NSF WORKSHOP ON "BEHAVIOR, COMPUTATION AND NETWORKS IN HUMAN SUBJECT EXPERIMENTATION," DEL MAR, CA, July 31-August 1, 2008

STUDYING STRATEGIC THINKING BY MONITORING SEARCH FOR HIDDEN PAYOFF INFORMATION AND INTERPRETING THE DATA IN THE LIGHT OF ALGORITHMS THAT LINK COGNITION, SEARCH, AND DECISIONS

Vincent P. Crawford, University of California, San Diego

Based on joint work with Miguel Costa-Gomes, University of Aberdeen, and Bruno Broseta, Red de Institutos Tecnológicos de la Comunidad Valenciana:

Costa-Gomes, Crawford, and Broseta, "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica* 2001 ("CGCB").

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

Costa-Gomes and Crawford, "Studying Cognition via Information Search in Two-Person Guessing Game Experiments," in preparation; see http://dss.ucsd.edu/~vcrawfor/12Jan07NYUCognitionSearchMain.pdf.

See also:

Johnson, Camerer, et al., "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining," *Journal of Economic Theory* 2002 ("CJ").

Wang, Spezio, and Camerer, "Pinocchio's Pupil: Using Eyetracking and Pupil Dilation To Understand Truth-telling and Deception in Games" ("WSC").

Overview

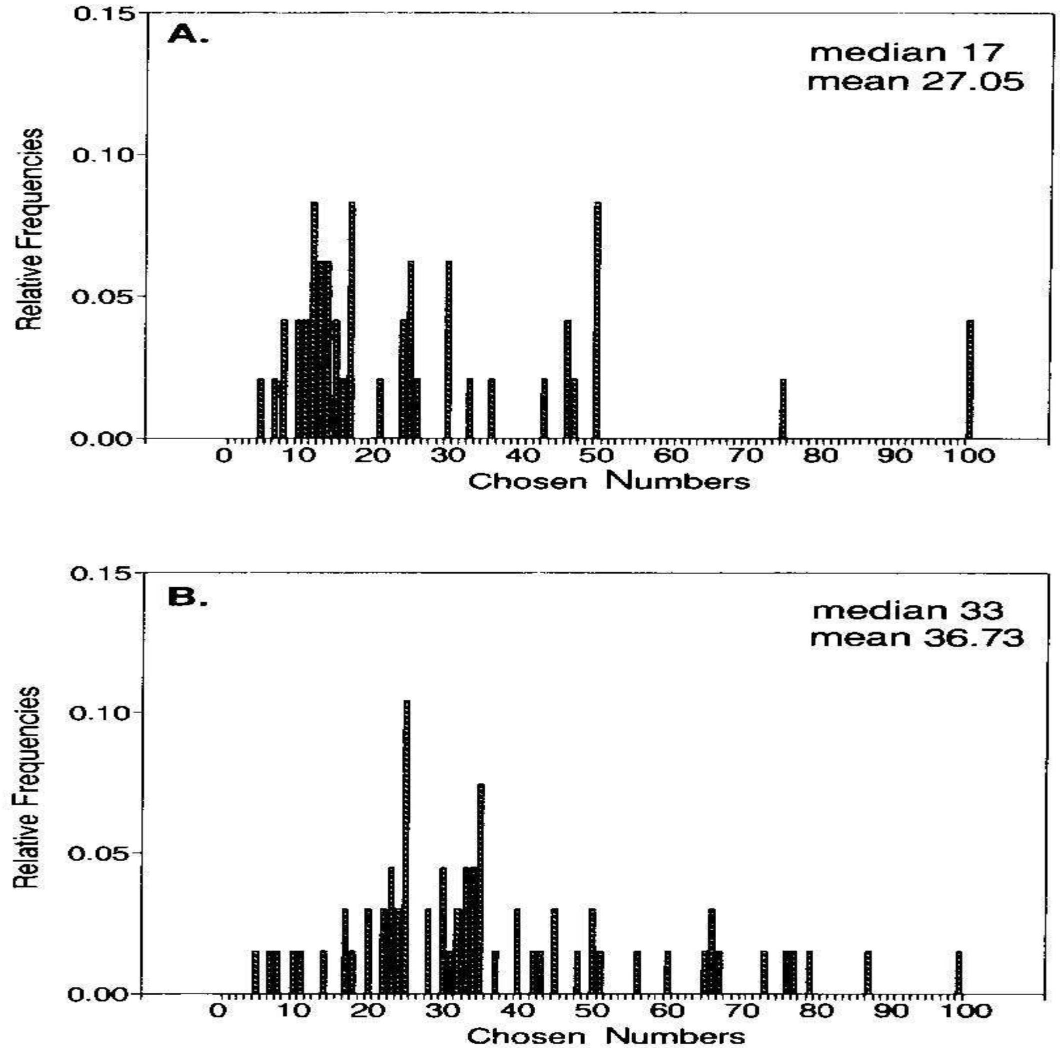
Starting with Nagel (1995 *AER*) and Stahl and Wilson (1995 *GEB*; "SW"), experimental evidence suggests that subjects' strategic thinking, as revealed by their initial responses to games, often deviates systematically from equilibrium.

Nonetheless, the deviations seem to have a structure that is stable across a range of games, and which is discrete and individually heterogeneous.

In Nagel's "guessing games", for example, 15-18 subjects simultaneously guess between limits [0,100], with the subject whose guess is closest to a target p (= 1/2 or 2/3) times the group average guess winning a prize, say $50.

These games are dominance-solvable, so the equilibrium can be found by iteratively eliminating stupid guesses. For example, if $p = ½$ it's stupid to guess more than 50 (1/2 × 100); unless you think other people are stupid, it's also stupid to guess more than 25; and so on. Thus the games have a unique equilibrium, in which all players guess 0.

Yet Nagel's subjects never guessed 0 initially; their guesses were heterogeneous, respecting 0 to 3 rounds of iterated dominance (first picture p = 1/2; second p = 2/3):



Spikes are clearly visible (amid the noise) at $50p^k$ for target p and low integers k.

Other studies of initial responses to games have since found evidence of systematic deviations from equilibrium with a similarly discrete, individually heterogeneous structure.

These results suggest the possibility of a general, structural non-equilibrium model of initial responses to games that can out-predict equilibrium in some games.

(Such a model would presumably mimic equilibrium in some games but deviate systematically in others, thus allowing us to predict both when equilibrium will be a reliable guide to behavior and what is likely to happen when it is not.)

The spikes in Nagel's data and similar subsequent results leave open some important questions about the structure of strategic thinking, which need to be resolved, at least approximately, before a model can be specified with confidence.

For example, the spikes at $50p^k$ might have come from subjects doing k–1 rounds of iterated dominance (as theorists often assume when they see her data) and then best responding to a uniform prior over the remaining guesses.

Such subjects would guess $p([0+100p^{k-1}]/2)$, following a type later called *Dk*-1.

Or, the spikes might have come from subjects starting with a uniform prior over feasible guesses and iterating best responses k times.

Such subjects would guess $p^k[(0+100)/2]$, following a type called *Lk*.

Although the difference between these interpretations of subjects' behavior matters a lot in some interesting games, in Nagel's design *Lk* and *Dk*-1 are not separated, and the decisions they imply are only weakly separated in most other designs.

One way to make further progress is to create designs that enhance the separation of the decisions implied by different modes of strategic thinking, and this has been done.

But another way—potentially of greater interest to this group—is to create designs using an interface that hides some payoff-relevant information but makes it freely searchable.

With careful design, the leading alternative models of strategic thinking imply different patterns of search as well as different decision patterns, and the process data provides a second view of subjects' cognition and behavior, at comparatively low additional cost.

Further, creating such designs and analyzing the search data seem to require an algorithmic view of how subjects process payoff information into decisions.

This view looks likely to yield deeper insight into the structure of strategic thinking than the conventional approach of theorizing about decisions as if the stork brought them.

In the rest of these slides I briefly describe three examples of this approach, highlighting modes of analysis and how search data changes the view of strategic thinking:

CJ's analysis of extensive-form alternating- offers bargaining games

CGCB's analysis of matrix games

CGC's analysis of normal-form two-person guessing games

These analyses each begin with a model in which each subject follows one of a pre-specified set of behavioral decision rules or “types” in each of the games he plays.

A subject’s type determines his information search, and his type and search (the latter possibly including errors) then determine his decision.

Each type is naturally associated with algorithms that process information into decisions.

The analysis uses these algorithms as models of cognition, deriving search implications under simple, empirically motivated assumptions about how cognition determines search.

