Master Class on Structural Nonequilibrium Models of Strategic Thinking: Theory, Measurement, and Applications

Cemmap (Center for Microdata Methods and Practice)
University College London, 5-6 March 2009

Economics 201C Segment on Behavioral Game Theory: Strategic Thinking

University of California, San Diego, 30 March-27 April 2009

Mini-Course on Limited Cognition, Strategic Thinking, and Learning in Games, University of Bonn, Graduate School of Economics Summer School, 20-24 July 2009

Vincent P. Crawford, University of California, San Diego
based on joint work with
Miguel A. Costa-Gomes, University of Aberdeen, and
Nagore Iriberri, Universitat Pompeu Fabra

Thanks also to S. Nageeb Ali, B. Douglas Bernheim, Tore Ellingsen, Navin Kartik, Botond Kőszegi, Tamar Kugler, Zvika Neeman, Robert Östling, Ady Pauzner and Joel Sobel.

Last revised 26 April 2009.

Outline of Lectures

1. Introduction: Why Study Strategic Thinking?
2. Nine "Folk Game Theory" Quotations (Keynes's Beauty Contest, Graham's Mr. Market, Kahneman's Entry Magic, Lake Wobegon, Huarongdao, October Surprise, Bank Runs, Poe's Outguessing Game)
3. Leading Models of Strategic Thinking (Equilibrium Plus Noise, Finitely Iterated (Strict) Dominance and k-Rationalizability, Quantal Response Equilibrium ("QRE") and Logit QRE ("LQRE"), Level-k Models, Cognitive Hierarchy Models, Noisy Introspection ("NI") Models)
4. Experimental Evidence (Nagel's Design and Results, Costa-Gomes and Crawford's Design and Results and Data Analysis)
5. Lessons from the Experiments for Modeling Strategic Behavior (Level-*k* versus CH Models, Level-*k* versus Equilibrium Plus Noise or LQRE Models, Level-*k* versus NI Models, Observations about the Models' Cognitive Requirements)
6. Illustration of Level-*k* Analyses of Matrix Games with Unique Mixed-Strategy Equilibria: M. M. Kaye's The Far Pavilions

7. Kahneman’s Entry Magic: Asymmetric Coordination via Structure in Entry Games
8. Bank Runs: Symmetric Coordination via Structure
9. Structural Alternatives to “Incomplete” Models
10. Yuschenko and Lake Wobegon: Framing Effects in Zero-Sum Two-Person Games (Evaluating the Model’s Explanation: Overfitting and Portability)
11. Chicago Skyscrapers: Framing Effects and Miscoordination in Schelling-Style Coordination Games
12. Huarangdao and D-day: Preplay Communication of Intentions in Zero-Sum Two-Person Games with Possibly Sophisticated Players
13. Preplay Communication of Intentions in Coordination Games
14. Experimental Evidence on Communication of Private Information in Sender-Receiver Games
15. October Surprise: Communication of Private Information in Zero-Sum Two-Person Games
16. Overbidding in Independent-Private-Value and Common-Value Auctions
17. Behaviorally Optimal Auction Design

1. Introduction: Why Study Strategic Thinking?

Strategic thinking is an essential part of human interaction, so much so that children must be taught to look both ways before crossing one-way streets.

(Once children develop enough “theory of mind” to distinguish other people as independent decision makers, they seem to become instinctively overoptimistic about using rationality to predict others’ decisions.)

Yet from a behavioral point of view, the importance of strategic thinking has been downplayed in economics and game theory.

Most applications of game theory in economics rely on Nash equilibrium.

However, although equilibrium can be viewed as a model of strategic thinking, we will see that there are many potential applications of game theory for which it is not an adequate model of behavior.

Players' strategies will be in equilibrium if two conditions are satisfied:

- Players are rational (in the sense of best responding to some beliefs).
- Players have the same beliefs about each other's strategies.

Accepting rationality for the sake of argument, there are two possible justifications for the assumption that players have the same beliefs:

- Thinking: If players have perfect models of each other's decisions, strategic thinking will lead them to have the same beliefs immediately, and so play an equilibrium even in their initial responses to a game.

  (Note that in this case the usual "as if" justification for equilibrium is unavailable: if players' models do not accurately reflect other players' cognition, equilibrium is unlikely to predict their decisions accurately.)

- Learning: Even without perfect models, if players repeatedly play perfectly analogous games (and their interaction patterns do not foster repeated-game effects or strategic teaching), experience may eventually allow them to predict each others' decisions well enough to play an equilibrium (in the game that is repeated) in the limit.

In many applications of game theory, the theoretical conditions for learning to converge to equilibrium are approximately satisfied.

In such settings experimental evidence and field data tend to support assuming that players' steady-state strategies are in equilibrium.

If only long-run outcomes matter, and if equilibrium is unique or if there are multiple equilibria but equilibrium selection does not depend on the details of learning, such applications can safely rely entirely on equilibrium.

Because in such settings the cognitive requirements for learning to converge to equilibrium are mild, there is then no need to study strategic thinking.

However, many other applications involve games played without clear precedents, so that the learning justification for equilibrium is unavailable.

In other applications eventual convergence to equilibrium is assured, but initial as well as limiting outcomes matter (e.g. the FCC Spectrum auction).

In still other applications convergence is assured and only long-run outcomes matter, but the equilibrium is selected from multiple possibilities via history-dependent learning dynamics.

All such applications depend on reliably predicting initial responses to games, which may require a non-equilibrium model of strategic thinking.

Aside on Equilibrium Selection via History-dependent Learning

In Van Huyck, Cook, and Battalio's 1997 *JEBO* experiment, seven subjects chose simultaneously and anonymously among efforts from 1 to 14, with each subject's payoff determined by his own effort and a summary statistic, the median, of all players' efforts.

After subjects chose their efforts, the group median was publicly announced, subjects chose new efforts, and the process continued.

The relation between a subject's effort, the median effort, and his payoff was publicly announced via a table as on the next slide.

The payoffs of a player's best responses to each possible median are highlighted in bold in the table as displayed here (but not to subjects).

The payoffs of the (symmetric, pure-strategy) equilibria "all–3" and "all–12" are highlighted in large bold.

Continental divide game payoffs

	Median Choice													
Your Choice	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	45	49	52	55	56	55	46	-59	-88	-105	-117	-127	-135	-142
2	**48**	53	58	62	65	66	61	-27	-52	-67	-77	-86	-92	-98
3	**48**	**54**	**60**	**66**	70	74	72	1	-20	-32	-41	-48	-53	-58
4	43	51	58	65	**71**	**77**	80	26	8	-2	-9	-14	-19	-22
5	35	44	52	60	69	**77**	**83**	46	32	25	19	15	12	10
6	23	33	42	52	62	72	82	62	53	47	43	41	39	38
7	7	18	28	40	51	64	78	75	69	66	64	63	62	62
8	-13	-1	11	23	37	51	69	83	81	80	80	80	81	82
9	-37	-24	-11	3	18	35	57	88	89	91	92	94	96	98
10	-65	-51	-37	-21	-4	15	40	**89**	**94**	98	101	104	107	110
11	-97	-82	-66	-49	-31	-9	20	85	**94**	**100**	105	110	114	119
12	-133	-117	-100	-82	-61	-37	-5	78	91	99	**106**	**112**	**118**	**123**
13	-173	-156	-137	-118	-96	-69	-33	67	83	94	103	110	117	**123**
14	-217	-198	-179	-158	-134	-105	-65	52	72	85	95	104	112	120

There were ten sessions, each with its own separate group.

Half the groups happened to have an initial median of eight or above, and half happened to have an initial median of seven or below.

(The experimenters probably chose the design to make the initial median vary this way, but this kind of variation is not uncommon.)

The results are graphed on the next slide:

The median-eight-or-above groups converged almost perfectly to the all–12 equilibrium.

By contrast, the median-seven-or-below groups converged almost perfectly to the all–3 equilibrium.

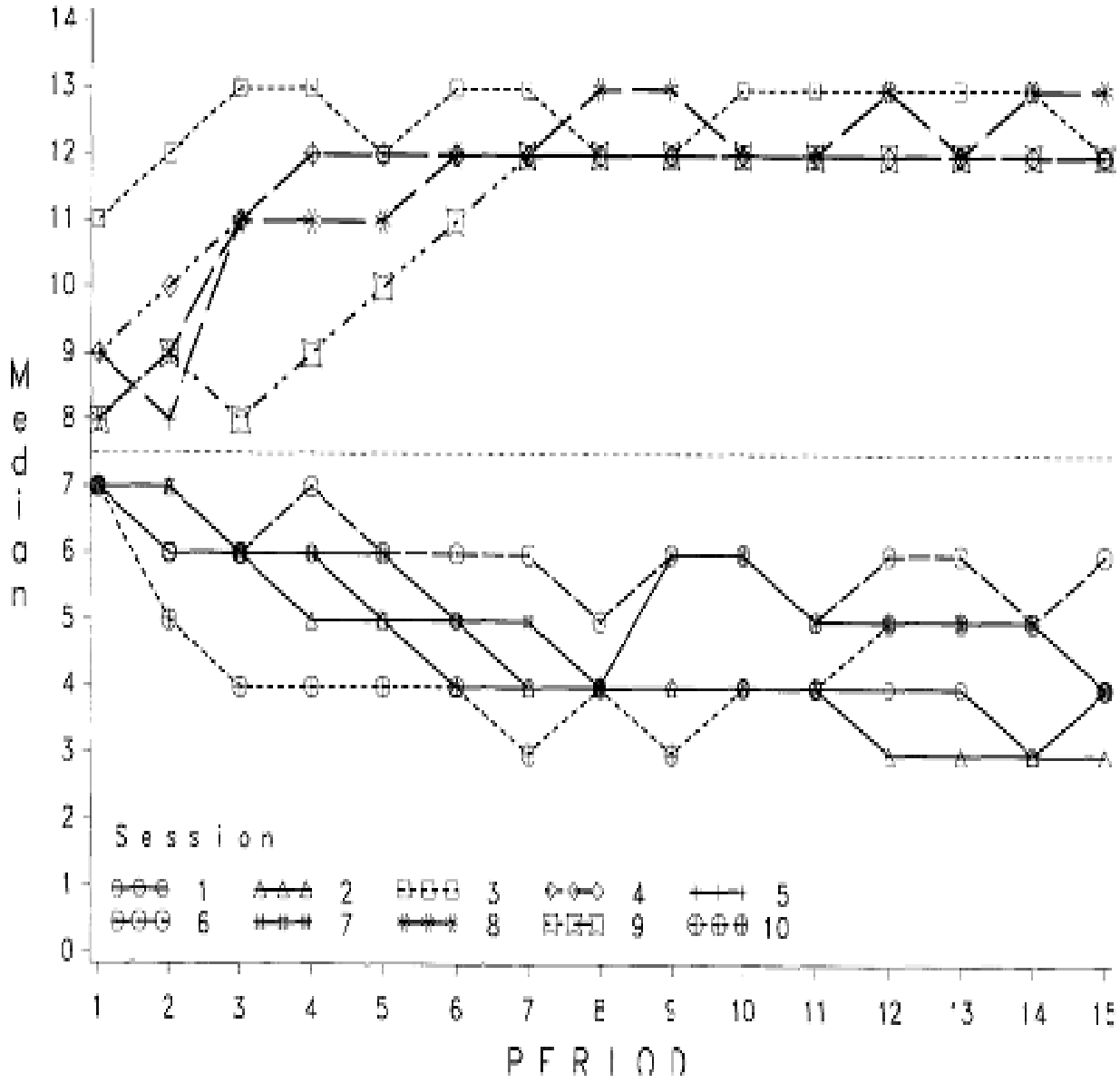


Fig. 3. Median choice in sessions 1 to 10 by period

Van Huyck, Cook, and Battalio's Figure 3

Thus, it's not enough to know that learning will eventually converge to some equilibrium, even if we are only interested in the final outcome.

Here we also need to know the prior probability distribution of the median initial response.

That distribution, together with a simple view of learning in which equilibrium selection is determined by which basin of attraction—defined by myopic best responses—subjects' initial responses fell into, seem adequate to determine the probability distribution of final outcomes in Van Huyck et al.'s experiment.

In other applications we may need to know more about the structure of subjects' learning rules as well as about their initial responses.

See for example Crawford, "Adaptive Dynamics in Coordination Games," 1995 *Econometrica*, and Crawford and Broseta, "What Price Coordination? The Efficiency-enhancing Effect of Auctioning the Right to Play," 1998 *AER*, which discuss Van Huyck, Battalio, and Beil's famous 1990 *AER*, 1991 *QJE*, 1993 *GEB* coordination experiments.

(End of aside)

Applications of game theory usually assume equilibrium even when its learning justification is unavailable.

This practice seems to be due to two factors:

- Fear that equilibrium is the only possible basis for analysis (rationalizability seldom yields predictions specific enough to be useful).

- Hope that equilibrium will still yield accurate predictions, on average.

But except in simple games, assuming equilibrium thinking in people's initial responses may be behaviorally far-fetched.

Even people who are capable of equilibrium thinking may doubt that others are capable, and therefore be unwilling to play their part of an equilibrium.

Moreover, there is a growing body of evidence—mostly experimental—that initial responses to novel or complex games often deviate systematically from equilibrium, especially if it requires thinking that is not straightforward.

Fortunately, the evidence also suggests that there are simple and tractable structural non-equilibrium models of strategic thinking that can explain a large fraction of people's deviations from equilibrium initial responses.

Those models allow equilibrium behavior, but do not assume equilibrium in all games.

Instead they assume that players follow strategic but non-equilibrium decision rules, which yield decisions that mimic equilibrium in simple games, but may deviate systematically in more complex games.

The models thereby provide a way to predict, in a given game, whether players' responses are likely to deviate from equilibrium, and if so, how.

Thus the hope that equilibrium yields predictions that are accurate on average is not well founded.

But neither is the fear that equilibrium is the only possible basis for analysis.

Modeling strategic thinking more accurately promises several benefits:

- It can establish the robustness of conclusions based on equilibrium in games where empirically reliable rules mimic equilibrium.

- It can challenge the conclusions of applications to games where equilibrium is implausible without learning.

- It can resolve empirical puzzles by explaining the deviations from equilibrium that some games evoke.

- It can also elucidate the structure of learning, where assumptions about cognition determine which analogies between current and previous games players recognize and also distinguish reinforcement from beliefs-based and more sophisticated rules.

Overview

The rest of these lectures are organized as follows. I focus on normal-form games, including extensive-form games mainly to study communication, with other kinds of extensive-form games left for future discussion.

- I begin with nine “folk game theory” quotations to illustrate the need for non-equilibrium models of strategic thinking, the issues successful models must address, and the range of potential applications.

- I next give a brief summary of the leading models of strategic thinking.

- I then summarize the experimental evidence on strategic thinking.

- I then discuss some theoretical and econometric lessons for modeling strategic behavior and critique the models in light of the evidence.

- The discussion is interwoven with illustrative applications of the level-*k* models that are suggested by the evidence, which take up some of the strategic issues raised by the quotations.

2. Nine “Folk Game Theory” Quotations

This section gives nine “folk game theory” quotations to illustrate the need for non-equilibrium models of strategic thinking, the issues successful models must address, and the range of potential applications.

Why study folk game theory instead of “real” game theory?

Folk game theory is only an imperfect reflection of traditional game theory, just as folk physics is an imperfect reflection of real physics.

But unlike folk physics, folk game theory has a direct and important influence on its observable counterpart, namely the part of behavioral game theory that concerns strategic thinking and initial responses to games.

I will argue below that the lessons regarding strategic thinking from folk game theory are largely confirmed by experiments designed to study strategic thinking in more conventional ways.

This correspondence is powerful evidence for a particular class of structural non-equilibrium models of strategic thinking.

Keynes's Beauty Contest: "...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 prize being awarded to the competitor whose choice most nearly corresponds to the average preferences of the competitors as a whole; so that each competitor has to pick, not those faces which he himself finds prettiest, but those which he thinks likeliest to catch the fancy of the other competitors, all of whom are looking at the problem from the same point of view. It is not a case of choosing those which, to the best of one's judgment, are really the prettiest, nor even those which average opinion genuinely thinks the prettiest. We have reached the third degree where we devote our intelligences to anticipating what average opinion expects the average opinion to be. And there are some, I believe, who practice the fourth, fifth and higher degrees."—John Maynard Keynes, *The General Theory of Employment, Interest, and Money*

(I suspect that the last sentence was Keynes's coy reference to himself.)

A simultaneous-move zero-sum *n*-person "outguessing" game, possibly with multiple equilibria. The key issue is anticipating others' strategic responses to a "landscape" of personal judgments about prettiness which is otherwise payoff-irrelevant. We will find that equilibrium alone is not very helpful in describing how people do this. The quotation suggests a thought process in which players "anchor" beliefs in instinctive reactions to the faces and then iterate best responses a finite number of times.

Graham's Mr. Market: "…imagine you are partners in a private business with a man named Mr. Market. Each day, he comes to your office or home and offers to buy your interest in the company or sell you his [the choice is yours]. The catch is, Mr. Market is an emotional wreck. At times, he suffers from excessive highs and at others, suicidal lows. When he is on one of his manic highs, his offering price for the business is high as well…. His outlook for the company is wonderful, so he is only willing to sell you his stake in the company at a premium. At other times, his mood goes south and all he sees is a dismal future for the company. In fact… he is willing to sell you his part of the company for far less than it is worth. All the while, the underlying value of the company may not have changed - just Mr. Market's mood."—Warren Buffett's intellectual hero Benjamin Graham (of Graham and Dodd's *Security Analysis*), in Graham's *The Intelligent Investor* (thanks to Steven Scroggin of Virginia Polytechnic Institute for the reference).

A simultaneous-move two-person game, possibly with multiple equilibria. Again the key issue is outguessing others' judgments, equilibrium alone is not very helpful, and the quotation suggests a thought process like Keynes's, but anchored in the psychology of a representative uninformed investor's reaction to news.

(In context, however, Graham's main goal in this passage was actually to keep readers from becoming too emotionally involved with their own portfolios.)

Kahneman's Entry Magic: "…to a psychologist, it looks like magic."—Kahneman 1988, quoted in Camerer, Ho, and Chong 2004 *QJE*.

Here Kahneman refers to the fact that subjects in his market-entry experiments (see also Rapoport and Seale 2002), structured like *n*-person Battle of the Sexes games, achieve better ex post coordination (number of entrants closer to market capacity) than in the natural symmetric mixed-strategy equilibrium benchmark.

(Thus Kahneman should have said "…to a *game theorist*, it looks like magic.")

The key issue here is breaking the symmetry of players' roles as required for efficient coordination. Equilibrium and refinements are not very helpful.

The same strategic issues arise in less abstractly framed, asymmetric field settings, exemplified by Roger Myerson's "Ware Medical Corporation" case (http://www.kellogg.northwestern.edu/faculty/weber/DECS-452/index.htm or http://dss.ucsd.edu/~vcrawfor/Ware.htm): A company is considering introducing a new product, which will be profitable only if its only competitor introduces a related product. The competitor's profits are determined qualitatively (not quantitatively) in the same way as the company's are. Both companies must decide, simultaneously and irreversibly, whether to begin development. In addition, there may be opportunities for commitment, signaling, and/or deceptive announcements. See also Goldfarb and Yang 2009 *Journal of Marketing Research*.

Yushchenko: “Any government wanting to kill an opponent…would not try it at a meeting with government officials.”—comment, quoted in Chivers 2004, on the poisoning of Ukrainian presidential candidate—now president—Viktor Yushchenko.

A simultaneous-move zero-sum two-person game with a unique mixed-strategy equilibrium. The players are a government assassin choosing one of several occasions at which to try to poison Yuschenko, only one of which is linked to the government; and an investigator who has the resources to check only one occasion.

Here the key issue is how players react to framing of decisions that is non-neutral but does not directly affect payoffs. Equilibrium in zero-sum two-person games leaves no room for such framing to affect outcomes, but people often react to it anyway.

The thinking reflected by the quotation is plainly strategic, but non-equilibrium: Any game theorist worth his salt would respond, “If that’s what people think, a meeting with government officials is exactly where *I* would try to poison Yushchenko.”

We will see that the quotation can be understood as a thought process in which a player anchors his beliefs in an instinctive reaction to the salience of the dinner with government officials and then iterates best responses a small number of times.

Lake Wobegon: "…in Lake Wobegon, the correct answer is usually 'c'."—Garrison Keillor 1997 on multiple-choice tests (quoted in Attali and Bar-Hillel 2003 *Journal of Educational Measurement*).

A simultaneous-move two-person zero-sum game with a unique mixed-strategy equilibrium. The players are a test designer deciding where to hide the correct answer and a clueless test-taker trying to guess the hiding place.

Again the key issue is how players react to the non-neutral framing, and the thinking reflected by the quotation is plainly strategic, but non-equilibrium.

Although there is nothing as uniquely salient as Yushchenko's dinner with government officials, psychologists like Christenfeld 1995 *Psychological Science* and Tversky (in Rubinstein, Tversky, and Heller 1996) think that with four possible answers, both the a and d end locations and location c are inherently salient (with the jury still out on which is more salient).

Again the quotation can be understood as a thought process in which a player anchors beliefs in an instinctive reaction to salience and iterates best responses a small number of times.

Huarongdao:

General Kongming: “Have you forgotten the tactic of ‘letting weak points look weak and strong points look strong’?”

General Cao Cao: “Don’t you know what the military texts say? ‘A show of force is best where you are weak. Where strong, feign weakness.’”

—Luo Guanzhong’s historical novel, *Three Kingdoms* (thanks to Duozhe Li of Chinese University of Hong Kong for the reference).

A two-person zero-sum game with complete information and one-sided preplay communication of intentions via cheap talk.

In the story, set around 200 A.D., fleeing general Cao Cao chose between two escape routes, the easier Main Road and the awful Huarong Road, trying to avoid capture by pursuing General Kongming.

Kongming (the sender in this example) waited in ambush along the Huarong Road and set campfires there, thus sending a deceptively truthful message.

Cao Cao (the receiver), misjudging Kongming’s communication strategy, inverted the truthful message and was caught by Kongming.

Huarongdao continued

The key issues here are how Kongming should choose his message and how Cao Cao—knowing Kongming is choosing strategically, trying to anticipate Cao Cao's interpretation—should interpret Kongming's message.

In real settings like this, a receiver's thinking often assigns a prominent role to the literal meanings of messages, without necessarily taking them at face value; and a sender's thinking takes this into account.

But a standard equilibrium analysis precludes a role for the literal meanings of payoff-irrelevant messages (Crawford and Sobel 1982 *Econometrica*; see however Farrell's 1993 *GEB* neologism-proofness refinement, which depends on meanings).

Moreover, there is no equilibrium (refined or not) in a zero-sum two-person game in which cheap talk conveys information or the receiver responds to the message.

In such an equilibrium, if there was information in the sender's message that the receiver found it optimal to respond to, the receiver's response would help him and so hurt the sender, who would then prefer to make his message uninformative.

We will see that the quotation can be understood as a thought process in which players anchor beliefs in an instinctive reaction to the literal meanings of messages and then iterate best responses a small number of times.

October Surprise: "…The news that day was the so-called 'October Surprise' broadcast by bin Laden. He hadn't shown himself in nearly a year, but now, four days before the [2004 presidential] election, his spectral presence echoed into every American home. It was a surprisingly complete statement by the al Qaeda leader about his motivations, his actions, and his view of the current American landscape. He praised Allah and, through most of the eighteen minutes, attacked Bush,…. At the end, he managed to be dismissive of Kerry, but it was an afterthought in his 'anyone but Bush' treatise….

Inside the CIA…the analysis moved on a different [than the presidential candidates' public] track. They had spent years, as had a similar bin Laden unit at FBI, parsing each expressed word of the al Qaeda leader…. What they'd learned over nearly a decade is that bin Laden speaks only for strategic reasons…. Today's conclusion: bin Laden's message was clearly designed to help the President's reelection."—Suskind, *The One Percent Doctrine*, 2006, pp. 335-6 (quoted in Jazayerli 2008 http://www.fivethirtyeight.com/2008/10/guest-column-will-bin-laden-strike.html).

October Surprise continued

A zero-sum two-person game with incomplete information and one-sided preplay communication of private information via cheap talk. Only bin Laden knows which candidate he wants; and, talk being cheap, he will say what it takes to help his candidate win. A representative American voter knows only that he wants whichever candidate bin Laden doesn't want.

The key issues are how bin Laden should relate his statement to what he really wants and how the American should interpret bin Laden's statement, knowing that bin Laden is choosing the message strategically.

Once again, the literal meanings of messages are likely to play a prominent role in applications, but equilibrium analysis precludes such a role.

There is again no equilibrium in which cheap talk conveys information, or in which the receiver responds to the sender's message.

We will see that the quotation can be understood as a thought process in which players anchor beliefs in an instinctive reaction to the literal meanings of messages and then iterate best responses a small number of times.

Bank Runs: “A crude but simple game, related to Douglas Diamond and Philip Dybvig’s 1983 *JPE* celebrated analysis of bank runs, illustrates some of the issues involved here. Imagine that everyone who has invested $10 with me can expect to earn $1, assuming that I stay solvent. Suppose that if I go bankrupt, investors who remain lose their whole $10 investment, but that an investor who withdraws today neither gains nor loses. What would you do? Each individual judgment would presumably depend on one's assessment of my prospects, but this in turn depends on the collective judgment of all of the investors.

Suppose, first, that my foreign reserves, ability to mobilize resources, and economic strength are so limited that if any investor withdraws I will go bankrupt. It would be a Nash equilibrium (indeed, a Pareto-dominant one) for everyone to remain, but (I expect) not an attainable one. Someone would reason that someone else would decide to be cautious and withdraw, or at least that someone would reason that someone would reason that someone would withdraw, and so forth. This…would likely lead to large-scale withdrawals, and I would go bankrupt. It would not be a close-run thing. …Keynes’s beauty contest captures a similar idea.

