Strategic Sophistication: An Experimental Test

Sotiris Georganas (OSU)
P.J. Healy (OSU)
Roberto Weber (CMU)

April 22nd, 2009
Harvard Business School

Two Warnings

1. Very much a work in progress
   - Feedback ~~appreciated~~ encouraged
2. Potentially boring background information

Elevator Pitch

1. Deeper exploration of the 'Level-k' model
2. Use a novel game to identify levels
3. Make 'small' changes to the game to see how levels change
4. Find: stable for game changes, not for new types of games
5. Rules of thumb, but not best-response levels
6. Takeaway: Call for '2nd-generation' model

Origin of this Paper

Tucson, Arizona
October 20, 2007

Behavioral Game Theory

- Goal: descriptive model of play in games
- Experiments ⇒ Nash equilibrium fails
- But wait... do we care?

"...it is crucial that the social scientist recognize that game theory is not <u>descriptive</u>, but rather (conditionally) <u>normative</u>. It states neither how people do behave nor how they should behave in an absolute sense, but how they should behave if they wish to achieve certain ends."

"Every indication now is that... the game model will have to be made more complicated if experimental data are to be handled adequately."

–Luce & Raiffa 1957

Changing The Game

On the Failure of Nash Equilibrium

1. Learning
   - Theory: Evolutionary, reinforcement, belief, hybrid (EWA), imitation...
   - Evidence: Horse races, statistical biases, overfitting?
2. Other-Regarding Preferences
   - Evidence: Dictator, ultimatum, PD, public goods games
   - Theory: Altruism, warm-glow, inequality-aversion, intentions-based models...
3. One-Shot Play of 'Well-Specified' Games
   - Theory: QRE or Cognitive Hierarchy
   - Evidence: Mixed?

QRE vs. Cognitive Hierarchy

	Accurate Beliefs	Best Response
Nash Equilibrium	✓	✓
Nash Equilibrium + Noise	✓*	✓*
QRE	✓	X
Cognitive Hierarchy	X	✓
HQRE	X	X

Quantal Response Equilibrium (QRE)

McKelvey & Palfrey 1995

- All players play a mixed strategy σ_i with full support
- Logit version:

$$\sigma_i(a_j) = \frac{\exp(\lambda E u_i(a_j, \sigma_{-i}))}{\sum_k \exp(\lambda E u(a_k, \sigma_{-i}))}$$

- Higher-payoff mistakes are more likely
- $\lambda \to \infty \Rightarrow$ Nash equilibrium
- Consistent beliefs, but not best response.

Cognitive Hierarchy/Level-k

Basic Version:

- Level 0 plays uniformly
  - Or: focal strategy, mix over undominated strategies,...
  - May or may not exist
- For all $k \in \{1, 2, \ldots\}$, Level k best responds to Level $k-1$.

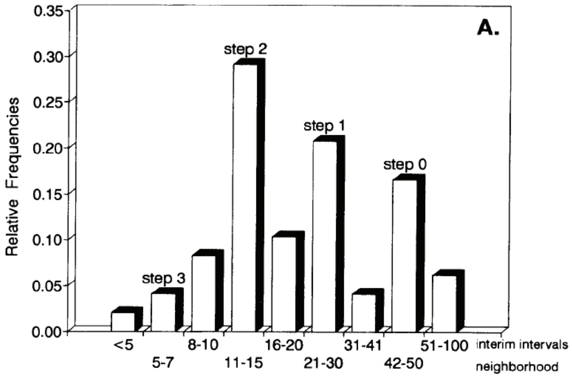
  Level k: Estimate frequencies of each level

  Cog. Hi.: Level distribution is Poisson, estimate λ
- Other types might be thrown in...
  - Nash type
  - Sophisticated type
  - Altruist type
  - Whatever-Fits-The-Data type :)

Nagel 1995 *AER*

Beauty contest game: $n > 2$ players pick $x_i \in [0, 100]$.
Winner is closest to $p * \sum_i x_i / n$ (typically $p < 1$).