Types’ derived implications provide a kind of basis for the huge space of possible decision and search sequences, which imposes enough structure to describe behavior in a comprehensible way and make it meaningful to ask how decisions and search are related.

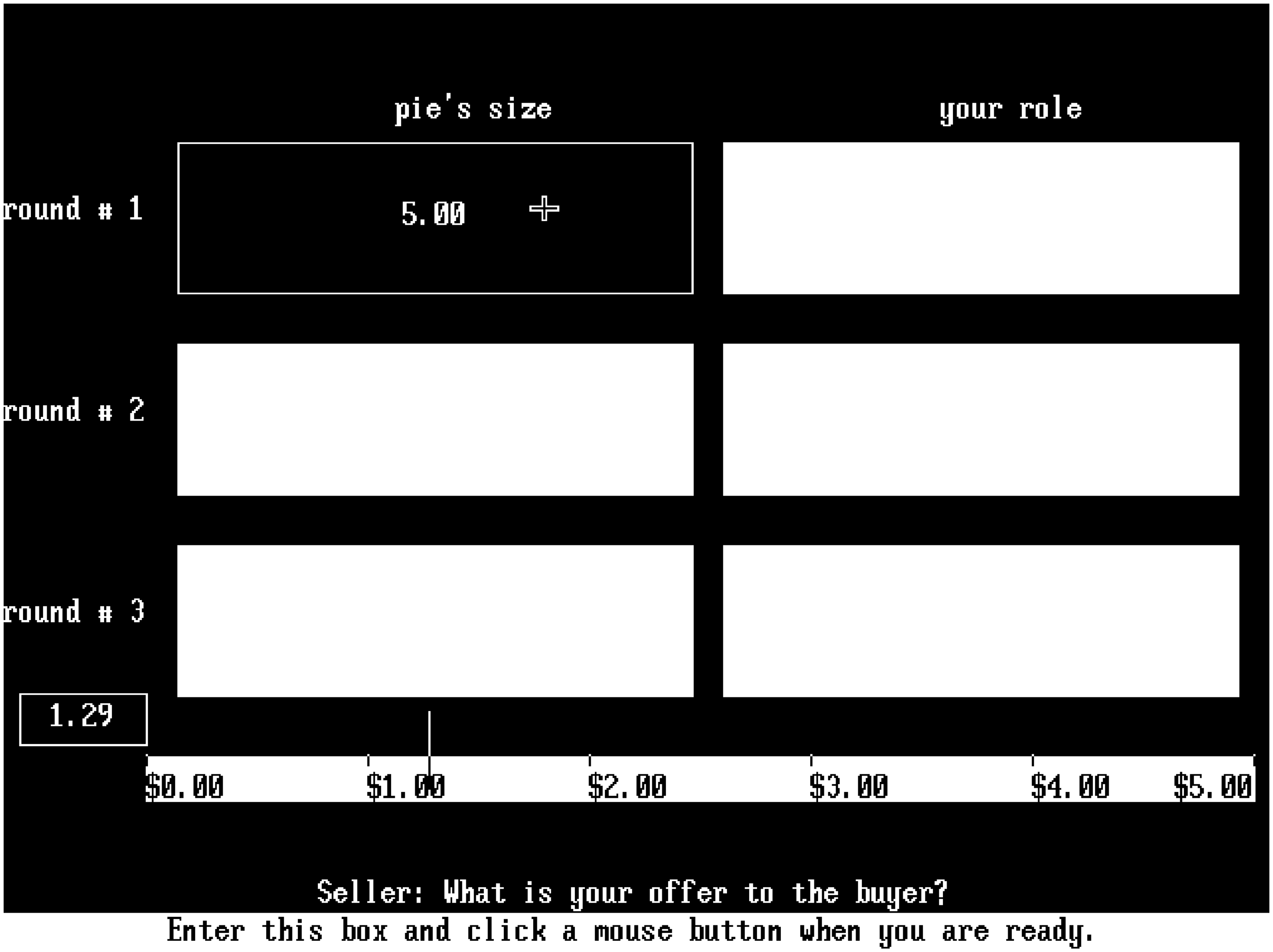
CJ's Extensive-Form Alternating-Offers Bargaining Games

CJ's subjects played series of alternating-offers bargaining games, in extensive form.

In these games subjects systematically deviated from the subgame-perfect equilibrium strategies, with proposers usually making offers more generous than equilibrium predicts, and responders often rejecting offers of positive amounts.

At the time the experiments were designed, there was controversy about whether the deviations were due to revenge motives (rejecting "unfair" offers) or cognitive limitations (subjects being unable to compute subgame-perfect equilibrium strategies).

Within a publicly announced structure, CJ presented a game as a series of “pies” via MouseLab, which normally concealed the pies but allowed subjects to look them up as often as desired, one at a time. (Subjects were not allowed to write down the pies, and the frequencies with which they looked them up made clear that they did not memorize them.)



CJ's Figure 1. MouseLab Screen Display

CJ argued that the backward induction that is the easiest way to compute the subgame-perfect or sequential equilibrium (with or without revenge motives) is naturally associated with search patterns in which subjects first look at the last-period pie, then back and forth between the past and second-last pie, then between the second-last and first pie.

CJ's robot/trained subjects (playing against a computer they were told was programmed to play the subgame-perfect equilibrium) came close to this search pattern.

But CJ's Baseline subjects, playing without training against other baseline subjects, deviated substantially from backward-induction search:

> About 10% of the subjects never looked at the last-period pie (so even if they had played subgame-perfect equilibrium strategies, we would not expect it to persist beyond sample).
>
> Many other subjects deviated from backward-induction search in more subtle ways.

CJ identified a weak but positive correlation between individual subjects' deviations from equilibrium search and their deviations from equilibrium decisions:

Subjects whose searches were further from backward induction made decisions further from subgame-perfect equilibrium decisions.

CJ also found evidence of a mixture of "types" in the subject population:

Level-0 types who treat the first round as an ultimatum game

Level-1 types who look one round ahead but truncate beyond that, and

Level-2 types who look two rounds ahead as subgame-perfect equilibrium requires, hence in this game are functionally equivalent to *Equilibrium* types.

Level-2, *Level-1*, and *Level-0* types deviate progressively more and more from subgame-perfect equilibrium in search as well as decisions, so that a mixture of types implies a positive correlation between search and decision deviations.

Overall, CJ's analysis suggests that the deviations from subgame-perfect equilibrium are due roughly half to revenge motives and half to cognitive limitations.

CJ's and other analyses of search must address the problem that it is *logically* possible for subjects to scan and memorize the hidden information and then retreat into their brains to decide what to do, in which case their search patterns reveal nothing about their thinking.

Inspecting actual searches suggests that there are strong regularities in search behavior, and as a result subjects' searches contain a lot of information about cognition.

(Something like these regularities persist for eyetracking data as in WSC, with the additional advantage that the tracker measures attention more directly; in MouseLab, by contrast, a subject can open a box and then gaze out the window without attending to it.)

The goal in search analysis is to add enough assumptions to make it possible to extract the signal from the noise in subjects' look-up sequences; but not so many assumptions that they distort the signal's meaning.

Given the noisiness and heterogeneity of subjects' searches, this is a delicate balance.

CJ, CGCB, and CGC stylize the search evidence by imposing empirically motivated restrictions on subjects' search patterns, which CGCB called:

> Occurrence (if your type's decision depends on a particular piece of hidden information, you must have looked at it at least once), and

> Adjacency (if the most basic operations your type's decision requires involve two pieces of hidden information, they "must" be adjacent in your look-up sequence).

CJ implicitly invoke Occurrence and Adjacency, but also use other measures.