Now suppose that my fundamental situation were such that everyone would be paid off as long as no more than one-third of the investors chose to withdraw. What would you do then? Again, there are multiple equilibria: everyone should stay if everyone else does, and everyone should pull out if everyone else does, but the more favorable equilibria seems much more robust.”—Lawrence Summers, “International Financial Crises: Causes, Prevention, and Cures,” 2000 *AER*.

An *n*-person coordination game with Pareto-ranked equilibria. Summers presumes that some equilibrium will emerge, but his model of the influence of fragility on equilibrium selection may implicitly invoke initial responses to shocks followed by adaptive learning (although he cites Morris and Shin’s 1998 *AER* non-adaptive “global games” analysis).

Poe's Outguessing Game: "…But he perpetually errs by being too deep or too shallow, for the matter in hand; and many a schoolboy is a better reasoner than he. I knew one about eight years of age, whose success at guessing in the game of 'even and odd' attracted universal admiration. This game is simple, and is played with marbles. One player holds in his hand a number of these toys, and demands of another whether that number is even or odd. If the guess is right, the guesser wins one; if wrong, he loses one. The boy to whom I allude won all the marbles of the school. Of course he had some principle of guessing; and this lay in mere observation and admeasurement of the astuteness of his opponents. For example, an arrant simpleton is his opponent, and, holding up his closed hand, asks, 'are they even or odd?' Our schoolboy replies, 'odd,' and loses; but upon the second trial he wins, for he then says to himself, 'the simpleton had them even upon the first trial, and his amount of cunning is just sufficient to make him have them odd upon the second; I will therefore guess odd'; --he guesses odd, and wins. Now, with a simpleton a degree above the first, he would have reasoned thus: 'This fellow finds that in the first instance I guessed odd, and, in the second, he will propose to himself upon the first impulse, a simple variation from even to odd, as did the first simpleton; but then a second thought will suggest that this is too simple a variation, and finally he will decide upon putting it even as before. I will therefore guess even' guesses even, and wins. Now this mode of reasoning in the schoolboy, whom his fellows termed 'lucky,'—what, in its last analysis, is it?"

'It is merely,' I said, 'an identification of the reasoner's intellect with that of his opponent.'"—Edgar Allan Poe, *The Purloined Letter* (http://xroads.virginia.edu/%7EHYPER/POE/purloine.html)

Poe's Outguessing Game continued

A finite-horizon extensive-form zero-sum two-person "outguessing" game with complete information.

The key issue here is outguessing one's opponents' response to the commonly observed history.

Equilibrium is unhelpful because in a finitely repeated zero-sum game it implies no response to payoff-irrelevant history, via a standard backward-induction argument.

However, in real settings like this, players' strategic thinking often assigns a central role to the history. (After 9/11, should Americans have started worrying about their country's other skyscrapers, or instead about some completely different kind of attack? At the time, both views were expressed in the press.)

The quotation can be understood as a thought process in which players anchor beliefs in an instinctive reaction to the history ("history repeats itself") and then iterates best responses (for the "lucky" schoolboy, the more times, the higher his assessment of his opponent's intellect: once for an "arrant simpleton," twice for a "simpleton a degree above the first," and so on).

Common Features of the Quotations

- They all concern games played without completely clear precedents.

- They all reflect coherent, clearly identified models of strategic thinking.

- But the thinking is systematically different from equilibrium thinking (or for Bank Runs and Outguessing, at least goes beyond equilibrium).

- The thinking tends to start with beliefs anchored in an instinctive reaction to the game, and then to iterate best responses a small number of times. In this respect the thinking resembles that in the "level-*k*" or "cognitive hierarchy" ("CH") models described below. The resemblance is not self-evident for Entry Magic, but as explained below, Camerer, Ho, and Chong 2004 *QJE* explain Kahneman's results via a CH model.

- The instinctive reactions follow different principles, each plausible in its setting, such as uniform randomness, salient labels, or truthfulness.

- Finite iteration of best responses is common across all settings, although the number of iterations may vary across individuals or even settings.

These common features are representative of folk game theory:

- One can also find quotations reflecting one or two steps of iterated (strict or weak) dominance in the normal form, or one or two steps of iterated (weak) dominance reflecting forward or backward induction in the extensive form.

- But it is difficult (counterexamples welcome) to find quotations involving more than one or two steps of iterated dominance.

- And it is at least as difficult (impossible? counterexamples welcome) to find quotations that illustrate the fixed-point reasoning that underlies equilibrium in games without dominance.

In Selten's 1998 *European Economic Review* words (but generalizing about experimental results, not folk game theory):

> "Basic concepts in game theory are often circular in the sense that they are based on definitions by implicit properties…. Boundedly rational strategic reasoning seems to avoid circular concepts. It directly results in a procedure by which a problem solution is found."

To paraphrase:

> "Real people don't use fixed-point reasoning to decide what to do."

This is not to say that with enough experience in a sufficiently stationary setting, learning can't make real people converge to steady states that an *analyst* would need fixed-point reasoning to characterize.

Selten's point is simply that when equilibrium requires fixed-point reasoning, it may not be a good behavioral model of people's cognition.

3. Leading Models of Strategic Thinking

The leading models of strategic thinking all allow players' strategies to be in equilibrium, but do not assume equilibrium in all games. They include:

- Adding noise to equilibrium predictions ("equilibrium plus noise"), plus refinements such as risk- or payoff-dominance or "global games".
- Finitely iterated strict dominance and *k*-rationalizability (Bernheim 1984 *Econometrica*, Pearce 1984 *Econometrica*; the two notions are equivalent in the two-person games I mostly focus on here).
- Quantal response equilibrium ("QRE") and its leading special case, logit QRE ("LQRE") (McKelvey and Palfrey 1995 *GEB*).
- "Level-*k*" models (Nagel 1995 *AER*; Stahl and Wilson 1994 *JEBO*, 1995 *GEB*; Ho, Camerer, and Weigelt 1998 *AER* ("HCW"); Costa-Gomes, Crawford, and Broseta 2001 *Econometrica* ("CGCB"); and Costa-Gomes and Crawford 2006 *AER* ("CGC").
- Cognitive hierarchy ("CH") models (Camerer, Ho, Chong 2004 *QJE*).
- Noisy introspection ("NI") models (Goeree and Holt 2004 *GEB*).

A. Equilibrium Plus Noise

Equilibrium plus noise adds noise with a specified payoff-sensitive error distribution (usually logit) and an estimated precision parameter to equilibrium predictions.

Although a player's error distribution is sensitive to the payoff costs of errors, those costs are evaluated assuming (unlike in most other models discussed here) that other players play their equilibrium strategies without errors.

Equilibrium plus noise often describes observed behavior well but sometimes misses systematic patterns in subjects' deviations from equilibrium.

B. Finitely Iterated (Strict) Dominance and *k*-Rationalizability

Finitely iterated strict dominance and *k*-rationalizability (Bernheim 1984 *Econometrica*, Pearce 1984 *Econometrica*) are set-valued restrictions on individual players' strategies.

(Equilibrium and QRE, by contrast, restrict the *relationship* among players' strategies. Level-*k*, CH, and NI models, by contrast, normally make unique (though possibly probabilistic) predictions conditional on the behavioral parameters, as does equilibrium plus noise when suitably "completed" by adding a refinement as discussed below.)

(Finitely iterated strict dominance and *k*-rationalizability are equivalent in two-person games; their differences in *n*–person games are unimportant here.)

Informally, a 1-rationalizable strategy (the sets R1 on the next slide) is one for which there is a profile of others' strategies that make it a best response; a 2-rationalizable strategy (the sets R2) is one for which there exists a profile of others' 1-rationalizable strategies that make it a best response; and so on.

The more familiar notion of rationalizability is equivalent to *k*-rationalizability for all *k*.

Rationalizability reflects the implications of common knowledge (sometimes replaced in the modern literature by common belief) of rationality (with no further restrictions on beliefs).

k-rationalizability reflects the implications of finite numbers of levels of mutual knowledge (or sometimes, belief) of rationality.

		R1,R2 L	R1,R2,R3,R4 C	R
R1,R2,R3	T	0 7	5 0	3 0
R1,R2,R3,R4	M	0 5	2 2	0 5
R1	B	7 0	5 0	3 7

Dominance-solvable game

		Rk for all k L	Rk for all k C	Rk for all k R
Rk for all k	T	0 7	5 0	7 0
Rk for all k	M	0 5	2 2	0 5
Rk for all k	B	7 0	5 0	0 7

Unique equilibrium without dominance

Each game has a unique equilibrium (M,C). In the first game M and C are the only rationalizable strategies; in the second all strategies are rationalizable.

Equilibrium reflects the implications of common knowledge of rationality *plus* common beliefs: Any equilibrium strategy is *k*-rationalizable for all *k*, but not all combinations of rationalizable strategies are in equilibrium.

In games that are dominance-solvable in *k* rounds, *k*-rationalizability implies that players have the same beliefs—with a qualification for mixed-strategy equilibrium that is not important here—so any combination of *k*-rationalizable strategies is in equilibrium, as in the first game on the above slide.

In other games, *k*-rationalizability and rationalizability allow deviations from equilibrium, as in the second game, where there is a "tower" of beliefs, consistent with common knowledge of rationality, to support any outcome.

(But except for the equilibrium beliefs (M, C) the beliefs differ across players.)

As we will see, finitely iterated dominance and *k*-rationalizability are often consistent with systematic patterns in subjects' deviations from equilibrium.

Finitely iterated dominance and *k*-rationalizability could be combined with an econometric error structure as in Aradillas-Lopez and Tamer 2008 *JBES*, but usually have not been.

C. Quantal Response Equilibrium ("QRE") and Logit QRE ("LQRE")

To capture the payoff-sensitivity of deviations from equilibrium that equilibrium plus noise sometimes misses, McKelvey and Palfrey 1995 *GEB* proposed the notion of Quantal Response Equilibrium or "QRE".

In a QRE, as in equilibrium plus noise, players' decisions are noisy, with the probability density of each decision increasing in its expected payoff.

But unlike in equilibrium plus noise—or in level-*k* and CH models, discussed below—the payoff costs of deviations are 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 others' distributions.

As the distributions' precision increases, QRE approaches equilibrium; and as their precision 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.

In applications LQRE’s precision is estimated econometrically or calibrated from previous analyses.

Like equilibrium plus noise, QRE is a general model of strategic behavior, applicable to any game, with a small number of behavioral parameters.

With estimated precision, LQRE’s sensitivity to out-of-equilibrium payoffs often allows it to fit subjects’ initial responses better than an equilibrium plus noise model.

D. Level-*k* Models

An alternative way to describe or explain the payoff-sensitivity of deviations from equilibrium is to treat the deviations as an integral part of the structure rather than as responses to errors.

Although the number of logically possible non-equilibrium structures seems daunting, both folk game theory and experimental evidence support a particular class of models called level-*k* or cognitive hierarchy (CH) models.

Level-*k* models allow behavior to be heterogeneous, but assume that each player follows a rule drawn from a common distribution over a particular hierarchy of decision rules or *types* (as they are called in this literature; no relation to “types” as realizations of private information variables).

Type *Lk* anchors its beliefs in a nonstrategic *L0* type, which is meant to describe *Lk*’s model of others’ instinctive reactions to the game.

The instinctive reactions may follow one of several principles depending on the setting, such as uniform randomness, salience, or truthfulness.

Lk then adjusts its beliefs 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 in that they choose best responses to beliefs, with perfect models of the game.

Lk's only departure from equilibrium is in replacing its perfect model of others' decisions with simplified models that avoid the complexity of equilibrium.

In applications it is usually assumed that *L1* and higher types make errors, which are often taken to be logit with estimated precision as in LQRE.

Thus the probability density of each type's decision is increasing in its expected payoff, evaluated using the type's model of others' decisions: *L2*, for example, makes errors whose distribution is sensitive to the payoff costs of deviations, evaluated assuming that other players' decisions are *L1*.

Unlike LQRE, *Lk* types do not respond to the noisiness of others' decisions.

Even so, the deterministic structure of a level-*k* model captures the sensitivity of players' deviations from equilibrium to out-of-equilibrium payoffs.

The population type frequencies are treated as behavioral parameters, to be estimated from the data or translated or extrapolated from previous analyses.

The estimated type distribution is typically fairly stable across games, with most weight on *L1*, *L2*, and perhaps *L3*.

The estimated frequency of the anchoring *L0* type is usually small.

Thus, *L0* “exists” mainly as *L1*’s model of others, *L2*’s model of *L1*’s model of others, and so on.

Low frequencies of *L0* are an important sign of health for a level-*k* model, in that high frequencies of *L0* would reduce the model to a parameterized distribution of responses, thus describing the data rather than explaining it.

Only when the strategic iteration of best responses plays a role can the model yield a useful explanation of the data.

Even though *L0* normally has a low frequency, its specification is the main issue in defining a level-*k* model and the key to its explanatory power.

As illustrated below, *L0* needs to be adapted to the setting, and there is an emerging consensus about how to do this in particular applications.

By contrast, the definition of *L1*, *L2*, and *L3* via iterated best responses allows a simple, reliable explanation of behavior across different settings.

Like equilibrium plus noise and QRE, level-*k* models are general models of strategic behavior, with small numbers of behavioral parameters.

Like CH models, discussed below, level-*k* models make point predictions that depend only on *L0* and the estimated type distribution.

L1 and higher types make undominated decisions, and *Lk* complies with *k* rounds of iterated dominance and *k*–rationalizability (thanks to Robert Östling of Stockholm University for clarifying this relationship).

Thus, a distribution of *Lk* types realistically concentrated on low levels of *k* mimics equilibrium in games that are dominance-solvable in a few rounds.

But such a distribution deviates systematically from equilibrium in some more complex games, in predictable ways.

These features allow level-*k* models to capture the sensitivity of deviations from equilibrium to out-of-equilibrium payoffs.

As a result, like LQRE, level-*k* (and CH) models often fit initial responses better than equilibrium plus noise.

E. Cognitive Hierarchy ("CH") Models

In Camerer, Ho, and Chong's 2004 *QJE* cognitive hierarchy ("CH") model, a close relative of level-*k* models, *Lk* best responds not to *Lk-1* alone but to an estimated mixture of lower-level types; and the type frequencies are not unrestricted, but instead are treated as a parameterized Poisson distribution.

For an outside observer modeling behavior econometrically, this estimated-mixture specification seems more natural than the level-*k* specification.

But which specification better describes people's strategic thinking remains an empirical question (on which the jury is still not completely in).

A CH *L1* is the same as a level-*k* *L1*, but CH *L2* and higher types may differ.

A CH *L1* and higher types make undominated decisions, but unlike level-*k* types, but a CH *Lk* might not comply with *k* rounds of iterated dominance and *k*–rationalizability.

Unlike in a level-*k* model, in a CH 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 *L1* and higher types.

A CH model makes point predictions that depend only on *L0* and the estimated Poisson parameter.

In some applications the Poisson constraint, imposed as a simplifying restriction, is not very restrictive and the CH model fits as well as a level-*k* model; but in others the Poisson constraint is strongly binding.

F. Noisy Introspection ("NI") Models

Although LQRE has so far been the most popular model of initial responses, not all researchers consider it suitable for that purpose.

McKelvey and Palfrey 1995 *GEB* suggest using LQRE for both initial responses and limiting outcomes, in the latter case with precision increasing over time as a reduced-form model of learning.

But Goeree and Holt 2004 *GEB* suggest using LQRE for limiting outcomes, instead proposing a Noisy Introspection ("NI") model for initial responses.

NI relaxes LQRE's equilibrium assumption while maintaining its assumption that players best respond to a probability distribution of others' responses:

Players form beliefs by iterating best responses roughly as in a level-*k* model, but with higher-order beliefs reflecting increasing amounts of noise.

For a given noise distribution, NI 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 extreme cases of NI mimic level-*k* types:

  If the noise jumps immediately to ∞, NI beliefs are *L0*.

  If the noise is zero for one iteration and then jumps immediately to ∞, NI beliefs are *L1*; and so on.

- In applications GH assume that the noisiness of higher-order beliefs grows geometrically with the number of iterations, which yields a range of possible decisions depending on the noise level and its rate of growth.

4. Experimental Evidence from Normal-Form Games

Level-*k* and CH models are now supported by a large body of experimental evidence on initial responses to games with various structures:

- Stahl and Wilson 1994 *JEBO*, 1995 *GEB*; Costa-Gomes, Crawford, and Broseta 2001 *Econometrica* ("CGCB"); Crawford and Iriberri 2007 *AER*; Costa-Gomes and Weizsäcker 2008 *RES* (normal-form matrix games).
- Nagel 1995 *AER*; Ho, Camerer, and Weigelt 1998 *AER* ("HCW"); Costa-Gomes and Crawford 2006 *AER* ("CGC") (normal-form guessing games).
- Camerer, Ho, and Chong 2004 *QJE* ("CHC") (normal-form matrix games, **e**ntry games, and incomplete-information zero-sum betting games).
- Crawford and Iriberri 2007 *Econometrica* (incomplete-information auctions).
- Cai and Wang 2006 *GEB*; Wang, Spezio, and Camerer 2009 *AER*; Kawagoe and Takizawa 2009 *GEB* (extensive-form communication games).
- Johnson, Camerer, Sen and Rymon 2002 *JET* (extensive-form bargaining).
- Kawagoe and Tazikawa 2009 (extensive-form centipede games).

Here I focus on two representative experiments with normal-form games:

- Nagel's 1995 *AER* experiments, which were directly inspired by Keynes's Beauty Contest, and which provide a simple introduction to the evidence and the class of models that it suggests.

- CGC's 2006 *AER* experiments, which use a more powerful design to identify subjects' strategic thinking more precisely.

CGC's conclusions are fully consistent with the conclusions of other studies of initial responses to abstract normal-form games, just more precise.

(CGC's Introduction and Section II.D summarize the evidence from Stahl and Wilson 1994 *JEBO* and 1995 *GEB*; HCW; CGCB; and CHC.)

With adjustments described below, CGC's conclusions are also consistent with those of the studies of the other kinds of games mentioned above.

I illustrate this later by discussing Wang, Spezio, and Camerer's 2009 *AER* experiments on communication games.

A. Nagel's Design and Results

In Nagel's *n*-person guessing game design:

- 15-18 subjects simultaneously guessed between [0,100].

- The subject whose guess was closest to a target *p* (= 1/2 or 2/3, say), times the group average guess wins a prize, say $50.

- The structure was publicly announced.

If you have not already done so, please take a moment to decide what you would guess, in a group of non-game-theorists:

- if *p* = 1/2,

- if *p* = 2/3.

Nagel's games have a unique equilibrium, in which all players guess 0.

The games are dominance-solvable, so the equilibrium can be found by iteratively eliminating dominated guesses.

For example, if $p = 1/2$:

- It's dominated to guess more than 50 (because $1/2 \times 100 \leq 50$).
- Unless you think that other people will make dominated guesses, it's also dominated to guess more than 25 (because $1/2 \times 50 \leq 25$).
- And so on, down to 12.5, 6.25, 3.125, and eventually to 0.

The rationality-based argument for this “all–0” equilibrium is stronger than many equilibrium arguments, because it depends only on iterated knowledge of rationality, not on the assumption that players have the same beliefs.

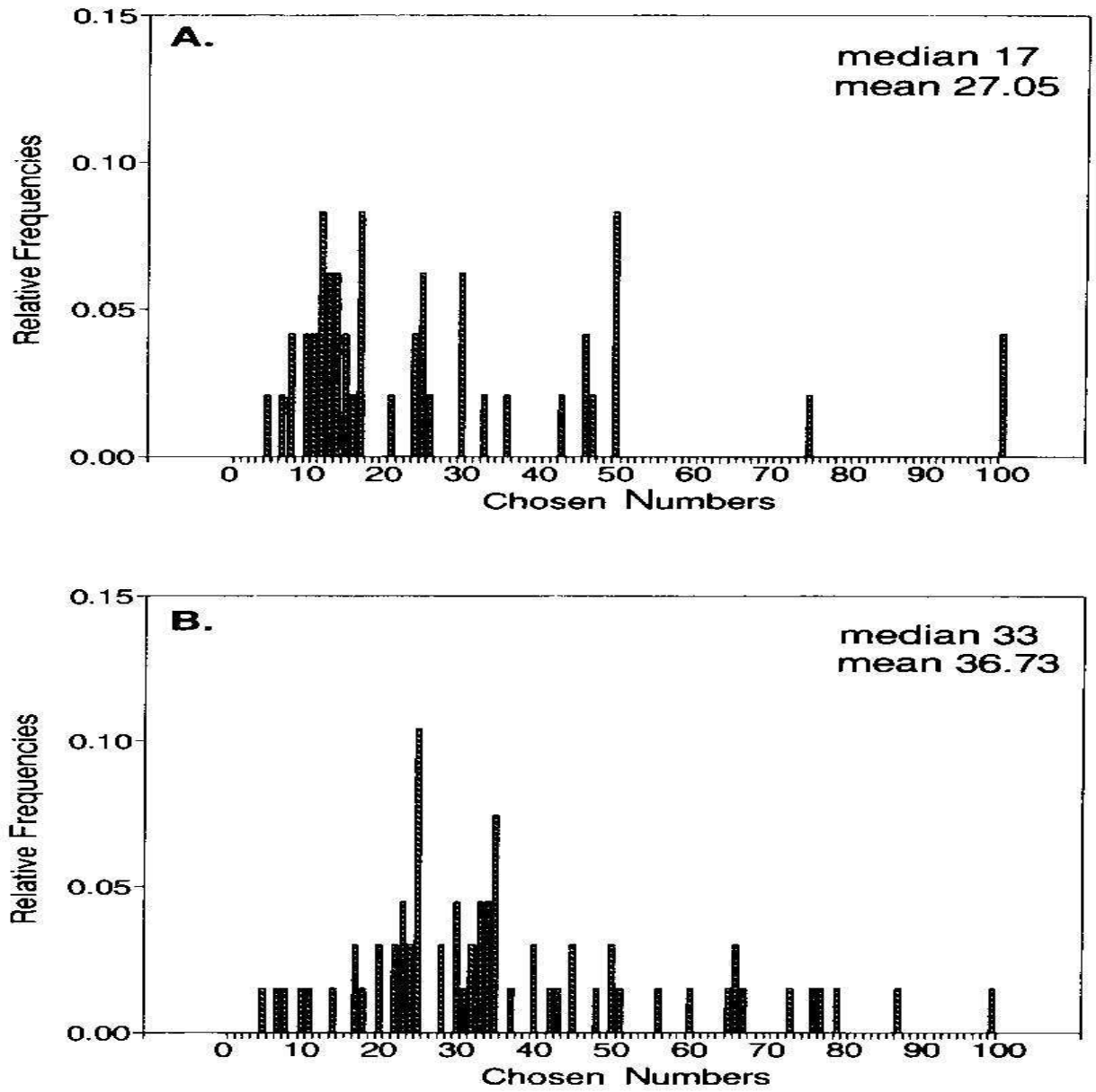
However, even people who are rational are seldom certain that others are rational, or that others believe that others are rational.

Thus, they won’t (and shouldn’t) guess 0. But what do (should) they do?

Nagel’s subjects played these games repeatedly, but we can view their initial guesses as responses to games played as if in isolation if they treated their influences on the future as negligible, which is plausible in groups of 15 to 18.

Nagel’s subjects never played their equilibrium strategies initially, and their responses deviated systematically from equilibrium.

Instead there were spikes that suggest a distribution of discrete thinking “types,” respecting 0 to 3 rounds of iterated dominance in each treatment (next slide).



Part of Nagel's Figure 1: top of figure $p = 1/2$, bottom of figure $p = 2/3$.

The spikes' locations and how they vary across treatments are roughly consistent with two plausible interpretations:

- In one interpretation, called *Dk*, a player does *k* rounds of iterated dominance for some small number, *k* = 1 or 2, and then best responds to a uniform prior over other players' remaining strategies (thus "completing" *k*-rationalizability by adding a specific selection as discussed below).

- In another interpretation, "level-*k*" or "*Lk*," a player starts with a naïve prior *L0* over others' strategies reflecting people's instinctive reactions to the game, and then iterates best responses *k* times, with *k* = 1, 2, or 3.

  In abstractly framed games like Nagel's, *L0* is usually taken to be a uniform random distribution, reflecting a player's understanding of the payoff function before he tries to model others' decisions.
  (In games without dominance this makes *Dk*, *k* = 1,2,… coincide with *L1*.)

  (Although in these lectures I focus mainly on two-person games, in *n*-person games it matters whether *L0* is independent across players or correlated, and the limited evidence (HCW, Costa-Gomes, Crawford, and Iriberri 2009 *JEEA*) suggests that most people have highly correlated models of others. Here I take *L0* to model all others' average guess.)

In many games *Dk* and *Lk*+1 respond similarly to dominance, yielding *k*-rationalizable strategies. (The difference in indices is only a quirk of notation.)

With a uniform random *L0*, in Nagel's games *Dk*'s and *Lk*+1's guesses are perfectly confounded, both tracking the spikes in Nagel's data across her treatments (which had different subject groups):

- *Dk* guesses $([0+100p^k]/2)p$.
- *Lk*+1 guesses $[(0+100)/2]p^{k+1}$.

Either way, one aspect of the message is already clear: Subjects do not rely on indefinitely iterated dominance or indefinitely iterated best responses; instead their decisions respect *k*-rationalizability for at most small values of *k*.

Despite the lack of separation of *Dk*'s and *Lk*+1's guesses, many theorists interpret Nagel's results as evidence that subjects explicitly performed finitely iterated dominance, the way we teach students to solve such games.

In HCW's and CGCB's experiments, *Dk*'s and *Lk*+1's guesses are weakly separated, and the results are inconclusive on this point; but in CGC's experiments *Dk*'s and *Lk*+1's guesses are strongly separated, and we will see that the results very clearly favor *Lk* over *Dk* rules.

B. CGC's Design and Results

In CGC's design, subjects were randomly and anonymously paired to play a series of 16 different two-person guessing games, with no feedback.

The design suppresses learning and repeated-game effects in order to elicit subjects' initial responses, game by game, studying strategic thinking "uncontaminated" by learning.