- Level k: $50p^k$
- For $p = 1/2$: $\{50, 25, 12.5, 6.25, \ldots\} \to 0$



Stahl & Wilson 1995 *GEB*

- 48 subjects play twelve 3×3 games
- No feedback, can 'go back through' the games
- Estimate 1 type/person

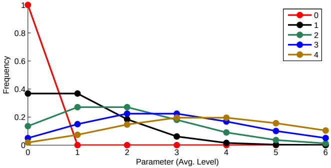
Type	Number
Level 0	6
Level 1	9
Level 2	1
Nash	5
Worldly*	17
Unclassified	10

- 44/48 subjs have likelihood ratio > 2.4

Camerer, Ho & Chong 2004 *QJE*

Proposes Cog.Hi. w/ Poisson specification

subject pool or game	source[1]	group size	data: mean	data: std dev	data: mode	fit of CH model: τ	fit of CH model: mean	fit of CH model: error	fit of CH model: std dev	fit of CH model: mode
PCC	CHC (new)	2	54.2	29.2	50	0.00	49.5	-4.7	29.5	0
p=0.9	HCW (98)	7	49.4	24.3	50	0.10	49.5	0.0	27.7	45
PCC	CHC (new)	3	47.5	29.0	50	0.10	47.5	0.0	28.6	26
Caltech board	Camerer	73	42.6	23.4	33	0.50	43.1	0.4	24.3	34
p=0.7	HCW (98)	7	38.9	24.7	35	1.00	38.8	-0.2	19.8	35
CEOs	Camerer	20	37.9	18.8	33	1.00	37.7	-0.1	20.2	34
German students	Nagel (95)	14-16	37.2	20.0	25	1.10	36.9	-0.2	19.4	34
70 yr olds	Kovalchik	33	37.0	17.5	27	1.10	36.9	-0.1	19.4	34
US high school	Camerer	20-32	32.5	18.6	33	1.60	32.7	0.2	16.3	34
econ PhDs	Camerer	16	27.4	18.7	N/A	2.30	27.5	0.0	13.1	21
1/2 mean	Nagel (98)	15-17	26.7	19.9	25	1.50	26.5	-0.2	19.1	25
portfolio mgrs	Camerer	26	24.3	16.1	22	2.80	24.4	0.1	11.4	26
Caltech students	Camerer	17-25	23.0	11.1	35	3.00	23.0	0.1	10.6	24
newspaper	Nagel (98)	3696, 1460, 2728	23.0	20.2	1	3.00	23.0	0.0	10.6	24
Caltech	CHC (new)	2	21.7	29.9	0	0.80	22.2	0.6	31.6	0
Caltech	CHC (new)	3	21.5	25.7	0	1.80	21.5	0.1	18.6	26
game theorists	Nagel (98)	27-54	19.1	21.8	0	3.70	19.1	0.0	9.2	16
					mean	1.30				
					median	1.61				



The Crawford Papers

- Costa-Gomes, Crawford & Broseta 2001 *Emtca*
  - 18 normal-form games w/ lookups (9 types!!)
- Costa-Gomes & Crawford 2006 *AER*
  - 2-person guessing games w/ lookups
- Crawford & Iriberri 2007 *Emtca*
  - auctions
- Crawford & Iriberri 2007 *AER*
  - hide-and-seek games
- Crawford, Kugler, Neeman & Pauzner 2009 *JEEA*
  - mechanism design example
- Costa-Gomes, Crawford & Iriberri 2009 *JEEA*
  - analysis of old coordination game data

Back to Tucson...

Tucson, Arizona
October 20, 2007

Generality vs. Usefulness

1. Level-k is at least as good as QRE and NE
2. Estimating dist'n of types vs. Poisson specification
3. Most action is on L1, L2, and maybe L3. (Ignore dominance.)
4. Model needs to be adapted to the setting
   - L0 may be uniform, focal, anti-focal,...

"The applications illustrate the generality of the level-k approach and the kinds of adaptations needed to use it in different settings."

–Crawford, Oct 20, 2007

Crawford & Iriberri 2007 *AER*

Hide-and-Seek:

A	B	A	A

Crawford & Iriberri 2007 *AER*

Hide-and-Seek:

	A	B	A	A
Hiders:	16%	18%	44%	22%
Seekers:	16%	19%	54%	11%

- Hiders found 32.2% of the time! Overthinking it??
- Level-k story:
  1. **Suppose L0 favors B and 'outside' A's...**
  2. L1: Hide at central A, Seek away from central A
  3. L2: Hide at central A, Seek at central A
  4. L3: Hide away from central A, Seek at central A
- Result: Lot of action at central A, more hiders than seekers.
- But... other models fit better. **Non**-focal L0 does better.