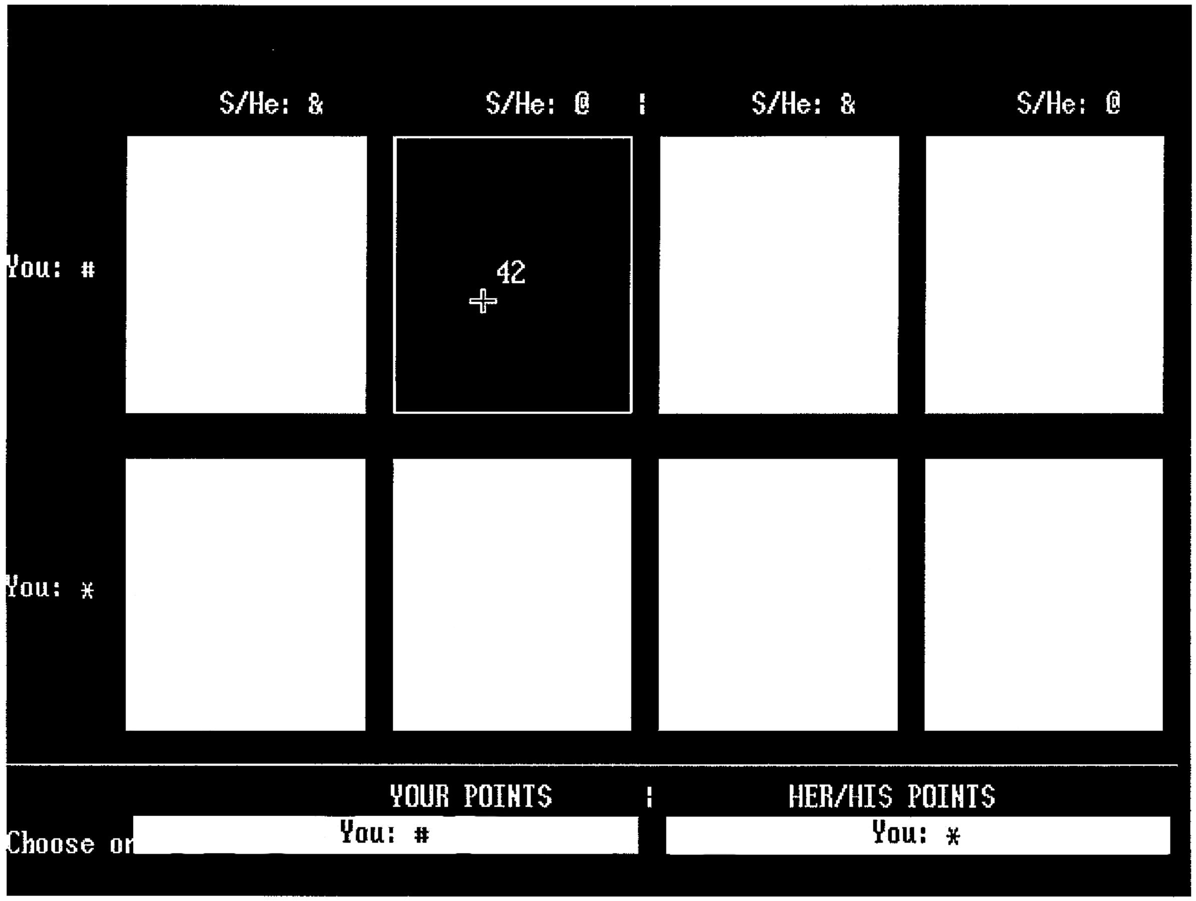
CGCB explicitly use Occurrence and Adjacency in their econometric analysis, and also use some other measures.

CGC use a more refined measure in the same spirit, based on measures of the density of a type's characteristic look-up sequence (characterized using requirements in the spirit of Occurrence and Adjacency) in a subject's observed sequence.

CGCB's Matrix Games

CGCB's subjects played a series of 18 normal-form games, with various patterns of iterated dominance or unique pure-strategy equilibria without dominance.

Within a publicly announced structure, each game was presented as a matrix with players' payoffs separated horizontally via MouseLab. (All subjects were framed as Row players.)



CGCB's Figure 1. MouseLab Screen Display (for a 2×2 game)

The 18 games were chosen to separate leading types' implications for decisions.

Subjects' decisions replicated most patterns in previous experiments.

In particular, there was evidence of a mixture of "types", with many apparent *L1* ("*Naïve*" in CGCB, distinct from CJ's extensive-form type *Level-1* above) and *L2* types, and apparent traces of an iterated dominance type *D1*.

But separation of some types was weak or nonexistent, leaving considerable ambiguity.

Analysis of subjects' search for hidden but freely accessible payoff information, assuming Occurrence and Adjacency, strengthens and refines this classification.

The multidimensionality of search makes it more informative—compare CJ's essentially one-dimensional searches—but also makes it necessary to impose structure on the huge space of possible search sequences, which is done by specifying a list of leading types.

The size of the space of possible decision sequences alone seems to make it necessary to impose a types-based structure.

But types play an additional role in search analysis, making it meaningful to ask whether a subject's searches deviated from equilibrium in the "same direction" as his decisions (by contrast, the directions are immediately apparent in CJ's one-dimensional search space).

One would like to dispense with the need to specify a list of types a priori, and my co-authors and I have spent considerable time trying to figure out how to do this by something like a multidimensional clustering analysis, without success.

One problem is computational feasibility; another is that standard clustering analyses depend on distance metrics that do not seem appropriate in this setting.

CGC address this problem indirectly by conducting a careful specification test, looking both for "underperforming" types and omitted types.

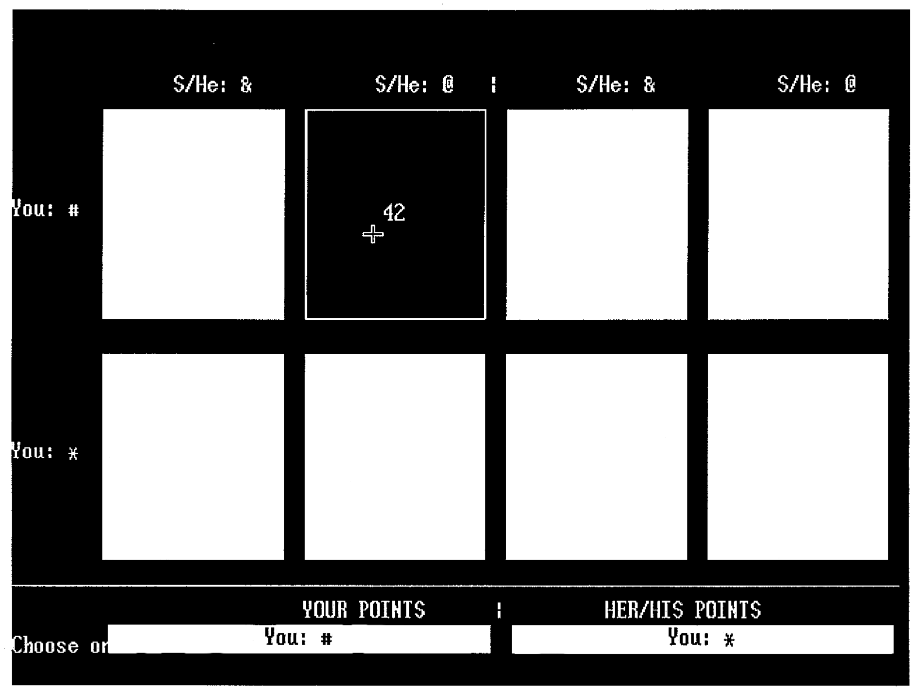
They find few or none of either, relative to their initially specified list; but this is not a completely satisfactory way to proceed.

In CGCB's display, a subject's searches can vary in three main dimensions:

The extent to which his transitions are up-down in his own payoffs, which under Occurrence and Adjacency is (for a Row player) associated with decision-theoretic rationality;

The extent to which his transitions are left-right in other's payoffs, which under Occurrence and Adjacency is associated with thinking about his partner’s incentives;

The extent to which he makes transitions from own to other's payoffs and back for the same decision combination, which under Occurrence and Adjacency is associated with interpersonal fairness or competitiveness comparisons.



Incorporating search compliance into the analysis shifts CGCB’s estimated type distribution toward *L1* (*Naïve*), at the expense of *Optimistic* (maximax) and *D1*.

Part of the shift occurs because *L1*’s search implications explain more of the variation in subjects’ searches and decisions than *Optimistic*'s, which are too unrestrictive to be useful in the sample.

Another part of the shift occurs because *L1*’s search implications explain more of the variation in subjects’ searches and decisions than *D1*’s, which are more restrictive, but very weakly correlated with subjects’ decisions.

D1 also loses some frequency to *L2*, even though their decisions are weakly separated in CGCB’s design, because *L2*’s search implications explain much more of the variation in subjects' searches and decisions.

Overall, CGCB's analysis of decisions and search yields a significantly different interpretation of behavior than their analysis of decisions alone.

The analysis including search suggests a strikingly simple view of behavior, with *L1* (*Naïve*) and *L2* making up 65-90% of the population, and *D1* making up 0% if one believes CGCB's model of search or 20% if not. (CGC’s subsequent work suggests 0% is right.)

CGC's Two-Person Guessing Games

CGC's subjects played a series of 16 different two-person guessing games.

In the games, each player has his own 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 (a trick to enhance separation of rules via search).

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

The targets and limits vary independently across players and 16 games, with the targets either both less than one, both greater than one, or mixed.

