CONFERENCE ON THE FOUNDATIONS OF POSITIVE AND NORMATIVE ECONOMICS
NEW YORK UNIVERSITY, 25-26 APRIL 2008

LOOK-UPS AS THE WINDOWS OF THE STRATEGIC SOUL: STUDYING COGNITION VIA INFORMATION SEARCH IN GAME EXPERIMENTS

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.

Introduction

In “The Case for Mindless Economics,” Gul and Pesendorfer (henceforth “GP”) argue that because economic theory was meant to explain only decisions, it should only be tested by observing decisions; and that the data should be analyzed via revealed preference.

GP’s argument tacitly assumes that the task of empirical economics is limited to estimating preferences as revealed by rational individual decisions.

Most critiques of GP’s argument also focus on individual decisions.

In “Revealed Preferences and Revealed Mistakes,” for instance, Kőszegi and Rabin argue that GP’s proposal is too narrow because some important economic questions rest on inferences about probabilistic judgment that cannot be drawn via revealed preference.

Here, instead, I explore the implications of GP’s proposal for testing theories of strategic behavior, focusing on strategic thinking as revealed by initial responses to games.

Adapting GP's proposal to games requires limiting the empirical task to estimating preferences.

This, in turn, requires assuming (and not testing) a theory of strategic behavior that makes unique predictions, presumably equilibrium (suitably refined if there are multiple equilibria).

But in recent experiments that elicit initial responses to games, such as Stahl and Wilson (*JEBO* 1994, *GEB* 1995); Nagel (*AER* 1995); Ho, Camerer, and Weigelt (*AER* 1998); CGCB, and CGC, most subjects deviated systematically from equilibrium.

CGC's experiments, for instance, elicited subjects' initial responses to a series of different but related guessing games with large strategy spaces.

Their subjects' decision patterns show clearly that their deviations from equilibrium can be attributed to non-equilibrium strategic decision rules that best respond to simplified models of other players, rather than to irrationality, risk aversion, altruism, spite, or confusion.

Just as Kőszegi and Rabin's analysis partly shifts the empirical focus in analyzing individual decisions from estimating preferences to inferences about probabilistic judgment that cannot be drawn via revealed preference alone, CGC's and previous analyses of initial responses to games partly shifts the focus from estimating preferences to inferences about strategic thinking that cannot be drawn via revealed preference alone.

CGC's analysis also allows a concrete assessment of the cost of limiting the analysis to decision data and revealed preference methods.

The evidence from CGC's, CGCB's, and previous experiments suggests that a large fraction of subjects' systematic deviations from equilibrium is well explained by a structural non-equilibrium model based on "level-*k* thinking"—or a "cognitive hierarchy" model, as Camerer, Ho, and Chong (*QJE* 2004) call their closely related model.

(The evidence also suggests that level-*k* models can out-predict "equilibrium with noise" models with payoff-sensitive error distributions, such as quantal response equilibrium.)

In principle, the structure of subjects' non-equilibrium decision rules (more precisely, their population distribution) could be estimated from decisions alone—though not via revealed preference—by generating enough data from a sufficiently powerful experimental design.

But CGC's design is already quite powerful from the standpoint of studying decisions alone; even so, it leaves open some questions regarding subjects' decision rules, as will be explained.

How should we respond to the demonstrated limitations of studying decisions alone?

If decision data were free, it might be a good research strategy to try to address open research questions just by gathering more decision data, perhaps in new environments.

But decision data are far from free, and existing methods for gathering them are fairly easily adapted to gather process data at the same time.

Given this, exclusive reliance on gathering more decision data seems unlikely to be optimal: Good research strategies seem more likely to leave the door open to process data, along with decision data, even if this requires non-revealed-preference methods.

Accordingly, following CGCB and earlier work by Camerer, Johnson, et al. (1993, *JET* 2002), CGC studied process data by using a MouseLab interface to monitor subjects' searches for hidden but freely accessible payoff information.

CGCB and CGC then used an explicit but rudimentary model of cognition to analyze subjects' information searches along with their decisions.

Their analyses illustrate how using process data along with decision data can help to identify the rules that govern subjects' decisions more precisely.

With a sufficiently powerful design, search data sometimes even directly reveals the algorithms subjects use to choose their decisions, making it possible to predict decisions without ever observing them.

Incorporating neural data into analyses of games and decisions is probably harder than incorporating search data.

But this demonstration of the usefulness of studying decision processes as well as decisions may bring us closer to agreement on how (or whether) to do neuroeconomics.

Aside: Why study strategic thinking or initial responses to games when even unthinking people will eventually converge to equilibrium anyway?

Equilibrium is often a reliable model of limiting behavior when people have had enough experience with analogous games to learn to predict each others' responses.

But many settings involve games without clear precedents, and in such games people's initial responses often deviate systematically from equilibrium.

Even when learning assures eventual convergence to equilibrium, when there are multiple equilibria initial responses often determine equilibrium selection.

Initial responses reveal strategic thinking in its purest form, "uncontaminated" by learning.

(By contrast, with enough feedback in a sufficiently stationary setting, even pigeons will eventually learn to mimic the decisions that follow from sophisticated strategic thinking.)

Studying initial responses allows us to identify structural non-equilibrium models of initial responses that can out-predict equilibrium or quantal response equilibrium.

It also yields information that is helpful in determining the structure of learning rules.

Outline

In the rest of the talk, I first describe CGC's design and their results for decisions.

I next discuss how CGC monitored subjects' searches for hidden but freely accessible information about payoffs.

I then discuss CGC's analysis of cognition and search, which rests on assumptions that stylize empirical regularities in how subjects search.

Finally, I highlight two puzzles left open by CGC's analysis of decisions (and likely to continue to resist analysis via decisions alone), and sketch routes to resolutions via analyzing search and other process data.

