STUDYING COGNITION VIA INFORMATION SEARCH IN TWO-PERSON GUESSING GAME EXPERIMENTS

Conference on Econometrics and Experimental Economics, Northwestern University, 28 April 2006

PRELIMINARY; DO NOT CIRCULATE

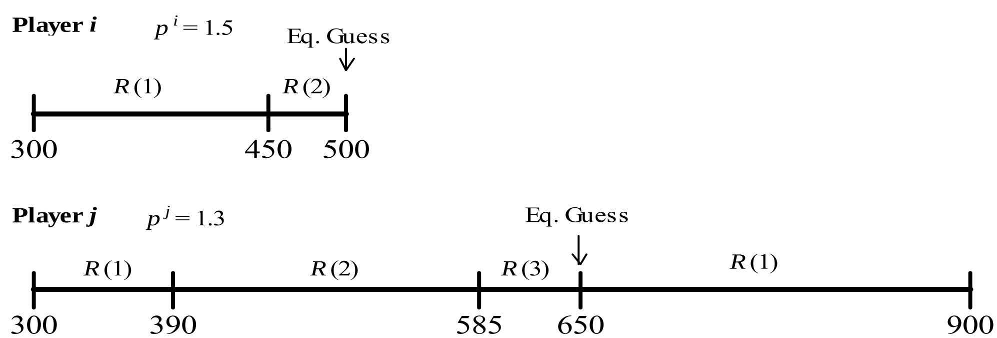
Based on Miguel Costa-Gomes, York, and Vincent Crawford, UCSD:
"Cognition and Behavior in Two-Person Guessing Games: An Experimental Study," *AER*, in press; and
"Studying Cognition via Information Search in Two-Person Guessing Game Experiments," manuscript in preparation.

TWO-PERSON GUESSING GAMES

Each player has a *lower* and *upper limit*, both strictly positive, but players are not required to guess between their limits. Guesses outside the limits are automatically adjusted up to the lower or down to the upper limit as necessary.

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

Example: *i*'s limits and target are [300, 500]; *j*'s are [300, 900] and 1.3. The equilibrium is essentially unique, with *i*'s adjusted guess at 500 and *j*'s at 650. (Guesses in *R*(*k*) are eliminated in round *k* of iterated dominance.)



Equilibrium is determined in this general way by players' targets and upper (lower) limits when the product of targets is greater (less) than one.

EXPERIMENTAL DESIGN

The focus is entirely on initial responses to games, a good place to starting studying strategic thinking: Subjects were randomly, anonymously paired to play a common series of 16 games without feedback, to suppress learning and repeated-game effects. Results should help us think about learning.

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

The games are generally asymmetric and dominance-solvable in 3 to 52 rounds, with essentially unique equilibria determined as in the example above.

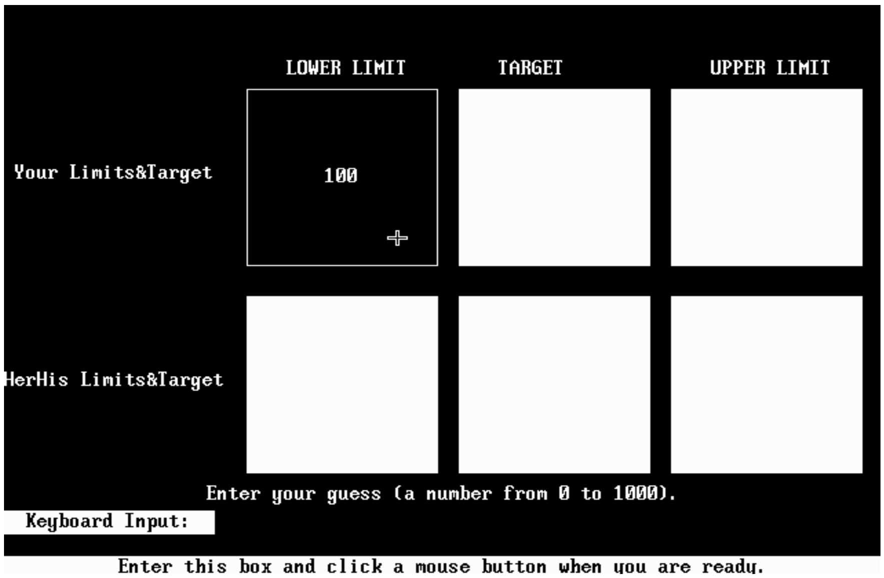
(The targets and limits in the previous guessing experiments of Nagel (1995, *AER*) and Ho, Camerer, and Weigelt (1998, *AER*) were always the same for both players, and varied either only across treatments or not at all.)

Table 3. Strategic Structures of the Games

Game *i j*	Order Played	Targets	Equilibrium	Rounds of Dominance	Pattern of Dominance	Dominance at Both Ends
1. α2β1	6	Low	Low	4	A	No
2. β1α2	15	Low	Low	3	A	No
3. β1γ2	14	Low	Low	3	A	Yes
4. γ2β1	10	Low	Low	2	A	No
5. γ4δ3	9	High	High	2	S	No
6. δ3γ4	2	High	High	3	S	Yes
7. δ3δ3	12	High	High	5	S	No
8. δ3δ3	3	High	High	5	S	No
9. β1α4	16	Mixed	Low	9	S/A	No
10. α4β1	11	Mixed	Low	10	S/A	No
11. δ2β3	4	Mixed	Low	17	S/A	No
12. β3δ2	13	Mixed	Low	18	S/A	No
13. γ2β4	8	Mixed	High	22	A	No
14. β4γ2	1	Mixed	High	23	A	Yes
15. α2α4	7	Mixed	High	52	S/A	No
16. α4α2	5	Mixed	High	51	S/A	No

MONITORING SEARCH FOR HIDDEN BUT FREELY ACCESSIBLE INFORMATION ABOUT PAYOFFS

The structure was publicly announced except for the targets and limits, to which subjects were given free access, game by game, via MouseLab:



Low search costs make the games' structures effectively public knowledge, inducing a series of 16 independent complete-information games. But varying targets and limits makes search a powerful tool for studying cognition.

Camerer, Johnson, Rymon, and Sen (1993; "CJ") and Johnson, Camerer, Rymon, and Sen (2002 *JET*) pioneered the use of MouseLab in games by studying backward induction in alternating-offers bargaining games in which subjects could look up the sizes of the “pies” to de divided in each period.

Costa-Gomes, Crawford, and Broseta (2001 *Econometrica*; "CGCB") used MouseLab to study two-person matrix games with unique equilibria in which subjects could look up the payoffs of each decision combination.

The current design combines the advantages of CJ's simple parametric structure and CGCB's high-dimensional search patterns, while making search implications of alternative decision rules almost independent of the game.

This often makes it possible to read a subject's decision rule directly from his search patterns, without even considering guesses, and makes search a powerful tool for studying cognition more generally (e.g. by identifying errors).

TYPES

The space of possible decision rules is enormous, and the search implications of a rule depend not only on what guesses you make but why you make them.

We organize the analysis around an a priori list of plausible "types," which provide a kind of basis for the space of possible rules, making it meaningful to ask if subjects' guesses and search are related in a coherent way:

L1 – best responds to uniform (between limits) random ***L0*** "anchoring type"

L2 – best responds to ***L1***

L3 – best responds to ***L2***

D1 – does one round of deletion of decisions dominated by pure decisions and then best responds to a uniform prior over other's remaining decisions

D2 – does two rounds of iterated deletion and then best responds to a uniform prior over other's remaining decisions

Equilibrium – plays its equilibrium decision

Sophisticated – best responds to the probabilities of other's decisions, estimated here from our subjects' observed frequencies (depends on data)

Our *Lk* definitions differ from Stahl and Wilson's (1995 *GEB*; "SW") and Camerer, Ho, and Chong's (2004 *QJE*; "CHC"). SW's *Lk* best responds to lower-level *Lk* types' decision noise, as in QRE. (All three allow types to make errors; the issue is whether mean decision responds to noise.) SW's and CHC's *Lk*s both best respond to an estimated mixture of lower-level *Lk* types.

Our design separates our *Lk* definitions from SW's, and our results strongly favor our "noiseless" definitions.

Our design does not separate our definitions from mixture definitions, but ours are simpler and we believe more plausible models of cognition.

Our design (unlike previous designs) strongly separates *Dk*-1 from *Lk*, which are both *k*-level rationalizable. (We show that *Lk* predominates. This suggests it's wrong to take Nagel's results—in her design *Dk*-1 and *Lk* make identical guesses—as evidence that subjects *explicitly* performed iterated dominance.)