Our Reactions

1. Lack of portability $\Rightarrow$ low predictive power
   - Too slippery... the model is *supposed* to bend
   - Maybe it's not *supposed* to be predictive...
2. The underlying cognitive model *must* be too simple...

The Cognitive Model

- What is the underlying invariant 'type' of a subject?
    - Not really made explicit
    - Clearly not meant to be an *as-if* model
- Strict interpretation: type = level
- Our interpretation:
    - invariant type = 'strategic sophistication'
    - level = f(type,game)
    - f(type,game) may not equal f(type,game')
    - 'game' may even include identities or properties of opponents
    - $\text{type}_i > \text{type}_j \Rightarrow \text{level}_i > \text{level}_j$ for all games
- Strict interpretation $\subset$ Our interpretation

Our Interpretation

- Hypotheses:
  1. *Uniform* shifts in levels between games
  2. Able to 'measure' a type with diagnostic tests
  3. Correlation between measured types and observed levels
  4. (Weakly) higher-level play against higher-type opponents
- NON-Hypotheses:
  1. *Stable* levels between games
  2. *Stable* levels across different opponents

Experiment Design Desiderata

A good experiment should use:

1. Quizzes believed to diagnose 'strategic sophistication'
2. Several 'one-shot' games
    - Multiple games with revision + no feedback
3. Symmetric games
4. No room for other-regarding preferences
    - Cost of being nice $\geq$ benefit to others
    - Zero-sum game, e.g.
    - Framing?
5. Unique action for each 'level' {L1,L2,L3,Nash}
    - But not *too* complicated!
6. Ability to play against opponents of different types
    - Use quiz scores as a proxy

The Quizzes

1. IQ test questions
2. ‘Eye Gaze’ test
3. Working Memory test
4. Cognitive Reflection Test (CRT)
5. One-Player Takeover game

IQ Test

- Controversial but possibly relevant
- *Some* correlations with success, income, patience, etc.
- Taken from Mensa exam

Examples:

1. If 2 typists can type 2 pages in 2 minutes, how many typists will it take to type 18 pages in 6 minutes?
2. Unscramble the following word: H C P R A A T E U
3. What number is missing from this pattern: 1 8 27 ? 125 216

Eye Gaze Test

☐ jealous ☐ panicked ☐ arrogant ☐ hateful

☐ aghast ☐ fantasizing ☐ impatient ☐ alarmed

Eye Gaze Test

- Developed by Simon Baron-Cohen
- Used to identify failures of theory-of-mind
    - Diagnostic of *adult* autism/Asperger's Syndrome
- Higher level $\Leftrightarrow$ better-developed theory of mind?
- Bruguier, Quartz & Bossaerts 2008
    - Higher score $\Rightarrow$ can identify insider traders from prices
    - (ToM areas of the brain light up when watching trades)

Working Memory Test

Movie3 Movie4 Movie5 Movie6 Movie7 Movie8 Movie9

- Wechsler digit span test (Walsch & Betz 1990)
- Devetag & Warglien working paper
  - Dom. solvable game, dirty faces game, centipede-like game
  - Play against ‘rational’ computer opponent (Nash)
  - Total score on games correlates with working memory

Cognitive Reflection Test

1. A bat and a ball cost \$1.10. The bat costs \$1.00 more than the ball. How much does the ball cost?
2. If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
3. A patch of lily pads doubles in size each day. If it takes 48 days to cover the lake, how many days would it take to cover half the lake?

CRT

Shane Frederick survey 2005

- Correlates with time preference
- Correlates with risk taking in gains...
- but risk aversion in losses
- Weak correlation with IQ measures
- Men score higher than women, controlling for SAT

One-Player Takeover Game

1. You are bidding to buy a company. To the seller it is worth $X \in \{\$0, \$30, \$60, \$90, \$120\}$, each with equal probability. The seller knows X but you do not. You are the only bidder and the company is worth $1.5X$ to you. The seller will only sell if you bid more than X. How much should you bid?