The 16 games subjects played are finitely dominance-solvable in 3 to 52 rounds, with essentially (because the only thing about a guess that matters is its adjusted guess) unique equilibria determined by the targets and limits in a simple way.

Consider a sample game where a player's own limits and target are [300, 900] and 1.5 and his partner's limits and target are [100, 900] and 0.5.

The product of targets $1.5 \times 0.5 < 1$, which is easily shown to imply that players' equilibrium adjusted guesses are determined (at least indirectly) by their lower limits.

The player's equilibrium adjusted guess equals his lower limit of 300, but his partner's equilibrium adjusted guess is above his lower limit at 150.

The way in which equilibrium is determined, by players' lower limits when the product of their targets is less than 1, or by players' upper limits when the product of their targets is greater than 1, enhances separation of *Equilibrium* from other types:

Equilibrium responds very strongly to small differences in the product of the targets, while other, empirically plausible types are almost completely unmoved by them.

That equilibrium is jointly determined by both players' payoff parameters also helps to separate the search implications of equilibrium and other rules.

The large strategy spaces and variation of targets and limits in CGC's design yields very strong separation of decisions, akin to strategic "fingerprinting":

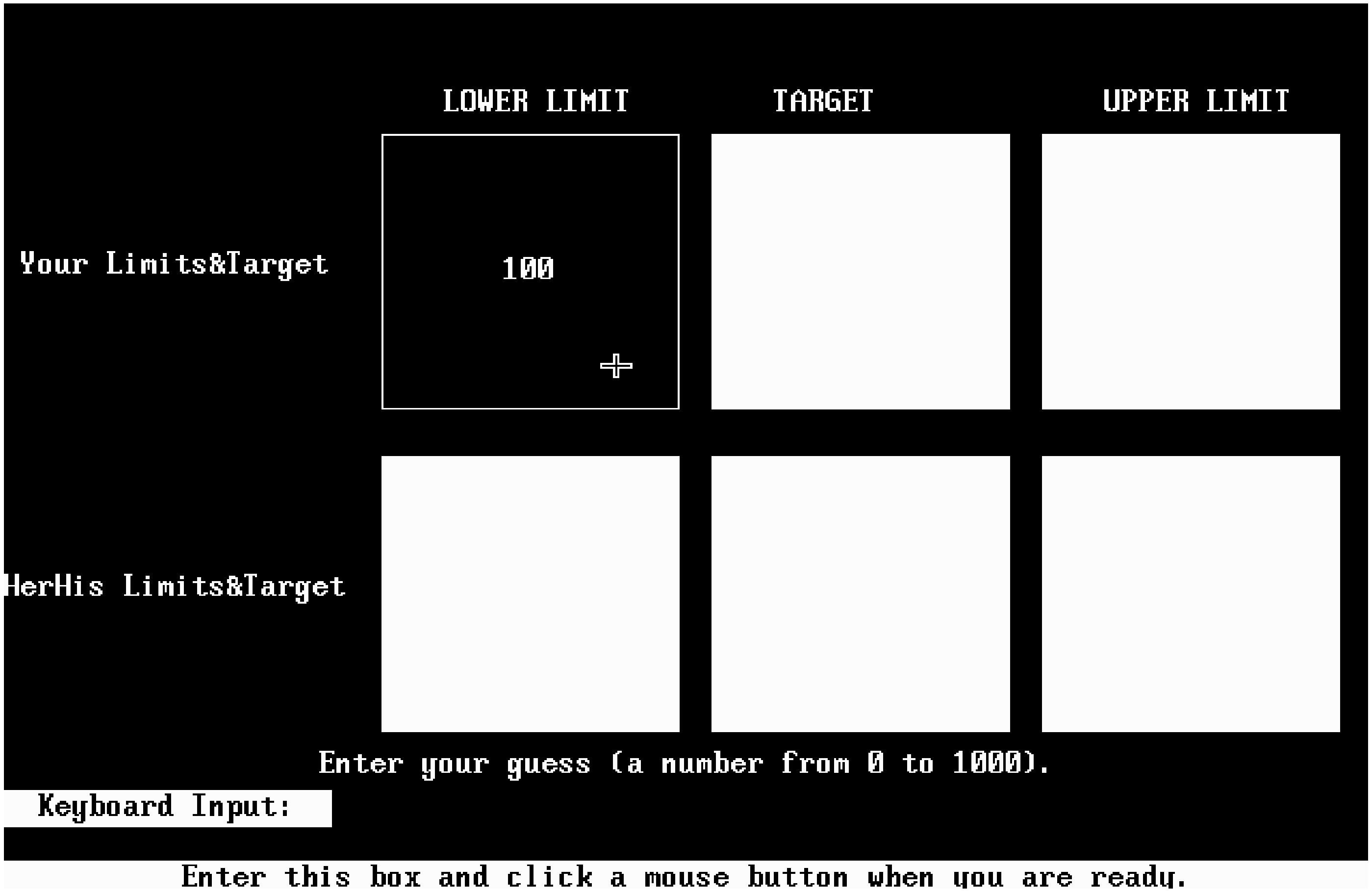
Types' guesses in the 16 games, in (randomized) order played

	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

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

CGC's other 45 subjects' types are less apparent from their guesses; but *L1*, *L2*, *L3*, and *Equilibrium* are still the only ones that show up in econometric estimates.

Following CJ and CGCB, within a publicly announced structure CGC presented each game to subjects via a MouseLab interface that normally concealed the targets and limits but allowed subjects to look them up as often as desired, one at a time.



CGC's Figure 6. Screen Shot of the MouseLab Display

Types as Models of Cognition, Search, and Decisions

As in CGCB's analysis, search is multidimensional, and a useful search analysis depends on using a pre-specified list of types to structure the space of possible searches and relate search to decisions.

Like CGCB's and CJ's, CGC's model of cognition, search, and decisions is based on a procedural/algorithmic view of decision-making, in which each subject follows one type in each game he plays, his type determines his search, and his type and search then determine his decision.

Each type is naturally associated with algorithms that process information into decisions.

The analysis uses these algorithms as models of cognition, deriving search implications under simple, empirically motivated assumptions about how cognition determines search.

Types' minimal search implications in CGC's games can be derived from their *ideal guesses*, those they would make if they had no limits. (With automatic rounding of guesses to fall within their limits, and quasiconcave payoffs, ideal guesses are all they need to know, and all that matters for minimal restrictions.)

Evaluating a formula for a type's ideal guess requires a series of *operations*, some of which are *basic* in that they logically precede any other operation.

CGC (and CGCB) derived types' search implications under the assumptions that subjects perform basic operations one at a time via adjacent look-ups, remember their results, and otherwise rely on repeated look-ups rather than memory.

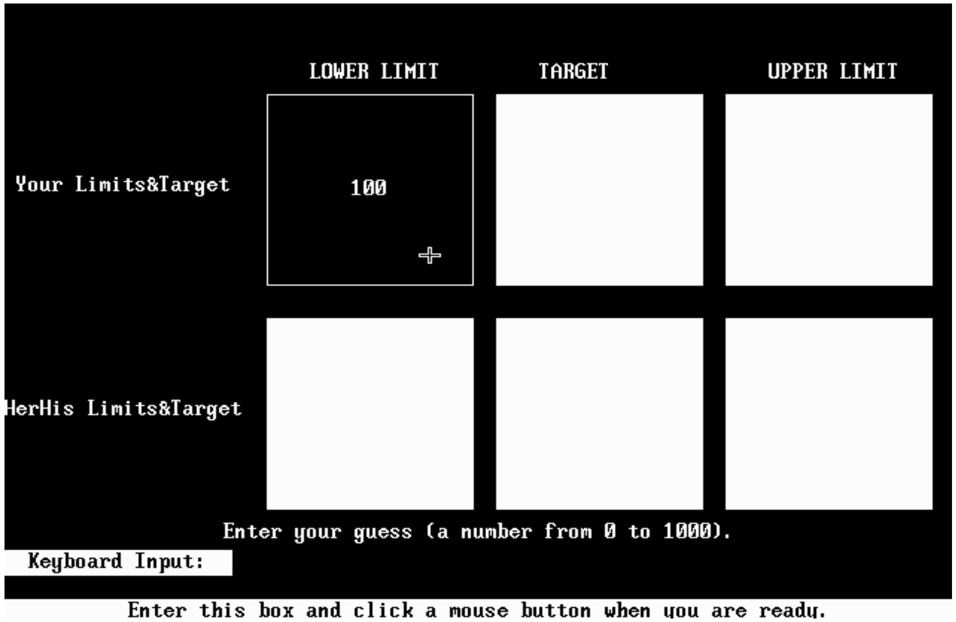
Basic operations are then represented by adjacent look-up pairs that can appear in either order, but cannot be separated by other look-ups.