TYPES' ADJUSTED GUESSES IN THE 16 GAMES
(games in the randomized order in which subjects played them)

Game	*ai*	*bi*	*pi*	*aj*	*bj*	*pj*	*L1*	*L2*	*L3*	*D1*	*D2*	*E*	*S*
1	100	900	1.5	300	500	0.7	600	525	630	600	611.25	750	630
2	300	900	1.3	300	500	1.5	520	650	650	617.5	650	650	650
3	300	900	1.3	300	900	1.3	780	900	900	838.5	900	900	900
4	300	900	0.7	100	900	1.3	350	546	318.5	451.5	423.15	300	420
5	100	500	1.5	100	500	0.7	450	315	472.5	337.5	341.25	500	375
6	100	500	0.7	100	900	0.5	350	105	122.5	122.5	122.5	100	122
7	100	500	0.7	100	500	1.5	210	315	220.5	227.5	227.5	350	262
8	300	500	0.7	100	900	1.5	350	420	367.5	420	420	500	420
9	300	500	1.5	300	900	1.3	500	500	500	500	500	500	500
10	300	500	0.7	100	900	0.5	350	300	300	300	300	300	300
11	100	500	1.5	100	900	0.5	500	225	375	262.5	262.5	150	300
12	300	900	1.3	300	900	1.3	780	900	900	838.5	900	900	900
13	100	900	1.3	300	900	0.7	780	455	709.8	604.5	604.5	390	695
14	100	900	0.5	300	500	0.7	200	175	150	200	150	150	162
15	100	900	0.5	100	500	0.7	150	175	100	150	100	100	132
16	100	900	0.5	100	500	1.5	150	250	112.5	162.5	131.25	100	187

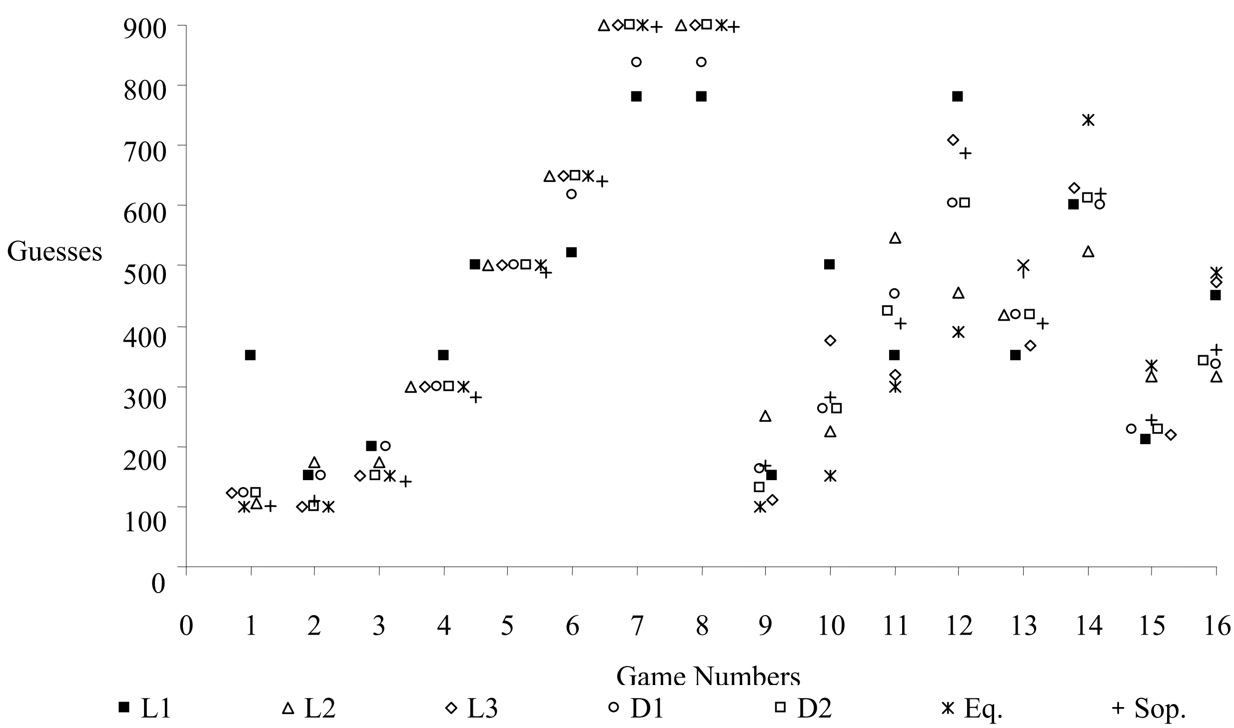


Figure 5. Separation of Types' Predicted Guesses Across Games

TREATMENTS

Baseline (B): Subjects randomly, anonymously paired to play the 16 games

Open Boxes (OB): identical to Baseline except targets and limits visible

(We find insignificant differences between B and OB guesses, suggesting that the need to look up parameters has no important effect on decisions.)

Robot/Trained Subjects (R/TS): identical to Baseline except (in six separate treatments) each subject is trained and rewarded as a specific decision rule or *type*: *L1*, *L2*, *L3*, *D1*, *D2*, or *Equilibrium* (defined below)

(R/TS treatments validate our proposed model of how cognition drives search, allow us to assess the cognitive difficulty of identifying leading types' guesses, and allow us to study how cognition varies with training in decision rules.)

SUBJECTS' GUESSES IN BASELINE AND OB TREATMENTS

43 of 88 Baseline and OB subjects made 7 to 16 of some type's exact guesses: 20 *L1*, 12 *L2*, 3 *L3*, and 8 *Equilibrium*. (No *Dk* or *Sophisticated*.)

For these 43 we can rule out alternative interpretations of behavior: The non-equilibrium behaviors of the 35 whose "fingerprints" are *Lk* are due to non-equilibrium beliefs, not irrationality, risk aversion, altruism, spite, or confusion.

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	34

: The far right-hand column includes 11 OB subjects classified by their econometric-from-guesses type estimates.