- Samuelson & Bazerman 1985
- Ball, Bazerman & Carroll 1990 *OBHDP*

Our New Game

"The undercutting game" (UG)

- Discrete version of beauty contest-type games
- 'Nearly' zero-sum
- 4 'serious' actions, one for each level
- 3 dominated actions

Variations

UG1: Baseline (7 actions)

UG2: Add 2 dominated actions (9 actions)

UG3: Add 2 'serious' actions (9 actions)

UG4: Compressed UG2 (7 actions)

Undercutting Game 1

	1	2	3	4	5	6	7
1	1 1	10 -10	0 0	0 0	0 0	0 0	-11 0
2	-10 10	0 0	10 -10	0 0	0 0	0 0	0 0
3	0 0	-10 10	0 0	10 -10	0 0	0 0	0 0
4	0 0	0 0	-10 10	0 0	10 -10	10 -10	10 -10
5	0 0	0 0	0 0	-10 10	0 0	0 0	0 0
6	0 0	0 0	0 0	-10 10	0 0	0 0	0 0
7	0 -11	0 0	0 0	-10 10	0 0	0 0	-11 -11

L1=4 L2=3 L3=2 Nash=1

Undercutting Game 2

	1	2	3	4	5	6	7	8	9
1	1 1	10 -10	0 0	0 0	0 0	0 0	0 0	0 0	-11 0
2	-10 10	0 0	10 -10	0 0	0 0	0 0	0 0	0 0	0 0
3	0 0	-10 10	0 0	10 -10	0 0	0 0	0 0	0 0	0 0
4	0 0	0 0	-10 10	0 0	10 -10	10 -10	10 -10	10 -10	10 -10
5	0 0	0 0	0 0	-10 10	0 0	0 0	0 0	0 0	0 0
6	0 0	0 0	0 0	-10 10	0 0	0 0	0 0	0 0	0 0
7	0 0	0 0	0 0	-10 10	0 0	0 0	0 0	0 0	0 0
8	0 0	0 0	0 0	-10 10	0 0	0 0	0 0	0 0	0 0
9	0 -11	0 0	0 0	-10 10	0 0	0 0	0 0	0 0	-11 -11

L1 (action 4) is more tempting

Undercutting Game 3

	1	2	3	4	5	6	7	8	9
1	1, 1	10, -10	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0	-11, 0
2	-10, 10	0, 0	10, -10	0, 0	0, 0	0, 0	0, 0	0, 0	0, 0
3	0, 0	-10, 10	0, 0	10, -10	0, 0	0, 0	0, 0	0, 0	0, 0
4	0, 0	0, 0	-10, 10	0, 0	10, -10	0, 0	0, 0	0, 0	0, 0
5	0, 0	0, 0	0, 0	-10, 10	0, 0	10, -10	0, 0	0, 0	0, 0
6	0, 0	0, 0	0, 0	0, 0	-10, 10	0, 0	10, -10	10, -10	10, -10
7	0, 0	0, 0	0, 0	0, 0	0, 0	-10, 10	0, 0	0, 0	0, 0
8	0, 0	0, 0	0, 0	0, 0	0, 0	-10, 10	0, 0	0, 0	0, 0
9	0, -11	0, 0	0, 0	0, 0	0, 0	-10, 10	0, 0	0, 0	-11, -11

Higher levels available

Undercutting Game 4

	1	2	3	4	5	6	7
1	1 1	10 -10	0 0	0 0	0 0	0 0	-11 0
2	-10 10	0 0	10 -10	0 0	0 0	0 0	0 0
3	0 0	-10 10	0 0	10 -10	0 0	0 0	0 0
4	0 0	0 0	-10 10	0 0	30 -30	10 -10	10 -10
5	0 0	0 0	0 0	-30 30	0 0	0 0	0 0
6	0 0	0 0	0 0	-10 10	0 0	0 0	0 0
7	0 -11	0 0	0 0	-10 10	0 0	0 0	-11 -11

UG2 with 3 dominated strategies ‘compressed’ into 1

Different Opponents

In each game subjects submit:

1. Choice vs. randomly-matched opponent
2. Choice vs. highest-scoring subject on quiz (other than self)
3. Choice vs. lowest-scoring subject on quiz (other than self)

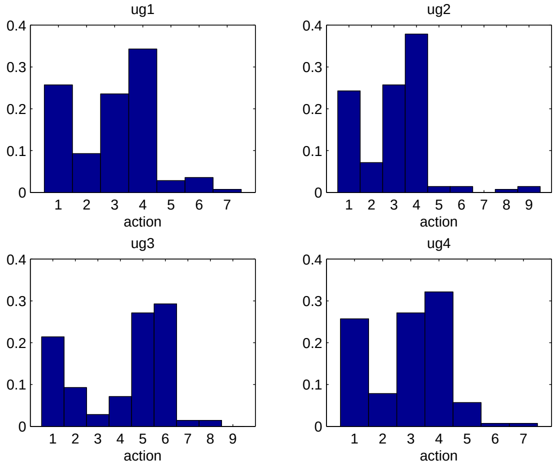
4 games, 3 choices per game.
2 choices randomly chosen for payment.

