

COMPARING MODELS OF STRATEGIC THINKING IN VAN HUYCK, BATTALIO, AND BEIL'S COORDINATION GAMES

By Miguel A. Costa-Gomes, Vincent P. Crawford, and Nagore Iriberrí¹

Revised 17 October 2008

Abstract: This paper compares the leading models of strategic thinking with subjects' initial responses to Van Huyck et al.'s (1990, 1991) coordination games. Among the refined "equilibrium plus noise" models we compare, payoff-dominant equilibrium performs better than risk-dominant or maximin equilibrium. Among the individualistic models we compare, level- k and cognitive hierarchy models usually fit better than logit quantal response equilibrium or noisy introspection models. In Van Huyck et al.'s games, payoff-dominant equilibrium usually fits better than level- k or cognitive hierarchy. The data favor versions of the models in which people model others as if they were perfectly correlated over the standard, independent versions.

Keywords: behavioral game theory, experimental economics, coordination games, equilibrium selection, risk- and payoff-dominance, level- k models, cognitive hierarchy models, quantal response equilibrium, noisy introspection

JEL codes: C72, C92, C51

¹Costa-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. Iriberrí: Departament d'Economia i Empresa, Universitat Pompeu Fabra, Barcelona, Spain, e-mail: nagore.iriberri@upf.edu. We are grateful to Colin Camerer, Tore Ellingsen, Jacob Goeree, Charles Holt, Jaimie Lien, Juanjuan Meng, Robert Östling, Thomas Palfrey, and Adam Sanjorjo for helpful discussions. Iriberrí 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”); see also 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 for learning to converge to equilibrium: Players must have perfectly coordinated beliefs, which without precedents generally requires players to have accurate models of each other’s decisions. It is easy to imagine strategic thinking this accurate in simple games. But the thinking required for equilibrium in more complex games is often behaviorally far-fetched, and even players who are capable of such thinking may doubt that others are capable, or doubt that others believe others are capable. 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 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 Iriberry 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, or a single-type level- k model (GH, Costa-Gomes and Weizsäcker 2008).

This paper provides a critical discussion of the cognitive requirements of equilibrium plus noise, LQRE, level- k /CH, and NI models and assesses their abilities to describe subjects’ initial responses to five of the games used in VHBB’s (1990, 1991) famous coordination experiments. The large strategy spaces of VHBB’s games (with seven decisions per player) allow more informative comparisons than the 2×2 or 3×3 games used in most previous work. Further, our analysis is the first to compare representatives of each leading model of strategic thinking, and the first to focus on the use of these noncooperative notions as *complete* models of coordination, comparing their ability to predict with those of equilibrium plus noise augmented by coordination refinements such as risk- or payoff-dominant equilibrium. Finally, our analysis is the first to fully address 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 or instead model others as if they were one perfectly correlated person, for which VHBB’s dataset is ideal.

The paper is organized as follows. Section 2 reviews the leading models of strategic thinking and discusses their cognitive requirements. Section 3 introduces VHBB’s games and assesses the models. Among our refined equilibrium plus noise models, payoff-dominant equilibrium performs better than risk-dominant or maximin equilibrium. Among our individualistic models, level- k and CH models usually fit better than LQRE or NI models. In VHBB’s games, payoff-dominant equilibrium plus noise usually fits better than level- k or CH. The data favor versions of the models in which people model others as if they were perfectly correlated over the standard, independent versions.

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 if probabilistic) prediction conditional on the values 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 an 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. We also consider maximin decisions, which VHBB gave a prominent role. In these symmetric games, maximin functions like an equilibrium refinement.

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 distributions, with each player's distribution a noisy best response to the others'. As the distributions' precision increases, QRE approaches equilibrium. 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 or "LQRE" implies error distributions that respond to out-of-equilibrium payoffs, often in plausible ways.² 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, 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).

² 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.

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 to errors. Although the number of possible non-equilibrium structures seems daunting, much experimental evidence supports a particular class of models called level- k or cognitive hierarchy (“CH”) models, which also 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 weakly dominance-solvable in infinite numbers of rounds, so that the reasoning required to identify equilibrium strategies is significantly simpler than usual. 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 iterated dominance rules. The rule called Dk does k rounds of iterated dominance for some small number, $k = 1$ or 2 , and then best responds to a uniform prior over its partner’s remaining strategies. In another interpretation, subjects follow “level- k ” rules. The rule called Lk starts with a uniform prior over others’ possible guesses and then iterates the best response mapping k times, with $k = 1, 2$, or perhaps 3 . In Nagel’s games Lk and $Dk-1$ 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) Lk and $Dk-1$ 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, Lk and $Dk-1$ are strongly separated by decisions as well as search, and the results strongly 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 footnote 8). 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 a 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.³ 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 removing it does not improve the fit of the CH model; but in others (CGC, Crawford and Iriberri 2007ab) that constraint would be strongly binding.

McKelvey and Palfrey (1995) suggest using LQRE for initial responses and limiting outcomes both, with increasing precision as a reduced-form model of learning. But while LQRE has been the most popular model of initial responses, not all researchers consider it suitable for that purpose. GH suggest reserving LQRE for limiting outcomes, instead proposing an NI model to describe initial

³ Level- k models thus provide an evidence-based way to think about robustness of mechanisms. Mechanisms that implement desired outcomes in dominant strategies may have an advantage over more complex mechanisms that implement superior outcomes, but only in equilibrium.

responses. Their NI model relaxes LQRE's equilibrium assumption while maintaining its assumption that players respond to a nondegenerate probability distribution of other players' responses. Instead players form beliefs by iterating best responses as in a level- k model, but higher-order beliefs reflect increasing amounts of noise, converging to uniform randomness. For given noise distributions, the NI model makes probabilistic predictions that depend on how fast the noise grows. In the extreme case where 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 infinity, NI beliefs are like $L1$'s; if it is zero for one iteration and then jumps to infinity, NI beliefs are like $L2$'s, and so on.⁴

In applications GH assume that the noisiness of higher-order beliefs grows geometrically with iterations, which yields beliefs similar but by no means identical to Lk 's; slower noise growth is like a higher k . The resulting NI model is more flexible than LQRE, and cognitively less taxing because it requires no 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. And in fact the evidence from Nagel's and more recent experiments suggests that neither the iteration of best responses beyond two or three rounds nor the assumed homogeneity of strategic thinking is realistic.