("Eureka!" learning was possible, but it was tested for and found to be rare.)

The design combines the variation of games of Stahl and Wilson's 1995 *GEB* design with the large strategy spaces of Nagel's 1995 *AER* design.

This greatly enhances its power, and the profile of a subject's guesses in the 16 games forms a "fingerprint" that helps to identify his strategic thinking more precisely than is possible by observing his responses to a series of games with small strategy spaces or a single game with large strategy space.

In CGC's guessing games, each player has his own lower and upper limit, both strictly positive, implying finite dominance-solvability.

(Players are not actually required to guess between their limits. Instead guesses outside the limits are automatically adjusted up to the lower limit or down to the upper limit as necessary: a trick to enhance separation of information search implications, not important for this discussion.)

Each player also has his own target, and his payoff increases with the closeness of his guess to his target times the other's guess.

The targets and limits vary independently across players and games, with targets both less than one, both greater than one, or "mixed".

(In Nagel's and HCW's previous guessing experiments, the targets and limits were always the same for both players, and they varied at most across treatments with different subject groups.)

CGC's guessing games have essentially unique equilibria ("essentially" due to the automatic adjustment), determined (not always directly) by players' lower (upper) limits when the product of targets is less (greater) than one.

The discontinuity of the equilibrium correspondence when the product of targets equals one stress-tests equilibrium, which responds much more strongly to the product of the targets than alternative decision rules do; and enhances the separation of equilibrium from alternative rules.

(It also reveals other interesting patterns, only briefly mentioned below; see Crawford, "Look-ups as the Windows of the Strategic Soul" at http://dss.ucsd.edu/~vcrawfor/#Search.)

Consider a game in which players' targets are 0.7 and 1.5, the first player's limits are [300, 500], and the second's are [100, 900].

The product of targets is 1.05 > 1, and it can be shown that the equilibrium is therefore determined by players' upper limits. (When the product of targets is < 1, the equilibrium is determined by their lower limits in a similar way.)

In equilibrium the first player guesses his upper limit of 500, but the second player guesses 750 (= 500 × his target 1.5), below his upper limit of 900.

No guess is dominated for the first player, but any guess outside [450, 750] is dominated for the second player.

Given this, any guess outside [315, 500] is iteratively dominated for the first player.

Given this, any guess outside [472.5, 750] is dominated for the second player, and so on until the equilibrium at (500, 750) is reached after 22 rounds of iterated dominance.

CGC’S Data Analysis

As suggested by previous work, CGC’s data analysis assumed that each subject’s guesses were determined, up to logit errors, by a single decision rule, or “type” as they are called in this literature (no relation to the use of “type” for the realization of a private information variable), in all 16 games.

This assumption was tested and found reasonable for almost all subjects.

Most of CGC's data analysis restricted attention to a list of behaviorally plausible types whose relevance was suggested by previous work:

- *L0*, *L1*, *L2*, and *L3*, with *L0* uniform random between a player's limits, *L1* best responding to *L0*, *L2* to *L1*, and so on.
- *D1* and *D2*, which does one round (respectively, two) of iterated dominance and then best responds to a uniform prior over its partner's remaining decisions (making a specific selection from *k*-rationalizable strategies).
- *Equilibrium*, which makes its equilibrium decisions.

  (Note that because CGC's games are all (finitely) dominance-solvable, traditional equilibrium refinements are not relevant in them.)
- *Sophisticated*, which best responds to the probability distributions of others' decisions, estimated from the observed frequencies.

  (*Sophisticated* is an ideal, included to learn if any subjects have an understanding of others' decisions that transcends mechanical rules.)

The restriction to this list was also tested as explained below, and found to be a reasonable approximation to the support of subjects' decision rules.

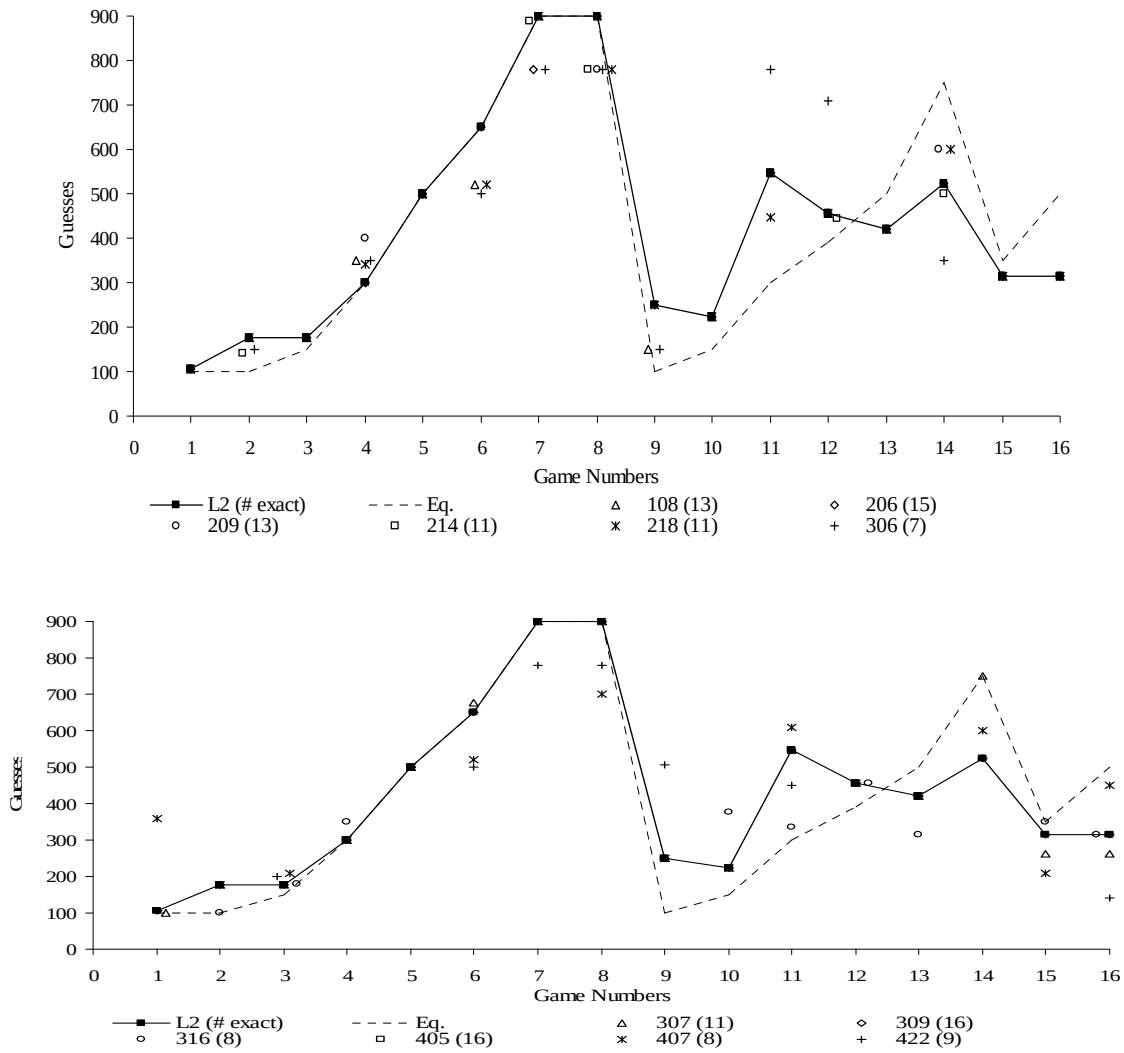
CGC's large strategy spaces and the independent variation of targets and limits across games greatly enhance the separation of types' implications, to the point where many subjects' types can be precisely identified from their guessing "fingerprints":

Types' guesses in the 16 games, in (randomized) order played

	L1	*L2*	*L3*	*D1*	*D2*	*Eq.*	*Soph.*
1	**600**	**525**	**630**	**600**	**611.25**	**750**	**630**
2	**520**	**650**	**650**	**617.5**	**650**	**650**	**650**
3	**780**	**900**	**900**	**838.5**	**900**	**900**	**900**
4	**350**	**546**	**318.5**	**451.5**	**423.15**	**300**	**420**
5	**450**	**315**	**472.5**	**337.5**	**341.25**	**500**	**375**
6	**350**	**105**	**122.5**	**122.5**	**122.5**	**100**	**122**
7	**210**	**315**	**220.5**	**227.5**	**227.5**	**350**	**262**
8	**350**	**420**	**367.5**	**420**	**420**	**500**	**420**
9	**500**	**500**	**500**	**500**	**500**	**500**	**500**
10	**350**	**300**	**300**	**300**	**300**	**300**	**300**
11	**500**	**225**	**375**	**262.5**	**262.5**	**150**	**300**
12	**780**	**900**	**900**	**838.5**	**900**	**900**	**900**
13	**780**	**455**	**709.8**	**604.5**	**604.5**	**390**	**695**
14	**200**	**175**	**150**	**200**	**150**	**150**	**162**
15	**150**	**175**	**100**	**150**	**100**	**100**	**132**
16	**150**	**250**	**112.5**	**162.5**	**131.25**	**100**	**187**

Of the 88 subjects in CGC's main treatments, 43 made guesses that complied *exactly* (within 0.5) with one type's guesses in from 7 to 16 of the games (20 *L1*, 12 *L2*, 3 *L3*, and 8 *Equilibrium*).

For example, CGC's Figure 2 (next slide) shows the "fingerprints" of the 12 subjects whose guesses conformed most closely to *L2*'s; 72% of their guesses were exact *L2* guesses; only their deviations are shown.



CGC's Figure 2. "Fingerprints" of 12 Apparent *L2* Subjects
(Only deviations from *L2*'s guesses are shown.
Of these subjects' 192 guesses, 138 (72%) were exact *L2* guesses.)

The size of CGC's strategy spaces, with 200 to 800 possible exact guesses in each of 16 different games, makes exact compliance powerful evidence for the type whose guesses are tracked: If a subject chooses 525, 650, 900 in games 1-3, intuitively and econometrically we already "know" he's an *L2*.

(By contrast, there are usually many possible reasons for choosing one of the strategies in a small matrix game; and even in Nagel's large strategy spaces, rules as cognitively disparate as *Dk* and *Lk*+1 yield identical decisions.)

Further, because CGC's definition of *L2* builds in risk-neutral, self-interested rationality, we also know that a subject's deviations from equilibrium are "caused" not by irrationality, risk aversion, altruism, spite, or confusion, but by his simplified model of others.

(Even so, doubts remain about the subjects with high exact compliance with *Equilibrium*, who appear to be following hybrid types that only mimic equilibrium in the games with targets both less than one or both greater than one; see Crawford, "Look-ups as the Windows of the Strategic Soul".)

That the level-*k* model is *directly* suggested by these subjects' data (rather than via data-fitting exercises) is an important advantage over alternatives.

CGC's other 45 subjects made guesses that conformed less closely to one of CGC's types, but econometric estimates of their types are concentrated on *L1*, *L2*, *L3*, and *Equilibrium*, in roughly the same proportions.

TABLE 1—SUMMARY OF BASELINE AND OB SUBJECTS' ESTIMATED TYPE DISTRIBUTIONS

Type	Apparent from guesses	Econometric from guesses	Econometric from guesses, excluding random	Econometric from guesses, with specification test	Econometric from guesses and search, with specification test
L1	20	43	37	27	29
L2	12	20	20	17	14
L3	3	3	3	1	1
D1	0	5	3	1	0
D2	0	0	0	0	0
Eq.	8	14	13	11	10
Soph.	0	3	2	1	1
Unclassified	45	0	10	30	33

Note: The far-right-hand column includes 17 OB subjects classified by their econometric-from-guesses type estimates.

For those 45 subjects, there is some room for doubt about whether CGC's specification omits relevant types and/or overfits by including irrelevant types.

To test for this, CGC conducted a specification test, which suggests that the types estimated to be in the population are relevant and that any omitted types are at most 1-2% of the population, hence not worth modeling.

Aside on CGC's Specification Test

To test for overfitting and omission of relevant types, CGC conducted a specification test, which compares the likelihood of each subject's econometric type estimate with the likelihoods of estimates based on 88 *pseudotypes*, each constructed from one subject's guesses in the 16 games.

With regard to overfitting, for a subject's type estimate to be credible it should have higher likelihood than at least as many pseudotypes as it would at random: with 8 types, assuming approximately i.i.d. likelihoods, this makes $87/8 \approx 11$.

Some subjects' type estimates do not pass this test, and so are left unclassified in columns 5 and 6 of CGC's Table 1.

With regard to omitted types, imagine that CGC had omitted a relevant type, say for concreteness *L2*.

The pseudotypes of CGC's estimated *L2* subjects would then outperform the non-*L2* types estimated for them and make approximately the same guesses.

Finding such a *cluster*, CGC diagnosed an omitted type, and studied what its subjects' guesses had in common to reveal its decision rule.

CGC found five small clusters involving 11 of the 88 subjects, and the subjects in these clusters were also left unclassified in Table 1.

The paper and its web appendix discuss what these 11 subjects seemed to be doing; most of it appears quite idiosyncratic.

Because a cluster must contain at least two subjects, it is reasonable to anticipate finding more than the five CGC found in a larger sample.

But because any such clusters did not reach the two-subject threshold in CGC's sample of 88, they are probably at most 2% of any larger sample.

(End of aside)

5. Lessons from the Experiments for Modeling Strategic Behavior

First, Nagel's 1995 *AER* subjects' initial guesses resembled neither equilibrium plus noise nor QRE for any reasonable distribution.

(Distributional assumptions are crucial here: Haile, Hortacsu, and Kovenock 2008 *AER* show that with an unrestricted distribution, QRE can "explain" any given dataset. Thus the power of QRE comes mainly from its distributional assumptions. But the use of the logit distribution in almost all applications has been guided by fit and custom rather than evidence.)

Nagel's results also suggest that even rationalizability is too strong: most subjects' guesses respected *k*–rationalizability only for small values of *k*.

Finally, Nagel's results call into question the common assumption that strategic thinking is homogeneous in the population. No model that imposes homogeneity, as equilibrium plus noise, QRE, and NI do, will do full justice to subjects' behavior. And allowing heterogeneity of strategic thinking is essential for the explanations of Kahneman's Entry Magic, Yushchenko, Lake Wobegon, and Huarongdao proposed below.

CGC's 2006 *AER* analysis significantly sharpens Nagel's conclusions, confirming by direct and econometric evidence and a specification test that a level-*k* model with a uniform random *L0* and only *L1*, *L2*, *L3*, and, possibly, *Equilibrium* subjects explains a large fraction of subjects' deviations from equilibrium in their games. In particular:

- There are no *Dk* subjects. CGC's subjects respect iterated dominance to the extent that *Lk* types do, not because they explicitly perform it.

  (This is reinforced by CGC's data on subjects' searches for hidden payoff information (Crawford, "Look-ups as the Windows of the Strategic Soul") and by their data on "robot/trained subjects," where 7 of 19 subjects, who were trained and rewarded to follow type *D1* and passed an understanding test in which *L2* answers were incorrect, then "morphed" into *L2* (*D1*'s closest *Lk* relative) in the guesses for which they were paid. Aside from the one of 19 robot/trained *D2* subjects who morphed into *L3*, this was the only kind of morphing that occurred. Although by standard measures *Dk*'s cognitive requirements are close to *Lk*+1's, and these treatments also show that most subjects were capable of learning to follow *Dk*, the morphing suggests that subjects find iterated dominance far less natural than the iterated best responses that underlie *Lk* rules.)

- Although level-*k* subjects make decisions that, via the iterated best responses that govern their strategic thinking) respect *k*-rationalizability, their presence is limited to small values of *k*, so even the *Lk* types respect *k*-rationalizability for at most small values of *k*.

- There are no *Sophisticated* subjects. Even the most sophisticated subjects seem to favor rules of thumb over less structured strategic thinking.

  (The jury is still out on the extent to which this conclusion generalizes.)

- Although about half of CGC's subjects' deviations from equilibrium remain unexplained by their proposed level-*k* model, CGC's specification test suggests that those deviations have little or no discernable structure; thus it may still be optimal to treat the remaining unexplained deviations as errors.

- CGC's evidence and analysis are more precise than previous studies of initial responses to normal-form games, but their conclusions are fully consistent with the results of earlier studies as well as folk game theory.

I now give more detailed comparisons of level-*k* versus alternative models.

Level-*k* versus CH Models

By a quirk of CGC's design (CGC's footnotes 34 and 36, p. 1763), level-*k* types' decisions are not separated from their CH counterparts' decisions:

- CHC's *L1* is identical to *CGC's L1*, and
- by fn. 34's "median voter" result (which stems from the piecewise linearity and symmetry of the payoff function), for empirically plausible type distributions, CHC's *L2* and *L3* are both identical to CGC's *L2* (fn. 36).

However, to fit the data CHC's Poisson parameter τ (roughly, the average *k*) must be approximately 1.5, which constrains the frequency of *L0* to 0.22.

By contrast, CGC's and other unconstrained estimates almost always assign *L0* a far lower frequency, usually 0 (and this, I argued above, is a good sign).

Thus the Poisson constraint is strongly binding in CGC's dataset, and with comparable error structures (though possibly not with the structure often assumed for CH, in which a uniform random *L0* doubles as the error structure for higher types), level-*k* will have an advantage in fit over CH.

Further, CGC's data on subjects' searches for hidden payoff information (Crawford, "Look-ups as the Windows of the Strategic Soul") are much more consistent with the search implications of the level-*k* model than with those of a CH model, which blurs the implications of some important types.

Finally, as illustrated below, estimating an unconstrained type distribution as in a level-*k* model provides a useful diagnostic:

If the data can only be fitted by a weird type distribution—non-hump-shaped (in a homogeneous population) or with implausibly high frequencies of higher types—then the explanation is not credible.

I conclude that the evidence is at least as favorable to level-*k* as to CH.

Level-*k* versus Equilibrium Plus Noise or LQRE Models

CGC's (footnote 34, p. 1763) "median voter" result shows that in CGC's games, except for small asymmetries in the payoff function due to automatic adjustment to the limits, equilibrium plus logit noise coincides with LQRE.

CGC's results, like Nagel's, call into question equilibrium plus noise and LQRE's assumption that strategic thinking is homogeneous in the population.

CGC's econometric analysis allows heterogeneity, with equilibrium plus noise represented by the *Equilibrium* type, with logit errors.

Only 11 of the 88 subjects in CGC's main treatments are estimated to be *Equilibrium* subjects, and there is clear evidence that even those subjects are following a rule or rules that only mimics *Equilibrium*, and that only in some of the games (Crawford, "Look-ups as the Windows of the Strategic Soul").

For CGC's remaining 77 subjects, equilibrium plus logit noise and LQRE both miss clear patterns in the data.

But these subjects' "errors" neither center on 0 nor usually exhibit the sensitivity to deviation costs assumed in a logit specification.

(We believe this is because the errors are cognitive or structural, reflecting misspecification rather than a trade-off between effort cost and accuracy.)

Instead the errors have a clear deterministic structure, which is well described by the level-*k* model that emerges from CGC's estimates.

Aside on CGC's "Near-Equilibrium" Subjects

I focus on the eight subjects whose fingerprints are closest to equilibrium.

Order the games by strategic structure, with the eight games with mixed targets (one > 1, one < 1) on the right.

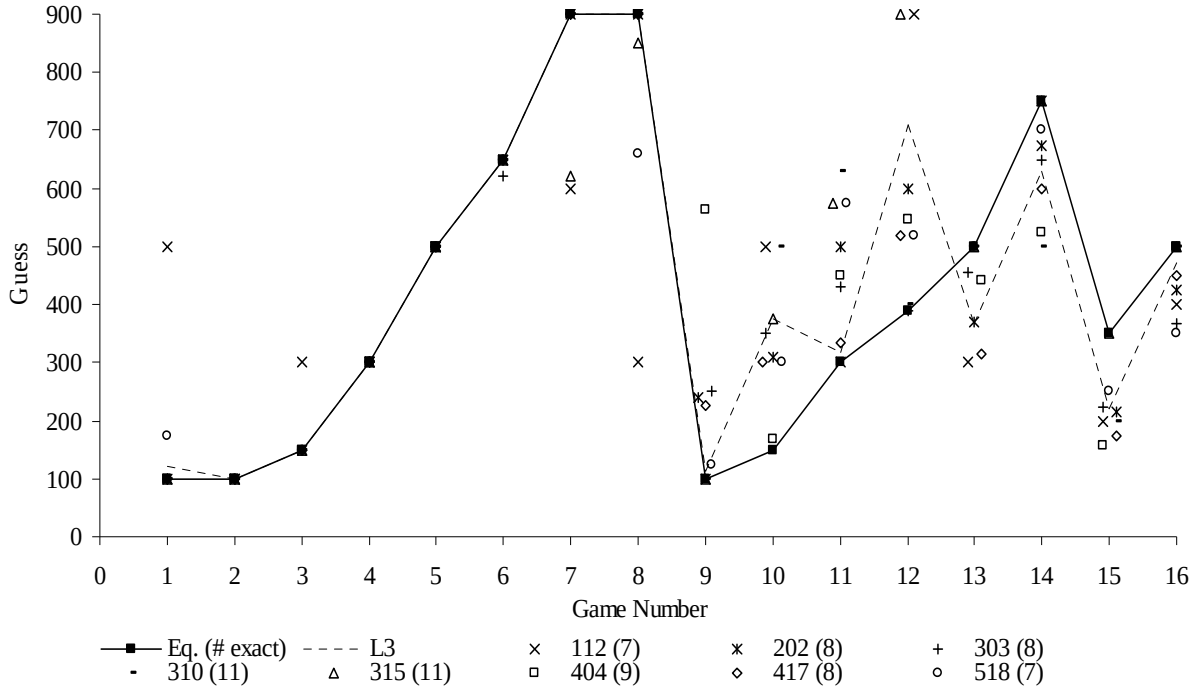
CGC's Figure 4 (next slide) then shows that these subjects' deviations from equilibrium almost always occur with mixed targets.

Thus it is (nonparametrically) clear that these subjects, whose equilibrium compliance is off the scale by normal standards, are actually following a rule that only mimics *Equilibrium*, and that only in games without mixed targets.

Yet all the ways we teach people to identify equilibria (best-response dynamics, equilibrium checking, and iterated dominance) work equally well with and without mixed targets.

Thus, whatever these subjects are doing, it's something we haven't thought of yet.

(And their debriefing questionnaires don't tell us what it is.)



CGC's Figure 4. "Fingerprints" of 8 Apparent Baseline *Equilibrium* Subjects

(Only deviations from *Equilibrium*'s guesses are shown.
69 (54%) of these subjects' 128 guesses were exact *Equilibrium* guesses.)

(End of aside)

Level-*k* versus NI Models

Recall that in an NI model, players form beliefs by iterating best responses, with higher-order beliefs reflecting increasing amounts of noise.

In Goeree and Holt's favored specification, in which the noisiness of higher-order beliefs grows geometrically, the highest-order relevant beliefs are uniform random, which I take to mean within the limits in CGC's games.

Because the NI decision is a continuous function of the noise level and its rate of growth, varying those parameters yields a range of possible decisions.

Preliminary calculations assuming geometric growth suggest that in CGC's games that range is wide, spanning level-*k* types as well as *Equilibrium*.

If so, NI is overparameterized for applications to a single game, unlike equilibrium plus noise, LQRE, level-*k*, and CH models.

NI may therefore risk overfitting even in datasets that span multiple games.

(Costa-Gomes, Iriberri, and I are comparing the models in CGC's dataset.)

Observations about the Models' Cognitive Requirements

Recalling that to work well, models of strategic thinking must accurately reflect people's cognition, comparing players' cognitive requirements in alternative theories may help to explain the prevalence of level-*k* thinking.

An equilibrium player must find his equilibrium decision via one of several methods, of which the easiest in CGC's games is iterating best responses; but in some of CGC's games this requires as many as 52 iterations.

In other games, equilibrium reasoning may be even more complex.

An LQRE player must not only respond to a complex 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, then LQRE reasoning is doubly taxing.

(From the point of view of the analyst, the complexity of LQRE means it must usually be solved for computationally and is not easily adapted to analysis.)

A level-*k* player begins with an instinctive reaction to the game, and then iterates best responses a few times, which is easy in most games.

Except for *L1*'s response to a random *L0*, which is straightforward, a level-*k* player need not respond to the noisiness of others' decisions.

These observations apply equally well to a CH player, except that he needs to respond not to a single lower type's response but a distribution of them, in proportions determined by an estimated but somehow known parameter.

Unlike equilibrium and LQRE players, level-*k* and CH players need not find fixed points.

Instead level-*k* and CH models have a simple recursive structure, which 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.

Finally, an NI model is cognitively less taxing than LQRE because it does not require fixed-point reasoning, but more taxing than a level-*k* or CH model because decisions are indefinitely iterated best responses to noisy higher-order beliefs (in applications, however, GH truncate iterations to ten rounds).

6. Illustration of Level-*k* Analyses of Matrix Games with Unique Mixed-Strategy Equilibria: M. M. Kaye's *The Far Pavilions*

I now give a simple example that illustrates applications of level-*k* models.

In M. M. Kaye's novel *The Far Pavilions*, the main male character, Ash, is trying to escape from his Pursuers along a North-South road.

Ash and his Pursuers have *strategically simultaneous* choices between North and South—although their choices are time-sequenced, the Pursuers must make their choice irrevocably before they learn Ash's choice.

If the Pursuers catch Ash, they gain 2 and he loses 2. But South is warm, and North is the Himalayas with winter coming. Thus both Ash and the Pursuers gain an extra 1 for choosing South, whether or not Ash is caught:

		Pursuers	
		South (*q*)	**North**
Ash	**South (*p*)**	**3** **-1**	**0** **1**
	North	**1** **0**	**2** **-2**

***Far Pavilions* Escape!**

Escape! has a unique equilibrium in mixed strategies, in which:

$$3p + 1(1 - p) = 0p + 2(1 - p) \text{ or } p = 1/4, \text{ and}$$

$$-1q + 1(1 - q) = 0q - 2(1 - q) \text{ or } q = 3/4.$$

This equilibrium responds to the payoff asymmetry between South and North in a decision-theoretically intuitive way for Pursuers (because $q = 3/4 >$ the 1/2 of equilibrium without the payoff asymmetry) but counterintuitively for Ash (because $p = 1/4 < 1/2$).