Results

Aggregate level distributions

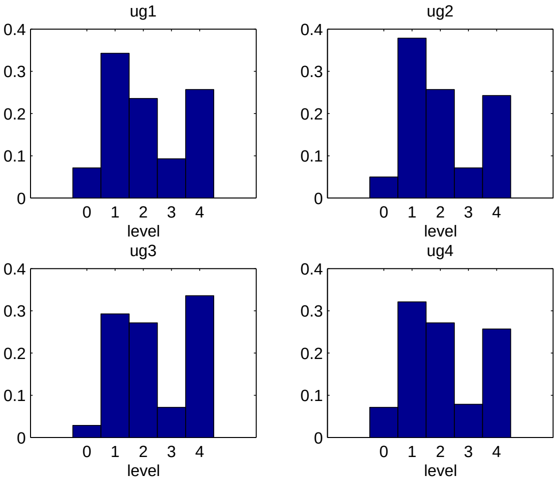
Action Distributions

Versus Random Opponent:



Level Distributions

Versus Random Opponent:

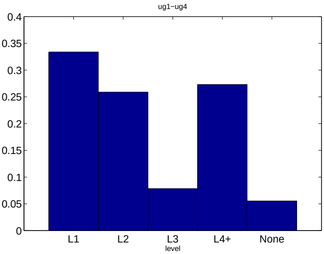


Level Distributions

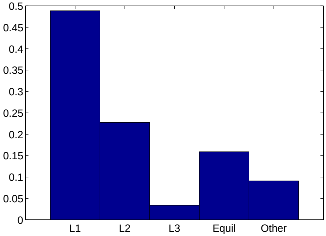
	Dominated	L1	L2	L3	L4-Nash
UG1	7.1%	34.3%	23.6%	9.3%	25.7%
UG2	5.0%	37.9%	25.7%	7.1%	24.3%
UG3	2.9%	29.3%	27.1%	7.1%	33.6%
UG4	7.1%	32.1%	27.1%	7.9%	25.7%
ALL	5.5%	33.4%	25.9%	7.9%	27.3%

K-S Tests	UG1	UG2	UG3	UG4
UG1	-	1.000	0.562	1.000
UG2	-	-	0.378	1.000
UG3	-	-	-	0.764
UG4	-	-	-	-

Level Distribution vs. CGC06



Our Data



Costa-Gomes & Crawford 2006

Level-switching

Level Switching: Aggregate

Occurrences of level-switching between games:

From↓ To→	Dom.	L1	L2	L3	L4+
Dom.	30	22	12	10	19
L1	22	**352**	88	25	74
L2	12	88	**264**	39	32
L3	10	25	39	26	32
L4+	19	74	32	32	**302**

L1, L2, Nash are persistent. L3 & Dominated are not.

Level Switching: Aggregate

Markov Transition Matrix:

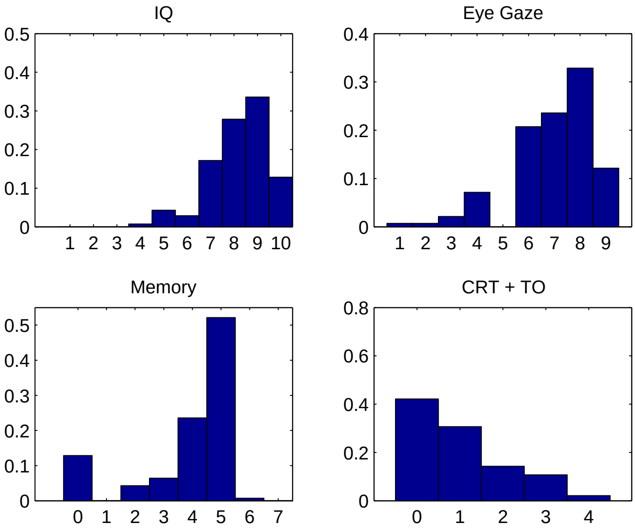
From↓ To→	Dom.	L1	L2	L3	L4+
Dom.	32.3%	23.7%	12.9%	10.8%	20.4%
L1	3.9%	62.7%	15.7%	4.5%	13.2%
L2	2.8%	20.2%	60.7%	9.0%	7.4%
L3	7.6%	18.9%	29.5%	19.7%	24.2%
L4+	4.1%	16.1%	7.0%	7.0%	65.8%

L3's randomize over undominated strategies?