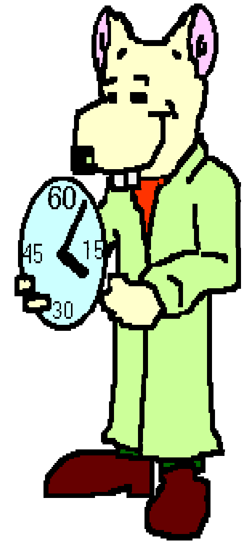
E.g. $[a^j+b^j]$ (averaging the partner's limits) is the only basic operation for *L1*'s ideal guess, $p^i[a^j+b^j]/2$. Such pairs are grouped within square brackets, as in $\{[a^j, b^j], p^i\}$ for *L1*.

Other operations can appear in any order and their look-ups can be separated. They are represented by look-ups grouped within curly brackets or parentheses.

It is easier to use this and other types' derivations to interpret the search data by translating them from CGC's notation into the box numbers MouseLab records:

SPEAK RODENT LIKE A NATIVE IN ONE EASY LESSON!





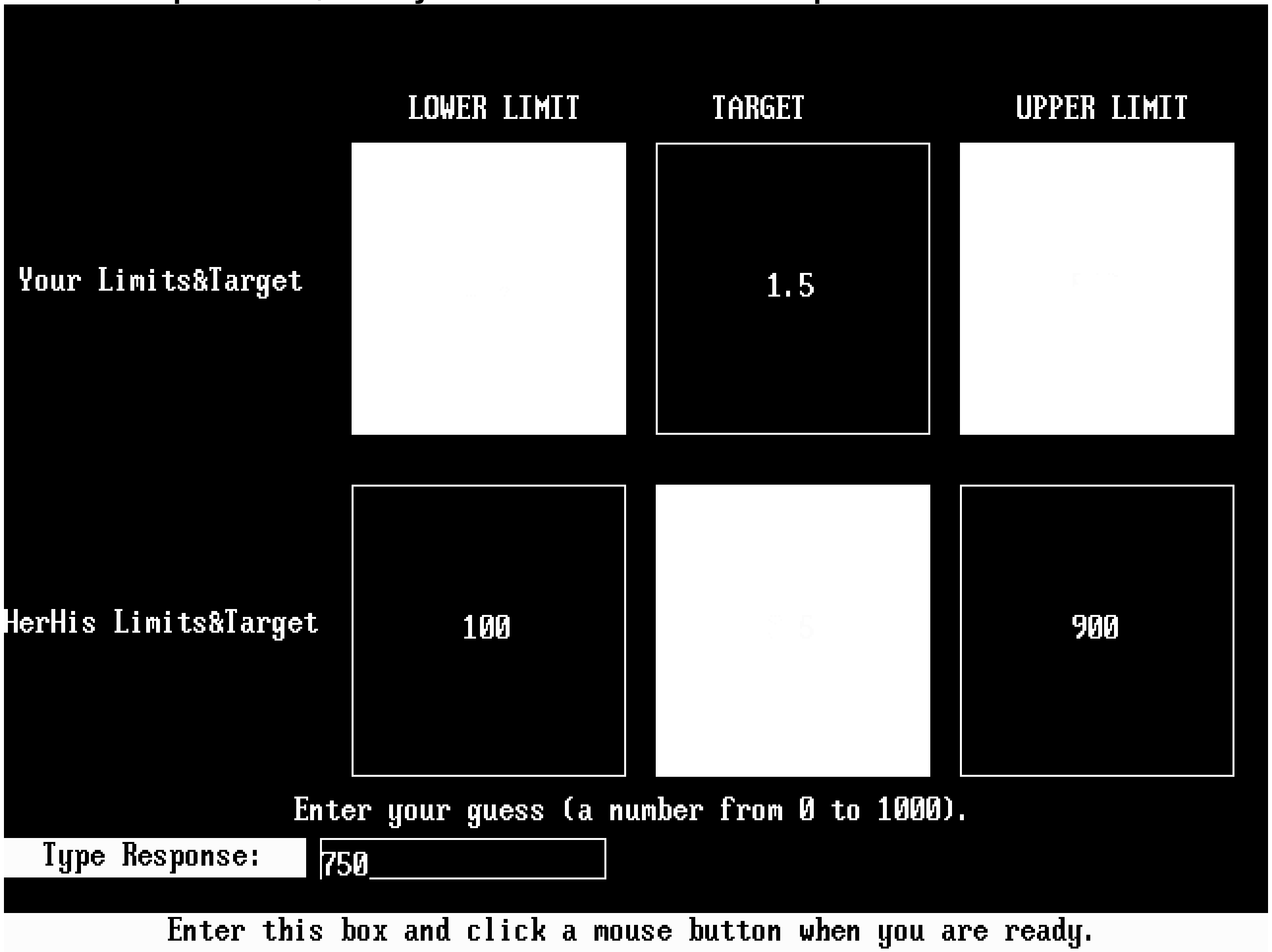
	a	***p***	***b***
You (*i*)	1	2	3
S/he (*j*)	4	5	6

MouseLab Box Numbers

For example, *Equilibrium*'s ideal guess is p^ia^j if $p^ip^j < 1$ or p^ib^j if $p^ip^j > 1$, and its search implications are $\{[p^i,p^j],a^j\} \equiv \{[2, 5], 4\}$ if $p^ip^j < 1$ or $\{[p^i,p^j],b^j\} \equiv \{[2, 5], 6\}$ if $p^ip^j > 1$: in this design, *Equilibrium*'s search implications are theoretically simpler than all but *L1*'s.

L1's search implications

(Note: Unlike in this picture, subjects could never open more than one box at a time.)

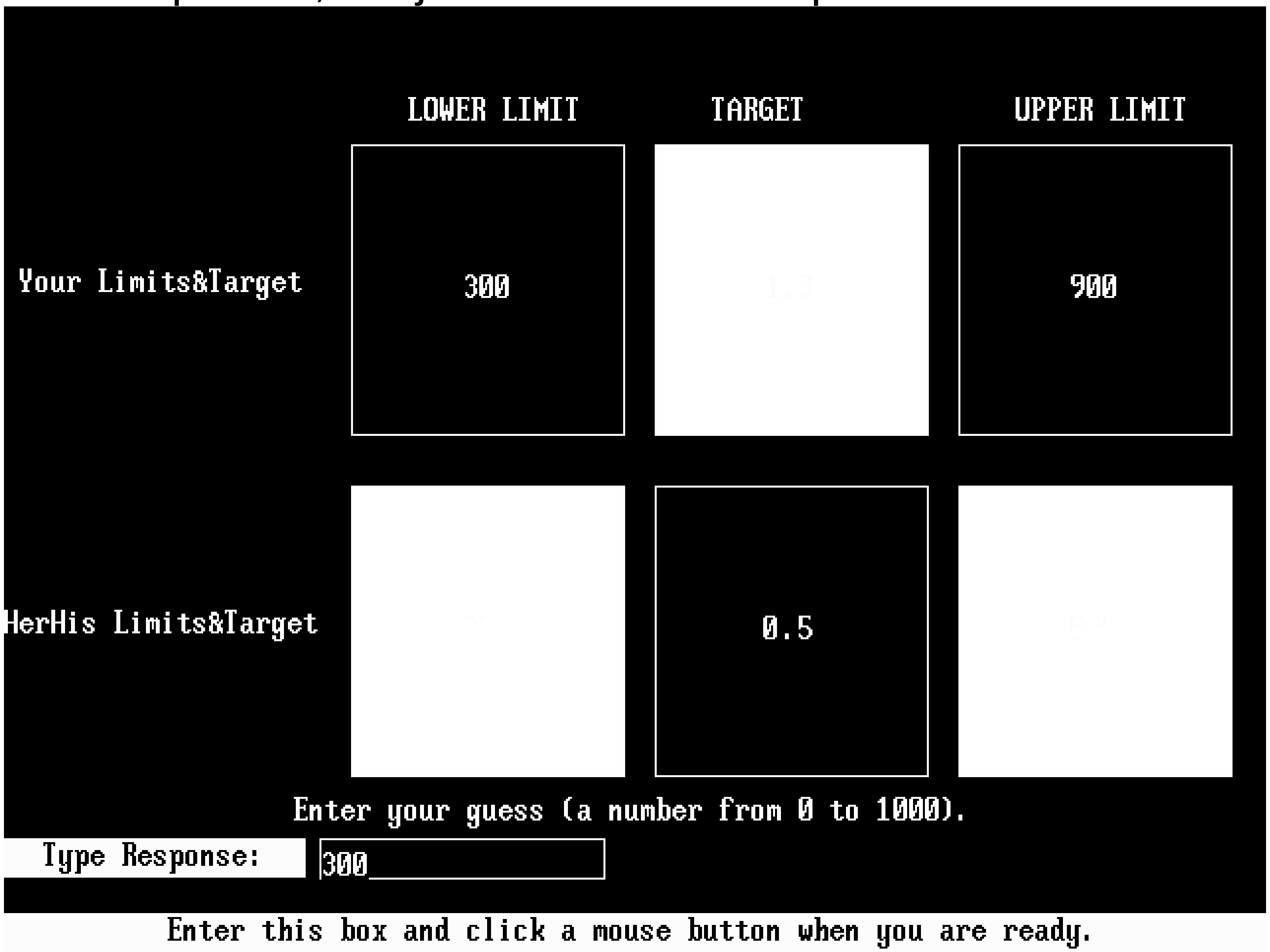


L1's ideal guess: $p^i[a^j+b^j]/2 = 750$. *L1*'s search implications: $\{[a^j, b^j], p^i\} \equiv \{[4, 6], 2\}$.