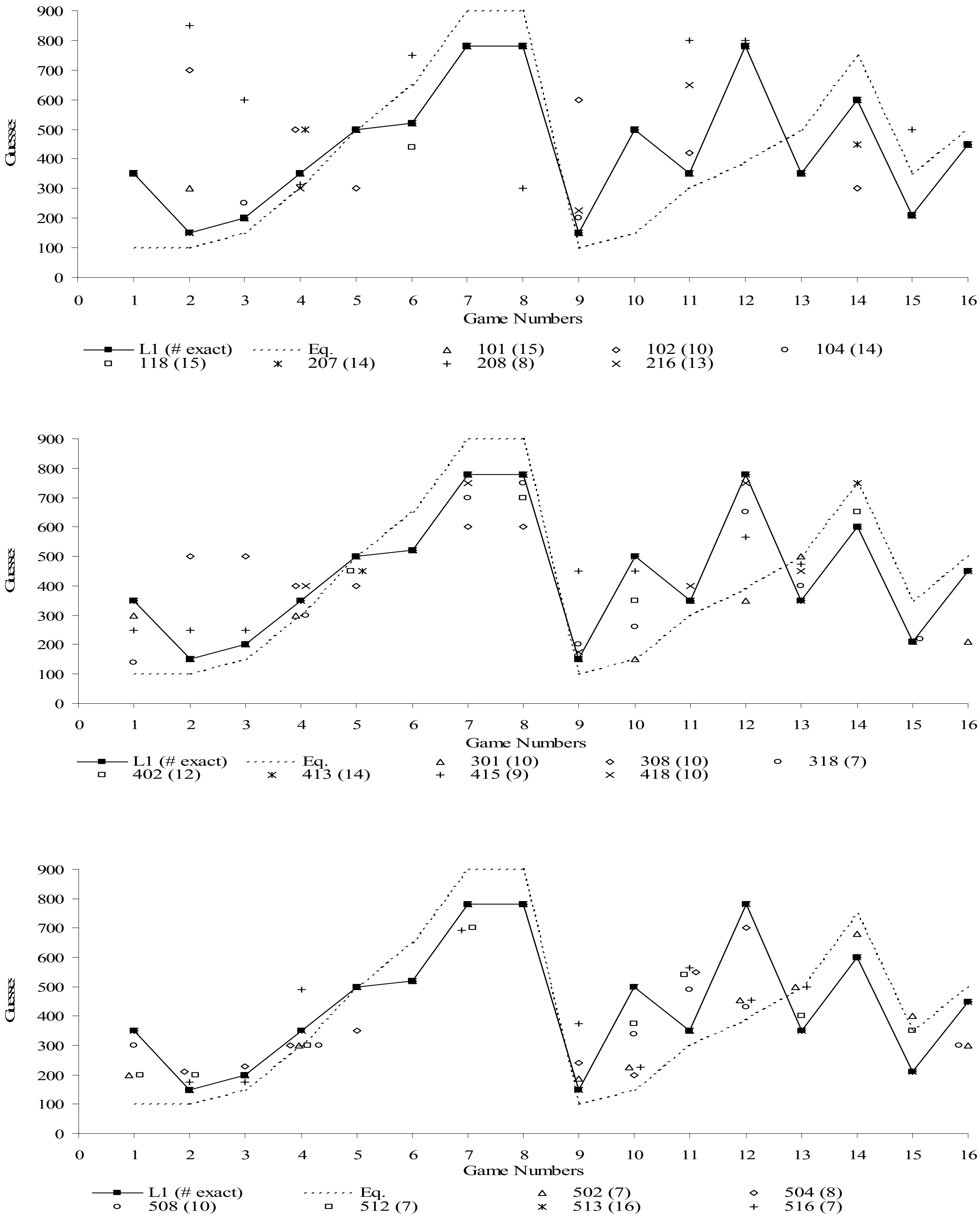
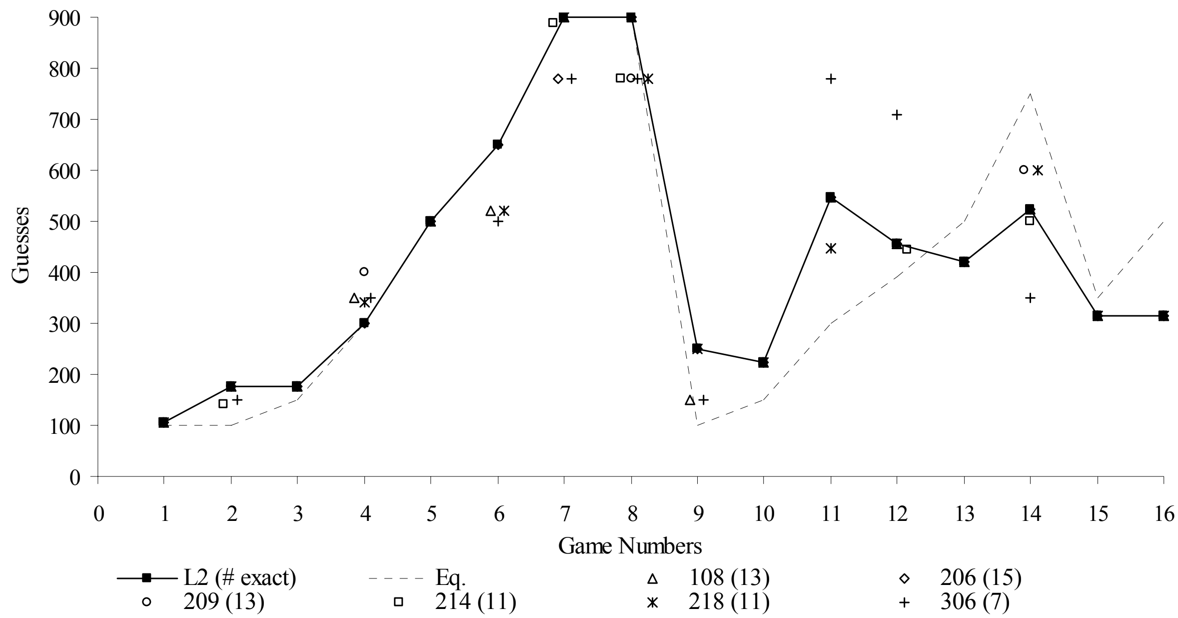


Figure 1. "Fingerprints" of 20 Apparent *L1* Subjects

(Only deviations from *L1*'s guesses are shown; games with mixed targets on right.)



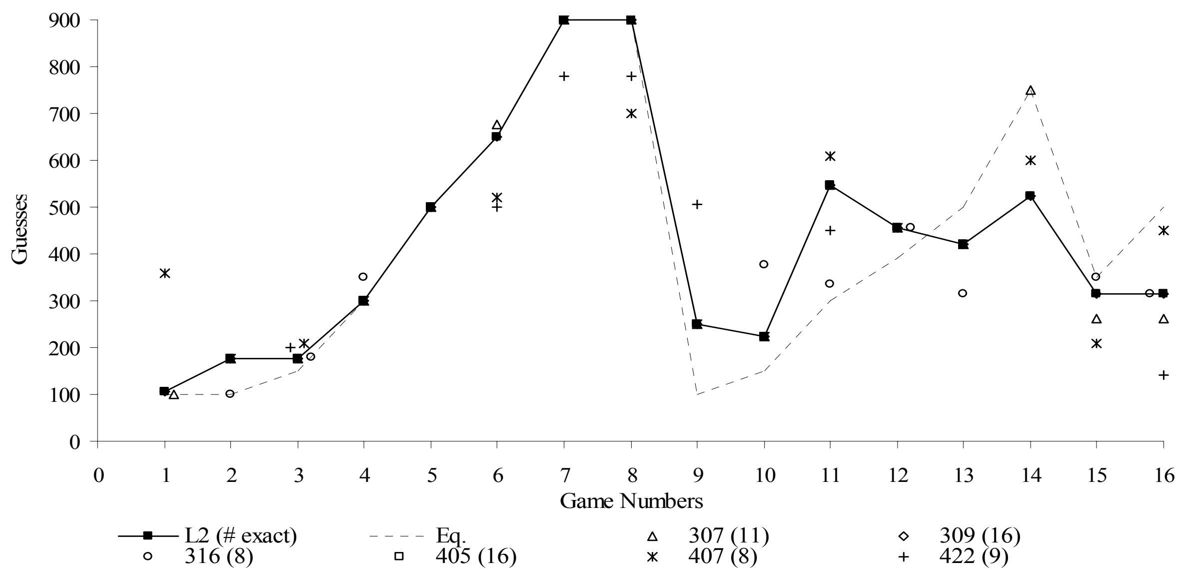


Figure 2. "Fingerprints" of 12 Apparent *L2* Subjects

(Only deviations from *L2*'s guesses are shown; mixed targets on right.)

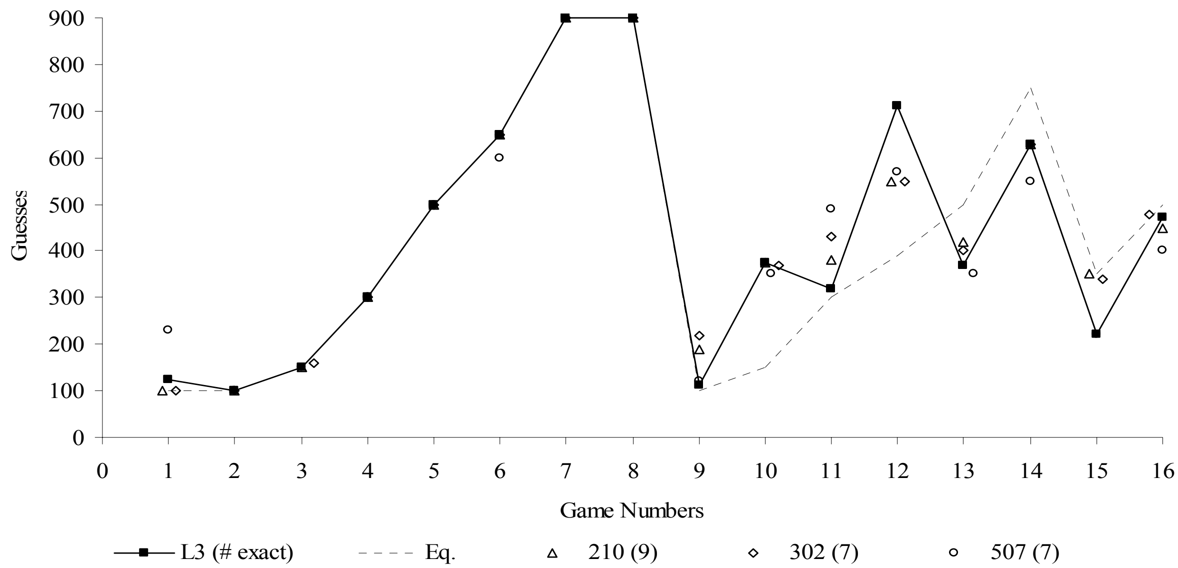


Figure 3. "Fingerprints" of 3 Apparent *L3* Subjects

(Only deviations from *L3*'s guesses are shown; mixed targets on right.)