Results

Quiz scores vs. levels

Quiz Scores

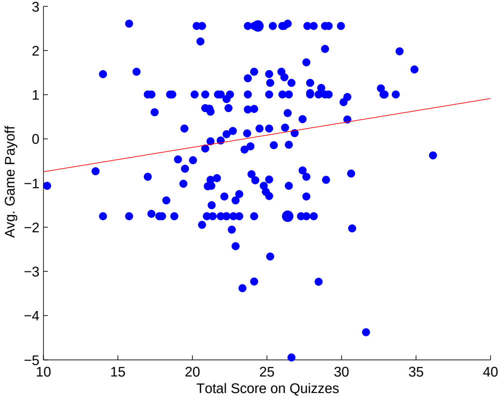

IQ
0.5
0.4
0.3
0.2
0.1
0
1 2 3 4 5 6 7 8 9 10
Eye Gaze
0.4
0.3
0.2
0.1
0
1 2 3 4 5 6 7 8 9
Memory
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7
CRT + TO
0.8
0.6
0.4
0.2
0
0 1 2 3 4

Quizzes vs. Quizzes

Correlation	IQ	Eye Gaze	Memory	CRT+TO
IQ		-0.016	**0.198**	**0.182**
Eye Gaze			0.015	-0.059
Memory				0.101
CRT+TO				

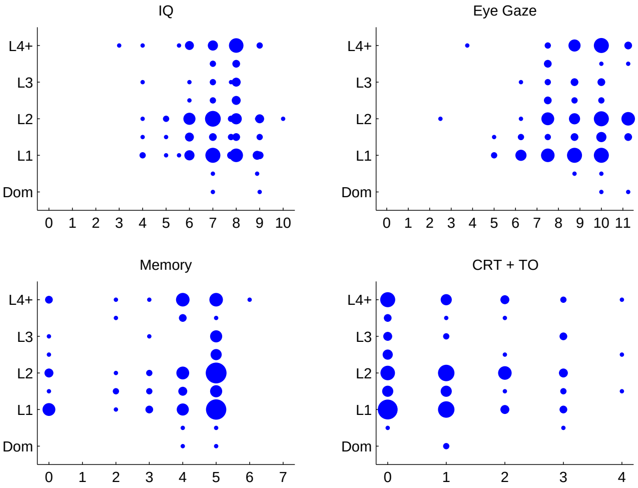
p-values	IQ	Eye Gaze	Memory	CRT+TO
IQ		0.849	**0.019**	**0.031**
Eye Gaze			0.862	0.489
Memory				0.233
CRT+TO				

Total Quiz Score vs. Earnings



Correlation: 0.154 *p*-value: 0.070

Quizzes vs. Median Levels



Quizzes vs. Median Levels

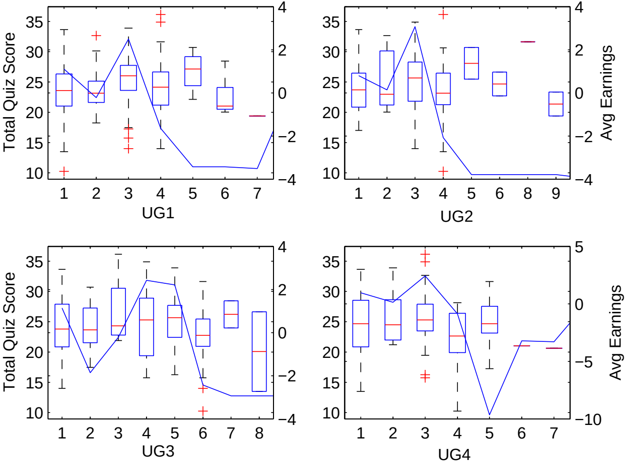
Which quizzes predict the median level?

$$\text{MedianLevel} = \beta_0 + \beta_1 S_{\text{IQ}} + \beta_2 S_{\text{EyeGaze}} + \beta_3 S_{\text{Memory}} + \beta_4 S_{\text{CRT+TO}}$$

	Const.	**IQ**	**EyeGaze**	**Memory**	**CRT+TO**
Est.	1.585	-0.057	0.086	0.021	0.018
p-**val.**	0.018	0.228	0.065	0.299	0.305

Higher Level = Smarter?