Although the equilibrium does not fully reflect intuition, experimental data from such games suggest that real people's decisions often do reflect it.

E.g., Camerer reports informally gathered data for a perturbed Matching Pennies game (see also Rosenthal, Shachat, and Walker 2003 *IJGT*):

	L (33%)	**R (67%)**
T (72%)	**0** **2**	**1** **0**
B (28%)	**1** **0**	**0** **1**

Perturbed Matching Pennies

The equilibrium mixed-strategy probabilities are Pr{T} = Pr{B} = 0.5 for Row and Pr{L} = 0.33 and Pr{R} = 0.67 for Column.

Although Column players are "right on" the equilibrium mixture, Row players overplay their superficially more attractive strategy T, not realizing that this allows a sophisticated Column to neutralize Row's advantage.

(Perhaps unsurprisingly: that realization may require fixed-point reasoning.)

Meanwhile, back in the novel, Ash overcomes his fear of freezing and goes North. The Pursuers—unimaginatively—go South, Ash escapes, and the novel continues…

Meanwhile, back in the novel, Ash overcomes his fear of freezing and goes North. The Pursuers—unimaginatively—go South, Ash escapes, and the novel continues…romantically…

Meanwhile, back in the novel, Ash overcomes his fear of freezing and goes North. The Pursuers—unimaginatively—go South, Ash escapes, and the novel continues…romantically…for 900 more pages.

In equilibrium the observed outcome {Ash North, Pursuers South} has probability $(1 - p)q = 9/16$: a fit much better than random.

But try a level-*k* model with a uniformly random *L0*:

Types	**Ash**	**Pursuers**
L0	uniformly random	uniformly random
L1	South	South
L2	North	South
L3	North	North
L4	South	North
L5	South	South

***Lk* types' decisions in *Far Pavilions* Escape!**

The level-*k* model precisely and correctly predicts the outcome provided that Ash is either *L2* or *L3* and the Pursuers are either *L1* or *L2*.

How do we know Ash's type? One advantage of using fiction as data is that the narrative sometimes reveals cognition as well as decisions:

Ash's mentor (Koda Dad, played by Omar Sharif in the HBO miniseries) gives Ash the following advice (p. 97 of the novel):

"…ride hard for the north, since they will be sure you will go southward where the climate is kinder…").

Koda Dad's advice reflects the belief that the Pursuers think Ash is *L1*, so that Ash will go south because it's "kinder" and that (assuming the Pursuers are uniform random *L0*) the Pursuers are no more likely to catch him there.

Thus Koda Dad must think the Pursuers are *L2*.

Hence Koda Dad advises Ash to think like an *L3*, and go North.

L3 ties my personal best *k* for a clearly explained level-*k* type in fiction. I suspect even postmodern fiction may have no *Lk*s higher than *L3*: they wouldn't be credible. I also doubt that one can find fixed-point reasoning.

Of course, most applications don't come with an omniscient author identifying characters' strategic thinking types for us.

But if the game is clearly defined and we have enough data, we can specify a level-*k* model, derive its implications, and use them to estimate the population frequency distribution of types and their precisions.

CGCB, CGC, Crawford and Iriberri 2007 *AER*, and Crawford and Iriberri 2007 *Econometrica*, discussed below, show how this is done in datasets from normal-form game experiments, in settings like Yuschenko and Lake Wobegon, and in sealed-bid auction experiments.

Alternatively, we can calibrate the model using previous estimates for similar applications.

Crawford and Iriberri 2007 *AER* illustrate how this is done in settings like Yuschenko and Lake Wobegon.

Returning to Camerer's experiment, for example, an *L1* Row plays T and an *L1* Column plays L and R with equal probabilities (for logit or alternative payoff-driven error structures). An *L2* Row plays T and an *L2* Column plays R. An *L3* Row plays B and an *L3* Column plays R.

	L (33%)	**R (67%)**
T (72%)	**0** **2**	**1** **0**
B (28%)	**1** **0**	**0** **1**

Perturbed Matching Pennies

With a plausible mixture of 50% *L1*s, 30% *L2*s, and 20% *L3*s in both player roles—it's natural to impose symmetry when roles are filled randomly from the same population—the level-*k* model's predicted choice frequencies are 80% T for Row and 25% L for Column: Not a perfect fit, but reasonable.

The outcome resembles a "purified" mixed-strategy equilibrium.

But the level-*k* model predicts choice frequencies that deviate from the equilibrium probabilities for Row Pr{T} = Pr{B} = 0.5 in the intuitive direction.

Similarly, in *Far Pavilions* Escape!, even though *Lk* types don't normally randomize, the heterogeneity of thinking reflected by the estimated distribution implies a mixture of decisions that reflects strategic uncertainty.

		Pursuers: South (*q*)	Pursuers: North
Ash: South (*p*)		3 -1	0 1
Ash: North		1 0	2 -2

***Far Pavilions* Escape!**

Suppose, for example, that each player role is filled from a 50-50 mixture of *L1*s and *L2*s and there are no errors.

Then Ash goes South with probability 0.5 > 1/4 (the equilibrium probability) and the Pursuers go South with probability 1 > 3/4 (the equilibrium probability).

Although the implied mixture of decisions again somewhat resembles a "purified" equilibrium, the model again deviates from equilibrium in the direction that intuition suggests: this time for both player roles.

7. Kahneman's Entry Magic: Asymmetric Coordination via Structure in Entry Games

I now use a simple level-*k* model to suggest an explanation of Kahneman's Entry Magic.

The analysis illustrates the importance of the structured heterogeneity of strategic thinking a level-*k* model allows.

I begin by recapitulating Kahneman's results and his reaction to them.

I then simplify Camerer, Ho, and Chong's 2004 *QJE*, Section III.C, CH analysis of *n*–person entry games to a level-*k* analysis of two-person Battle of the Sexes games, which are like two-person market-entry games with capacity one, and which makes the central points as simply as possible.

(Goldfarb and Yang 2008 *Journal of Marketing Research* give a CH analysis of field data on analogous technology adoption games.)

In market-entry experiments, *n* subjects choose simultaneously between entering (“In”) and staying out (“Out”) of a market with given capacity.

In yields a given positive profit if no more subjects enter than a given market capacity; but a given negative profit if too many enter.

For simplicity, Out yields 0 profit, no matter how many subjects enter.

Because players have no way to distinguish their symmetric roles, it is not sensible to predict systematic differences in behavior across roles.

Thus, the natural equilibrium benchmark prediction is the symmetric mixed-strategy equilibrium, in which each player enters with a given probability that makes all players indifferent between In and Out.

This mixed-strategy equilibrium yields an expected number of entrants approximately equal to market capacity, but there is a positive probability that either too many or too few will enter.

Even so, subjects in market-entry experiments regularly have better ex post coordination (number of entrants stochastically closer to market capacity) than in the symmetric equilibrium.

This led Kahneman to remark, "…to a psychologist, it looks like magic."

(But no one would be at all surprised by this unless he believed in equilibrium, so Kahneman should have said, "…to a *game theorist*….")

	In	Out
In	**0** **0**	**1** ***a***
Out	***a*** **1**	**0** **0**

Battle of the Sexes

In the simplified two-person Battle of the Sexes model studied here, with $a > 1$, the unique symmetric equilibrium is in mixed strategies, with $p \equiv \Pr\{\text{In}\} = a/(1+a)$ for both players.

The equilibrium expected coordination rate is $2p(1-p) = 2a/(1+a)^2$.

Players' equilibrium expected payoffs are $a/(1+a)$.

With $a > 1$ these expected payoffs $a/(1+a) < 1$: worse for each player than his worst pure-strategy equilibrium.

Consider a level-*k* model in which each player follows one of four types, *L1*, *L2*, *L3*, or *L4*, with each role filled by a draw from the same distribution.

Assume for simplicity that the frequency of *L0* is 0, and that *L0* chooses its action uniformly randomly, with Pr{In} = Pr{Out} = 1/2.

*L1*s mentally simulate *L0*s' random decisions and best respond, thus, with *a* > 1, choosing In; *L2*s choose Out; *L3*s choose In; and *L4*s choose Out.

	In	**Out**
In	**0** **0**	**1** ***a***
Out	***a*** **1**	**0** **0**

Battle of the Sexes

Type pairings	***L1***	***L2***	***L3***	***L4***
L1	**In, In**	**In, Out**	**In, In**	**In, Out**
L2	**Out, In**	**Out, Out**	**Out, In**	**Out, Out**
L3	**In, In**	**In, Out**	**In, In**	**In, Out**
L4	**Out, In**	**Out, Out**	**Out,In**	**Out, Out**

The predicted outcome distribution is determined by the outcomes of the possible type pairings and the type frequencies.

If both roles are filled from the same distribution, players have equal ex ante payoffs, proportional to the expected coordination rate.

L3 behaves like *L1*, and *L4* like *L2*. Lumping *L1* and *L3* together and letting v denote their total probability, and lumping *L2* and *L4* together, the expected coordination rate is $2v(1 - v)$.

This is maximized at $v = ½$, where it takes the value ½.

Thus for v near ½, which is behaviorally plausible, the coordination rate is close to ½. (For more extreme values the rate is worse, $\to 0$ as $v \to 0$ or 1.)

By contrast, the mixed-strategy equilibrium expected coordination rate, $2a/(1 + a)^2$, is maximized when $a = 1$, where it takes the value ½.

As $a \to \infty$, $2a/(1 + a)^2 \to 0$ like $1/a$. Even for moderate values of a, the level-k coordination rate is higher than the equilibrium rate.

The level-*k* model, and the closely related CH model CHC used to explain Kahneman's results, yield a completely different view of asymmetric coordination via structure than a traditional refined-equilibrium model:

- Neither equilibrium nor refinements play any role in players' thinking.

- Coordination, when it occurs, is an accidental (though statistically predictable) by-product of players' non-equilibrium decision rules.

- Even though decisions are simultaneous and there is no communication or observation of the other's decision, the predictable heterogeneity of strategic thinking allows more sophisticated players such as *L2*s to mentally simulate the decisions of less sophisticated players such as *L1*s and accommodate them, just as Stackelberg followers would.

- This mental simulation doesn't work perfectly, because an *L2* is as likely to be paired with another *L2* as an *L1*. Neither would it work if strategic thinking were homogeneous. But it's very surprising that it works at all.

8. Bank Runs: Symmetric Coordination via Structure

Reconsider Summers's Bank Runs example. The game he describes can be represented by a payoff table (not a payoff matrix!) as follows:

		Summary statistic: In	Summary statistic: Out
Representative player	In	1	-10
	Out	0	0

Bank Runs

The summary statistic is a measure of whether or not the required number of investors stays In. In Summers's first example, all investors must stay In to prevent the bank from collapsing, so the summary statistic takes the value In if and only if all (but the representative player) stay In. In his second example two-thirds of the investors need to stay In, so the summary statistic takes the value In if and only if (adding in the representative player) this is the case.

In each example there are two pure-strategy equilibria: "all-In" and "all-Out". (There is also a mixed-strategy equilibrium in which the probability that the summary statistic equals In just balances the benefits of In and Out; but this equilibrium is behaviorally implausible.)

What will happen? In this example the coordination refinement of payoff-dominance uniquely favors the all-In equilibrium, for any value of the population size *n*. This again seems behaviorally unlikely even for small *n*.

The basic idea of risk-dominance (the precise formalization is controversial, and is fully agreed on only in two-action games) is to choose the equilibrium with the largest "basin of attraction" in beliefs space.

In 2x2 symmetric two-person games, this amounts to selecting the equilibrium that results if each player best responds to a uniform random prior over the other's strategies (just as *L1* does when *L0* is uniform random).

Thus for population size 2, risk-dominance favors the all-Out equilibrium.

In 2x2 symmetric games for population $n > 2$, risk-dominance again favors the equilibrium with the larger basin of attraction in beliefs space. Assuming independence, with Summers's payoff assumptions risk-dominance favors the all-Out equilibrium for any $n > 2$, even if only two-thirds need to stay In.

A global games analysis (Carlsson and Van Damme 1993 *Ecma*, Morris and Shin 1998 *AER*) yields the same conclusion as risk-dominance here.

Now consider a level-*k* model.

In this context an *L0* in the style of “Graham’s Mr. Market” is behaviorally plausible, but that would require a complex discussion of market psychology.

To illustrate how the model works, I assume instead a uniform random *L0*.

Recall that in *n*-person games it is also possible to define a level-*k* model in which *L0* is correlated across players instead of independent.

(Risk-dominance is usually defined assuming independence, but correlation is possible there too. Correlation is irrelevant in defining payoff-dominance.)

In Summers’s first example, where the summary statistic takes the value In only when all stay In, *L1*s decision is Out with independent or correlated *L0*.

In Summers’s second example, where the summary statistic takes the value In when two-thirds or more stay In, *L1*s decision is still Out in either case.

In all cases *L2* and higher types also stay Out, so if the frequency of *L0* is 0, the outcome is observationally equivalent to the all-Out equilibrium.

Now consider an example like Bank Runs in which the summary statistic takes the value In when *one-third* or more of the investors stay In.

If, say, $n = 6$, then given a choice of In by the representative player himself, the summary statistic will be In unless all five other players stay Out.

If *L0* is independent, *L1* assigns all others staying Out probability $1/2^5 \approx 0.03$.

If *L0* is correlated, *L1* assigns all others staying Out probability ½.

In the former case, *L1* and therefore all higher *Lk* types stay In, and the outcome is observationally equivalent to the all-In equilibrium.

In the latter case, *L1* and therefore all higher *Lk* types stay Out, and the outcome is observationally equivalent to the all-Out equilibrium.

In each of these symmetric coordination games, the level-*k* model derives the outcome from strategic responses to instinctive reactions to the game.

Unlike traditional coordination refinements, the level-*k* approach is easy to combine with richer models of market psychology, via an *L0* in the style of "Graham's Mr. Market."

And because such an *L0* is a psychological rather than a strategic concept, it is easier to extrapolate its specification across games, as illustrated below.

Again neither equilibrium nor refinements play any role in players' thinking.

And coordination, when it occurs, is again an accidental by-product of players' non-equilibrium, level-*k* decision rules.

Because in these symmetric coordination games *L1* responses to a uniform random *L0* are in equilibrium, there is no “magic”:

The level-*k* model reduces to an equilibrium selection device, which coincides here with risk-dominance, but need not do so in general.

In 2×2 symmetric coordination games *L1* responses to a uniform random *L0* also coincide with the equilibrium selected by a global games analysis.

Selecting an equilibrium via *L1* responses seems empirically more promising, because *L1* responses are less cognitively taxing and are directly suggested by experimental evidence.

By contrast, a global games analysis relies on indefinitely iterated dominance in a game with payoff uncertainty artificially grafted onto its structure in a particular way; and the empirical support even for finitely iterated dominance is weak.

9. Structural Alternatives to “Incomplete” Models

How might the availability of structural non-equilibrium models that reliably describe initial responses to games change the way we think about data?

Although some empirical applications concern games that are dominance-solvable in small numbers of rounds, many involve games that are not, and many others involve games with multiple equilibria.

In games that are not sufficiently dominance-solvable, finitely iterated dominance and *k*-rationalizability are “incomplete” (my term, not standard) in that they do not specify a unique (though possibly probabilistic) prediction conditional on the value of the behavioral parameters.

In games with multiple equilibria, equilibrium plus noise but without refinements is incomplete in the same general sense.

In the empirical literature, such incompleteness has been dealt with in one of two ways:

- By accepting a theory's set-valued restrictions as the only implications of the model and testing them (e.g. for *k*-rationalizability—which they call "level-*k* rationality"—in Aradillas-Lopez and Tamer 2008 *JBES*; or for unrefined equilibria in Echenique and Komunjer 2009 *Econometrica*).

- By estimating an unrestricted probability distribution over the set of equilibria (Bresnahan and Reiss 1991 *Journal of Econometrics*).

However, just as experimental results suggest that equilibrium is too strong to be descriptive of people's responses to novel or complex games, it also suggests that *k*-rationalizability and even rationalizability are too weak.

Rationalizability sometimes agnostically allows beliefs that are behaviorally outlandish, even though consistent with common knowledge of rationality (recall Section 3's unique equilibrium without dominance example).

Because CGC's experimental results suggest that to the extent that people respect *k*-rationalizability, they do so not because they perform finitely iterated dominance leading to a set of *k*-rationalizable decisions, but because they follow a level-*k* decision rule that selects a specific such decision, it seems behaviorally natural to replace *k*-rationalizability (and equilibrium) by a structural level-*k* model that has the advantage of being complete.

In settings where this can be done without risking serious misspecification, it seems likely to yield significantly more useful econometric models.

Aradillas-Lopez and Tamer 2008 *JBES* provide some indirect evidence on the potential benefits of structural non-equilibrium models by comparing the identification powers of equilibrium and *k*-rationalizability in two-person entry games without or with privately observed payoff perturbations; and in first-price auctions with incomplete information and independent private values.

In entry games attention centers on identification and estimation of payoff parameters, which are normally unobservable in the field.

In auctions attention centers on identification and estimation of bidders' value distributions, which are again normally unobservable in the field.

The standard approach assumes equilibrium and shows that the parameters of interest are identified (parametrically or nonparametrically).

Aradillas-Lopez and Tamer show that weakening equilibrium to *k*-rationalizability implies weaker identifying restrictions—sometimes much weaker, for low values of *k*—and that individuals' *k*'s are not fully identified.

In entry games, 1-rationalizability only slightly restricts the payoff parameters:

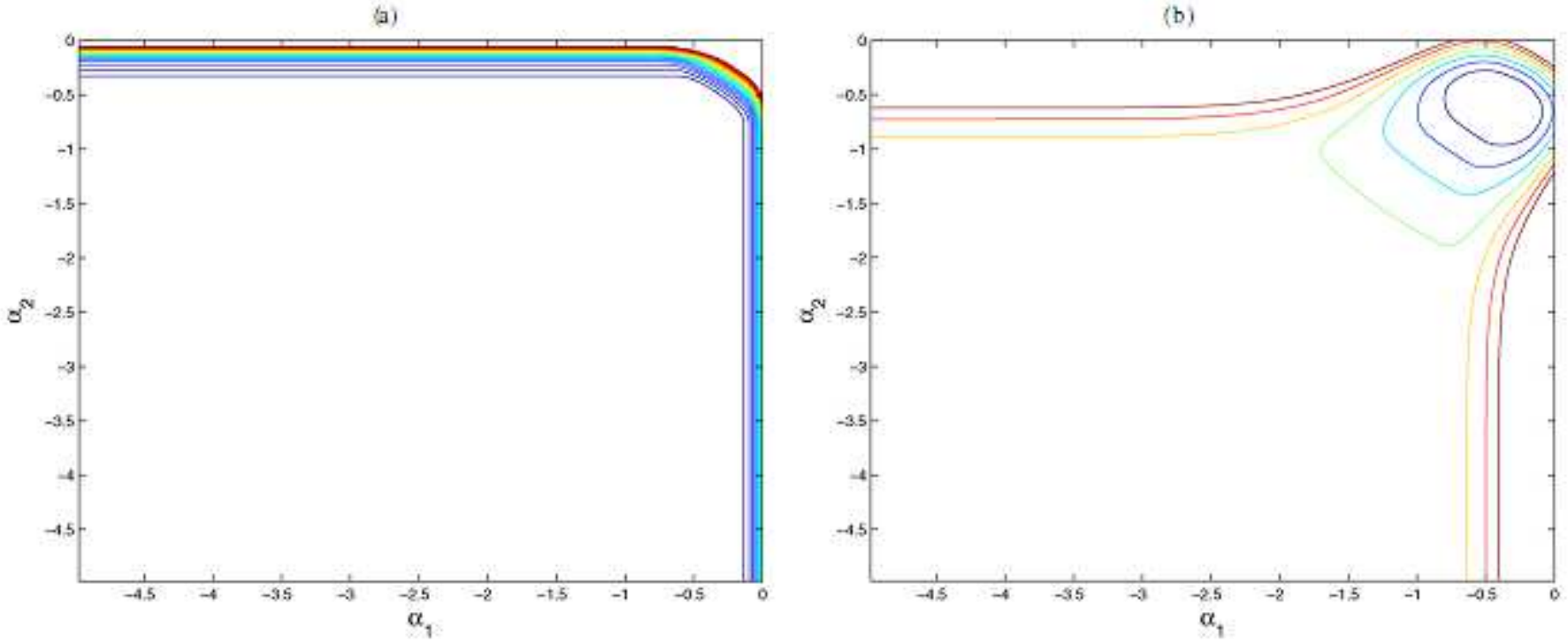


Figure 3. Identification set under Nash and 1-level rationality. Shown the identified regions for (α_1, α_2) under $k = 1$ rationality (a) and Nash (b). We set in the underlying model $(\alpha_1, \alpha_2) = (-.5, -.5)$. The model was simulated assuming Nash with (0, 1) selected with probability one in regions of multiplicity. Note that in (a), the model only places upper bounds on the alphas, whereas in (b) (α_1, α_2) are constrained to lie a much smaller set (the inner "circle").

Aradillas-Lopez and Tamer's Figure 3

In first-price auctions Aradillas-Lopez and Tamer note (following Battigalli and Siniscalchi 2003 *GEB*) that *k*-rationalizability implies only a weak upper bound on bids, which shrinks with *k* but for any *k* allows bids both above and below equilibrium; with correspondingly weak bounds on value distributions.

Benjamin Gillen, "Identification of Level-*k* Auctions," UCSD 2009 provides additional evidence on the benefits of structural non-equilibrium models.

He shows that in a level-*k* model (based on Crawford and Iriberri's 2007 *Econometrica* model, discussed below), under a reasonable (but not completely unrestrictive) assumption on the separation of different types' (*k*s') bidding functions, both the value distributions and individual bidders' *k*s are identified, parametrically or nonparametrically.

The difference arises because Gillen's level-*k* model "completes" Aradillas-Lopez and Tamer's *k*-rationalizability model, which with enough data makes it theoretically possible to estimate the level-*k* model's additional structural parameters along with bidders' value distributions.

CGC's footnote 42, p. 1766, makes a similar point in a different way.

CGC note that in their maximum likelihood estimation of a model of subjects' guesses and searches for hidden payoff information, the guess part of the log-likelihood is nearly six times larger than the search part.

This occurs because their theory of subjects' decisions makes very precise predictions of a subject's decisions, conditional on his type.

By contrast, CGC's theory of cognition and search imposes (via filters described in the paper) only weak, set-valued restrictions on a subject's searches, conditional on his type:

Although CGC's theory of decisions is complete, their theory of search is incomplete.

As a result, the search restrictions are much more likely than the decision restrictions to be satisfied by chance, which causes the disparity in likelihood weights.

Turning to games with multiple equilibria, the freedom that assuming rationalizability or estimating an unrestricted probability distribution over the set of equilibria can yield severe overfitting and/or very weak tests.

Costa-Gomes, Crawford, and Iriberri ("CGCI") 2009 *JEEA* address this issue for Van Huyck, Battalio, and Beil's ("VHBB") 1990 *AER*, 1991 *QJE* coordination games, in which any of the seven pure strategies is both rationalizable and consistent with one of the seven pure-strategy equilibria.

Using VHBB's data to estimate an unrestricted probability distribution over equilibria yields good fits, but it also yields estimates that vary incoherently across games and don't inspire confidence for beyond-sample prediction.

CGCI "complete" equilibrium plus noise by adding coordination refinements risk- or payoff-dominance (in turn), to put it on a more equal footing with LQRE, level-*k*, CH, and NI models, which are already complete.

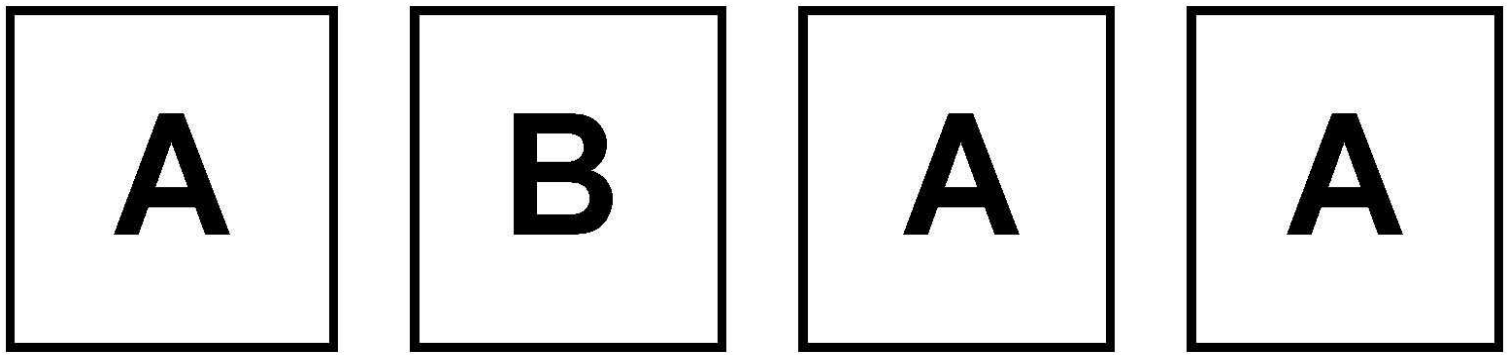
10. Yuschenko and Lake Wobegon: Framing Effects in Zero-Sum Two-Person Games

Consider Rubinstein, Tversky, and Heller's 1993, 1996, 1998-99 ("RTH") experiments with zero-sum, two-person "hide-and-seek" games with non-neutral framing of locations, analyzed by Crawford and Iriberri 2007 *AER*.

(See also Östling, Wang, Chou, and Camerer's 2008 CH analysis of field and lab data on lowest unique positive integer ("LUPI") games.)