(*L1* does not need to look up its own limits because it can enter its ideal guess and rely on automatic adjustment to ensure that its adjusted guess is optimal. Thus this design even separates *L1* from a *Solipsistic* type that only looks up its own parameters.)

L2's search implications: first step

(Note: Unlike in this picture, subjects could never open more than one box at a time.)

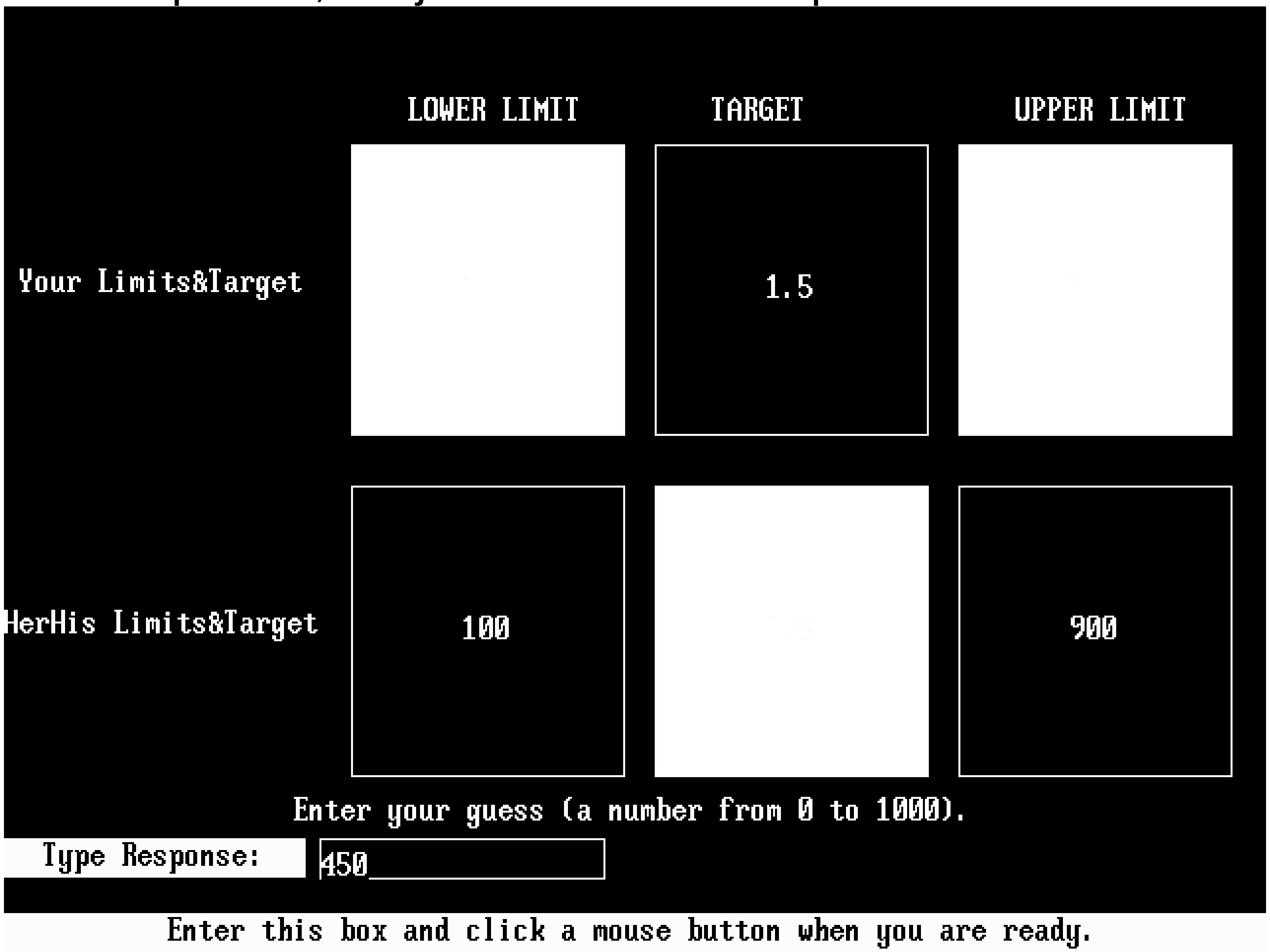


L2's model of its partner's *L1* guess: $p^j[a^j+b^j]/2 = 300$.
Search implications: $\{[a^i,b^i],p^j\} \equiv \{[1,3],5\}$.

(*L2* needs to look up its own limits only to predict its partner's *L1* guess; like *L1* it can enter its ideal guess and rely on automatic adjustment to ensure its adjusted guess is optimal.)

L2's search implications: second step

(Note: Unlike in this picture, subjects could never open more than one box at a time.)



L2's ideal guess: $p^iR(a^j,b^j;\, p^j[a^i+b^i]/2) = 450$.

L2's search implications: $\{([a^i,b^i],p^j),a^j,b^j,p^i\} \equiv \{([1, 3], 5), 4, 6, 2\}$.

(*L2* needs to look up its partner's limits $a^j = 4$ and $b^j = 6$ to predict its partner's *L1 adjusted* guess.)

Search data for representative R/TS and Baseline subjects, chosen for high compliance with their type's guesses, not for their compliance with any theory of search, suggest that:

There is little difference between the look-up sequences of R/TS and Baseline subjects of a given type (assigned for R/TS, apparent for Baseline).

The relevant sequences for a type (Table 4) are unusually dense in the sequences of subjects of that type (assigned or apparent), at least for the simpler types.

We can quickly learn to read the algorithms many subjects are using directly from the data, to the point where their types can be reliably identified from search alone.

The *Equilibrium* and *D2* subjects are clearly stressed out, yet they usually "get it right".

For some subjects, search is an important check on decisions; e.g. Baseline subject 309, with 16 exact *L2* guesses, misses some of *L2*'s relevant look-ups, avoiding deviations from *L2* only by luck (even without feedback, s/he later has a Eureka! moment between games 5 and 6, and from then on complies perfectly; reminiscent of CJ's subjects who never looked at the last period pie and so could not have been performing the backward induction needed to identify subgame-perfect equilibrium).

Perhaps more instructive is Baseline subject 415 (not shown), who is plainly an *L1* who fails Adjacency because he can comfortably remember three numbers at a time (CGC fn. 43), and is therefore misclassified as a very noisy *D1* in CGC's search analysis: he is the only clear failure of CGC's model of cognition and search among 71 Baseline subjects.

Table 10.2. Selected Robot/Trained Subjects' Information Searches.

Subject	Type/Alt[a]	Game 1[b]	Game 2[b]
904	L1 (16)	1234564623	1234564321
1716	L1 (16)	14646213464623	46246213
1807	L1 (16)	462513	46213225
1607	L2 (16)	1354621313	1354613546213
1811	L2 (16)	1344465213*46	13465312564231356252
2008	L2 (16)	1113131313135423	13131356622333
1001	L3 (16)	46213521364*24623152	4621356425622231462562*62
1412	L3 (16)	1462315646231	462462546231546231
805	D1 (16)	1543564232132642	51453561536423
1601	D1 (16)	25451436231	5146536213
804	D1 (3)/L2 (16)	1543465213	5151353654623
1110	D2 (14)	1354642646*313	135134642163451463211136 414262135362*146546
1202	D2 (15)	246466135464641321342462 4226461246255*1224654646	123645132462426262241356 462*135242424661356462
704	DEq (16)	123456363256565365626365 6526514522626526	123456525123652625635256 262365456
1205	Eq (16)	1234564246525625256352*465	123456244565565263212554 14666265425144526*31
1408	Eq (15)	12312345644563213211	123456456123635241
2002	Eq (16)	142536125365253616361454 61345121345263	1436253614251425236256563

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

Table 10.3. Selected Baseline Subjects' Information Searches.