Should you play a high level?



Quizzes & Level 1

Which quizzes predict Level 1 play?

$$Pr(\mathrm{L1}_{\geq 3}) = \phi(\beta_0 + \beta_1 \mathrm{S}_{\mathrm{IQ}} + \beta_2 \mathrm{S}_{\mathrm{EyeGaze}} + \beta_3 \mathrm{S}_{\mathrm{Memory}} + \beta_4 \mathrm{S}_{\mathrm{CRT+TO}})$$

	Const.	**IQ**	**EyeGaze**	**Memory**	**CRT+TO**
Est.	1.307	0.152	**-0.276**	**-0.171**	-0.124
***p*-val.**	0.429	0.379	**0.023**	**0.036**	0.151

Quizzes & Higher Levels

But quizzes don't predict L2 or L4+...

L2	**Const.**	**IQ**	**EyeGaze**	**Memory**	**CRT+TO**
Est.	-3.169	-0.052	0.153	0.091	0.062
***p*-val.**	0.080	0.770	0.276	0.383	0.433

L4+	**Const.**	**IQ**	**EyeGaze**	**Memory**	**CRT+TO**
Est.	-2.849	-0.114	0.229	-0.006	0.018
***p*-val.**	0.121	0.526	0.138	0.949	0.837

Decision Time vs. Level

We monitor how long players take to click ‘OK’...

$$\text{Time}_{ig} = \beta_0 + \beta_1 \text{I}(L1_{ig}) + \beta_2 \text{I}(L2_{ig}) + \beta_3 \text{I}(L3_{ig}) + \beta_4 \text{I}(L4_{ig}) + \varepsilon_{ig}$$

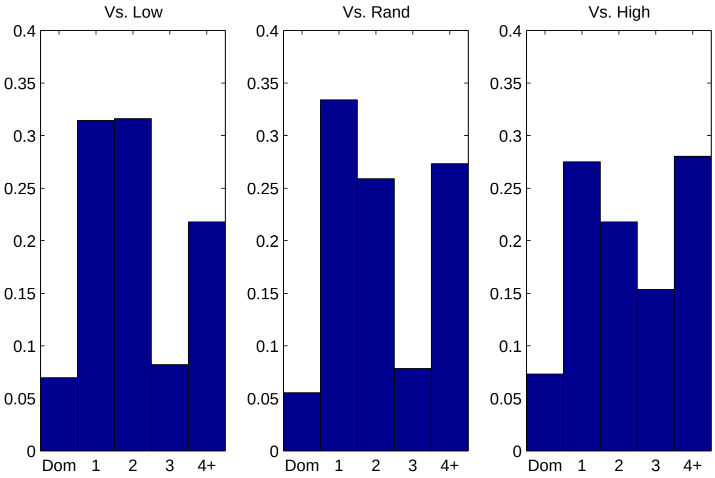
	β_0	β_1	β_2	β_3	β_4
Est.	88.86	-8.33	-1.24	-3.15	-7.82
p-val.	0.000	0.194	0.450	0.392	0.212

Note: omitted dummy is ‘dominated strategy’

Results

Levels vs. Different Opponents

Level Distributions by Opponent



Level Distributions by Opponent

K-S tests for differences of distributions:

p-Values	vs. Low	vs. Random	vs. High
vs. Low		0.348	$< \mathbf{0.001}$
vs. Random			**0.0432**
vs. High			

Players play differently against high-scoring opponents.

Moving Up or Down

When moving from Situation A to Situation B, we say you *move up* in level if

1. min level in B $\geq$ max level in A, and
2. levels in B $\neq$ levels in A.

Fraction of subjects moving up:

From ↓ To →	vs. Low	vs. Random	vs. High
vs. Low	-	11.43%	20.71%
vs. Random	7.14%	-	16.43%
vs. High	4.29%	7.14%	-

Results

Model Comparison

Model Comparisons

Estimate fit of The Big Three models:

N=560	Level-k	QRE	Nash + Noise
λ	0.47	0.49	0.084
LL	-952.39	-987.19	-1159.6
BIC	1930.092	1980.708	2325.528
L1	52%	17%	0%
L2	15%	22%	0%
L3	11%	13%	0%
Nash	22%	29%	100%

QRE consistently predicts more L2 than L1 here...