A typical seeker's instructions (a hider's instructions are analogous):

Your opponent has hidden a prize in one of four boxes arranged in a row. The boxes are marked as shown below: A, B, A, A. Your goal is, of course, to find the prize. His goal is that you will not find it. You are allowed to open only one box. Which box are you going to open?



RTH's framing of the hide-and-seek game is non-neutral in two ways:

- The "*B*" location is distinguished by its label.
- The two "*end A*" locations may be inherently focal.

This gives the "*central A*" location its own brand of uniqueness as the "least salient" location.

Mathematically this "negative" uniqueness is analogous to the "positive" uniqueness of "*B*".

However, Crawford and Iriberri's 2007 *AER* analysis shows that its psychological effects are completely different.

RTH's design is important as a tractable abstract model of a non-neutral cultural or geographic frame, or "landscape."

Hide-and-seek games are often played on such landscapes, even though traditional game theory rules out any influence of the landscape by fiat.

This is well illustrated by the Yuschenko and Lake Wobegon quotations:

"Any government wanting to kill an opponent…would not try it at a meeting with government officials."

"…in Lake Wobegon, the correct answer is usually 'c'."

Yuschenko's meeting with government officials is analogous to RTH's B, although in that example there's nothing like RTH's end locations.

With four possible choices arrayed left to right in the zero-sum game between a test designer deciding where to hide the correct answer and a clueless test-taker trying to guess where it is, the Lake Wobegon example is very close to RTH's design.

RTH's hide-and-seek game has a clear equilibrium prediction, which leaves no room for framing to systematically influence the outcome.

The traditional theory of zero-sum two-person games is the strongpoint of noncooperative game theory, where the arguments for playing equilibrium strategies are immune to most of the usual counterarguments.

Yet framing has a strong and systematic effect in RTH's experiments, qualitatively the same around the world, with *Central A* (or its analogs in other treatments, as explained in the paper) most prevalent for hiders (37% in the aggregate) and even more prevalent for seekers (46%).

In this game any strategy, pure or mixed, is a best response to equilibrium beliefs. Thus one might argue that deviations do not violate the theory.

However, systematic deviations of aggregate choice frequencies from equilibrium probabilities must (with very high probability) have a cause that is partly common across players. They are therefore symptomatic of systematic deviations from equilibrium.

TABLE 1—AGGREGATE CHOICE FREQUENCIES IN RTH'S TREATMENTS

RTH-4	A	*B*	A	A
Hider (53; $p = 0.0026$)	9 percent	36 percent	40 percent	15 percent
Seeker (62; $p = 0.0003$)	13 percent	31 percent	45 percent	11 percent
RT-AABA–Treasure	A	A	*B*	A
Hider (189; $p = 0.0096$)	22 percent	35 percent	19 percent	25 percent
Seeker (85; $p = 9E\text{-}07$)	13 percent	51 percent	21 percent	15 percent
RT-AABA–Mine	A	A	*B*	A
Hider (132; $p = 0.0012$)	24 percent	39 percent	18 percent	18 percent
Seeker (73; $p = 0.0523$)	29 percent	36 percent	14 percent	22 percent
RT-1234–Treasure	*1*	*2*	3	4
Hider (187; $p = 0.0036$)	25 percent	22 percent	36 percent	18 percent
Seeker (84; $p = 3E-05$)	20 percent	18 percent	48 percent	14 percent
RT-1234–Mine	*1*	*2*	3	4
Hider (133; $p = 6E\text{-}06$)	18 percent	20 percent	44 percent	17 percent
Seeker (72; $p = 0.149$)	19 percent	25 percent	36 percent	19 percent
R-ABAA	A	*B*	A	A
Hider (50; $p = 0.0186$)	16 percent	18 percent	44 percent	22 percent
Seeker (64; $p = 9E\text{-}07$)	16 percent	19 percent	54 percent	11 percent

Notes: Sample sizes and p-values for significant differences from equilibrium in parentheses; salient labels in italics; order of presentation of locations to subjects as shown.

Crawford and Iriberri's Table 1

RTH's results raise several puzzles:

- Hiders' and seekers' responses are unlikely to be completely non-strategic in such simple games. So if they aren't following equilibrium logic, what are they doing?

- On average hiders are as smart as seekers, so hiders tempted to hide in *central A* should realize that seekers will be just as tempted to look there. Why do hiders allow seekers to find them 32% of the time when they could hold it down to 25% via the equilibrium mixed strategy?

- Further, why do seekers choose *central A* (or its analogs) even more often (46% in Table 3 below) than hiders (37%)?

Note that although the payoff structure of RTH's game is asymmetric, QRE ignores labeling and (logit or not) coincides with equilibrium in the game, and so does not help to explain the asymmetry of choice distributions.

The role asymmetry in subjects' behavior and how it is linked to the game's payoff asymmetry points strongly in the direction of a level-*k* or CH model, and is a mystery from the viewpoint of other theories I am aware of.

In constructing such a model, defining *L0* as uniform random would be unnatural, given the non-neutral framing of decisions and that *L0* describes others' instinctive responses.

(It would also make *Lk* the same as *Equilibrium* for $k > 0$.)

But a level-*k* model with a role-independent *L0* that probabilistically favors salient locations yields a simple explanation of RTH's results.

Assume that *L0* hiders and seekers both choose A, B, A, A with probabilities $p/2$, q, $1 - p - q$, $p/2$ respectively, with $p > ½$ and $q > ¼$.

L0 favors both the end locations and the B location, equally for hiders and seekers, but the model lets the data decide which is more salient.

For behaviorally plausible type distributions (estimated 0% *L0*, 19% *L1*, 32% *L2*, 24% *L3*, 25% *L4*—almost hump-shaped), a level-*k* model gracefully explains the major patterns in RTH's data, the prevalence of *central A* for hiders and its even greater prevalence for seekers:

- Given *L0*'s attraction to salient locations, *L1* hiders choose *central A* to avoid *L0* seekers and *L1* seekers avoid *central A* searching for *L0* hiders (the data suggest that end locations are more salient than B).

- For similar reasons, *L2* hiders choose *central A* with probability between 0 and 1 (breaking payoff ties randomly) and *L2* seekers choose it with probability 1.

- *L3* hiders avoid *central A* and *L3* seekers choose it with probability between zero and one (breaking payoff ties randomly).

- *L4* hiders and seekers both avoid *central A*.

TABLE 2—TYPES' EXPECTED PAYOFFS AND CHOICE PROBABILITIES IN RTH'S GAMES WHEN $p > 1/2$ AND $q > 1/4$

Hider	Expected payoff $p < 2q$	Choice probability $p < 2q$	Expected payoff $p > 2q$	Choice probability $p > 2q$
L0 (Pr. r)				
A	–	$p/2$	–	$p/2$
B	–	q	–	q
A	–	$1-p-q$	–	$1-p-q$
A	–	$p/2$	–	$p/2$
L1 (Pr. s)				
A	$1-p/2 < 3/4$	0	$1-p/2 < 3/4$	0
B	$1-q < 3/4$	0	$1-q < 3/4$	0
A	$p+q > 3/4$	1	$p+q > 3/4$	1
A	$1-p/2 < 3/4$	0	$1-p/2 < 3/4$	0
L2 (Pr. t)				
A	1	1/3	1/2	0
B	0	0	1	1/2
A	1	1/3	1	1/2
A	1	1/3	1/2	0
L3 (Pr. u)				
A	1	1/3	1	1/3
B	1	1/3	1	1/3
A	0	0	0	0
A	1	1/3	1	1/3
L4 (Pr. v)				
A	2/3	0	1	1/2
B	1	1	1/2	0
A	2/3	0	1/2	0
A	2/3	0	1	1/2
Total	$p < 2q$		$p > 2q$	
A	$rp/2+(1-\varepsilon)[t/3+u/3]+(1-r)\varepsilon/4$		$rp/2+(1-\varepsilon)[u/3+v/2]+(1-r)\varepsilon/4$	
B	$rq+(1-\varepsilon)[u/3+v]+(1-r)\varepsilon/4$		$rq+(1-\varepsilon)[t/2+u/3]+(1-r)\varepsilon/4$	
A	$r(1-p-q)+(1-\varepsilon)[s+t/3]+(1-r)\varepsilon/4$		$r(1-p-q)+(1-\varepsilon)[s+t/2]+(1-r)\varepsilon/4$	
A	$rp/2+(1-\varepsilon)[t/3+u/3]+(1-r)\varepsilon/4$		$rp/2+(1-\varepsilon)[u/3+v/2]+(1-r)\varepsilon/4$	

Seeker	Expected payoff $p < 2q$	Choice probability $p < 2q$	Expected payoff $p > 2q$	Choice probability $p > 2q$
L0 (Pr. r)				
A	–	$p/2$	–	$p/2$
B	–	q	–	q
A	–	$1-p-q$	–	$1-p-q$
A	–	$p/2$	–	$p/2$
L1 (Pr. s)				
A	$p/2 > 1/4$	0	$p/2 > 1/4$	1/2
B	$q > 1/4$	1	$q > 1/4$	0
A	$1-p-q < 1/4$	0	$1-p-q < 1/4$	0
A	$p/2 > 1/4$	0	$p/2 > 1/4$	1/2
L2 (Pr. t)				
A	0	0	0	0
B	0	0	0	0
A	1	1	1	1
A	0	0	0	0
L3 (Pr. u)				
A	1/3	1/3	0	0
B	0	0	1/2	1/2
A	1/3	1/3	1/2	1/2
A	1/3	1/3	0	0
L4 (Pr. v)				
A	1/3	1/3	1/3	1/3
B	1/3	1/3	1/3	1/3
A	0	0	0	0
A	1/3	1/3	1/3	1/3
Total	$p < 2q$		$p > 2q$	
A	$rp/2+(1-\varepsilon)[u/3+v/3]+(1-r)\varepsilon/4$		$rp/2+(1-\varepsilon)[s/2+v/3]+(1-r)\varepsilon/4$	
B	$rq+(1-\varepsilon)[s+v/3]+(1-r)\varepsilon/4$		$rq+(1-\varepsilon)[u/2+v/3]+(1-r)\varepsilon/4$	
A	$r(1-p-q)+(1-\varepsilon)[t+u/3]+(1-r)\varepsilon/4$		$r(1-p-q)+(1-\varepsilon)[t+u/2]+(1-r)\varepsilon/4$	
A	$rp/2+(1-\varepsilon)[u/3+v/3]+(1-r)\varepsilon/4$		$rp/2+(1-\varepsilon)[s/2+v/3]+(1-r)\varepsilon/4$	

TABLE 3—PARAMETER ESTIMATES AND LIKELIHOODS FOR THE LEADING MODELS IN RTH'S GAMES

Model	Ln L	Parameter estimates	Observed or predicted choice frequencies					MSE
			Player	A	B	A	A	
Observed frequencies (624 hiders, 560 seekers)			H	0.2163	0.2115	0.3654	0.2067	–
			S	0.1821	0.2054	0.4589	0.1536	
Equilibrium without perturbations	−1641.4		H	0.2500	0.2500	0.2500	0.2500	0.00970
			S	0.2500	0.2500	0.2500	0.2500	
Equilibrium with restricted perturbations	−1568.5	$e_H \equiv e_S = 0.2187$	H	0.1897	0.2085	0.4122	0.1897	0.00084
		$f_H \equiv f_S = 0.2010$	S	0.1897	0.2085	0.4122	0.1897	
Equilibrium with unrestricted perturbations	−1562.4	$e_H = 0.2910, f_H = 0.2535$	H	0.2115	0.2115	0.3654	0.2115	0.00006
		$e_S = 0.1539, f_S = 0.1539$	S	0.1679	0.2054	0.4590	0.1679	
Level-k with a role-symmetric $L0$ that favors salience	−1564.4	$p > 1/2$ and $q > 1/4, p > 2q$, $r = 0, s = 0.1896, t = 0.3185$, $u = 0.2446, v = 0.2473, \varepsilon = 0$	H	0.2052	0.2408	0.3488	0.2052	0.00027
			S	0.1772	0.2047	0.4408	0.1772	
Level-k with a role-asymmetric $L0$ that favors salience for seekers and avoids it for hiders	−1563.8	$p_H < 1/2$ and $q_H < 1/4$, $p_S > 1/2$ and $q_S > 1/4$, $r = 0, s = 0.66, t = 0.34$, $\varepsilon = 0.72$; $u \equiv v \equiv 0$ imposed	H	0.2117	0.2117	0.3648	0.2117	0.00017
			S	0.1800	0.1800	0.4600	0.1800	
Level-k with a role-symmetric $L0$ that avoids salience	−1562.5	$p < 1/2$ and $q < 1/4, p < 2q$, $r = 0, s = 0.3636, t = 0.0944$, $u = 0.3594, v = 0.1826, \varepsilon = 0$	H	0.2133	0.2112	0.3623	0.2133	0.00006
			S	0.1670	0.2111	0.4549	0.1670	

Crawford and Iriberri's Table 3

Note that only a heterogeneous population with substantial frequencies of *L2* and *L3* as well as *L1* (estimated 0% *L0*, 19% *L1*, 32% *L2*, 24% *L3*, 25% *L4*) can reproduce the aggregate patterns in the data.

(Even though there is a nonnegligible estimated frequency of *L4*s, they don't really matter here because they never choose *central A* (Table 2 above), hence they are not implicated in the major aggregate patterns.

For the same reason, their frequency is not well identified in the estimation.)

For example, Crawford and Iriberri estimate (Table 3 above, row 5) that the salience of an end location is greater than the salience of the *B* ($p > 2q$).

Given this, a 50-50 mix of *L1*s and *L2*s in both player roles would imply (Table 2 above, right-most columns in each panel) 75% of hiders but only 50% of seekers choosing *central A*, in contrast to the 37% of hiders and 46% of seekers who did choose *central A*.

In Crawford and Iriberri's analysis of RTH's data, the role asymmetry in aggregate behavior follows naturally from the asymmetry of the game's payoff structure, via hiders' and seekers' asymmetric responses to *L0*'s *role-symmetric* choices.

Allowing *L0* to vary across roles as in Bacharach and Stahl 2000 *GEB*, although it yields a small improvement in fit (Table 3), would beg the question of why subjects' responses were so role-asymmetric.

Crawford and Iriberri's analysis, discussed below, also suggests that allowing *L0* to vary across roles leads to overfitting.

RTH took the main patterns in their data as evidence that their subjects did not think strategically:

● “The finding that both choosers and guessers selected the least salient alternative suggests little or no strategic thinking.”

● “In the competitive games, however, the players employed a naïve strategy (avoiding the endpoints), that is not guided by valid strategic reasoning. In particular, the hiders in this experiment either did not expect that the seekers too, will tend to avoid the endpoints, or else did not appreciate the strategic consequences of this expectation.”

RTH could have said the same thing about the Yuschenko quotation:

- "Any government wanting to kill an opponent…would not try it at a meeting with government officials",

to which a game theorist would (almost involuntarily) respond:

- "If that's what people think, a meeting with government officials is exactly where *I* would try to poison Yushchenko."

But strategic thinking need not be equilibrium thinking.

Crawford and Iriberri's analysis suggests that RTH's subjects were actually quite strategic and in fact more than usually sophisticated (with many *L3*s and even some *L4*s, even though in most settings *L1*s and *L2*s are more common)—they just didn't follow equilibrium logic.

Crawford and Iriberri's analysis suggests that the Yushchenko quotation simply reflects the reasoning of an *L1* poisoner, or equivalently of an *L2* investigator reasoning about an *L1* poisoner.

E**valuating the Model's Explanation: Overfitting and Portability**

Although prior intuitions about the likely hump shape and location of the type distribution impose some discipline in specifying a level-*k* model, the freedom to specify *L0* leaves room for doubts about overfitting and portability, the extent to which a model estimated from responses to one game can be extended to predict or explain responses to different games.

To see if the proposed level-*k* explanation of RTH's results is more than an after the fact "just-so" story, Crawford and Iriberri compared it on the overfitting and portability dimensions with the leading alternatives:

- Equilibrium with intuitive payoff perturbations (salience lowers hiders' payoffs, other things equal; while salience raises seekers' payoffs).
- LQRE with similarly intuitive payoff perturbations.
- Alternative level-*k* specifications (for example, with role-asymmetric *L0* or an *L0* that avoids salience, as in Table 3).

Crawford and Iriberri tested for overfitting by re-estimating each model separately for each of RTH's six treatments and using the re-estimated models to "predict" the choice frequencies of the other treatments.

Their favored level-*k* model, with a role-symmetric *L0* that favors salience, has a modest prediction advantage over equilibrium and LQRE with perturbations models, with mean squared prediction error 18% lower and better predictions in 20 of 30 comparisons.

LQRE with payoff perturbations (in different cases) either gets the patterns in the data qualitatively wrong or estimates an infinite precision and thereby turns itself back into an equilibrium model (Crawford and Iriberri's online Appendix).

A more challenging test regards portability.

Crawford and Iriberri tested for portability by using the leading alternative models, estimated from RTH's data, to "predict" subjects' initial responses in the two closest relatives of RTH's games in the literature:

- O'Neill's 1987 *PNAS* famous card-matching game, and
- Rapoport and Boebel's 1992 *GEB* closely related game.

These games both raise the same kinds of strategic issues as RTH's games, but with more complex patterns of wins and losses, different framing, and in the latter case five locations.

I focus here on Crawford and Iriberri's analysis of O'Neill's game.

In O’Neill’s card-matching game, players simultaneously and independently choose one of four cards: A, 2, 3, J.

One player, say the row player—but the game was presented to subjects as a story, not a matrix—wins if there is a match on J or a mismatch on A, 2, or 3; the other player wins in the other cases.

	A	**2**	**3**	**J**
A	1 0	0 1	0 1	1 0
2	0 1	1 0	0 1	1 0
3	0 1	0 1	1 0	1 0
J	1 0	1 0	1 0	0 1

O’Neill’s card-matching game

O’Neill’s game is like a hide-and-seek game, except that each player is a hider (h) for some locations and a seeker (s) for others.

A, 2, and 3 are strategically symmetric, and equilibrium (without payoff perturbations) has Pr{A} = Pr{2} = Pr{3} = 0.2, Pr{J} = 0.4.

	A (s)	**2 (s)**	**3 (s)**	**J (h)**
A (h)	1 0	0 1	0 1	1 0
2 (h)	0 1	1 0	0 1	1 0
3 (h)	0 1	0 1	1 0	1 0
J (s)	1 0	1 0	1 0	0 1

O’Neill’s card-matching game

The portability test directly addresses the issue of whether level-*k* models allow the modeler too much flexibility.

With regard to the flexibility of *L0*, first consider how to adapt our "psychological" specification of *L0* from RTH's to O'Neill's game.

Even Obama and McCain could agree on the right kind of *L0*:

- A and J, "face" cards and end locations, are more salient than 2 and 3, but the specification should allow either A or J to be more salient.

That the RTH estimates suggested that their end locations are more salient than the *B* label does *not* dictate whether A or J is more salient, though it does reinforce that they are both more salient than 2 and 3.

This is a psychological issue, but because it is "only" a psychological issue, it is easy to gather evidence on it from different settings, and such evidence is more likely to yield convergence than if it were partly a strategic issue.

Further, because all that matters about *L0* is what it makes *L1*s do in each role, the remaining freedom to choose *L0* allows only two models.

With regard to the flexibility of the type frequencies, empirically plausible frequencies often imply severe limits on what decision patterns a level-*k* model can generate.

Readers of the first version of Crawford and Iriberri 2007 *AER* often asked if the model could explain behavior in games other than RTH's.

O'Neill's game was the most natural choice in the experimental literature.

We did not have his data, but discussions of it (e.g. McKelvey and Palfrey 1995 *GEB*) had been dominated by an "Ace effect": aggregated over all 105 rounds, row and column players played A with frequencies 22.0% and 22.6%, significantly above the equilibrium 20%.

(O'Neill speculated that this was because "…players were attracted by the powerful connotations of an Ace".

But—we thought—what about the equally powerful connotations of the Joker and its unique payoff role? They seem to make it even more salient than Ace, but in the aggregate data row subjects chose Joker with frequencies of only 36%, and columns with frequencies of only 43%.)

We also knew that with an Obama-McCain specification of *L0* and the resulting types' decisions in O'Neill's game (Tables A3 and A4 from the paper's web appendix, on the next two slides), no behaviorally plausible level-*k* model could make a row player ("Player 1") play A more than the equilibrium 20%:

Tables A3 and A4 show that, excluding *L0*s (which normally have 0 estimated frequencies) and restricting attention to Player 1, when A is more salient ($3j - a < 1$) only *L4* chooses A, and that with probability at most 1/3 (Table A3); and that when A is less salient ($3j - a > 1$) only *L3* chooses A, and that with probability at most 1/3 (Table A4).

This is *logically* possible, but in the first case it would require a population of 60% or more *L4*s, and in the second case it would require 60% or more *L3*s: in each case behaviorally extremely unlikely on the available evidence.

Thus, despite the flexibility of the estimated type distribution, the level-*k* model's structure and the principles that guide the specification of *L0* imply a strong restriction: that row players play A less than the equilibrium 20%.

Table A3. Types' Expected Payoffs and Choice Probabilities in O'Neill's Game when $3j - a < 1$

Player 1	Exp. Payoff $A+2j<1$	Choice Pr. $a+2j<1$	Exp. Payoff $a+2j>1$	Choice Pr. $a+2j>1$	Player 2	Exp. Payoff $a+2j<1$	Choice Pr. $a+2j<1$	Exp. Payoff $a+2j>1$	Choice Pr. $a+2j>1$
***L0* (Pr. *R*)**					***L0* (Pr. *r*)**				
A	-	a	-	A	**A**	-	a	-	a
2	-	$(1-a-j)/2$	-	$(1-a-j)/2$	**2**	-	$(1-a-j)/2$	-	$(1-a-j)/2$
3	-	$(1-a-j)/2$	-	$(1-a-j)/2$	**3**	-	$(1-a-j)/2$	-	$(1-a-j)/2$
J	-	j	-	J	**J**	-	j	-	j
***L1* (Pr. *s*)**					***L1* (Pr. *s*)**				
A	$1-a-j$	0	$1-a-j$	0	**A**	$a+j$	0	$a+j$	1
2	$(1+a-j)/2$	1/2	$(1+a-j)/2$	1/2	**2**	$(1-a+j)/2$	0	$(1-a+j)/2$	0
3	$(1+a-j)/2$	1/2	$(1+a-j)/2$	1/2	**3**	$(1-a+j)/2$	0	$(1-a+j)/2$	0
J	J	0	J	0	**J**	$1-j$	1	$1-j$	0
***L2* (Pr. *t*)**					***L2* (Pr. *t*)**				
A	0	0	0	0	**A**	0	0	0	0
2	0	0	1	1/2	**2**	½	0	1/2	0
3	0	0	1	1/2	**3**	½	0	1/2	0
J	1	1	0	0	**J**	1	1	1	1
***L3* (Pr. *u*)**					***L3* (Pr. *u*)**				
A	0	0	0	0	**A**	1	1/3	0	0
2	0	0	0	0	**2**	1	1/3	1/2	0
3	0	0	0	0	**3**	1	1/3	1/2	0
J	1	1	1	1	**J**	0	0	1	1
***L4* (Pr. *v*)**					***L4* (Pr. *v*)**				
A	2/3	1/3	0	0	**A**	1	1/3	1	1/3
2	2/3	1/3	0	0	**2**	1	1/3	1	1/3
3	2/3	1/3	0	0	**3**	1	1/3	1	1/3
J	0	0	1	1	**J**	0	0	0	0
Total	$a+2j<1$		$a+2j>1$		**Total**	$a+2j<1$		$a+2j>1$	
A	$ra+(1-\varepsilon)[w/3]+(1-r)\,\varepsilon/4$		$ra+(1-r)\,\varepsilon/4$		**A**	$ra+(1-\varepsilon)\,[u/3+v/3]+(1-r)\,\varepsilon/4$		$ra+(1-\varepsilon)\,[s+v/3]+(1-r)\,\varepsilon/4$	
2	$r(1-a-j)/2+(1-\varepsilon)\,[s/2+w/3]+(1-r)\,\varepsilon/4$		$r(1-a-j)/2+(1-\varepsilon)\,[s/2+t/2]+(1-r)\,\varepsilon/4$		**2**	$r(1-a-j)/2+(1-\varepsilon)\,[u/3+v/3]+(1-r)\,\varepsilon/4$		$r(1-a-j)/2+(1-\varepsilon)\,[v/3]+(1-r)\,\varepsilon/4$	
3	$r(1-a-j)/2+(1-\varepsilon)\,[s/3+v/3]+(1-r)\,\varepsilon/4$		$r(1-a-j)/2+(1-\varepsilon)\,[s/2+t/2]+(1-r)\,\varepsilon/4$		**3**	$r(1-a-j)/2+(1-\varepsilon)\,[u/3+v/3]+(1-r)\,\varepsilon/4$		$r(1-a-j)/2+(1-\varepsilon)\,[v/3]+(1-r)\,\varepsilon/4$	
J	$Rj+(1-\varepsilon)\,[t+u]+(1-r)\,\varepsilon/4$		$rj+(1-\varepsilon)\,[u+v]+(1-r)\,\varepsilon/4$		**J**	$rj+(1-\varepsilon)\,[s+t]+(1-r)\,\varepsilon/4$		$rj+(1-\varepsilon)\,[t+u]+(1-r)\,\varepsilon/4$	

Table A4. Types' Expected Payoffs and Choice Probabilities in O'Neill's Game when $3j - a > 1$

Player 1	Exp. Payoff	Choice Pr.	Player 2	Exp. Payoff	Choice Pr.
L0 (Pr. *R*)			*L0* (Pr. *r*)		
A	-	a	A	-	a
2	-	$(1-a-j)/2$	2	-	$(1-a-j)/2$
3	-	$(1-a-j)/2$	3	-	$(1-a-j)/2$
J	-	j	J	-	j
L1 (Pr. *S*)			*L1* (Pr. *s*)		
A	$1-a-j$	0	A	$a+j$	1
2	$(1+a-j)/2$	0	2	$(1-a+j)/2$	0
3	$(1+a-j)/2$	0	3	$(1-a+j)/2$	0
J	j	1	J	$1-j$	0
L2 (Pr. *T*)			*L2* (Pr. *t*)		
A	0	0	A	1	1/3
2	1	1/2	2	1	1/3
3	1	1/2	3	1	1/3
J	0	0	J	0	0
L3 (Pr. *U*)			*L3* (Pr. *u*)		
A	2/3	1/3	A	0	0
2	2/3	1/3	2	1/2	0
3	2/3	1/3	3	1/2	0
J	0	0	J	1	1
L4 (Pr. *V*)			*L4* (Pr. *v*)		
A	0	0	A	1/3	0
2	0	0	2	1/3	0
3	0	0	3	1/3	0
J	1	1	J	1	1
Total			Total		
A	$Ra+(1-\varepsilon)[u/3]+(1-r)\,\varepsilon/4$		A	$ra+(1-\varepsilon)[s+t/3]+(1-r)\,\varepsilon/4$	
2	$r(1-a-j)/2+(1-\varepsilon)[t/2+u/3]+(1-r)\,\varepsilon/4$		2	$r(1-a-j)/2+(1-\varepsilon)[t/3]+(1-r)\,\varepsilon/4$	
3	$R(1-a-j)/2+(1-\varepsilon)[t/2+u/3]+(1-r)\,\varepsilon/4$		3	$r(1-a-j)/2+(1-\varepsilon)[t/3]+(1-r)\,\varepsilon/4$	
J	$Rj+(1-\varepsilon)[s+v]+(1-r)\,\varepsilon/4$		J	$rj+(1-\varepsilon)[u+v]+(1-r)\,\varepsilon/4$	

We decided to get O'Neill's data and test the model on it anyway, speculating, based on the level-*k* model's success in RTH's and other games, that his subjects' *initial* responses must not have had an Ace effect.