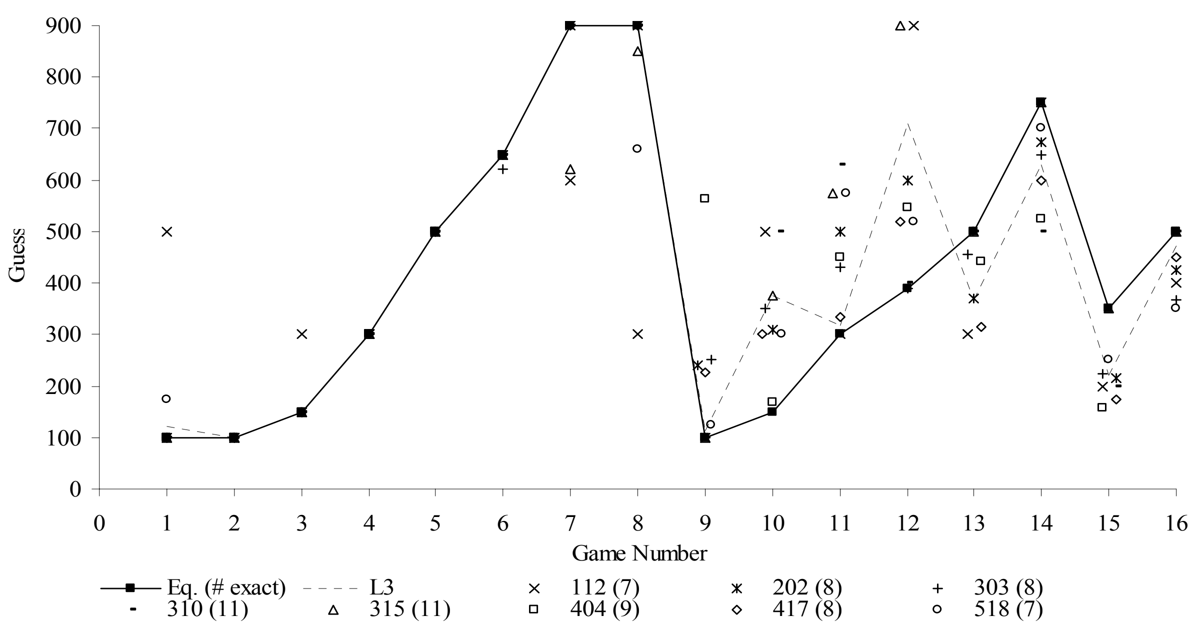


Figure 4. "Fingerprints" of 8 Apparent *Equilibrium* Subjects

(Only deviations from *Equilibrium*'s guesses are shown; mixed targets on right.)

GUESSMETRICS

Econometric analysis of guesses allows us to classify more subjects.

Our approach builds on Harless and Camerer (1994 *Econometrica*), El-Gamal and Grether (1995 *JASA*), Stahl and Wilson (1994 *JEBO*, 1995 *GEB*), CGCB.

Estimate subject by subject, using maximum-likelihood error-rate model with a "spike-logit" error structure: in each game *g* a subject *i* makes his type's guess exactly (within 0.5) with probability 1- ε and otherwise makes logit errors.

(Estimating a mixture model as in CGCB and most other studies is often theoretically superior; but given that we try to err by including rather than excluding types, parameter estimates are normally on the boundary of the parameter space, which eliminates the theoretical advantage. In an exploratory study of cognition, estimating subject by subject seems safer and, comparing CGCB with earliest version, probably yields similar estimates.)

Subject *i*'s guesses-only log-likelihood reduces to:

$$(7) \qquad (G-n^{ik})\ln(G-n^{ik})+n^{ik}\ln(n^{ik})+\sum_{g\in N^{ik}}\ln d_g^k(R_g^i(x_g^i),\lambda)-G\ln G,$$

where $d_g^k(R_g^i(x_g^i),\lambda)$ is a standard logit term for non-exact guesses and λ is the logit precision. (Deviation costs are measured using each type's beliefs.)

The maximum likelihood estimate of ε is n^{ik}/G, the sample frequency of subject *i*'s non-exact guesses for type *k*. The maximum likelihood estimate of λ is the standard logit precision, restricted to non-exact guesses.

The maximum likelihood type estimate maximizes (7) over *k*, given estimated ε and λ, trading off the count of exact guesses against logit cost of deviations.

This yields types estimates as in Table 1: 43 *L1*, 20 *L2*, 3 *L3*, 5 *D1*, 14 *Equilibrium*, and 3 *Sophisticated* (some questioned by specification tests).

$\varepsilon = 1$ is rejected for all but for 7 subjects, so the spike is necessary.

$\lambda = 0$ is rejected for 34, so payoff-sensitive logit errors significantly improve the fit over a spike-uniform model like CGCB's for only 39% of the subjects, which suggests to us that most errors are cognitive, or the result of misspecification.

$\{\lambda = 0$ and $\varepsilon = 1\}$ is rejected at the 5% level for all but 10 subjects (6 *L1*, 2 *D1*, 1 *Equilibrium*, and 1 *Sophisticated*), so the model does significantly better than a random model of guesses for 89% of the subjects.

ECONOMETRIC PUZZLE:

Our estimates could be sensitive to our a priori specification, which might err by omitting relevant types and/or overfitting by including irrelevant types.

Is there any reasonable way to estimate the distribution of subjects' decision rules without imposing an a priori list of possible types?

We want the types to be general decision rules (not just lists of predicted guesses in the 16 games) for at least two reasons:

(a) Types should be meaningful in other classes of games

(b) A type's search implications depend not only on what guesses it implies in our games, but why; so using search to study cognition seems to require general decision rules even within our class of games

But the space of possible decision rules is enormous, and it has very little mathematical structure; to avoid ruling out equilibrium, may have to allow all—including discontinuous—piecewise linear functions of the targets and limits.

Defining what it means for subjects' choices to be close is also problematic: usual notions are based on Euclidean distance, but that seems arbitrary here.

Intuitively, qualitative and possibly structure-dependent patterns of deviation from a reference pattern—such as the tendency of our *Equilibrium* subjects with the clearest fingerprints to deviate much more often in games with mixed targets, and always in the direction of *L3*—seem more relevant here; our analysis of clusters below gives them more weight.

SPECIFICATION TEST

Pending a solution to the problem of estimating without a priori specification of types, we start with the above inclusive list of plausible types and then refine our econometric analysis of guesses via a specification test.

The test is based on *pseudotype*s, each constructed from one Baseline or OB subject's guesses in the 16 games. It compares the likelihood of a subject's estimated type to those of the 87 other subjects' pseudotypes.

Suppose e.g. that we had omitted *L2*. The pseudotypes of subjects now estimated to be *L2* would then outperform the non-*L2* types estimated for them, and those subjects would make approximately the same (*L2*) guesses.

Define a *cluster* as a group of 2 or more subjects such that each subject's pseudotype has higher likelihood than the estimated types for other subjects in the group; and subjects' pseudotypes make "sufficiently similar guesses."

Finding a cluster should lead us to diagnose an omitted type. We find 5, with 3, 2, 2, 2, and 3 subjects: not much evidence of important omitted types.

We test for overfitting by asking whether a subject's estimated type performs at least as well against the pseudotypes as it would, on average, at random.

These tests leave us with 27 *L1*, 17 *L2*, 1 *L3*, 1 *D1*, 11 *Equilibrium*, and 1 *Sophisticated* subject, with 30 of 88 subjects unclassified (Table 1).

The only identifiable systematic non-equilibrium behavior is *L1*, *L2*, or *L3*: rational but with simplified models of others, yielding a surprisingly simple structural model of non-equilibrium behavior in initial responses to games.

SW, Nagel, CGCB, and CHC find similar type distributions, less definitively.

(We don't think subjects form beliefs first and then best respond; we think they use rules of thumb that happen to have decision-theoretic interpretations.)

ECONOMETRIC PUZZLE: Are there better ways to do specification tests?

OTHER PUZZLES LEFT OPEN BY OUR ANALYSIS OF GUESSES

What are our 8 subjects with near-*Equilibrium* fingerprints actually doing?

Their deviations from equilibrium almost always occur in games with mixed targets, and are always (when *Equilibrium* and *L3* are separated) in the direction of *L3* (sometimes beyond). Yet the ways that the cognoscenti use to identify equilibria all work equally well with or without mixed targets.

What are the 3 subjects with near-*L3* fingerprints actually doing? Their deviations from *L3* are also almost always in games with mixed targets.