UG Conclusions

1. Level distributions are fairly stable, similar to old data
2. L1, L2, and L4/Nash are quite stable types
3. Eye Gaze + Memory quizzes may be useful in predicting types
4. Players react to expectations about opponents
    - 'Opponents' should be part of the description of the game
5. Level-k fits best of 1^{st}-generation models

A Second Family of Games

- How stable are levels moving to a ‘similar’ family of games?
- Do our results translate to other games?

Costa-Gomes & Crawford 2006 *AER*: 2-Person Guessing Games

	Lower Limit	Upper Limit	Target
Your Parameters	215	815	1.40
Other’s Parameters	0	650	0.90

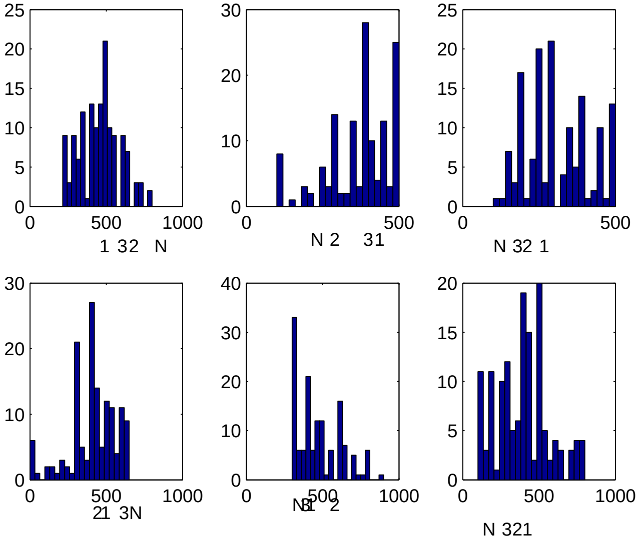
34 subjects, 6 plays (3 games, 2 roles each)
2 games from CGC06, 1 new game

2PGG

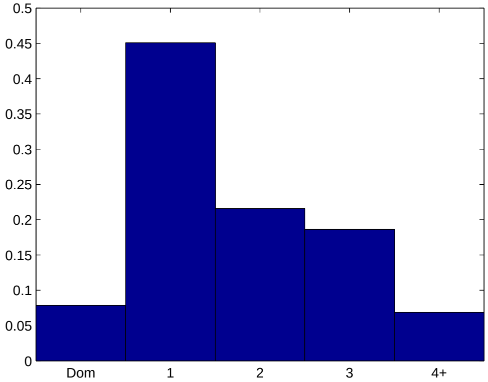
	Lower Limit	Upper Limit	Target
Your Parameters	215	815	1.40
Other's Parameters	0	650	0.90

- L0: $1.4(650-0)/2 = 455$ (Other: 463.5)
- L1: $1.4(463.5) = 649$ (Other: 409.5)
- L2: $1.4(409.5) = 573$ (Other: 584)
- L3: $1.4(584.01) = 818 \rightarrow 815$ (Other: 516)
- Nash: 815 (Other: 650)

Action Distributions



Level Distribution



Nash play and L2/L1 much lower. More L1. Similar to CGC06.

Switching

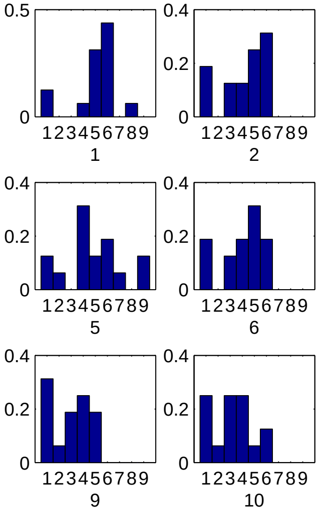
Randomly draw 2 players, one UG, and one 2PGG.

Event	Probability
Neither changes level:	4.64%
Only one changes level:	33.9%
Both change in same direction:	34.4%
Both change i opposite directions:	27.1%

Conclusions

1. Level-k model reasonably stable within very similar games
2. Across different games stability is weak
3. Definite heterogeneity
4. ‘Rules of thumb’
5. Ability to predict behavior?

Future: Dynamics



THE END