There was in fact no Ace effect for initial responses.

Instead there was a Joker effect, a full order of magnitude stronger (but to our knowledge never before mentioned in the literature):

- 8% A, 24% 2, 12% 3, 56% J for rows, and
- 16% A, 12% 2, 8% 3, 64% J for columns.

(An order of magnitude stronger because (56 - 40)% and (64 - 40)% are respectively roughly ten times larger than (22 - 20)% and (22.6 - 20)%.)

Moreover, unlike the putative Ace effect, the Joker effect and the other observed frequencies *can* be gracefully explained by a level-*k* model with an Obama-McCain *L0* that probabilistically favors the salient A and J cards.

The analysis also suggests that the Ace effect in the time-aggregated data was an accidental by-product of how subjects learned, not of salience at all.

TABLE 5—COMPARISON OF THE LEADING MODELS IN O'NEILL'S GAME

Model	Parameter estimates	Observed or predicted choice frequencies					MSE
		Player	A	2	3	J	
Observed frequencies (25 Player 1s, 25 Player 2s)		1	0.0800	0.2400	0.1200	0.5600	–
		2	0.1600	0.1200	0.0800	0.6400	–
Equilibrium without perturbations		1	0.2000	0.2000	0.2000	0.4000	0.0120
		2	0.2000	0.2000	0.2000	0.4000	0.0200
Level-k with a role-symmetric *L0* that favors salience	$a > 1/4$ and $j > 1/4$ $3j - a < 1, a + 2j < 1$	1	0.0824	0.1772	0.1772	0.5631	0.0018
		2	0.1640	0.1640	0.1640	0.5081	0.0066
Level-k with a role-symmetric *L0* that favors salience	$a > 1/4$ and $j > 1/4$ $3j - a < 1, a + 2j > 1$	1	0.0000	0.2541	0.2541	0.4919	0.0073
		2	0.2720	0.0824	0.0824	0.5631	0.0050
Level-k with a role-symmetric *L0* that avoids salience	$a < 1/4$ and $j < 1/4$	1	0.4245	0.1807	0.1807	0.2142	0.0614
		2	0.1670	0.1807	0.1807	0.4717	0.0105
Level-k with a role-asymmetric *L0* that favors salience for locations for which player is a seeker and avoids it for locations for which player is a hider	$a_1 < 1/4, j_1 > 1/4;$ $a_2 > 1/4, j_2 < 1/4$ $3j_1 - a_1 < 1,$ $a_1 + 2j_1 < 1, 3a_2 + j_2 > 1$	1	0.1804	0.2729	0.2729	0.2739	0.0291
		2	0.1804	0.1804	0.1804	0.4589	0.0117

Crawford and Iriberri's Table 5

Equilibrium or LQRE with perturbations are well-defined for O'Neill's game, but they both fit significantly worse than our favored level-*k* model.

As explained in the paper, equilibrium or LQRE with perturbations are not even well-defined for Rapoport and Boebel's game.

A level-*k* model is well-defined, and explains some but by no means all of the patterns in Rapoport and Boebel's data.

Importantly, Crawford and Iriberri's analysis traces the superior portability of the level-*k* model to the fact that *L0* is psychological rather than strategic, and that it is based on simple and universal intuition and evidence.

If *L0* were strategic, it would interact with the strategic structure in new ways in each new game, and it would be a rare event when one could extrapolate a specification from one game to another as Crawford and Iriberri did from RTH's games to O'Neill's.

Thus, the definition of *L0* as an instinctive, nonstrategic response is more that a convenient cognitive categorization: it is important for portability.

11. Chicago Skyscrapers: Framing Effects and Miscoordination in Schelling-Style Coordination Games

Perhaps the most famous examples of framing effects in economics are Schelling's (1960) classic "meeting in New York City" experiments.

Crawford, Gneezy, and Rottenstreich ("CGR") 2008 *AER* randomly paired subjects to play games with commonly observed, non-neutral decision labels like Schelling's, but except for a game with the payoff symmetry of Schelling's, CGR used payoff-asymmetric games like Battle of the Sexes.

In unpaid pilots run in Chicago, CGR used naturally occurring labels, pitting the world-famous Sears Tower versus the little-known AT&T Building across the street.

Sears Tower with the AT&T Building in the background on its left (the AT&T Building is actually almost as tall as Sears Tower)

		P2	
		Sears	AT&T
P1	Sears	100,100	0,0
	AT&T	0,0	100,100

Symmetric

		P2	
		Sears	AT&T
P1	Sears	100,101	0,0
	AT&T	0,0	101,100

Slight Asymmetry

		P2	
		Sears	AT&T
P1	Sears	100,110	0,0
	AT&T	0,0	110,100

Moderate Asymmetry

Chicago Skyscrapers

The salience of Sears Tower makes it easy and, in principle, obvious for subjects to coordinate on the “both-Sears” equilibrium; and they almost all do this in the symmetric version of the game.

Since Schelling’s experiments with symmetric games, people have assumed that slight payoff asymmetry would not interfere with this.

However, even with slight payoff asymmetry, the game poses a new strategic problem because both-Sears is one player’s favorite way to coordinate but not the other player’s.

Just as in a society of men and women playing Battle of the Sexes, in which Ballet is more salient than Fights, there is a tension between the “label salience” of Sears and the “payoff-salience” of a player’s favorite way to coordinate:

Payoff salience reinforces label salience in one player role (P2s) but opposes it for players in the other (P1s).

This tension may lead players to respond asymmetrically, which in this game is bad for coordination.

As CGR suspected, although the Chicago Skyscrapers results replicated Schelling's results in the symmetric version of the game, there was a substantial decline in coordination with even slight payoff asymmetry.

		P2 (90% Sears)	
		Sears	**AT&T**
P1 (90% Sears)	**Sears**	**100,100**	**0,0**
	AT&T	**0,0**	**100,100**

Symmetric

		P2 (58% Sears)	
		Sears	**AT&T**
P1 (61% Sears)	**Sears**	**100,101**	**0,0**
	AT&T	**0,0**	**101,100**

Slight Asymmetry

		P2 (47% Sears)	
		Sears	**AT&T**
P1 (50% Sears)	**Sears**	**100,110**	**0,0**
	AT&T	**0,0**	**110,100**

Moderate Asymmetry

Chicago Skyscrapers

To investigate the reasons for the decline in coordination, CGR conducted more formal, paid treatments using abstract decision labels, pitting X against Y, with X presumed (and shown) to be more salient than Y.

		P2 X	P2 Y
P1	X	5,5	0,0
	Y	0,0	5,5

Symmetric

		P2 X	P2 Y
P1	X	5,5.1	0,0
	Y	0,0	5.1,5

Slight Asymmetry

		P2 X	P2 Y
P1	X	5,6	0,0
	Y	0,0	6,5

Moderate Asymmetry

		P2 X	P2 Y
P1	X	5,10	0,0
	Y	0,0	10,5

Large Asymmetry

Like the salience of Sears Tower, the salience of the X label makes it obvious for subjects to coordinate on the “both-X” equilibrium; and they again do this in the symmetric version of the game.

But with payoff asymmetry there is again a tension between the “label salience” of X and the “payoff-salience” of a player’s favorite way to coordinate: Payoff salience again reinforces label salience for P2s but opposes it for P1s.

This tension again had a large and surprising effect:

		P2 (76% X)	
		X	Y
P1 (76% X)	X	5,5	0,0
	Y	0,0	5,5

Symmetric

		P2 (28% X)	
		X	Y
P1 (78% X)	X	5,5.1	0,0
	Y	0,0	5.1,5

Slight Asymmetry

		P2 (61% X)	
		X	Y
P1 (33% X)	X	5,6	0,0
	Y	0,0	6,5

Moderate Asymmetry

		P2 (60% X)	
		X	Y
P1 (36% X)	X	5,10	0,0
	Y	0,0	10,5

Large Asymmetry

Even tiny payoff asymmetries caused a large drop in the expected coordination rate, from 64% (0.64 = 0.76×0.76 + 0.24×0.24) in the symmetric game to 38%, 46%, and 47% in the asymmetric games.

Perhaps more surprisingly (and unlike in the unpaid Chicago Skyscrapers treatment), the pattern of miscoordination reversed as asymmetric games progressed from small to large payoff differences:

- With slightly asymmetric payoffs, most subjects in both roles favored their partners' payoff-salient decisions.

- But with moderate or large asymmetries, most subjects in both roles switched to favoring their own payoff-salient decisions.

There are two things to explain here:

- Why didn't subjects in the asymmetric games ignore the payoff asymmetry, which cannot be used to break the symmetry as required for coordination, and use the salience of Sears Tower to coordinate?

- Why did the pattern of miscoordination reverse as the asymmetric games progressed from small to large payoff differences?

Standard notions such as equilibrium plus noise with refinements and QRE ignore labeling, and so cannot help.

A level-k model can gracefully explain the patterns in the data, but again it's important to have an *L0* that realistically describes people's beliefs about others' instinctive reactions to the tension between label- and payoff-salience that seems to drive the results.

CGR assume that *L0* is the same in both player roles, and that it responds instinctively to both label and payoff salience; but with a "payoffs bias" that favors payoff over label salience, other things equal:

- In symmetric games *L0* chooses X with some probability greater than ½.
- In any asymmetric game, (for simplicity only) whether or not label-salience opposes payoff-salience, *L0* chooses its payoff-salient decision with probability $p > ½$.

(These assumptions are consistent with Crawford and Iriberri's 2007 *AER* *L0* assumptions, because their games had no payoff-salience.

However, there remain some unresolved issues about how to generalize it.)

Under these assumptions about *L0*, *L1*'s and *L2*'s choices in roles P1 and P2 are completely determined by *p*, the extent of *L0*'s payoff bias.

Except in symmetric games, even though *L0*'s choice probabilities are the same for P1s and P2s, they imply *L1* and *L2* choice probabilities that differ across player roles due to the asymmetric relationships between label and payoff salience for P1s and P2s.

Simple calculations (CGR's Table 3, reproduced next slide) show that a level-*k* model can track the reversal of the pattern of miscoordination between the slightly asymmetric game and the games with moderate or large payoff asymmetries if (and only if) 0.505 (= 5.1/[5.1+5]) < *p* < 0.545 (= 6/[6+5]), so that *L0* has only a modest payoff bias.

If *p* falls into this range and the population frequency of *L1* is 0.7 and that of *L2* is 0.3, close to most previous estimates, the model's predicted choice frequencies differ from the observed frequencies by more than 10% only in the symmetric game, where the model somewhat overstates the homogeneity of CGR's subject pool (Table 3).

	Symmetric Labeled (SL)	Asymmetric Slight Labeled (ASL)	Asymmetric Moderate Labeled (AML)	Asymmetric Large Labeled (ALL)
Payoffs for coordinating on "*X*"	$5, $5	$5, $5.10	$5, $6	$5, $10
Payoffs for coordinating on "*Y*"	$5, $5	$5.10, $5	$6, $5	$10, $5
Pr{X} for P1 *L0*	> ½	1-*p*	1-*p*	1-*p*
Pr{X} for P2 *L0*	> ½	*p*	*p*	*p*
Pr{X} for P1 *L1*	1	1	0	0
Pr{X} for P1 *L2*	1	0	1	1
Pr{X} for P2 *L1*	1	0	1	1
Pr{X} for P2 *L2*	1	1	0	0
Total P1 predicted Fr{X}	100%	100*q*%	100(1-*q*)%	100(1-*q*)%
Total P1 predicted Fr{X}|*q*=0.7	100%	70%	30%	30%
Total P1 observed Fr{X}	76%	78%	33%	36%
Total P2 predicted Fr{X}	100%	100(1-*q*)%	100*q*%	100*q*%
Total P2 predicted Fr{X}|*q*=0.7	100%	30%	70%	70%
Total P2 observed Fr{X}	76%	28%	61%	60%

Table 3. *L1*'s and *L2*'s choice probabilities in *X*-*Y* treatments when 0.505 < *p* < 0.545

CGR's Table 3

The details are as follows:

		P2 (76%) *X*	P2 (76%) *Y*
P1 (76%)	*X*	5,5	0,0
	Y	0,0	5,5

Symmetric

- In the symmetric game, with no payoff salience, *L0* favors the salience of X.

- *L1* P1s and P2s therefore both choose X.

- *L2* P1s and P2s do the same.

In this case the model predicts that 100% of P1s and P2s will choose X. Thus, here it makes the same prediction as equilibrium selection based on salience as in a Schelling focal point. This prediction is fairly accurate, but it overstates the homogeneity of the subject pool.

		P2 (28%)	
		X	***Y***
P1 (78%)	***X***	**5,5.1**	**0,0**
	Y	**0,0**	**5.1,5**

Slight Asymmetry

- In the slightly asymmetric game, with $p > 0.505$ (= 5.1/[5.1+5]), the payoff differences are small enough that *L1* P1s choose P2s' payoff-salient decision, X, because *L1* P1s think it is sufficiently likely that *L0* P2s will choose X that X yields them higher expected payoffs.
- *L2* P2s, who best respond to *L1* P1s, thus choose X as well.
- With $p > 0.505$, *L1* P2s choose P1s' payoff-salient decision, Y, because *L1* P2s think it sufficiently likely that *L0* P1s will choose Y.
- *L2* P1s thus choose Y.

In this case the model predicts that *L1* P1s choose X and *L2* P1s choose Y, while *L1* P2s choose Y and *L2* P2s choose X. Thus, when $q = 0.7$, the model predicts that 70% of P1s will choose X but only 30% of P2s will choose X, reasonably close to the observed 78% and 28%.

		P2 (61%) *X*	P2 (61%) *Y*
P1 (33%)	*X*	5,6	0,0
	Y	0,0	6,5

Moderate Asymmetry

		P2 (60%) *X*	P2 (60%) *Y*
P1 (36%)	*X*	5,10	0,0
	Y	0,0	10,5

Large Asymmetry

- In the games with moderate or large payoff asymmetries, *L0*'s payoffs bias is strong enough, but not too strong ($p < 0.545$ (= 6/[6+5])), that *L1* P1s and P2s both choose their own instead of their partners' payoff-salient decisions, Y for P1s and X for P2s.
- *L2* P1s choose X and *L2* P2s choose Y.

In this case the model predicts that *L1* P1s choose Y and *L2* P1s choose X, while *L1* P2s choose X and *L2* P2s choose Y. Thus, when $q = 0.7$, the model predicts that 30% of P1s will choose X but 70% of P2s will choose X, again close to the observed 33-36% and 61-60%.

12. Huarangdao and D-day: Preplay Communication of Intentions in Zero-Sum Two-Person Games with Possibly *Sophisticated* Players

Consider a simple perturbed Matching Pennies game as in Crawford 2003 *AER*, viewed as a model of the Allies' choice of where to invade Europe on D-Day (6 June 1944):

		Germans	
		Defend Calais	**Defend Normandy**
Allies	**Attack Calais**	1 -1	-2 2
	Attack Normandy	-1 1	1 -1

- Attacking an undefended Calais is better for the Allies than attacking an undefended Normandy, so better for them on average.
- Defending an unattacked Normandy is worse for the Germans than defending an unattacked Calais and so worse for them on average.

Now imagine that D-Day is preceded by a message from the Allies to the Germans regarding their intentions about where to attack, as in Operation Fortitude South (http://en.wikipedia.org/wiki/Operation_Fortitude).

Imagine further that the message is (approximately!) cheap talk.

A “Tank” from Operation Fortitude

In any equilibrium (refined or not) of a zero-sum game preceded by a cheap-talk message regarding intentions, the sender must make his message uninformative, and the receiver must ignore it.

If in equilibrium the receiver found it optimal to respond to the message, his response would benefit him and so hurt the sender, who would therefore do better by making the message uninformative.

Thus communication can have no effect in any equilibrium, and as a result the underlying game must be played according to its unique mixed-strategy equilibrium, as if there were no communication phase.

Yet intuition suggests that in many such situations:

- The sender's message and action are part of a single, integrated strategy.
- The sender tries to anticipate which message will fool the receiver and chooses it nonrandomly.
- The sender's action differs from what he would have chosen with no opportunity to send a message.

Moreover, in my stylized version of D-Day:

- The deception succeeded (the Allies faked preparations for invasion at Calais, the Germans defended Calais and left Normandy lightly defended, and the Allies then invaded Normandy).
- But the sender won in the less beneficial of the two possible ways.

D-Day is only one datapoint, if that (the model is greatly oversimplified).

But there's an ancient Chinese antecedent of D-Day, Huarongdao (http://en.wikipedia.org/wiki/Battle_of_Red_Cliffs), in which General Cao Cao chooses between two roads, the comfortable Main Road and the awful Huarong Road, trying to avoid capture by General Kongming.

Huarongdao

		Kongming: Main	**Kongming: Huarong**
Cao Cao: Main		3 -1	0 1
Cao Cao: Huarong		1 0	2 -2

- Cao Cao loses 2 and Kongming gains 2 if Cao Cao is captured.

- But both Cao Cao and Kongming gain 1 by taking the Main Road, whether or not Cao Cao is captured: It's important to be comfortable, even if (especially if?) if you think you're about to die.

In Huarongdao, essentially the same thing happened as in D-Day:

Kongming lit campfires on the Huarong road; Cao Cao was fooled by this into thinking Kongming would ambush him on the *Main Road*; and Kongming captured Cao Cao but only by taking Huarong Road.

(The ending however was happy: Kongming later let Cao Cao go.)

In what sense did the “essentially the same thing” happen?

In D-Day the message was literally deceptive but the Germans were fooled because they “believed” it (either because they were credulous or because they inverted the message one too many times).

Kongming's message was literally truthful—he lit fires on the Huarong Road and ambushed Cao Cao there—but Cao Cao was fooled because he misread Kongming’s message strategy and inverted the message.

The sender’s and receiver’s message strategies and beliefs were different, but the outcome—what happened in the underlying game—was the same: The sender won, but in the less beneficial of the two possible ways.

Why was Cao Cao fooled by Kongming's message?

One advantage of using fiction as data is that it can reveal cognition:

- *Three Kingdoms* gives Kongming's rationale for sending a deceptively truthful message: "Have you forgotten the tactic of 'letting weak points look weak and strong points look strong'?"
- It also gives Cao Cao's rationale for inverting Kongming's message: "Don't you know what the military texts say? 'A show of force is best where you are weak. Where strong, feign weakness.' "

Why was Cao Cao fooled by Kongming's message?

One advantage of using fiction as data is that it can reveal cognition:

- *Three Kingdoms* gives Kongming's rationale for sending a deceptively truthful message: "Have you forgotten the tactic of 'letting weak points look weak and strong points look strong'?"
- It also gives Cao Cao's rationale for inverting Kongming's message: "Don't you know what the military texts say? 'A show of force is best where you are weak. Where strong, feign weakness.' "

Cao Cao must have bought a used, out-of-date edition….

As we will see, with *L0* suitably adapted to this setting, Cao Cao's rationale resembles *L1* thinking; but Kongming's rationale resembles *L2* thinking.

We can now restate the puzzle more concretely, for both D-Day and Huarongdao:

- Why did the receiver allow himself to be fooled by a costless (hence easily faked) message from an *enemy*?

- If the sender expected his message to fool the receiver, why didn't he reverse it and fool the receiver in the way that would have allowed him to win in the *more* beneficial way? (Why didn't the Allies feint at Normandy and attack at Calais? Why didn't Kongming light fires and ambush Cao Cao on the Main Road?)

A level-*k* analysis suggests that it was more than a coincidence that the same thing happened in both cases.

Although *Sophisticated* subjects are rare in laboratory experiments, one hopes they are more common in field settings; and it is interesting to see whether a plausible model allows deception between *Sophisticated* players.

Accordingly, let Allies' and Germans' types be drawn from separate distributions, each including both level-*k* or *Mortal* types (as Crawford 2003 *AER* called them), and a fully strategically rational or *Sophisticated*, type.

Mortal types use step-by-step procedures that generically determine unique pure strategies, and avoid simultaneous determination of the kind used to define equilibrium; recall the Selten 1998 *EER* quote above.

Sophisticated types know everything about the game, including the distribution of *Mortal* types; and so play an equilibrium in a "reduced game" between possible *Sophisticated* players, taking *Mortals*' choices as given.

How should *L0* be adapted to an extensive-form game with communication?

Here a uniform random *L0* seems quite unnatural. For sender or receiver, the instinctive reaction to a message in a language one understands is surely to focus on its literal meaning, even if one ends up either lying or not taking the message at face value.

The level-*k* model therefore anchors *Mortal* types' messages and responses on *L0*s based on truthfulness for senders and credulity for receivers, just as in the informal literature on deception.

(The literature has not yet converged on whether *L0* receivers should be defined as credulous or uniform random—compare Ellingsen and Östling 2009—but the distinction is partly semantic because truthful *L0* senders imply that *L1* receivers are also credulous.)

Mortal Allied types' simplified models of other players make *L1* or higher *Mortal* Allied types always expect to fool the Germans, either by lying (like the Allies) or by telling the truth (like Kongming).

Given this, all *L1* or higher *Mortal* Allied types send a message they expect to make the Germans think they will attack Normandy, and then attack Calais.

If we knew the Allies and Germans were *Mortal*, we could now derive the model's implications from an estimate of the type frequencies of *Mortal* Allies who tell the truth or lie, and of *Mortal* Germans who believe or invert the Allies' message.

But the analysis must also take into account the possibility of *Sophisticated* Allies and Germans, who know everything about the game, including the distribution of *Mortal* types, and play an equilibrium in the resulting game.

To take into account the possibility of *Sophisticated* Allies and Germans, note that *Mortals* players' strategies are determined independently of each other's and *Sophisticated* players' strategies, and so can be treated as exogenous (even though they affect other players' payoffs).

Plug in the distributions of *Mortal* Allies' and Germans' independently determined behaviors to obtain a "reduced game" between *Sophisticated* Allies and *Sophisticated* Germans.

Because *Sophisticated* players' payoffs are influenced by *Mortal* players' decisions, the reduced game is no longer zero-sum, its messages are not cheap talk, and it has incomplete information.

The sender's message, ostensibly about his intentions, is in fact read by a *Sophisticated* receiver as a signal of the sender's type.

Thus, the possibility of *Mortal* players completely changes the character of the game between *Sophisticated* players, which is what gives the model the ability to explain the effectiveness of communication in a zero-sum game and the possibility of deception between *Sophisticated* players.

The equilibria of the reduced game are determined by the population frequencies of *Mortal* and *Sophisticated* senders and receivers.

There are two leading cases, with different implications:

- When *Sophisticated* Allies and Germans are common—not behaviorally plausible—the reduced game has a mixed-strategy equilibrium whose outcome is virtually equivalent to D-Day's without communication.
- When *Sophisticated* Allies and Germans are rare, the game has an essentially unique pure-strategy equilibrium, in which *Sophisticated* Allies can predict *Sophisticated* Germans' decisions, and vice versa.

In the latter, pure-strategy equilibrium, *Sophisticated* Germans always defend Calais (because they know that *Mortal* Allies, who predominate when *Sophisticated* Allies are rare, will always attack Calais).

Sophisticated Allies send the message that fools the most likely kind of *Mortal* German (feinting at Calais or Normandy depending on whether more *Mortal* Germans believe than invert messages), and then attack Normandy.

Surprisingly, there never exists a pure-strategy equilibrium in which *Sophisticated* Allies feint at Normandy and then attack Calais.

In such an equilibrium any deviation from *Sophisticated* Allies' equilibrium message would "prove" to *Sophisticated* Germans that the Allies were *Mortal*, making it optimal for *Sophisticated* Germans to defend Calais and suboptimal for *Sophisticated* Allies to attack there.

If in the equilibrium, *Sophisticated* Allies feinted at Normandy and attacked Calais, then their message would fool only the most likely kind of *Mortal* German—*Sophisticated* Germans can never be fooled in a pure-strategy equilibrium, and a given message cannot fool both *Mortal* Germans who believe and *Mortal* Germans who invert messages—with expected payoff gain equal to the frequency of the most likely kind of *Mortal* German times the payoff of attacking an undefended Normandy.