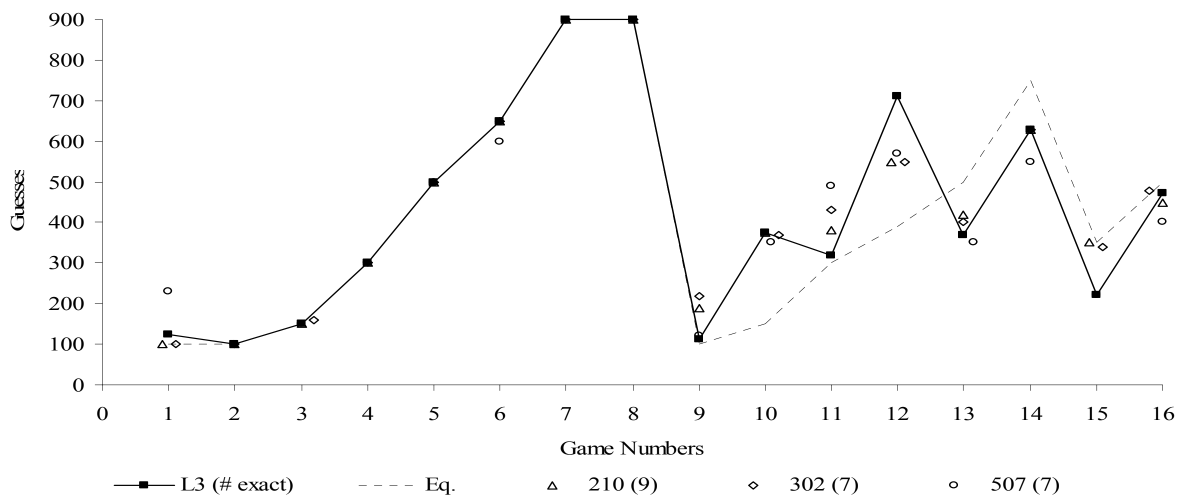


Figure 3. "Fingerprints" of 3 Apparent *L3* Subjects

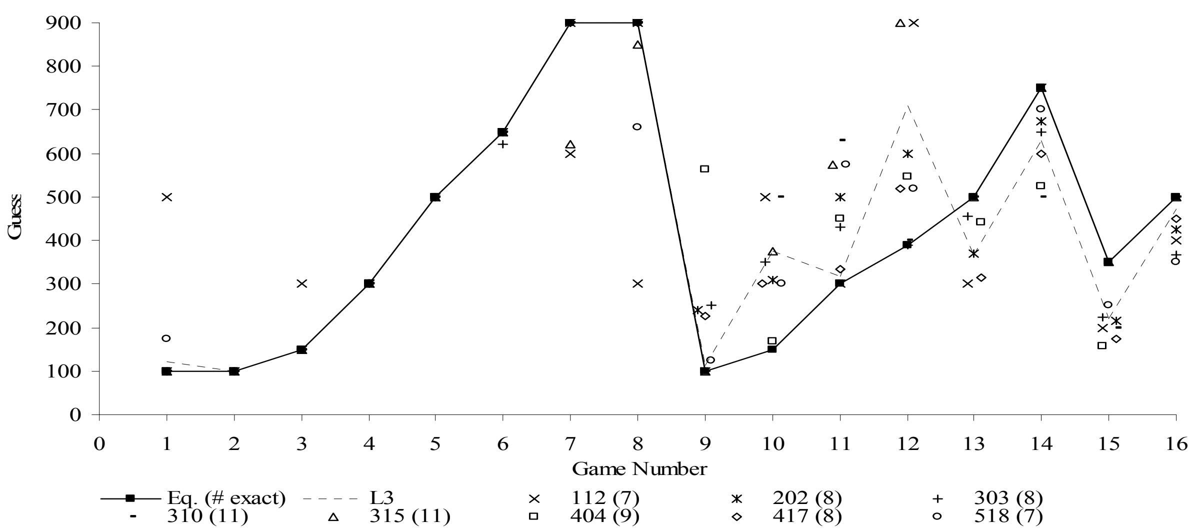


Figure 4. "Fingerprints" of 8 Apparent *Equilibrium* Subjects

Potential sources of answers:

Is there any difference in these subjects' (near-*Equilibrium* or -*L3*) search patterns with and without mixed targets?

How do these differences compare with the analogous differences for *L1* subjects (whose compliance does not differ in games with and without mixed targets, and for whom the distinction is theoretically irrelevant)?

(Can check compliance with types' search implications and re-estimate econometric model of search below, subject by subject, separately for games with and without mixed targets.)

Can we tell which of the standard methods—best-response dynamics, equilibrium checking, or iterated dominance—Baseline near-*Equilibrium* subjects are using; or if not, what else they are using?

(The absence of *Dk* subjects is strong evidence against iterated dominance. We bet on best-response dynamics, perhaps truncated after one or two rounds. To check we need to refine the characterization of *Equilibrium* search implications below and use it to re-do the searchmetrics below.)

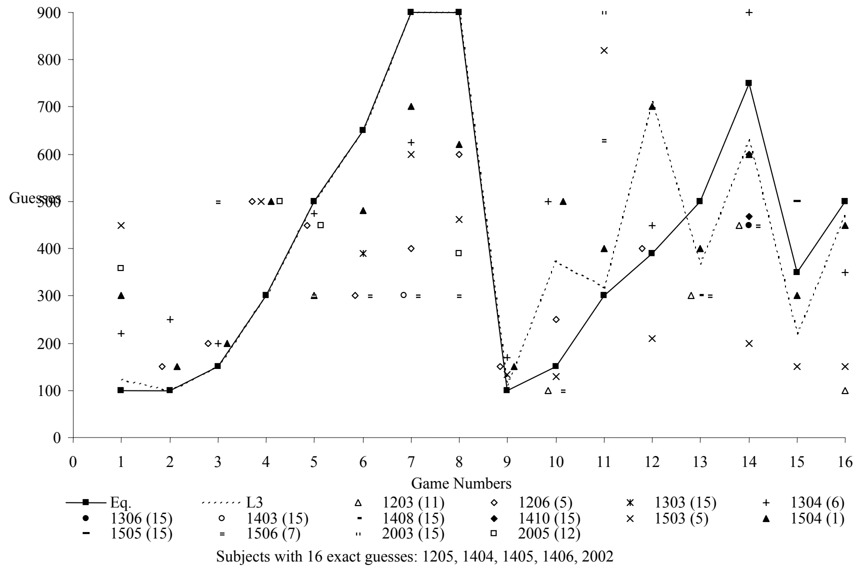
Is there any difference between Baseline and R/TS subjects' (*Equilibrium* or *L3*) patterns of deviations from *Equilibrium* or *L3* guesses across games (see next two pages for R/TS *Equilibrium* and *L3* subjects' deviations)?

Which of the standard methods do successful R/TS *Equilibrium* subjects use?

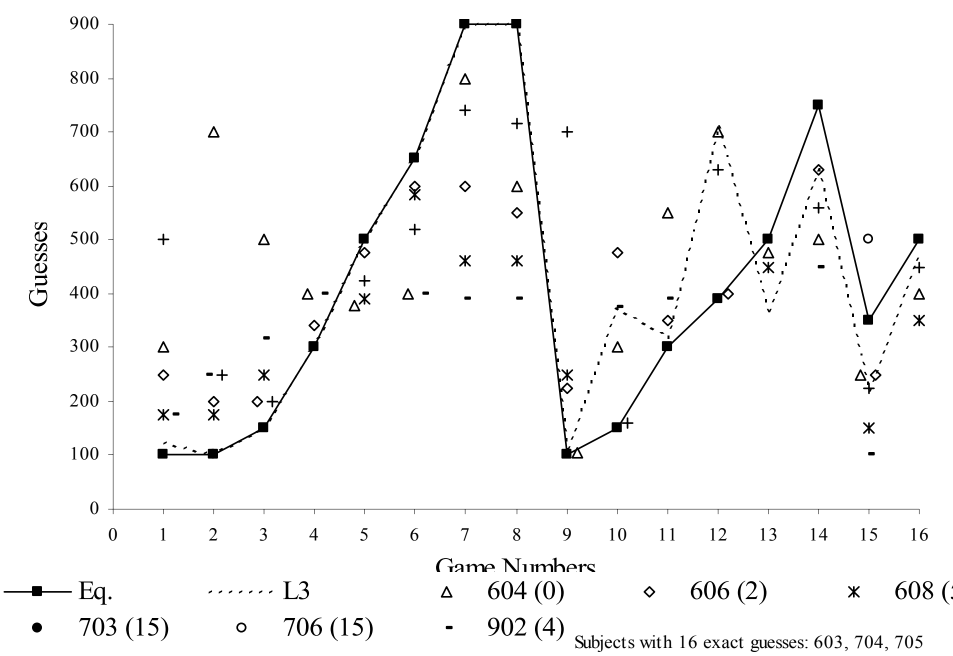
(Our training tries to be neutral, but something must come first, and we taught the methods in the order: equilibrium checking, best-response dynamics, and iterated dominance. To the extent that R/TS *Equilibrium* subjects use these methods, it explains why they have equal guess compliance with and without mixed targets. Examining the differences between their and Baseline near-*Equilibrium* subjects' searches may help identify what the Baseline subjects are doing and why it doesn't "work" with mixed targets.)