3. 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. 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 $\{1, \dots, 7\}$: 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 $x_i \neq M$.⁵

⁴ Compare Camerer et al. (2007), who also nest generalized variants of LQRE and CH models.

⁵ Anderson et al. (2001) compare LQRE as precision approaches infinity with limiting outcomes

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 more extreme 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 (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).⁶ We define maximin, RDE and PDE plus noise, LQRE, level- k types, and NI with logit errors, each with estimated precision.⁷

In specifying the models for these n -person games, one important issue is whether players take the independence of others' decisions 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' decisions, 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 LO , which is the channel by which the correlation influences players' choices, through the higher-level types, in those models. Correlated maximin and PDE are the same as the independent ones.

in VHBB's minimum games, and Yi (2003) does the same for their median games. An appendix at <http://dss.ucsd.edu/~vcrawfor/#VHBB> discusses LQRE as a model of limiting outcomes.

⁶ The subject groups were large enough for subjects to treat their own influences on future choices as negligible, so that their initial responses to each game can be viewed as responses to the game as if played in isolation. There was some evidence of order effects in later treatments, particularly C_d , which was run last; but these are beyond the scope of this paper's analysis.

⁷ Because maximin does not (and cannot) have rational beliefs, we evaluate its expected deviation costs using the beliefs of the associated equilibrium.

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 log-likelihoods of the empirical frequencies and random frequencies, which provide upper and lower bounds on the attainable likelihoods for any model. (The upper bound is not 0, as is usual for a perfect fit, because the estimated models 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 maximin, 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 log-likelihoods without reporting type frequencies or other parameter estimates.⁸

The results in Table 1 suggest several conclusions. First, the correlated versions of the models almost always do as well or better than the independent versions (the exceptions are level- k in B and level- k , LQRE, and NI in Γ): Despite its importance in treatment A, few subjects' thinking reflects the independence of others' decisions. Second, among the equilibrium selection criteria maximin, PDE, and RDE, PDE always fits at least as well as the others, and often better; this is surprising given previous findings for 2×2 games that favor RDE. 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 and ties in 8. Comparing PDE, the best selection criterion, against level- k and CH, the best individualistic models, PDE wins in 7 comparisons and loses in 3.

Thus, the structural non-equilibrium models considered here, level- k and CH, are promising alternatives to LQRE, NI, RDE, and PDE; but the choice

⁸ For level- k we allow only types $L1$ and $L2$; the estimated frequency of $L0$ is usually zero in other settings, and in VHBB's games, higher types would be identical to $L2$. 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, 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. An appendix at <http://dss.ucsd.edu/~vcrawfor/#VHBB> provides additional detail on the estimates.

among models must be guided by more than VHBB's data. It is noteworthy that level- k and CH models completely change our view of coordination: Unlike in 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. This makes level- k and CH models of coordination closer to our models of decisions in other games and decision problems, and in our view, behaviorally more plausible.

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." In Rami Zwick and Amnon Rapoport (Eds.), *Experimental Business Research: 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: Saliency, 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.
- Haile, Philip, Ali Hortaçsu, and Grigory Kosenok (2008). "On the Empirical Content of Quantal Response Equilibrium." *American Economic Review*, 98, 180-200.
- 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). "Strategic Thinking and Learning in the Field and the Lab: Evidence from Poisson LUPI Lottery Games." 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)	-208.2124 (1-7)	-208.2124 (1-7)	-208.2124 (1-7)	-208.2124 (1-7)
					-207.8228 (4)	-208.1302 (4)	-207.8228 (4)	-207.9439 (4)	-208.1302 (4)
B	-63.8718 (7)	-177.0778 (1-7)	- 177.0778 (1-7)	-100.3950 (7)	-100.3950 (7)	-172.0179 (4,5-7)	-69.7289 (7)	-67.6081 (7)	-172.0179 (4,5-7)
					-100.3950 (7)	-111.8437 (7)	-98.0386 (7)	-67.6081 (7)	-111.8437 (7)
C _d	-49.3084 (7)	-58.3773 (1-7)	-58.3773 (1)	-57.8714 (7)	-58.3773 (4)	-58.3773 (1-7)	-58.3773 (1-7)	-58.3108 (4)	-58.3773 (1-7)
					-58.3773 (4)	-58.3773 (1-7)	-58.3773 (1-7)	-58.3108 (4)	-58.3773 (1-7)
Γ	-41.0777 (5)	-52.5396 (1-7)	-52.5396 (3)	-46.8985 (7)	-46.8985 (7)	-44.1974 (5)	-48.3459 (4)	-50.4512 (4)	-44.1808 (5)
					-46.8985 (7)	-49.8153 (4)	-49.8153 (4)	-50.4512 (4)	-49.8153 (4)
Ω	-28.9699 (7)	-52.5396 (1-7)	-52.5396 (1-7)	-41.9893 (7)	-41.9893 (7)	-52.5396 (1-7)	-52.5396 (1-7)	-52.5396 (1-7)	-52.5396 (1-7)
					-41.9893 (7)	-41.0017 (7)	-37.6399 (7)	-41.9894 (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.