But such *Sophisticated* Allies could reverse both their message and attack location, again fooling the most likely kind of *Mortal* German, but now with expected payoff gain equal to the frequency of that kind of German times the higher payoff of attacking an undefended Calais, a contradiction.

In the pure-strategy equilibrium that exists when *Sophisticated* Allies and Germans are rare, the Allies' message and action are part of a single, integrated strategy; and the probability of attacking Normandy is much higher than if no communication was possible.

The Allies choose their message nonrandomly, the deception succeeds most of the time, but it allows the Allies to win in the less beneficial way.

Nonetheless, *Sophisticated* players in either role do strictly better than their *Mortal* counterparts; their advantage comes from the ability to avoid being fooled and/or to choose which *Mortal* type(s) to fool.

Thus for plausible parameter values, with no unexplained difference in the sophistication of Allies and Germans, the model explains why *Sophisticated* Germans might allow themselves to be "fooled" by a costless message from a *Sophisticated* enemy: It is an unavoidable cost of exploiting mistakes by *Mortal* enemies, who are much more common.

In a weaker sense (resting on a preference for pure-strategy equilibria and deterministic predictions), the model also explains why *Sophisticated* Allies don't feint at Normandy and attack Calais, even though this would be more profitable if it succeeded.

In the mixed-strategy equilibrium that prevails when *Sophisticated* Allies and Germans are common, *Sophisticated* players' equilibrium mixed strategies offset each other's gains from fooling *Mortal* Receivers, and in each role *Sophisticated* and *Mortal* players have equal expected payoffs.

This suggests that in an adaptive analysis of the dynamics of the type distribution, as in Conlisk 2001 *AER*, the frequencies of *Sophisticated* types will grow until the population is in or near (depending on costs) the region of mixed-strategy equilibria in which types' expected payoffs are equal.

Thus *Sophisticated* and *Mortal* players can coexist in long-run equilibrium.

13. Preplay Communication of Intentions in Coordination Games

If level-*k* models allow preplay communication of intentions to affect the outcomes of zero-sum games, it should come as no surprise that they also allow effective communication in coordination games.

Ellingsen and Östling 2009 and Crawford 2007, not discussed in detail here, adapt Crawford's 2003 *AER* approach to study different aspects of preplay communication of intentions in coordination and other games.

Ellingsen and Östling 2009 use a level-*k* model to study the effectiveness of a single round of one- or two-sided preplay communication in games where communication of intentions plays various roles.

Crawford 2007 uses a level-*k* model to study the effectiveness of one- or multi-round two-sided communication in games like Battle of the Sexes, building on Farrell's 1987 *RAND J* and Rabin's 1994 *JET* analyses.

In each case the power of the analysis stems from the use of a model that does not assume equilibrium, which is question-begging in this context; but which imposes a realistic structure less agnostic than rationalizability.

14. Experimental Evidence on Communication of Private Information in Sender-Receiver Games

I now discuss some experimental evidence on communication of private information in discretized versions of Crawford and Sobel's 1982 *Econometrica* Sender-Receiver Games, from Wang, Spezio, and Camerer 2009 *AER*, who built on the experiments of Cai and Wang's 2006 *GEB*.

Sender observes state S = 1, 2, 3, 4, or 5, sends message M = 1, 2, 3, 4, or 5. Receiver observes message, chooses action A = 1, 2, 3, 4, or 5.

The Receiver's choice of A determines the welfare of both:

- The Receiver's ideal outcome is A = S.
- The Sender's ideal outcome is A = S + b.

The Receiver's von Neumann-Morgenstern utility function is $110 - 20|S - A|^{1.4}$, and the Sender's is $110 - 20|S + b - A|^{1.4}$.

The difference in preferences varied across treatments: b = 0, 1, or 2.

Crawford and Sobel's theoretical analysis characterized the possible equilibrium relationships between Sender's observed S and Receiver's choice of A, which determines the informativeness of communication.

They showed, for a class of models with continuous state and action spaces that generalizes Wang et al.'s examples (except for discreteness), that all equilibria are "partition equilibria", in which as illustrated below, the Sender partitions the set of states into contiguous groups and tells the Receiver, in effect, only which group his observation lies in.

For any given difference in Sender's and Receiver's preferences (b), there is a range of equilibria, from a "babbling" equilibrium with one partition element to more informative equilibria that exist when b is small enough.

Under reasonable assumptions there is a "most informative" equilibrium, which has the most partition elements and gives the Receiver the highest ex ante (before the Sender observes the state) expected payoff.

As the preference difference decreases, the amount of information transmitted in the most informative equilibrium increases (measured either by the correlation between S and A or the Receiver's expected payoff).

The unambiguous part of Crawford and Sobel's characterization of equilibrium concerns the possible relationships between S and A.

Because messages have no direct effect on payoffs ("cheap talk"), there is nothing to tie down their meanings in equilibrium.

As a result, any equilibrium relationship between S and A can be supported by any sufficiently rich language, with the meanings of messages determined by players' equilibrium beliefs.

(By contrast, in Tom Stoppard's play "Dogg's Hamlet", the actors speak a language called "Dogg", which consists of ordinary English words but with meanings completely different from their normal meanings. This creates a lot of amusing confusion when they interact with true English speakers—confusion that would not arise if Dogg did not sound so much like English.)

Behaviorally, however, in experiments like Wang et al.'s with a clear correspondence between state and message—S = 1, 2, 3, 4, or 5 and M = 1, 2, 3, 4, or 5—or where communication is in a common natural language, the interpretations of messages are dictated by their literal meanings.

Thus messages are always understood—even if not always believed.

Wang et al.'s data analysis therefore fixes the meanings of Sender subjects' messages at their literal values.

Even with this restriction, when b = 0 or 1 in their design (Sender's and Receiver's preferences are close enough) there are multiple equilibria.

Wang et al.'s analysis then focuses on the "most informative" equilibrium.

When b = 0, the most informative equilibrium has M = S and A = S: perfect truth-telling, credulity, and information transmission, as is intuitively plausible when Sender and Receiver have identical preferences.

When b = 2, the most informative equilibrium has Senders sending a completely uninformative message M = {1, 2, 3, 4, 5} for any value of S; and Receivers ignoring it, hence choosing A = 3, which is optimal given their prior beliefs, for any value of M.

(A babbling equilibrium also exists when b = 0 or 1, but then it is not the most informative equilibrium.)

When b = 1, the most informative equilibrium has Senders sending M = 1 when S = 1 but M = {2, 3, 4, 5} when S = 2, 3, 4, or 5; and Receivers choosing A = 1 when M = 1 and A = 3 or 4 when M = {2, 3, 4, 5}.

(The Sender's message M = {2, 3, 4, 5} is the simplest way to implement the intentional vagueness of this partition equilibrium. Another way would be for the Sender to randomize M uniformly on {2, 3, 4, 5} when S = 1.)

Thus, when b = 1 the difference in preferences causes noisy information transmission even in the most informative equilibrium.

Importantly, however, the Receiver's beliefs on hearing the Sender's message M are necessarily an unbiased—though noisy—estimate of S:

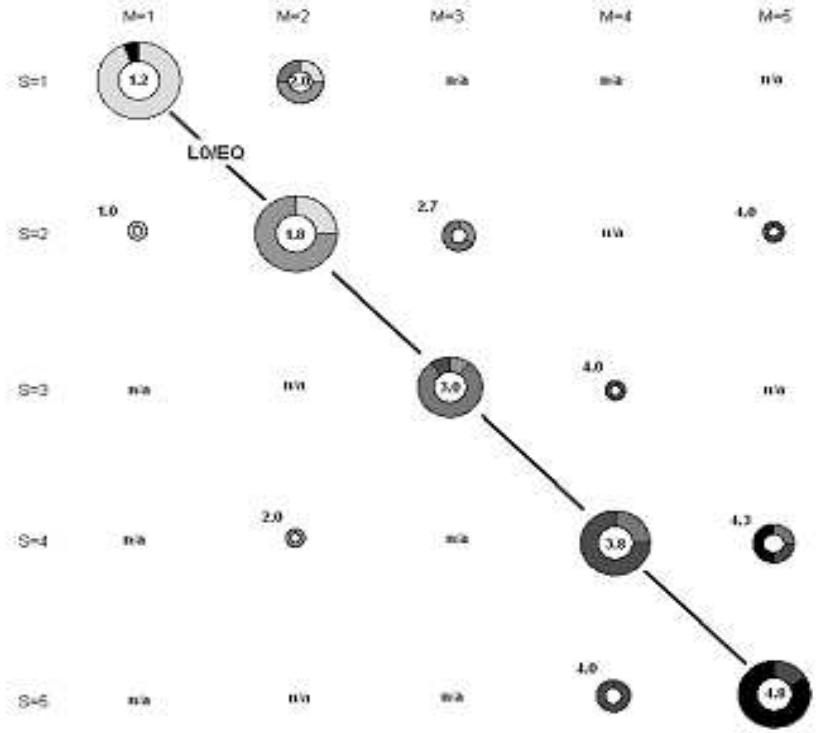
In equilibrium there is no lying or deception, only intentional vagueness.

(When b = 1, there's another, more informative equilibrium, found by David Eil, in which Senders send M = {1, 2} when S = 1 or 2 but M = {3, 4, 5} when S = 3, 4, or 5; and Receivers choose A = 2 when M = {1, 2} and A = 4 when M = {3, 4, 5}. But this equilibrium is not "robust", in that Senders who observe S = 2 are indifferent between M = {1, 2} and M = {3, 4, 5}.)

Turning to Wang et al.'s results, when b = 0 Senders almost always set M = S and Receivers almost always set A = M: The result is near the perfect information transmission predicted by the most informative equilibrium.

Figure 1 shows the Sender's message frequencies and the Receiver's action frequencies as functions of the observed state S: A circle's size shows the Sender's message frequencies. A circle's darkness and the poorly visible numbers inside show the Receiver's action frequencies.

Figure 1: Raw Data Pie Charts (b=0)
(Hidden Bias-Stranger)



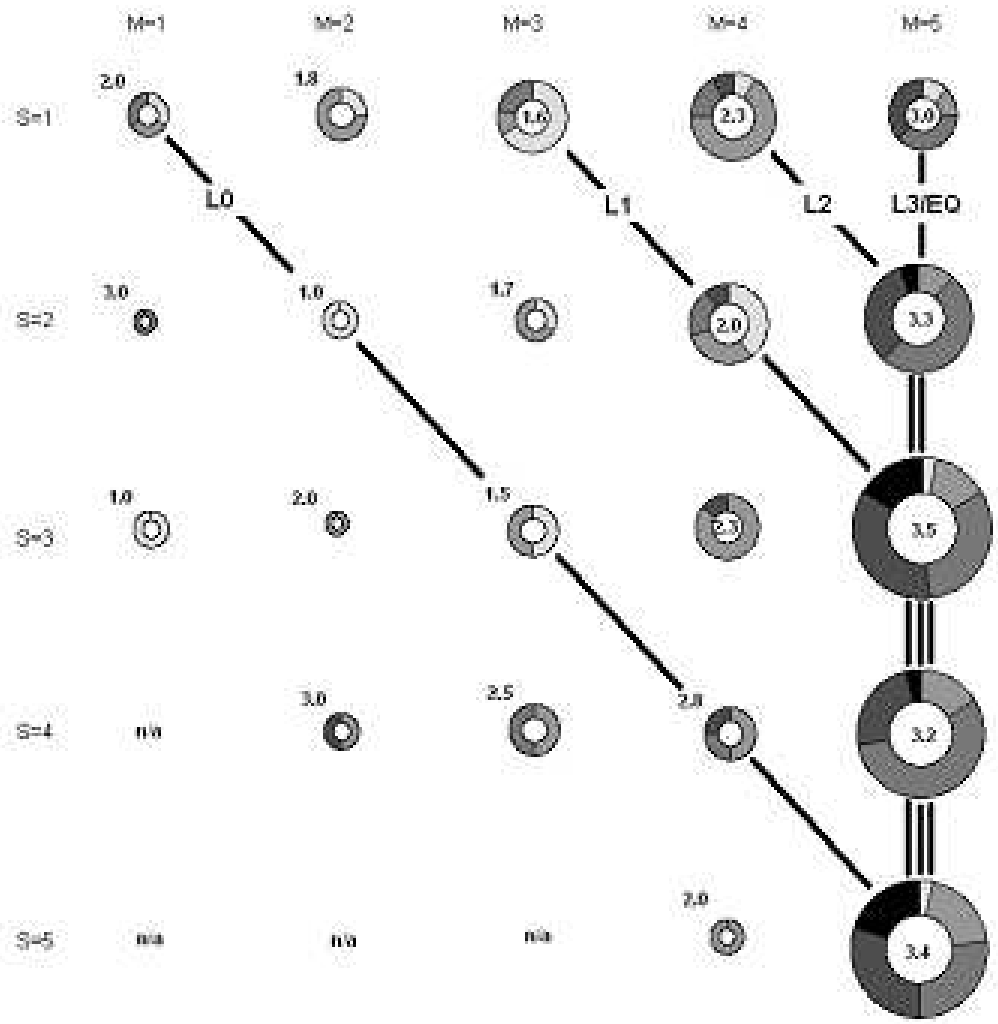
As b increases to b = 1 or b = 2, the amount of information transmitted decreases as predicted by Crawford and Sobel's equilibrium comparative statics, but there are also systematic deviations from the most informative (or any) equilibrium, and lying and successful deception occur.

In Figure 3 (next slide; b = 2 omitted from Wang et al.'s label by accident), in the essentially unique, most informative equilibrium M = {1, 2, 3, 4, 5}, so equilibrium message distributions would look the same for all five rows; and equilibrium actions would be concentrated on A = 3.

However, although the observed actions are fairly close to A = 3, message distributions shift rightward as S increases (going down in the table); thus:

- Most Senders exaggerate the truth (most messages above the diagonal), apparently trying to move Receivers from Receivers' ideal action A = S toward Senders' ideal action A = S + 2 (or 5, whichever is smaller).
- Even so, there is some information in Senders' messages (message distributions shift rightward going down in the table, so messages are positively correlated with the state).
- Receivers are usually deceived to some extent (average A usually > S).

Figure 3: Raw Data Pie Chart, (Hidden Bias-Stranger)



When b = 1, in the most informative robust equilibrium, the Sender's message is M = 1 when S = 1 and M = {2, 3, 4, 5} when S = 2, 3, 4, or 5; and the Receiver chooses A = 1 when M = 1 and A = 3 or 4 when M = {2, 3, 4, 5}. Thus, in equilibrium the distributions of messages and actions would be the same for S = 2, 3, 4, or 5.

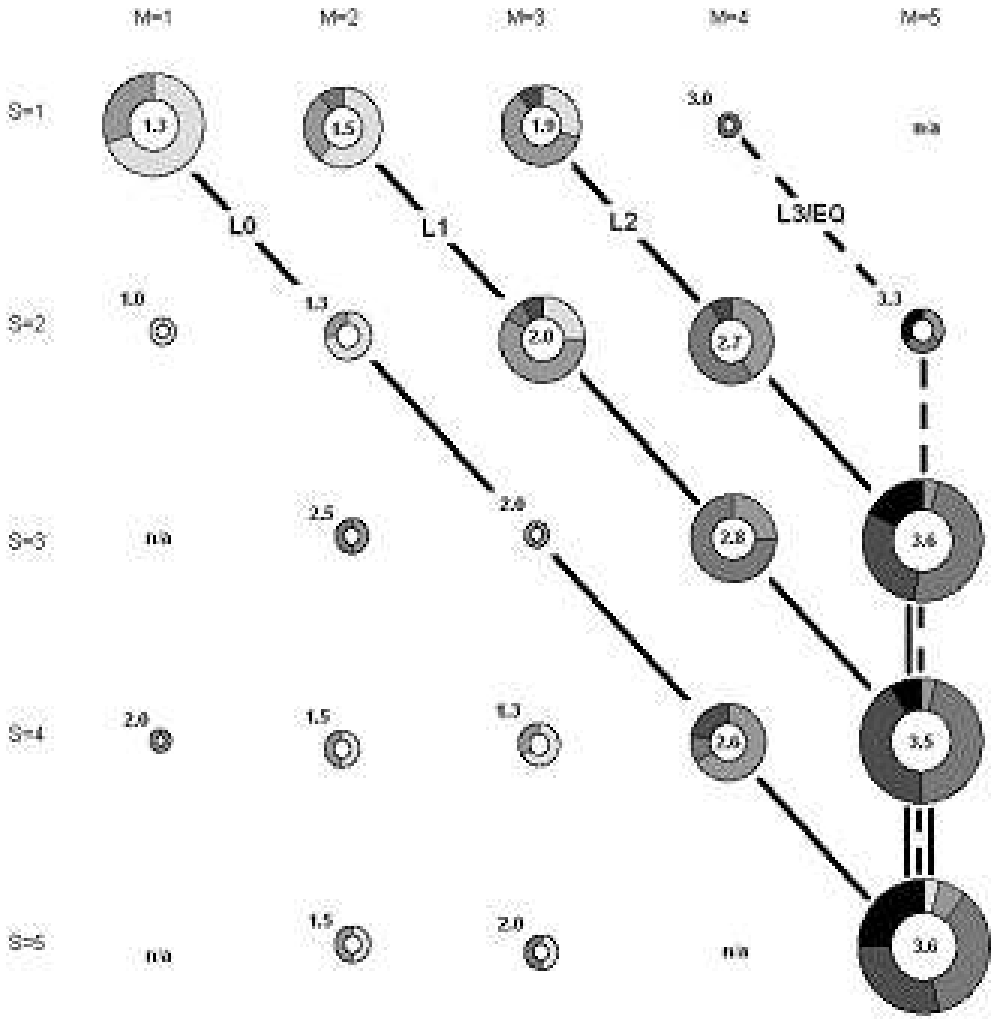
By contrast, turning to Figure 2 (b = 1; next slide):

- Senders almost always exaggerate the truth (messages above the diagonal), apparently trying to move Receivers from Receivers' ideal action A = S toward Senders' ideal action A = S + 1.

- Even so, there is some information in Senders' messages (message distributions shift rightward going down in the table, so messages are positively correlated with the state).

- Receivers are usually deceived to some extent (average A usually > S).

Figure 2: Raw Data Pie Chart (b=1)
(Hidden Bias-Stranger)



What kind of model can explain results like this? Wang et al., following Cai and Wang 2006 *GEB*, propose a level-*k* explanation based on Crawford's 2003 *AER* analysis of preplay communication of intentions (see also Kartik et al. 2007 *JET*):

Anchor beliefs in a truthful Sender *L0*, which sets M = S; and a credulous Receiver *L0* (which also best responds to an *L0* Sender), setting A = M.

L1 Senders best respond to *L0* Receivers by inflating their messages by b: M = S + b (up to M = 5), so that *L0* Receivers will choose S + b, yielding the Sender's ideal action given S.

L1 Receivers (as defined by Wang et al.; the numbering is a convention) best respond to *L1* Senders by discounting the message, normally setting A = M – b, yielding Receivers' ideal action given M = S + b of S.

The qualification "normally" reflects Wang et al.'s assumption that *L1* Receivers take into account that when b = 2, *L1* senders with S = 3, 4, or 5 all send M = 5, with the result that *L1* Receivers, knowing that S is equally likely to be 3, 4, or 5, choose A = 4 instead of A = M – 2b = 3.

L2 Senders best respond to *L1* Receivers by inflating their messages by 2b: M = S + 2b (up to M = 5), so that *L1* Receivers will set A = M – b = S + b, yielding Senders' ideal action given S.

L2 Receivers best respond to *L2* Senders by discounting the message, normally setting A = M – 2b, yielding Receivers' ideal action given M = S + 2b of S.

The qualification "normally" reflects Wang et al.'s assumption that *L2* Receivers take into account that when b = 1, *L2* senders with S = 3, 4, or 5 all send M = 5, with the result that *L2* Receivers, knowing that S is equally likely to be 3, 4, or 5, choose A = 4 instead of A = M – 2b = 3.

L2 Receivers also take into account that when b = 2, *L2* senders with S = 2, 3, 4, or 5 send M = 5, with the result that *L2* Receivers, knowing that S is equally likely to be 2, 3, 4, or 5, choose A = 4 instead of A = M – 2b = 3.

Figure 2: Raw Data Pie Chart (b=1)
(Hidden Bias-Stranger)

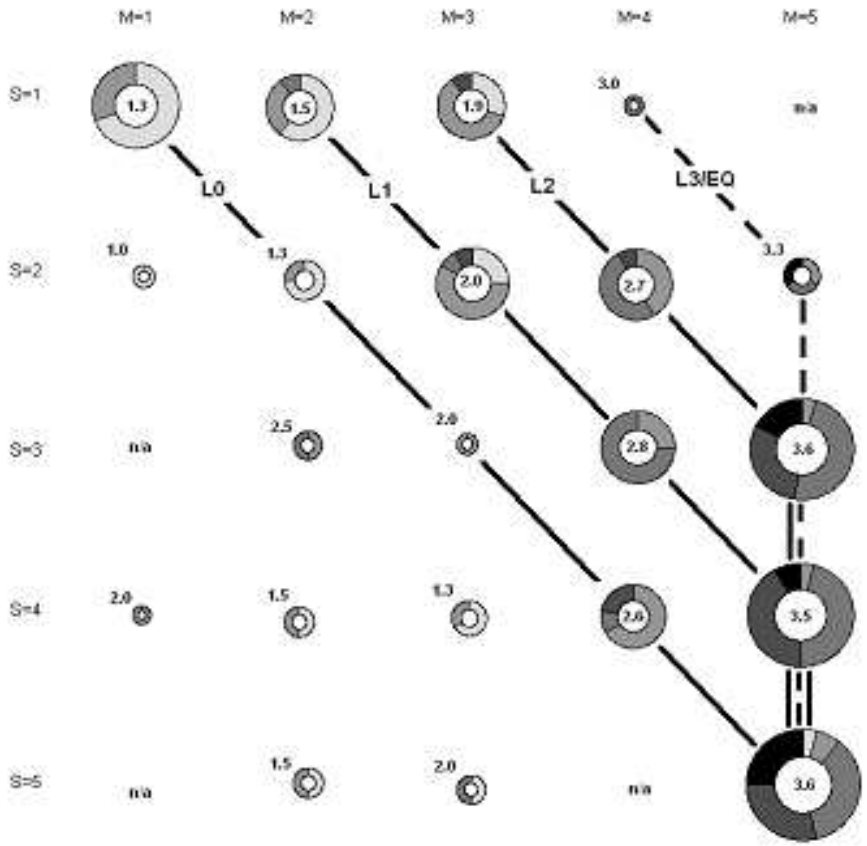
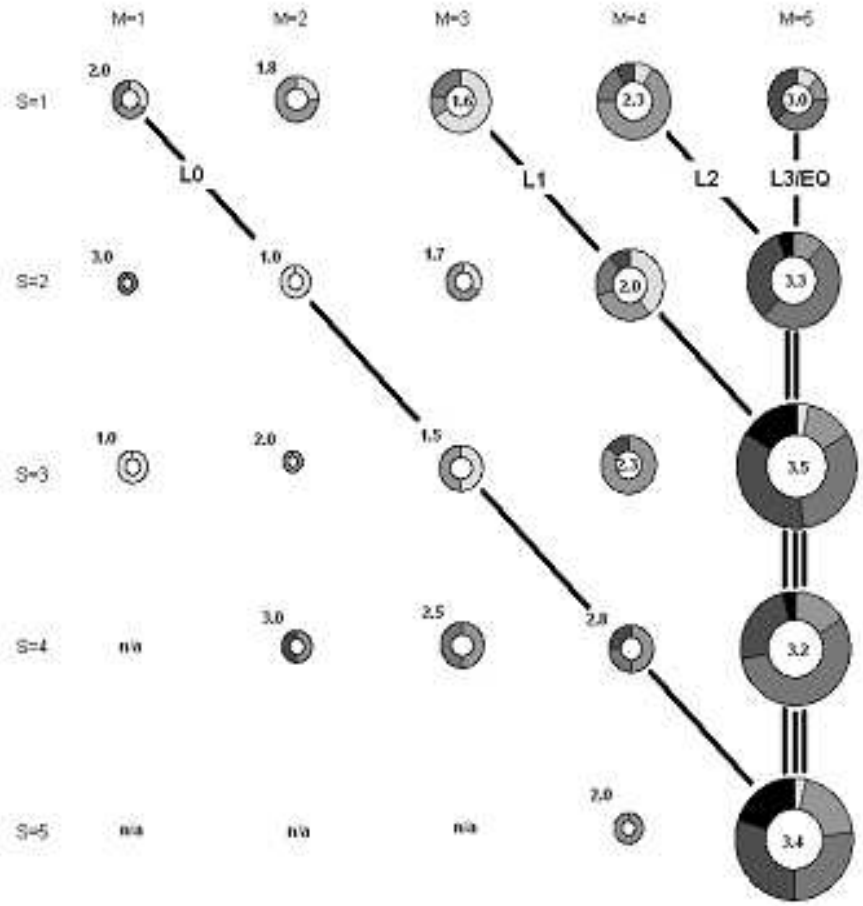


Figure 3: Raw Data Pie Chart, (Hidden Bias-Stranger)



Note that when b = 1, *L1*, *L2*, and *Eq* all predict M = 5 when S = 4 or 5; and when b = 2, *L1*, *L2*, and *Eq* all predict M = 5 when S = 3, 4, or 5.

Econometric estimation classifies 18% of 16 Sender subjects as *L0*, 25% as *L1*, 25% as *L2,* 14% as *Sophisticated*, and 18% as *Equilibrium* (not implausible, but note different type definitions).

15. October Surprise: Communication of Private Information in Zero-Sum Two-Person Games

Crawford's 2003 *AER* approach to preplay communication of intentions via cheap talk is easily adapted, as in Wang, Spezio, and Camerer's 2009 *AER* analysis of communication of private information, to model the CIA's conclusion in October Surprise that bin Laden's October 2004 verbal attack on George W. Bush was intended to aid Bush's reelection.

Assume that only bin Laden knows which candidate he wants to win; and, talk being cheap, that he will say whatever it takes to get it.