Does it help to know which UT2 questions R/TS *Equilibrium* or *L3* subjects missed?

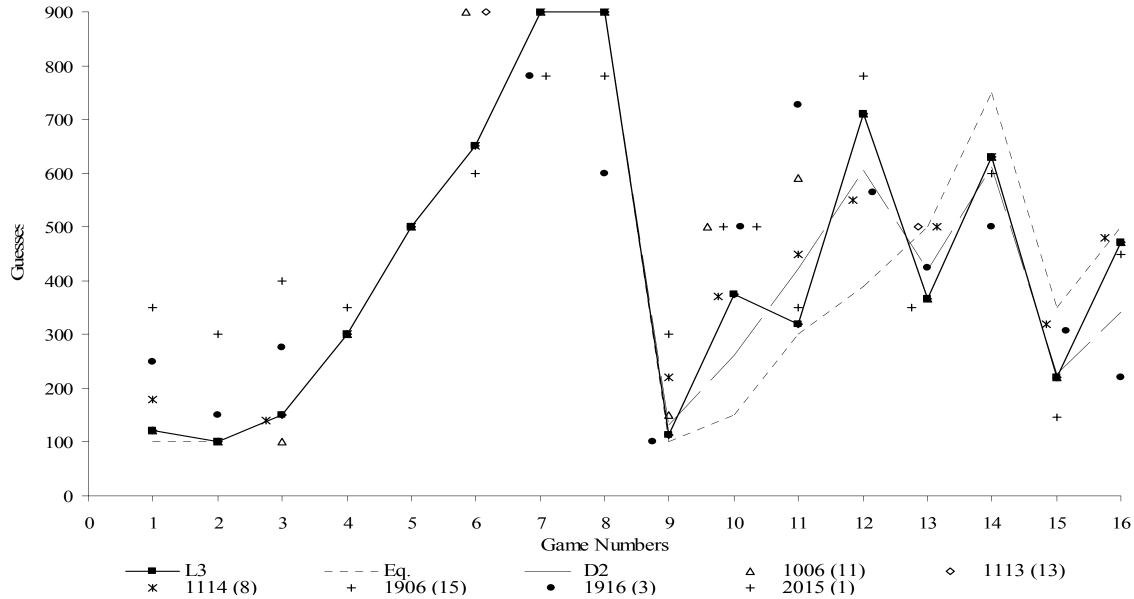
Fingerprints of 18 York Equilibrium R/TS Subjects
(only deviations from Eq.'s guesses are shown)



Fingerprints of 10 UCSD Equilibrium R/TS Subjects
(only deviations from Eq.'s guesses are shown)

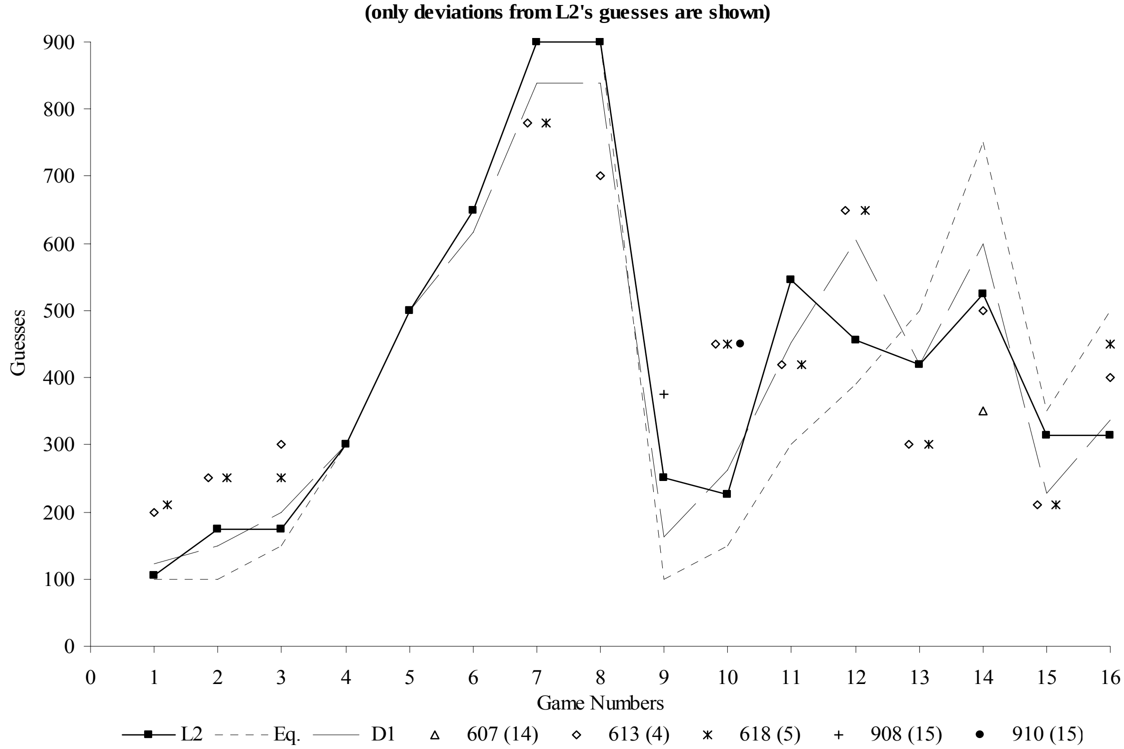


Fingerprints of 18 York L3 R/TS Subjects
(only deviations from L3's guesses are shown)



Subjects with 16 exact guesses: 1001, 1003, 1004, 1005, 1007, 1008, 1010, 1011, 1013, 1201,1216, 1412

Fingerprints of 9 UCSD L2 R/TS Subjects
(only deviations from L2's guesses are shown)



Subjects with 16 exact guesses: 615, 616, 909, 911

MORE PUZZLES

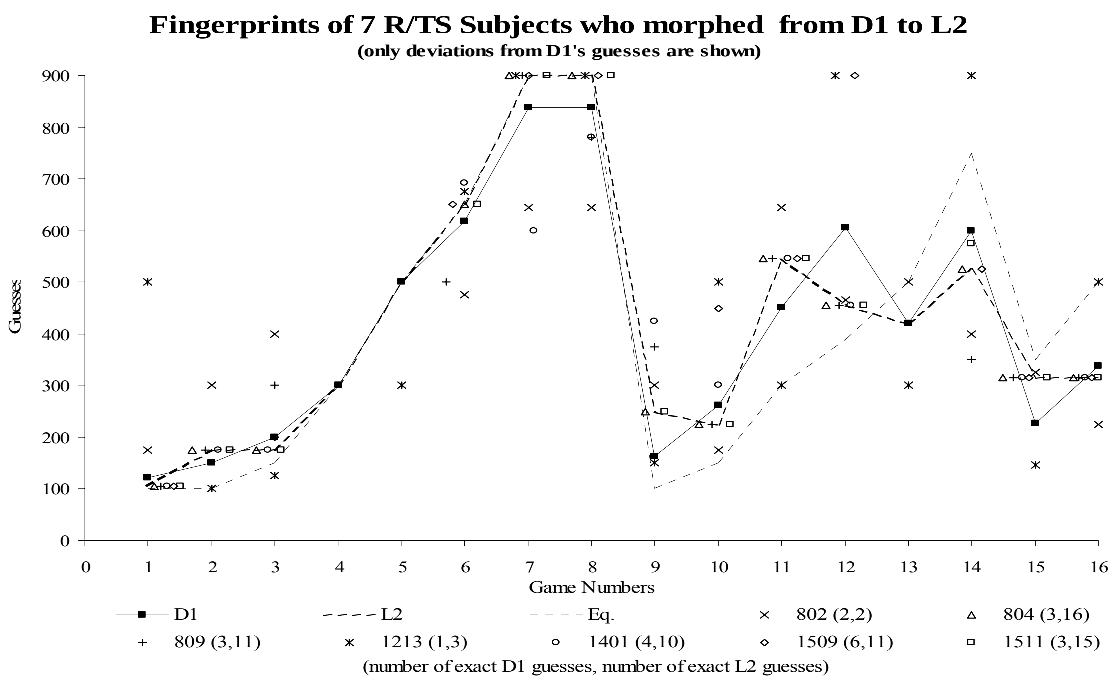
Why do *L1*, *L2*, and *L3* so strongly outnumber other non-equilibrium rules in the Baseline?

Potential sources of answers:

UT2 failure rates and R/TS subjects' compliance confirm that *Lk* types are easy for all *k*, *Dk* and *Eq.* types are hard or unnatural (monotonicity).

