NI. We therefore simplify by giving only the modal actions implied by each model in each treatment, and comparing fits by likelihoods without reporting type frequencies or other parameter estimates.[11]

The results in Table 1 suggest several conclusions. First, the correlated versions of the models almost always do as well or better than their independent counterparts (the exceptions are level-*k* in treatment B and level-*k*, LQRE, and NI in Γ). In these games few subjects' thinking reflects the independence of their partners' decisions, despite its importance in treatment A.

Second, among the equilibrium selection criteria Maximin, PDE, and RDE, PDE always fits at least as well as the others, and often better. Third, among the individualistic models LQRE, level-*k*, CH, and NI, level-*k* and CH perform comparably well: each wins 4 pairwise comparisons, ties 2, and loses 4. Level-*k* versus either NI or LQRE wins in 4 comparisons, ties in 5, and loses in 1. CH versus either NI or LQRE wins in 5 comparisons, ties in 2, and loses in 3. NI versus LQRE wins in 2 comparisons, one slightly, and ties in 8. Comparing PDE, the best of the selection criteria, against level-*k* and CH, the best of the individualistic models, PDE wins in 7 comparisons and loses in 3.

---

[11] For level-*k* we allow only types *L1* and *L2*; in VHBB's games, higher types would not be distinguished from *L2*. For CH we allow all types. For NI we truncate iterations at 10, as GH do. And we approximate LQRE by setting NI's telescoping parameter equal to one. Plainly these games are not well suited to identifying type distributions. It does not follow that the types are never identified. In the level-*k* model, because *L1* and *L2* have different beliefs their deviation costs are different, so their frequencies are usually identified via the logit error structure, but in our experience such identification is weak. In the CH model, because *L1* and higher types make identical predictions in VHBB's games, their frequencies are identified only by the estimated frequency of *L0* and the assumed Poisson type distribution, in which there is little independent reason for confidence. The Maximin, PDE, RDE, and LQRE models each have one estimated parameter, their precisions. The level-*k* model has two parameters, the population frequency of *L1* (versus *L2*) and the types' common precision. However, due to the low or nonexistent separation between *L1* and *L2* in VHBB's games, the level-*k* model has effectively one parameter. Given its use of *L0* to explain all errors, the CH model has one parameter, for its Poisson type distribution; and the NI model has two, its initial precision and a "telescoping" parameter measuring the rate at which precision declines with iterations.

Thus, both the structural non-equilibrium models considered here, level-*k* and CH, remain plausible alternatives to LQRE and NI; but the choice among models of strategic thinking must be guided by more than VHBB's data. It is noteworthy that level-*k* and CH models adopt a very different view of coordination than PDE or RDE: Players do not first identify the set of equilibria and then refine it. Instead they respond to coordination games using the same decision rules they use to respond to other games; and both equilibrium and equilibrium selection are by-products of how those rules interact with the game. These models completely change our view of coordination, bringing it closer to our view of decisions in other games and decision problems.

References

Anderson, Simon, Jacob Goeree, and Charles Holt (2001). "Minimum-Effort Coordination Games: Stochastic Potential and Logit Equilibrium." *Games and Economic Behavior*, 34, 177-199.

Bresnahan, Timothy, and Peter Reiss (1991). "Econometric Models of Discrete Games," *Journal of Econometrics*, 48, 57-81.

Camerer, Colin, Teck-Hua Ho, and Juin Kuan Chong (2004). "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, 119, 861-898.

Camerer, Colin, Thomas Palfrey, and Brian Rogers (2007), "Heterogeneous Quantal Response Equilibrium and Cognitive Hierarchies." Working Paper, California Institute of Technology.

Chong, Juin Kuan, Colin Camerer, and Teck-Hua Ho (2005). "Cognitive Hierarchy: A Limited Thinking Theory in Games." Chapter 9 in Rami Zwick and Amnon Rapoport (Eds.), *Experimental Business Research, Volume III: Marketing, Accounting and Cognitive Perspectives*. Kluwer Academic Press.

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

Costa-Gomes, Miguel, and Vincent Crawford (2006). "Cognition and Behavior in Two-Person Guessing Games: An Experimental Study." *American Economic Review*, 96, 1737-1768.

Costa-Gomes, Miguel and Georg Weizsäcker (2008). "Stated Beliefs and Play in Normal-Form Games." *Review of Economic Studies*, 75, 729-762.

Crawford, Vincent (1991). "An 'Evolutionary' Interpretation of Van Huyck, Battalio, and Beil's Experimental Results on Coordination," *Games and Economic Behavior*, 3, 25-59.

Crawford, Vincent (1995). "Adaptive Dynamics in Coordination Games." *Econometrica*, 63, 103-143,

Crawford, Vincent, and Nagore Iriberri (2007a). "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.