A representative American voter knows only that he wants the opposite of what bin Laden wants.

This yields a zero-sum two-person game with incomplete information, which like Wang et al.'s b = 2 treatment has only a babbling equilibrium.

Thus, there is no equilibrium in which bin Laden's cheap talk attack conveys information, or in which an American responds to it.

Consider, however, a level-*k* model in which *L0* is anchored on truthfulness for the sender (bin Laden) and credulity for the receiver (American).

(Or one could derive credulity for an *L1* receiver and start from there.)

An *L0* or *L1* American believes bin Laden's message, and therefore votes for whichever candidate bin Laden attacks.

An *L0* bin Laden who wants Bush to win attacks Kerry, but an *L1* (*L2*) bin Laden who wants Bush to win attacks Bush to induce *L0* (*L1*) Americans to vote for Bush.

Given bin Laden's attack on Bush, an *L0* or *L1* American ends up voting for Bush, and an *L2* American ends up voting for Kerry.

Note that bin Laden's message is always influential, but he needs to choose it to fool the most prevalent kind of American—believer or inverter—as in Crawford's 2003 *AER* analysis.

An *L2* bin Laden believes that Americans are *L1*, hence that "reverse psychology" will be effective.

16. Overbidding in Independent-Private-Value and Common-Value Auctions (time permitting, this section will be replaced in the lectures by a PowerPoint that covers the analysis in more detail)

Equilibrium predictions		
	First-Price	**Second-Price**
Independent-Private-Value Auctions	**Shaded Bidding**	**Truthful Bidding**
Common-Value Auctions	**Value Adjustment + Shaded Bidding**	**Value Adjustment**

Systematic overbidding (relative to equilibrium) has been observed in subjects' initial responses to all kinds of auctions (Goeree, Holt, and Palfrey 2002 *JET*; Kagel and Levin 1986 *AER*, 2000; Avery and Kagel 1997 *JEMS*; Garvin and Kagel 1994 *JEBO*).

(With independent private values, most of the examples that have been studied experimentally do not separate level-*k* from equilibrium bidding strategies, hence our choice to study GHP's results.)

But the literature has proposed completely different explanations of overbidding for private- and common-value auctions:

- "Joy of winning" and/or risk-aversion for private-value auctions.
- Winner's curse for common-value auctions.

Crawford and Iriberri 2007 *Econometrica* propose a level-*k* analysis that provides a unified explanation of these results, without invoking joy of winning, which seems like an intellectual dead end, or risk-aversion.

Crawford and Iriberri's analysis extends Kagel and Levin's 1986 *AER* and Holt and Sherman's 1994 *AER* analyses of "naïve bidding".

It also builds on Eyster and Rabin's ("ER") 2005 *Econometrica* analysis of "cursed equilibrium" and CHC's 2004, Section VI CH analysis of zero-sum betting.

The analysis makes it possible to explore how to extend level-*k* models to an important class of incomplete-information games.

It also makes it possible to explore the robustness of equilibrium auction theory to failures of the equilibrium assumption.

Finally, it establishes a connection between a large body of auction experiments and a large body of experiments on strategic thinking.

The key issue is how to specify *L0*; there are two natural possibilities:

- *Random L0* bids uniformly on the interval between the lowest and highest possible values (even if above own realized value).

- *Truthful L0* bids its expected value conditional on its own signal (meaningful here, though not in all incomplete-information games).

In judging these specifications, bear in mind that *L0* describes only the instinctive starting point of a subject's strategic thinking about others; higher *Lk*s model the actual strategic thinking.

The model constructs separate type hierarchies on these *L0*s, and allows each subject to be one of the types, from either hierarchy.

Random (*Truthful*) *Lk* is *Lk* defined by iterating best responses from *Random* (*Truthful*) *L0*; and is not itself random or truthful.

Given a specification of *L0*, the optimal bid must take into account:

- Value adjustment for the information revealed by winning (only in common-value auctions).

- The bidding trade-off between the higher price paid if the bidder wins and the probability of winning (only in first-price auctions).

With regard to value adjustment, Random *L1* does not condition on winning because Random *L0* bidders bid randomly, hence independently of their values; Random *L1* is “fully cursed” (ER).

All other types do condition on winning, in various ways, but this conditioning tends to make bidders’ bids strategic substitutes, in that the higher others’ bids are, the greater the (negative) adjustment.

Thus, to the extent that Random *L1* overbids, Random *L2* tends to underbid (relative to equilibrium): if it’s bad news that you beat equilibrium bidders, it’s even worse news that you beat overbidders.

The bidding tradeoff, by contrast, can go either way.

The question, empirically, is whether the distribution of types' bids (for example, a mixture of Random *L1* overbidding and Random *L2* underbidding) fits the data better than alternative models.

In three of the four leading cases Crawford and Iriberri study, a level-*k* model does better than equilibrium plus noise, cursed equilibrium, and/or LQRE.

For the remaining case (Kagel and Levin's first-price auction), the most flexible cursed equilibrium specification has a small advantage.

Except in Kagel and Levin's second-price auctions, the estimated type frequencies are similar to those found in other experiments:

Random and Truthful *L0* have low or zero estimated frequencies, and the most common types are (in order of importance) Random *L1*, Truthful *L1*, Random *L2*, and sometimes *Equilibrium* or Truthful *L2*.

17. Behaviorally Optimal Auction Design

A number of recent papers reconsider core microeconomic questions taking a “behavioral” view of individual decisions or probabilistic judgment.

Most such papers focus on consumer behavior, but a few analyze questions in mechanism design (Glazer and Rubinstein 1998 *JET*; Neeman 2003 *GEB*; Eliaz and Spiegler 2006 *REStud*, 2007 *Econometrica*).

In those papers, however, the behavioral aspect is limited to decisions or judgment rather than beliefs: Despite the central role of equilibrium assumptions in the theory of mechanism design, there are very few analyses of design outside the equilibrium paradigm.

Taking a broader view of strategic behavior should increase the practical usefulness of mechanism design theory.

Design inherently involves the creation of new games, for which the learning justification for equilibrium may be weak or nonexistent.

Yet it is often important for an application to work the first time.

In the U.S. FCC spectrum auction, billions of dollars were at stake.

Partly because they believed sealed-bid designs would not yield outcomes close enough to equilibrium to ensure good results, the designers adopted a progressive, partly “open” design for which the theory was weaker but experimental results were more promising.

Further, assuming equilibrium may yield theoretically optimal designs that are too complex for confidence in equilibrium behavior, even if learning is possible.

Replacing equilibrium with a model that better describes people's responses to new and/or complex games—equilibrium responses in games that tend to elicit them and systematic deviations in games that don't—should allow us to design more effective mechanisms.

It also suggests a concrete, evidence-based way to assess the robustness of mechanisms, something previously left to intuition.

In a level-*k* analysis, a "robust" mechanism that implements desired outcomes in dominant strategies or is dominance-solvable in one or two rounds may have an actual advantage over a more complex mechanism that theoretically implements better outcomes, but only in equilibrium.

Crawford, Kugler, Neeman, and Pauzner 2009 *JEEA* conducted a level-*k* analysis of optimal auction design began to explore relaxing the equilibrium assumption in mechanism design.

They considered the leading case of an optimal (expected-revenue maximizing) single-object sealed-bid auction with two symmetric bidders who have independent private values, for which Myerson 1981 *Mathematics of Operations Research*) provides a complete equilibrium-based analysis.

To focus sharply on strategic behavior, they maintained the standard rationality assumptions regarding decisions and judgment.

They modeled strategic behavior via a level-*k* model that follows Crawford and Iriberri 2007 *Econometrica*.

They assumed that bidders are drawn from a given population of level-*k* types, known to the designer.

Because the question of optimal auctions with level-*k* bidders is a difficult one, most of their analysis was conducted in representative examples.

They considered what reserve prices are optimal and how much revenue they yield in first-price auctions.

They also consider the optimality of auction forms, and the use of exotic auctions that exploit bidders' non-equilibrium beliefs to exceed Myerson's revenue bound.

Equilibrium Analysis of Optimal Auctions

Consider single-object auctions with two risk-neutral bidders whose values are independently and identically distributed (“i.i.d.”).

Consider two examples, one with increasing and one with decreasing value density, which lead to different patterns of level-*k* deviations from equilibrium beliefs that together are representative of the possibilities.

In the increasing-density ("I") example values have the distribution function $F_{I\gamma}(v) = v^{\gamma}$ on [0,1], with the density $f_{I\gamma}(v) = \gamma v^{\gamma-1}$.

$\gamma > 0$ is required for $F_{I\gamma}(v)$ to be a valid distribution function, and we strengthen this to $\gamma > 1$ to make the density increasing.

Suppress γ below, writing F_I and f_I instead of $F_{I\gamma}$ and $f_{I\gamma}$.

Because

$$v - \frac{1 - F_I(v)}{f_I(v)} = v - \frac{1 - v^{\gamma}}{\gamma v^{\gamma-1}} = \frac{(\gamma+1)v}{\gamma} - \frac{1}{\gamma v^{\gamma-1}}$$

is increasing in v when $\gamma > 1$, F_I is "regular" in Myerson's sense.

Thus, Myerson's famous result establishes that among the optimal mechanisms in this environment is a first-price auction with suitably chosen reserve price.

In a first-price auction with reserve price r, the equilibrium bid for value $v \geq r$ can be shown to be

$$b^{E}_{I|r}(v) = \frac{r^{\gamma+1}}{(\gamma+1)v^{\gamma}} + \frac{\gamma}{\gamma+1}v,$$

which increases from r at $v = r$ to $\frac{r^{\gamma+1}}{\gamma+1} + \frac{\gamma}{\gamma+1}$ at $v = 1$.

The optimal reserve can be shown to be

$$r = \left[\frac{1}{\gamma+1}\right]^{\frac{1}{\gamma}}.$$

In the decreasing-density (“D”) example values have the distribution function

$$F_{D\alpha\beta}(v) = \frac{\beta}{\beta - \alpha}\left(1 - \frac{\alpha}{v}\right) \text{ on } [\alpha, \beta], \text{ with } \alpha > 0,$$

with well-defined density $\alpha\beta/[(\beta-\alpha)v^2]$ that is positive and continuous on $[\alpha, \beta]$.

Suppress α and β below, writing F_D and f_D instead of $F_{D\alpha\beta}$ and $f_{D\alpha\beta}$.

Because

$$v - \frac{1 - F_D(r)}{f_D(v)} = \frac{v^2}{\beta}$$

is increasing in v on $[\alpha, \beta]$, F_D is regular.

Thus, Myerson’s 1981 result again establishes that an optimal mechanism in this environment is a first-price auction, with reserve price α.

In a first-price auction with reserve $r \in [\alpha, \beta]$, the equilibrium bid for value $v \geq \max\{\alpha, r\}$ is

$$b^E_{D|r}(v) = r\frac{F_D(r)}{F_D(v)} + \frac{1}{F_D(v)}\int_r^v x f_D(x)dx = \frac{v}{v-\alpha}[r - \alpha + \alpha(\ln v - \ln r)].$$

Note that for this distribution, $b^E_{D|r}(v)$ is independent of β.

Further, although $b^E_{D|r}(v) \to \infty$ as $v \to \infty$, it increases extremely slowly, like ln v.

As a result, if $r = \alpha = 1$, for example, $b^E_{D|r}(1{,}000{,}000) = 13.82$, so that in equilibrium, a bidder who values the object at $1,000,000 bids only $13.82.

In a first-price auction with reserve price $r \leq \beta$, the seller's expected revenue in equilibrium is:

$$\Pi_{D\,\alpha\beta\,|r} = r\,(1 - G\,(r)) + \int_{r}^{\beta} (1 - H\,(x))\,dx\ ,$$

where $G(v) = F_D(v)^2$ and $H(v) = F_D(v)^2 + 2(1 - F_D(v))F_D(v)$ denote the cumulative distributions of the first and second order statistics, respectively, of bidders' valuations (Neeman 2003 *GEB*).

Algebra shows that for reserve price $r \geq \alpha$,

$$\Pi_{D\alpha\beta|r} = \frac{\alpha}{r}\frac{(\beta - r)(r(\beta - \alpha) + \beta(r - \alpha))}{(\alpha - \beta)^2} + \alpha^2 \frac{\int_r^\beta \frac{(v-\beta)^2}{v^2} dv}{(\alpha - \beta)^2},$$

and for reserve price $r \in [0, \alpha]$, $b^E_{D|r}(v)$ for $r = \alpha$ remains an equilibrium, so the seller's expected revenue is the same as when $r = \alpha$.

Thus, because bidders' virtual valuations are nonnegative, Myerson's analysis implies that any $r \in [0, \alpha]$ maximizes the seller's expected revenue. It follows that the expected revenue to the seller under the optimal auction is:

$$\Pi_{D\,\alpha\beta\,|\,r\,=\,\alpha} = \alpha + \alpha^2 \frac{\int_{\alpha}^{\beta} \frac{(\beta - v)^2}{v^2}\, dv}{(\beta - \alpha)^2}.$$

Note that $\Pi_{D\alpha\beta|r=\alpha}$ is bounded from above by $\frac{2\alpha\beta}{\beta - \alpha} \approx 2\alpha$ for $\beta >> \alpha$.

This fact is used below to compare the seller's equilibrium expected revenue to his level-*k* expected revenue.

Level-*k* Analysis of Optimal Reserves in First-Price Auctions

Recall that in a level-*k* model, bidders are drawn from a distribution of types. Type *Lk* anchors its beliefs in an *L0* type and adjusts them via iterated best responses: *L1* best responds to *L0*, *L2* to *L1*, and so on.

To complete the specification, define the *L0* type following Crawford and Iriberri 2007 *Econometrica*: either a "random" *L0* that bids uniformly over the natural range of bids (as in most previous level-*k* analyses) or a "truthful" *L0* that bids its private value (following Crawford 2003 *AER*).

Call the associated *L1*s or *L2*s "random" or "truthful" *L1*s or *L2*s.

Crawford and Iriberri estimated large frequencies (59-65%) of random *L1* bidders and much smaller but significant frequencies of random *L2* (4-9%), truthful *L1* (9-18%), and truthful *L2* (1-16%).

In this analysis, unlike Crawford and Iriberri's 2007 *Econometrica* analysis, reserve prices r above the lowest possible value are potentially important.

This creates an ambiguity regarding the "natural range of bids," which for the I example could be either [0,1] or (truncating the value distribution) [r,1] and for the D example could be either [α, β] or [r, β].

Focus on the latter specifications, which experiments suggest are more descriptive of most subjects' bidding behavior.

Given this specification, for the I example all types decline to bid (or, equivalently, bid less than r) when $v < r$.

A random *L1*'s bid is given by

$$b_{I|r}^{1R}(v) \in \arg\max_{b \geq r} (v - b)\frac{b - r}{1 - r} = \frac{r + v}{2}.$$

A truthful *L1*'s bid is given by

$$b_{I|r}^{1T}(v) \in \arg\max_{b \geq r} (v - b)b^{\gamma} = \max\{ r, \frac{\gamma}{\gamma + 1} v\}.$$

A random *L2*'s bid is given by

$$b_{I|r}^{2R}(v) \in \arg\max_{b \geq r} (v - b)(2b - r)^{\gamma} = \max\{r, \frac{r}{\gamma + 1} + \frac{\gamma - 1}{\gamma + 1} v\}.$$

A truthful *L2*'s bid is the same as a truthful *L1*'s

$$b_{I|r}^{2T}(v) = \arg\max_{b \geq r} (v - b)(\frac{\gamma + 1}{\gamma})^{\gamma} b^{\gamma} = \max\{ r, \frac{\gamma}{\gamma + 1} v\}.$$

Given the above type specification, for the D example all types decline to bid (or, equivalently, bid less than *r*) when $v < r$.

A random *L1*'s bid is given by

$$b_{D|r}^{1R}(v) \in \arg\max_{r \le b \le \beta} (v-b)\frac{b-\max\{r,\alpha\}}{\beta-\max\{r,\alpha\}} = \frac{v+\max\{r,\alpha\}}{2}$$

if $v \ge \max\{r, \alpha\}$ and declines to bid otherwise.

A truthful *L1*'s bid is given by

$$b_{D|r}^{1T}(v) \in \arg\max_{r \le b \le \beta} (v-b)F(b) = \max\{r, (\alpha v)^{1/2}\}$$

if $v \ge r$ and declines to bid otherwise.

A random *L2*'s bid is given by

$$b_{D|r}^{2R}(v) \in \operatorname{argmax}_{r \le b \le \beta}(v-b)F((b_{Dr}^{1R})^{-1}(b)) = \operatorname{argmax}(v-b)F(\max\{r, 2b - \max\{r,\alpha\}\}.$$

Random *L2* believes it wins if it bids $b > \max\{r, [v+\max\{r,\alpha\}]/2\}$ if and only if the other bidder's value is less than $\max\{r, 2b - \max\{r,\alpha\}$.

If $r<b<\max\{r,\alpha\}]/2$, the probability of winning is independent of b and $b = r$ is optimal. Thus we need to compare $b = r$ with the $b > \max\{r,\alpha\}/2$.

Writing the first-order condition in the latter case and simplifying yields

$$b_{D|r}^{2R}(v) = \frac{\max\{r, \alpha\}}{2} + \frac{\alpha}{2}\left[\frac{2v - \max\{r, \alpha\}}{\alpha}\right]^{\frac{1}{2}}.$$

Thus, this expression is optimal whenever it is both larger than *r* and yields higher expected payoff than bidding *r*.

Finally, a truthful *L2*'s bid is given by

$$b_{D|r}^{2T}(v) = \arg\max_{r \le b \le \beta} (v - b) F\left(\left(b_{Dr}^{1T} \right)^{-1}(b) \right).$$

Truthful *L2* believes it wins if it bids $b > \max\{r, (\alpha v)^{1/2}\}$ if and only if the other bidder's value is less than b^2/α. Plugging this in, we get

$$b_{D|r}^{2T}(v) = \arg\max_{r \le b \le \beta} (v - b)\left[\beta/(\beta - \alpha)\right]\left[1 - \alpha^2 / b^2\right].$$

Writing the first-order condition and simplifying yields

$$b_{D|r}^{2T}(v) = \left[(\alpha^4 v^2 + \alpha^6/27)^{1/2} + \alpha^2 v\right]^{1/3} - \alpha^2/3\left[(\alpha^4 v^2 + \alpha^6/27)^{1/2} + \alpha^2 v\right]^{1/3},$$

which is optimal whenever it is larger than *r*; otherwise truthful *L2*'s bid equals *r*. Like truthful *L1*'s, truthful *L2*'s bid is independent of *r* when it is above *r*.

Note that for large values of *v*, $b_{D|r}^{1R}(v)$ is approximately linear in *v*, $b_{D|r}^{1T}(v)$ and $b_{D|r}^{2R}(v)$ are proportional to $v^{½}$, and $b_{D|r}^{2T}(v)$ is proportional to $v^{1/3}$.

By contrast, $b_{D|r}^{E}(v)$ is proportional to ln *v*, which is much smaller for large values of *v*. For example, if $r = \alpha = 1$, then $b_{D|r}^{1R}(1{,}000{,}000) = 500{,}000.5$ and $b_{D|r}^{1T}(1{,}000{,}000) = 1000$, much larger than $b_{D|r}^{E}(1{,}000{,}000) = 13.82$.

Given that in the D example, level-*k* bidders bid more aggressively than equilibrium bidders, a designer facing a known distribution of level-*k* bidders should be able to realize more expected revenue than is possible with equilibrium bidders.

One can derive a lower bound on this revenue by calculating it for two random *L1* bidders and reserve price $r = 0$ and then multiplying by the probability that two such bidders are drawn.

(It is feasible for the designer to design optimally for this contingency ignoring all others, and the optimal design can do no worse.)

When $G(v) = F_D(v)^2$ as above, the expected revenue from two random *L1* bidders and reserve price $r = 0$ is

$$\int_{\alpha}^{\beta} \frac{v}{2} dG(v) = \frac{\alpha\beta^2}{(\beta - \alpha)^2}\left(\ln \beta - \ln \alpha + \frac{\alpha}{\beta} - 1\right)$$

Notably, this expression $\to \infty$ (albeit slowly, like $\ln \beta$) as α is held fixed and $\beta \to \infty$.

If $\alpha = 1$ and $\beta = 2000$, the value of this expression is approximately 7.5, almost four times as large as the 2α that approximates the seller's expected revenue with equilibrium bidders.

Since the probability of drawing two random *L1*s is about ¼, it is clear that in at least some cases, the seller can realize more expected revenue than with equilibrium bidders.

As noted above, for the D example with equilibrium bidders, a second-price auction (or equivalently an English auction) with reserve $r \leq \alpha$ is optimal.

But because (with independent private values) a second-price auction makes the equilibrium bid a dominant strategy, level-*k* bids coincide with equilibrium bids, hence a second-price auction yields only the equilibrium expected revenue.

Thus the analysis shows that in the D example with level-*k* bidders, a first-price auction with suitable reserve yields higher expected revenue than the best second-price auction.

Trivially and unsurprisingly, it also shows that revenue-equivalence breaks down.

Finally, there are examples where the optimal reserve is large with equilibrium bidders but small with level-*k* bidders, and vice versa.

Interesting open questions are when a reserve induces more aggressive bidding for equilibrium than level-*k* bidders, and the extent to which this makes optimal level-*k* reserves higher than optimal equilibrium reserves.

Exotic Auctions That Exploit Level-*k* Bidders' Non-Equilibrium Beliefs

I now give an example in a slightly different environment that illustrates the fact that a designer can exploit level-*k* bidders' non-equilibrium beliefs to obtain very large expected revenues.

As before, consider a single-object auction with two risk-neutral bidders whose values are independently and identically distributed.

But now suppose that the values are uniformly distributed on the unit interval.

The maximum expected surplus (ignoring incentive constraints) for this environment is $E[\max\{v_1,v_2\} = 2/3$. Myerson showed that a second-price auction with reserve 0.5 is optimal, and such an auction can be shown to yield expected revenue 0.41667.

Consider the following exotic auction:

- Bidders submit simultaneous sealed bids b_1, b_2 ε [0,1].
- A bidder who bids 1 wins the object if the other bids less than 1.
- If both bid 1, the winning bidder is chosen randomly.
- A bidder who bids 1 pays 0.5 if the other bidder bids less than 1, and pays $M > 1$ if the other bidder bids 1.
- A bidder who bids less than 1 pays nothing, but cannot win the object.

For this auction, truthful and random *L0* bid uniformly randomly on the unit interval.

As a result, truthful and random *L1*, defined as above for an auction with no reserve, both believe that if they bid 1 they will win the object and pay 0.5. Because truthful and random *L0* bid uniformly randomly, truthful and random *L1* assign zero prior probability to the possibility that they will have to pay *M*.

Consequently, truthful and random *L1* are willing to participate, and both bid 1 if their value is 0.5 or higher, and 0 otherwise.

Given this behavior, both truthful and random *L2* believe that the other bidder will bid 1 with prior probability 0.5 and 0 otherwise.

Truthful and random *L2* therefore expect that bidding 1 will result in their winning the object with probability 0.75, paying 0.5 with probability 0.5, and paying $M > 1$ with probability 0.5. Hence they decline to bid (or equivalently, bid less than 1).

Truthful and random *L3*, *L5*, …, behave the same as truthful and random *L1*.

Truthful and random *L4*, *L6*, …, behave the same as truthful and random *L2*.

Thus when bidders both have *k* odd the seller's expected revenue is *M*, and when bidders both have *k* even it is 0.

Even a designer who does not know the type distribution can obtain very large expected revenue by setting *M* very large to exploit level-*k* bidders' non-equilibrium beliefs.

General Observations on Level-*k* Auction Design

This formulation of the design problem takes the level-*k* model's specification as given, independent of the auction design, just as the standard formulation assumes bidders will play an equilibrium for any design.

The specification is based on substantial experimental evidence and is general enough to apply to any game; but there is reason to doubt the exogeneity assumption, particularly for exotic auctions, which go beyond the evidence on which our specification is based.

For example, bidders might view the above exotic auction as having as reserve of 1, in which case the most natural *L0* specification (random or truthful) has a spike at 1, with the result that a value of *M* high enough to be profitable would make even truthful and random *L1* bidders decline to bid.

More generally, a level-*k* model can be expected to describe bidders' behavior for a reasonably wide class of standard auction designs (Crawford and Iriberri 2007 *Econometrica*), but the model's descriptive accuracy has not been evaluated for non-standard auctions.

An auction design that is optimal for a level-*k* specification when it is assumed to be independent of the design might not be an auction for which the level-*k* model describes behavior well.

A general formulation of the design problem must take a position on how the design influences the rules that describe bidders' behavior and develop new methods to deal mathematically with that influence.

It is therefore important to learn more about the link between auction design and behavior.

Even without influences like those just discussed, the heterogeneity of level-*k* beliefs and behavior greatly complicates the characterization of optimal auctions.

In the standard analysis there is no loss of generality in using the revelation principle to restrict attention to direct mechanisms because, if equilibrium is assumed (with a selection rule in case of multiple equilibria), a bidder's private value is all that is needed to predict his behavior.

Given the restriction to direct mechanisms, the design problem is well-behaved enough that it is guaranteed to have a solution.

The above exotic example shows that this is no longer the case with level-*k* bidders, even if their level-*k* types are all the same, and even if this is known to the designer.

With a heterogeneous population of types, even if it is known to the designer, the problem becomes more complex.

Bidders with the same private values but different level-*k* types have different beliefs and will generally behave differently.

Our preliminary analyses suggest that Myerson's methods can be used to characterize an optimal auction if the designer knows that the population is homogeneous, and knows its type; and if the class of possible designs is restricted to rule out those that are too exotic for an optimal auction to exist.

But if the population is heterogeneous the problem becomes multidimensional, and much more difficult; and the high-dimensional reporting mechanisms one would consider for this case complicate the specification of *L0* and the influence of design on behavior.

Behaviorally optimal auction design poses interesting challenges, and meeting them should increase the practical usefulness of design.