Subject	Type/Alt[a]	Game 1[b]	Game 2[b]	Game 3[b]
101	L1 (15)	146246213	46213	462*46
118	L1 (15)	24613462624132*135	2462622131	246242466413*426
413	L1 (14)	1234565456123463*	12356462213*	264231
108	L2 (13)	135642	1356423	1356453
206	L2 (15)	533146213	53146231	5351642231
309	L2 (16)	1352	1352631526*2*3	135263
405	L2 (16)	14465231331254623212512	13245625315645654546312315656262	3124565231*12365455233**513
210	L3 (9) Eq (9) D2(8)	123456123456213456 254213654	123456465562231654456*2	123455645 6123
302	L3 (7) Eq (7)	221135465645213213 45456*541	2135465662135454 6321*26654123	265413232145563214 563214523*654123
318	L1 (7) D1 (5)	1324564652521324 2* 1462	132465132*462	1346521323*4
417	Eq (8) L3 (7) L2 (5)	2525314646564465 31 6412524621213	25523662*3652435 63	5213636415265263* 652
404	Eq (9) L2 (6)	462135464655645515 21354*135462426256 35623413135464 5	4624613525242613 1 5463562	462135215634*52
202	Eq (8) D2 (7) L3 (7)	123456254613621342 *525	1234564456132554 6251356523	1234561235623
310	Eq (11)	123126544121565421 254362*21545 4*	1235462163262314 56*62	123655463213
315	Eq (11)	213465624163564121 325466	1346521246536561 213	13246554416 3*3625

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

Two puzzles we hope the search analysis will help to resolve:

What are Those Baseline "*Equilibrium*" Subjects Really Doing?

Consider CGC's eight Baseline subjects with near-*Equilibrium* fingerprints

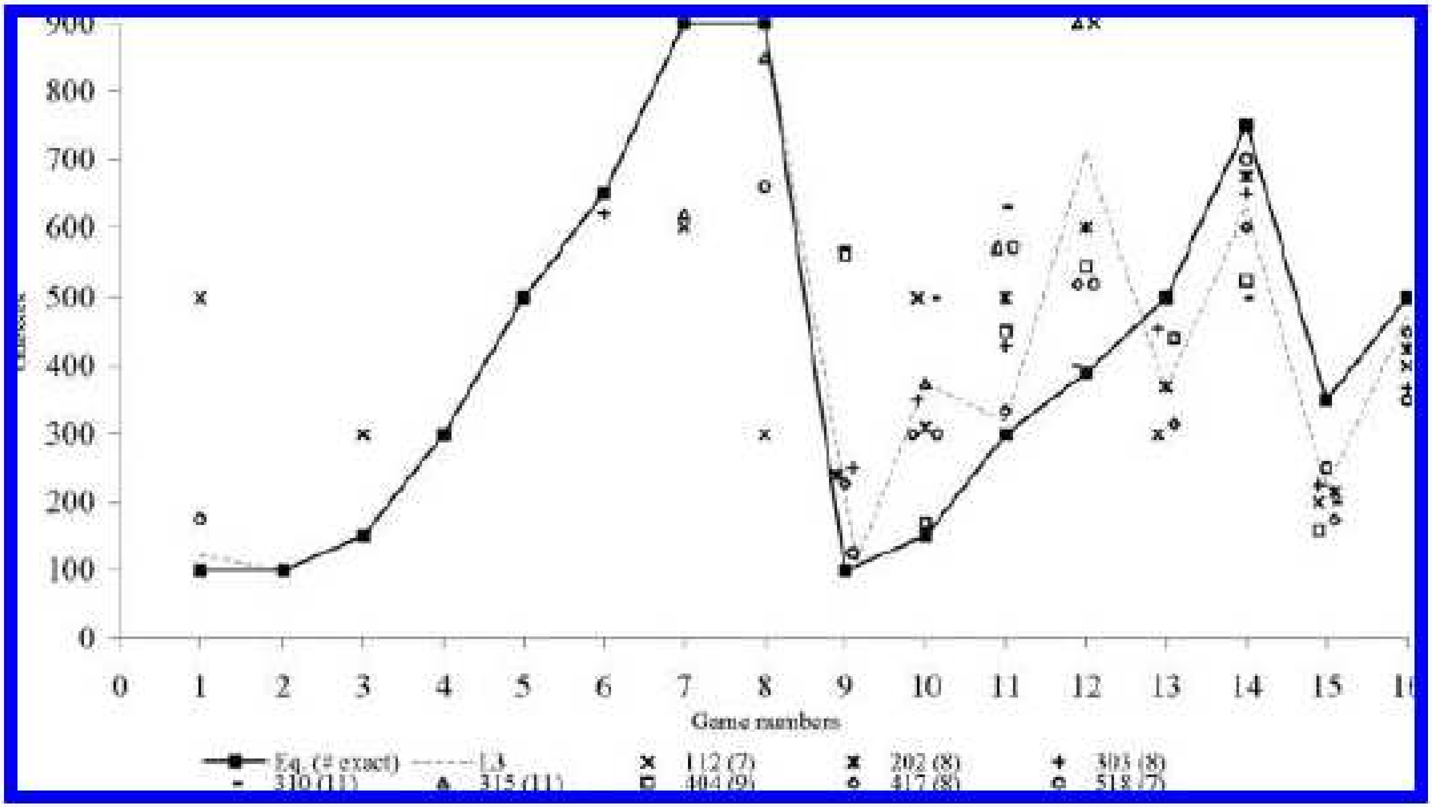


FIGURE 4. "FINGERPRINTS" OF EIGHT APPARENT *EQUILIBRIUM* SUBJECTS

Notes: Only deviations from *Equilibrium*'s guesses are shown. Of these subjects' 128 guesses, 69 (54 percent) were exact *Equilibrium* guesses.

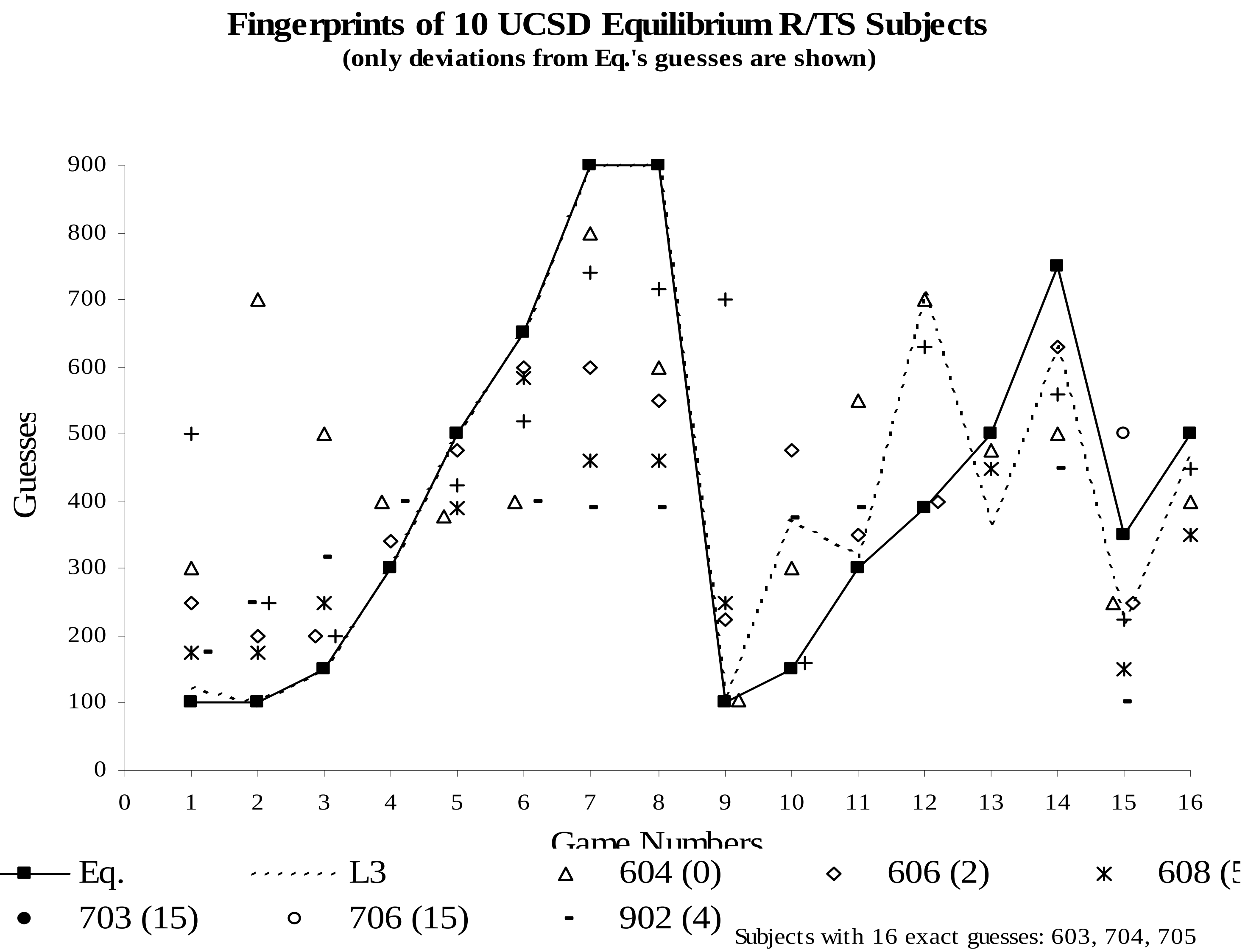
Ordering the games by strategic structure as in Figure 4, with the 8 games with mixed targets (one > 1, one < 1) on the right, shows that these subjects' deviations from equilibrium occur almost exclusively with mixed targets.