Crawford, Vincent, and Nagore Iriberri (2007b). “Fatal Attraction: Salience, Naivete, and Sophistication in Experimental Hide-and-Seek Games.” *American Economic Review*, 97, 1731-1750.
Goeree, Jacob, and Charles Holt (2004). “A Model of Noisy Introspection.” *Games and Economic Behavior*, 46, 365–382.
Goeree, Jacob, Charles Holt, and Thomas Palfrey (2002). “Quantal Response Equilibrium and Overbidding in Private-Value Auctions.” *Journal of Economic Theory*, 104, 247-272.
Goeree, Jacob, Charles Holt, and Thomas Palfrey (2005). “Regular Quantal Response Equilibrium,” *Experimental Economics*, 8, 347-367.
Haile, Philip, Ali Hortaçsu, and Grigory Kosenok (2008). “On the Empirical Content of Quantal Response Equilibrium.” *American Economic Review*, 98, 180-200.
Harsanyi, John, and Reinhard Selten (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.
Haruvy, Ernan, and Dale O. Stahl (2007). “Equilibrium Selection and Bounded Rationality in Symmetric Normal-Form Games,” *Journal of Economic Behavior and Organization*, 62, 98–119.
Ho, Teck-Hua, Colin Camerer, and Keith Weigelt (1998). “Iterated Dominance and Iterated Best Response in Experimental ‘*p*-Beauty Contests’,” *American Economic Review*, 88, 947-969.
McKelvey, Richard and Thomas Palfrey (1995). “Quantal Response Equilibria for Normal-Form Games.” *Games and Economic Behavior*, 10, 6-38.
Nagel, Rosemarie (1995). “Unraveling in Guessing Games: An Experimental Study.” *American Economic Review*, 85, 1313-1326.
Östling, Robert, Joseph Tao-yi Wang, Eileen Chou, and Colin Camerer (2008). “Field and Lab Convergence in Poisson LUPI Games.” Working Paper, Stockholm School of Economics.
Stahl, Dale and Paul Wilson (1995). “On Players’ Models of Other Players: Theory and Experimental Evidence." *Games and Economic Behavior*, 10, 218-254.
Van Huyck, John, Raymond Battalio, and Richard Beil (1990). “Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure.” *American Economic Review*, 80, 234-248.
Van Huyck, John, Raymond Battalio, and Richard Beil (1991). “Strategic Uncertainty, Equilibrium Selection Principles, and Coordination Failure in Average Opinion Games." *Quarterly Journal of Economics*, 106, 885-910.
Weizsäcker, Georg (2003). “Ignoring the Rationality of Others: Evidence from Experimental Normal-Form Games.” *Games and Economic Behavior*, 44, 145-171.
Yi, Kang-Oh (2003). “A Quantal Response Equilibrium Model of Order-Statistic Games.” *Journal of Economic Behavior & Organization*, 51, 413-425.

Table 1: Log-Likelihood Comparisons for Alternative Models

Model Treatment	Empirical Frequencies (Modal effort*)	Random Frequencies (Modal effort)	Maximin (Modal effort)	PDE (Modal effort)	Independent RDE (Modal effort) Correlated RDE (Modal effort)	Independent LQRE (Modal effort) Correlated LQRE (Modal effort)	Independent Level-*k* (Modal effort) Correlated Level-*k* (Modal effort)	Independent CH (Modal effort) Correlated CH (Modal effort)	Independent NI (Modal effort) Correlated NI (Modal effort)
A	-172.1785 (5)	-208.2124 (1-7)	- 208.2124 (1)	-186.9741 (7)	-208.2124 (1) -207.8228 (4)	-208.2124 (1-7) -208.1302 (4)	-208.2124 (1-7) -207.8228 (4)	-208.2124 (1-7) -207.9439 (4)	-208.2124 (1-7) -208.1302 (4)
B	-63.8718 (7)	-177.0778 (1-7)	- 177.0778 (1-7)	-100.3950 (7)	-100.3950 (7) -100.3950 (7)	-172.0179 (4,5-7) -111.8437 (7)	-69.7289 (7) -98.0386 (7)	-67.6081 (7) -67.6081 (7)	-172.0179 (4,5-7) -111.8437 (7)
C_d	-49.3084 (7)	-58.3773 (1-7)	-58.3773 (1)	-57.8714 (7)	-58.3773 (4) -58.3773 (4)	-58.3773 (1-7) -58.3773 (1-7)	-58.3773 (1-7) -58.3773 (1-7)	-58.3108 (4) -58.3108 (4)	-58.3773 (1-7) -58.3773 (1-7)
Γ	-41.0777 (5)	-52.5396 (1-7)	-52.5396 (3)	-46.8985 (7)	-46.8985 (7) -46.8985 (7)	-44.1974 (5) -49.8153 (4)	-48.3459 (4) -49.8153 (4)	-50.4512 (4) -50.4512 (4)	-44.1808 (5) -49.8153 (4)
Ω	-28.9699 (7)	-52.5396 (1-7)	-52.5396 (1-7)	-41.9893 (7)	-41.9893 (7) -41.9893 (7)	-52.5396 (1-7) -41.0017 (7)	-52.5396 (1-7) -37.6399 (7)	-52.5396 (1-7) -41.9894 (7)	-52.5396 (1-7) -37.8427 (7)

*The modal and median efforts are the same in all treatments, except C_d where the median is 4 and Γ where the median is 4 or 5.