	L1	***L2***	***L3***	***D1***	***D2***	***Eq.***
UCSD subjects	7	9	-	11	-	10
% Compliance	77.7	81.3	-	55.1	-	58.1
% Failed UT2	0.0	0.0	-	8.3	-	28.6
York subjects	18	18	18	19	19	19
% Compliance	80.9	95.8	84.4	66.1	55.3	76.6
% Failed UT2	0.0	0.0	0.0	0.0	5.0	13.6
UCSD + York subjects	25	27	18	30	19	29
% Compliance	80.0	91.0	84.4	62.1	55.3	70.3
% Failed UT2	0.0	0.0	0.0	3.2	5.0	19.4

7 of 19 R/TS *D1* subjects passed a UT2 in which *L2* answers were wrong and then "morphed" into *L2*s. (No significant morphing of any other kind.)



SEARCHMETRICS

How does search refine our guesses-only estimates of subjects' types?

The search behavior of Baseline subjects with clear fingerprints and of the analogous R/TS subjects show common patterns that can be understood using a simple theory of how cognition drives search.

Our initial econometric analysis focuses on the order of look-ups and ignores duration, following CJ and CGCB.

We view search for hidden payoff information as just another kind of decision—not the kind conventionally studied, but potentially also useful in helping to identify the decision rules that best describe subjects' behavior.

Standard assumptions imply that a rational subject looks up all freely available information that might affect its beliefs, and then best responds to his beliefs.

In our design each type is naturally associated with algorithms that process information about targets and limits into guesses. We take those algorithms as models of cognition, and infer a type's minimal search implications from them under conservative assumptions about how cognition affects search:

a. *Basic* operations are associated with adjacent look-ups, and can appear in any order.

b. Other operations can be separated, and can appear in any order.

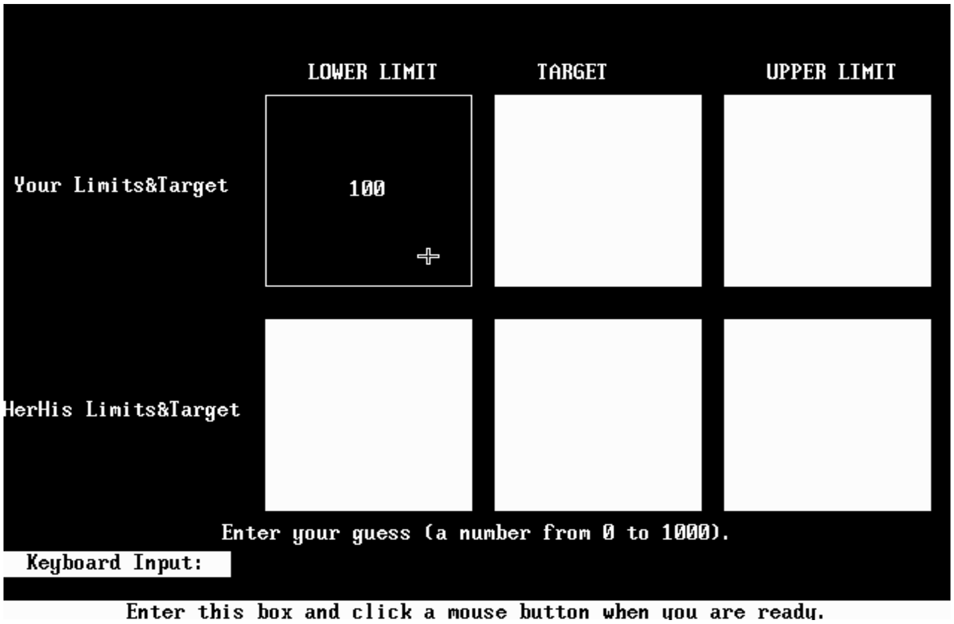
(Similar, but not identical, to CGCB's "Occurrence" and "Adjacency".)

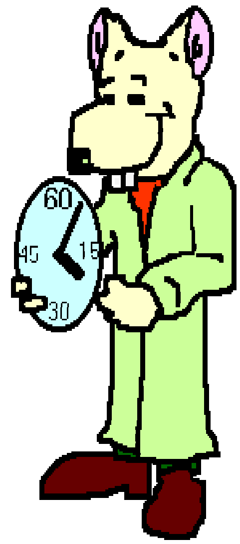
Motivated by limitations on working memory, "efficient" information processing; compare CJ's characterization of search implications of backward induction in extensive-form alternating-offers bargaining games.

The argument is essentially empirical: It's theoretically possible for a subject to scan and memorize the parameters in any order, then go into his brain and process them, in which case his search pattern yields no information about cognition (unless he fails to look at a parameter he needs to know).

Now for a look at the search data, but first…

SPEAK RODENT LIKE A NATIVE IN ONE EASY LESSON!





	a	***p***	***b***
You (*i*)	1	2	3
S/he (*j*)	4	5	6

Type	**Ideal guess**	**Relevant look-ups**
L1	$p^i[a^j+b^j]/2$	$\{[a^j,b^j],p^i\} \equiv \{[4, 6], 2\}$
L2	$p^iR(a^i,b^i; p^j[a^i+b^i]/2)$	$\{([a^i,b^i],p^j),a^j,b^j,p^i\} \equiv \{([1, 3], 5), 4, 6, 2\}$
L3	$p^iR(a^j,b^j; p^jR(a^i,b^i; p^i[a^j+b^j]/2))$	$\{([a^j,b^j],p^i),a^i,b^i,p^j\} \equiv \{([4, 6], 2), 1, 3, 5\}$
D1	$p^i(\max\{a^j,p^ja^i\} + \min\{p^jb^i,b^j\})/2$	$\{(a^j,[p^j,a^i]),(b^j,[p^j,b^i]),p^i\} \equiv \{(4,[5,1]),(6,[5,3]),2\}$
D2	$p^i[\max\{\max\{a^j,p^ja^i\},p^j\max\{a^i,p^ia^j\}\}$ $+\min\{p^j\min\{p^ib^j,b^i\},\min\{p^jb^i,b^j\}\}]/2$	$\{(a^i,[p^i,a^j]),(b^i,[p^i, b^j]),(a^j,[p^j,a^i]),(b^j,[p^j,b^i]),p^j,p^i\}$ $\equiv \{(1,[2,4]),(3,[2,6]),(4,[5,1]),(6,[5,3]),5,2\}$
Eq.	p^ia^j if $p^ip^j < 1$ or p^ib^j if $p^ip^j > 1$	$\{[p^i,p^j],a^j\} \equiv \{[2, 5], 4\}$ if $p^ip^j < 1$ or $\{[p^i,p^j],b^j\} \equiv \{[2, 5], 6\}$ if $p^ip^j > 1$
Soph.	[no closed-form expression; search implications are the same as *D2*'s]	$\{(a^i,[p^i,a^j]),(b^i,[p^i, b^j]),(a^j,[p^j,a^i]),(b^j,[p^j,b^i]),p^j,p^i\}$ $\equiv \{(1,[2,4]),(3,[2,6]),(4,[5,1]),(6,[5,3]),5,2\}$

Despite comparatively weak assumptions, the theory yields high resolution of cognition, with implications almost independent of the game, so that one can often see the algorithms subjects are using in the search data.

Selected R/TS Subjects' Information Searches and Assigned Types' Search Implications

MouseLab box numbers

	a	*p*	*b*
You (*i*)	1	2	3
S/he (*j*)	4	5	6

Types' Search Implications

L1	{[4,6],2}
L2	{([1,3],5),4,6,2}
L3	{([4,6],2),1,3,5}
D1	{(4,[5,1], (6,[5,3]),2}
D2	{(1,[2,4]),(3,[2,6]),(4,[5,1],(6,[5,3]),5,2}
Eq	{[2,5],4} if pr. tar.<1,{[2,5],6} if > 1

Subject	**904**	**1716**	**1807**	**1607**	**1811**	**2008**	**1001**	**1412**	**805**	**1601**	**804**	**1110**	**1202**	**704**	**1205**	**1408**	**2002**
Type(#rt.)	L1 (16)	L1 (16)	L1 (16)	L2 (16)	L2 (16)	L2 (16)	L3 (16)	L3 (16)	D1 (16)	D1 (16)	D1 (3)	D2 (14)	D2 (15)	Eq (16)	Eq (16)	Eq (15)	Eq (16)
Alt.(#rt.)											L2 (16)						
Est. style	late	often	early	often	early				early								
Game																	
1	123456 4623	146462 134646 23	462513	135462 1313	134446 5213*4 6	111313 131313 5423	462135 21364* 246231 52	146231 564623 1	154356 423213 2642	254514 36231	154346 5213	135464 2646*1 313	246466 135464 641321 342462 422646 124625 5*1224 654646	123456 363256 565365 626365 652651 452262 6526	123456 424652 562525 6352*4 65	123123 456445 632132 11	142536 125365 253616 361454 613451 213452 63
2	123456 4231	462462 13	462132 25	135461 354621 3	134653 125642 313562 52	131313 566622 333	462135 642562 223146 2562*6 2	462462 546231 546231	514535 615364 23	514653 6213	515135 365462 3	135134 642163 451463 211136 414262 135362 *14654 6	123645 132462 426262 241356 462*13 524242 466135 6462	123456 525123 652625 635256 262365 456	123456 244565 565263 212554 146662 654251 44526* 31	123456 456123 643524 1	143625 361425 142523 625656 3