Thus these subjects, whose exact compliance with *Equilibrium* guesses is off the scale by any normal standard, 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, iterated dominance) work just as well with mixed targets.

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

Equilibrium R/TS subjects' compliance is equally high with and without mixed targets, so training eliminates whatever the Baseline subjects are doing:



Why are *Lk* the only non-*Equilibrium* types that exist?

A careful analysis of CGC's decision data, including specification tests not described here, reveals many subjects of types *L1*, *L2*, *Equilibrium*, or hybrids of *L3* and/or *Equilibrium*, but no other types that do better than a random model of guesses for more than one subject.

Why, out of the enormous number of possibilities do these rules predominate?

Why, for instance, don't we get *Dk* rules, which are closer to what we teach?

Answering this question may be the key to a deeper theory of bounded rationality.

(i) Most R/TS subjects could reliably identify their type's guesses, even *Equilibrium* or *D2*. (These average rates are for exact compliance, and so are quite high. Individual subjects' compliance was usually bimodal within type, on very high and very low.)

R/TS Subjects' Exact Compliance with Assigned Type's Guesses and Duration

	L1	***L2***	***L3***	***D1***	***D2***	***Eq.(N/A)***
Number of subjects	25	27	18	30	19	29
% Compliance|Passed UT2	80.0	91.0	84.7	62.1	56.6	70.3
% Failed UT2	0.0	0.0	0.0	3.2	5.0	19.4
Duration (seconds)	45.4	54.9	79.2	77	120.5	96.3

(ii) But there are noticeable signs of differences in difficulty across types:

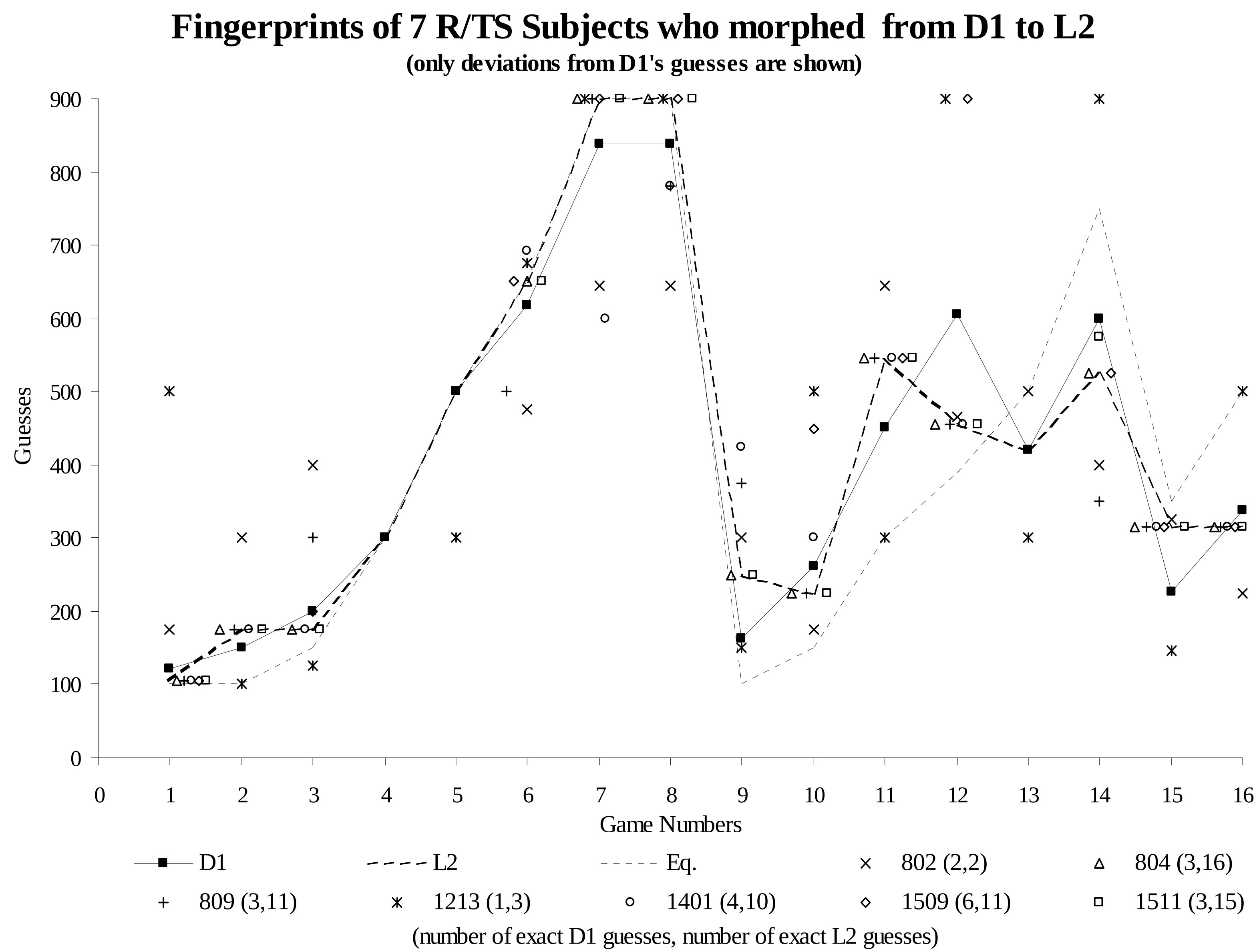
(a) No one ever failed an *Lk* Understanding Test, while some failed the *Dk* and many failed the *Equilibrium* Understanding Test.

(b) For those who passed, compliance was highest for *Lk* types, then *Equilibrium*, then *Dk* types. This suggests that *Dk* is even harder than *Equilibrium*, but could just be an artifact of the more stringent screening of the *Equilibrium* Test.

(c) Among *Lk* and *Dk* types, compliance was higher for lower *k* as expected, except *L1* was lower than *L2* or *L3* compliance. (We suspect that this is because *L1* best responds to a random *L0* robot, which some subjects think they can outguess; while *L2* and *L3* best respond to a deterministic *L1* or *L2* robot.)

(d) Remarkably, 7 of 19 R/TS *D1* subjects passed the *D1* Understanding Test, in which *L2* answers are wrong; and then "morphed" into *L2*s when making their guesses, significantly reducing their earnings. E.g. R/TS *D1* subject 804 made 16 exact *L2* (and so only 3 exact *D1*) guesses. (Recall that it is *L2* that is *D1*'s cousin.) This kind of morphing, in this direction, is the only kind that occurred. We view this as pretty compelling evidence that *Dk* types are unnatural.

Perhaps level-*k* thinking is the most workable effective model of others' decisions.


Fingerprints of 7 R/TS Subjects who morphed from D1 to L2
(only deviations from D1's guesses are shown)
Guesses
Game Numbers
D1
L2
Eq.
802 (2,2)
804 (3,16)
809 (3,11)
1213 (1,3)
1401 (4,10)
1509 (6,11)
1511 (3,15)
(number of exact D1 guesses, number of exact L2 guesses)