Notes: The subjects' frequencies of making their assigned types' (and when relevant, alternate types') exact guesses are in parentheses after the assigned type. A * in a subject's

Selected Baseline Subjects' Information Searches and Estimated Types' Search Implications

MouseLab box numbers

	a	*p*	*b*
You (*i*)	1	2	3
S/he (*j*)	4	5	6

Types' Search Implications

L1	{[4,6],2}
L2	{([1,3],5),4,6,2}
L3	{([4,6],2),1,3,5}
D1	{(4,[5,1], (6,[5,3]),2}
D2	{(1,[2,4]),(3,[2,6]),(4,[5,1],(6,[5,3]),5,2}
Eq	{[2,5],4} if pr. tar.<1,{[2,5],6} if > 1

Subject	**101**	**118**	**413**	**108**	**206**	**309**	**405**	**210**	**302**	**318**	**417**	**404**	**202**	**310**	**315**
Type(#rt.)	L1 (15)	L1 (15)	L1 (14)	L2 (13)	L2 (15)	L2 (16)	L2 (16)	L3 (9)	L3 (7)	L1 (7)	Eq (8)	Eq (9)	Eq (8)	Eq (11)	Eq (11)
Alt.(#rt.)								Eq (9)	Eq (7)	D1 (5)	L3 (7)	L2 (6)	D2 (7)		
Alt.(#rt.)								D2 (8)			L2 (5)		L3 (7)		
Est. style	early/late	early	late	early	early	early/late	early	early	early	early	early	early	early	early/late	early
Game															
1	146246 213	246134 626241 32*135	123456 545612 3463*	135642	533146 213	1352	144652 313312 546232 12512	123456 123456 213456 254213 654	221135 465645 213213 45456* 541	132456 465252 13242* 1462	252531 464656 446531 641252 462121 3	462135 464655 645515 21354* 135462 426256 356234 131354 645	123456 254613 621342 *525	123126 544121 565421 254362 *21545 4*	213465 624163 564121 325466
2	46213	246262 2131	123564 62213*	135642 3	531462 31	135263 1526*2 *3	132456 253156 456545 463123 156562 62	123456 465562 231654 456*2	213546 566213 545463 21*266 54123	132465 132*46 2	255236 62*365 243563	462461 352524 261315 463562	123456 445613 255462 513565 23	123546 216326 231456 *62	134652 124653 656121 3
3	462*46	246242 466413 *426	264231	135642 53	535164 2231	135263	312456 5231*1 236545 5233** 513	123455 645612 3	265413 232145 563214 563214 523*65 4123	134652 1323*4	521363 641526 5263*6 52	462135 215634 *52	123456 123562 3	123655 463213	132465 544163 *3625

SEARCHMETRICS CONTINUED

Combine above guessmetrics with a maximum-likelihood error-rate model of search as in CGCB (but subject-by-subject, not mixture model).

The main econometric problem is extracting signals from subjects' highly idiosyncratic, noisy look-up sequences, without a well-tested model that implies strong restrictions on how cognition drives search.

Subjects vary in the location of look-ups relevant to their types in their sequences. Filter this out via subject-specific nuisance parameter called style ("early" or "late"), assumed constant across games for each subject. (58 of 71 Baseline subjects' estimated styles are "early," 10 are "late," and 3 are tied.)

Quantify compliance with a type's search implications as the density of the type's relevant look-up sequence in the subject's look-up sequence. If style is early, start at the beginning of the sequence and continue until the type's relevant sequence is first completed. Compliance is the length of the relevant sequence divided by the length of the sequence that first completes it. This definition makes compliance meaningfully comparable across games, styles.

We assume that a subject's type and style determine his search and guess in a given game, each with error; and we further assume that, given type and style, errors in search and guesses are independent of each other and across games. This strong but useful simplifying assumption makes the log-likelihood separable across guesses and search, avoiding some complications in CGCB.

To avoid stronger distributional assumptions, we discretize compliance into three categories: $C_H \equiv [0.67, 1.00]$, $C_M \equiv [0.33, 0.67]$, and $C_L \equiv [0, 0.33]$.

Subject *i*'s guesses-and-search log-likelihood is:

$$\sum_c \left[m_c^{isk} \ln(\zeta_c) + (m_c^{isk} - n_c^{isk}) \ln(1-\varepsilon) + n^{isk} \ln(\varepsilon) + \sum_{g \in N_c^{isk}} \ln d_g^k (R_g^i(x_g^i), \lambda) \right] \equiv$$

$$(G-n^{ik})\ln(G-n^{ik}) + n^{ik}\ln(n^{ik}) + \sum_{g \in N^{ik}} \ln d_g^k (R_g^i(x_g^i), \lambda) - G\ln G + \sum_c \left[m_c^{isk} \ln m_c^{isk} \right] - 2G\ln G,$$

where m_c^{isk} is the number of games for which subject *i* has type-*k* style-*s* compliance *c*. (The search term is convex in the m_c^{isk}, and therefore favors types for which compliance varies less across games, because such types "explain" search behavior better. See CGCB, Section 4.D.)

The maximum-likelihood estimates of ε and ζ_c, given *k* and *s*, are n^{ik}/G and m_c^{isk}/G, the sample frequencies with which subject *i*'s adjusted guesses are non-exact for that *k* and *i* has compliance *c* for that *k* and *s*. The maximum likelihood estimate of *λ* is the standard logit precision.

The maximum likelihood estimate of subject *i*'s type *k* maximizes the above log-likelihood over *k* and *s*, given the estimated ε and *λ*.

The model favors such types without regard to whether compliance is high or low. This seems appropriate because compliance is neither meaningfully comparable across types nor guaranteed to be high for the "true" type (which could be cognitively very difficult). But it means we need to rule out estimates where a type wins simply because its compliance is very low in all games.

A few subjects' type estimates change (Table 1) when search is included:

For some subjects a tension between guesses-only and search-only type estimates is resolved in favor of the search estimates. (The search part of the likelihood has weight only about 1/6 of the guesses part, because our theory of search makes much less precise predictions than our theory of guesses—a necessary evil, given the noisiness and idiosyncrasy of search behavior.)

For other subjects the guesses-only type estimate has 0 search compliance in 8 or more games, and so we rule it out a priori.

ECONOMETRIC PUZZLE: Are there better ways to do this?

MISCELLANEOUS QUESTIONS

To what extent can Baseline subjects' guess "errors" be explained by a more detailed analysis of search?

What more can we say about subjects with high guess compliance but 0 search compliance in several games? (E.g. Baseline subject 415, who could remember 3 parameters at a time; and the "perfect-16" R/TS *L2* subject 2008, who missed *L2*'s search requirements in first 5 games.)

Can we separate the effects of training from the strategic-uncertainty-eliminating effects of robot treatments? Conditional on style, how does search differ between Baseline subjects with clear fingerprints (*Equilibrium*, *L1*, *L2*, or *L3*) and successful R/TS subjects of same type?

(Baseline subjects with clear fingerprints are, to the extent that we know their beliefs, like robot *un*trained subjects, which usually don't exist because you have to teach true robot subjects what the robot's decision rule is. Are there any systematic differences between the errors Baseline subjects with clear fingerprints make and the errors made by R/TS subjects of the same type? If so, what can the differences tell us about cognition?)

Can we divide decision rules into those that just don't occur to subjects, and those that (like *Dk*) are unnatural even after training?

Our search analysis has so far focused on the order of look-ups. (Compare CJ, "Thinking about Attention in Games: Backward and Forward Induction" and Rubinstein, "Instinctive and Cognitive Reasoning: A Study of Response Times.") Is there useful information in look-up durations? Can we say more about the difficulty of types using duration data?

Average time per guess according to subjects' estimated types (incomplete):

Baseline: *L1* (22): 63.7 seconds, *L2* (13): 82.1, *Eq.* (8): 117.2.

Baseline overall 74.5 seconds vs. OB overall 91.9.

Baseline overall 74.5 seconds vs. UCSD R/TS overall 65.2.

We also plan nonparametric analyses of search durations, transitions.