More detail is given in my paper for the conference volume, and in the slides for my 2006 conference talk, posted at http://dss.ucsd.edu/~vcrawfor/NYUGuessSearchTalk.pdf.

CGC's Two-Person Guessing Game Experiments

CGC's experiments randomly and anonymously paired subjects to play series of different but related twp-person guessing games, with no feedback between games.

The design suppresses learning from experience and repeated-game effects in order to elicit subjects' initial responses, game by game.

The goal is to focus on how people model others' decisions by studying strategic thinking "uncontaminated" by learning from experience.

"Eureka!" learning remains possible, but it can be tested for and is rare.

(The results yield insights into cognition that also help us think about how to model learning from experience, but that's another story.)

In CGC's guessing 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.

(In previous guessing experiments such as Nagel's (*AER* 1995) and Ho, Camerer, and Weigelt's (*AER* 1998), the targets and limits were always the same for both players, and they varied only across treatments, or not at all.)

The 16 games subjects played are finitely dominance-solvable in 3-52 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 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 here, 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, is general in CGC's guessing games.

CGC's design exploits the discontinuity of the equilibrium correspondence when the product of targets is 1 by including some games that differ mainly in whether the product is slightly greater, or slightly less, than 1.

Equilibrium responds very strongly to such differences, but empirically plausible non-equilibrium decision rules are almost completely unmoved by them.

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

Leading Strategic Decision Rules or *Types*

CGC's analysis of decisions, like Stahl and Wilson's (*JEBO* 1994, *GEB* 1995); Nagel's (*AER* 1995); Ho, Camerer, and Weigelt's (*AER* 1998); and CGCB's, uses a structural non-equilibrium model of initial responses in which each subject's decisions are determined by one of several decision rules or *types* (as they are called in this literature).

The leading types play a central role in CGC's and CGCB's model of cognition, search, and decisions, which takes a procedural view of decision-making, in which a subject's type determines his search and his type and search determine his decision.

CGC's types, which all build in risk-neutrality and rule out social preferences, include:

L1, which best responds to a uniform random *L0* "anchoring type".

(*L0* is meant to represents a subject's instinctive, nonstrategic reaction to the game—more precisely, other subjects' models of a subject's instinctive—and usually has zero estimated population frequency.)

L2 (*L3*), which best responds to *L1* (*L2*).

(*Lk* for $k > 0$ is rational, but deviates from equilibrium because it uses a simplified model of others' decisions. It is *k*-level rationalizable so coincides with equilibrium in games that are *k*-dominance solvable. With plausible type frequencies this yields an inverse relationship between strategic complexity and equilibrium compliance, as is often observed.)

CGC's types also include:

D1 (*D2*), which does one round (two rounds) of deletion of dominated decisions and then best responds to a uniform prior over the other's remaining decisions.

(By a quirk of our notation, *L2* is *D1*'s cousin, and *L3* is *D2*'s. Those pairs' guesses are perfectly confounded in Nagel's (*AER* 1995) games; and in two-person games *Lk* guesses are *k*-rationalizable, just as *Dk-1*'s are.)

Equilibrium, which makes its equilibrium decision.

Sophisticated, which best responds to the probabilities of other's decisions, proxied in CGC's analysis by their subjects' observed frequencies.

CGC also conducted a comprehensive specification test for excluded types, finding few or none.

CGC's Results for Decisions

The large strategy spaces and independent variation of targets and limits in CGC's design greatly enhance separation of types' implications for decisions, to the point where many subjects' types can be precisely identified from decisions alone. From CGC's Table 5:

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 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*).

Given how strongly types' guesses are separated, and that they could take 200-800 different rounded values, subjects' compliance is far higher than could occur by chance.

Further, because the types specify precise, well-separated guess sequences in a very large space of possibilities, their high exact compliance rules out (intuitively or econometrically) alternative interpretations of their behavior.

In particular, because the types build in risk-neutral, self-interested rationality and perfect models of the game, the deviations from equilibrium of the 35 subjects whose apparent types are *L1*, *L2*, or *L3* can be attributed to non-equilibrium beliefs rather than irrationality, risk aversion, altruism, spite, or confusion.

(In previous designs with small strategy spaces, even a perfect fit does not distinguish a subject's best-fitting type from nearby omitted types; and in previous guessing designs, with large strategy spaces but each subject playing one game, the ambiguity is worse.)

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.

Unlike the often-suggested interpretation of previous guessing results—that subjects are performing finitely iterated dominance—separating *Lk* from *Dk-1* reveals that *Dk* types don't exist in any significant numbers, at least in this setting.

Subjects' decisions respect finitely iterated dominance not because they are explicitly performing it, but because they are following *Lk* rules that implicitly respect it.

Sophisticated, which in this design is clearly separated from *Equilbrium*, also doesn't exist in significant numbers.

CGC's data also strongly resist an "equilibrium plus noise" or QRE interpretation.

Instead subjects' "errors" usually appear to be structural or cognitive, without the payoff-sensitivity a QRE interpretation requires.

Aside: Other applications of level-*k* models

Level-*k* models have been used to analyze data from experiments by Cai and Wang (*GEB* 2006); Wang, Spezio, and Camerer (2006); Sánchez-Pagés and Vorsatz (*GEB* 2007); Costa-Gomes and Weizsäcker (*RES* 2008); and Kawagoe and Tazikawa (*GEB* in press).

Level-*k* models have also been used in a number of applications, e.g.:

Crawford, "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions," *American Economic Review* 2003 (misleading communication via cheap talk in zero-sum two-person games)

Camerer, Ho, and Chong, "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics* 2004 (tacit coordination via structure and "magical" ex post coordination in market-entry games; speculation and zero-sum betting; money illusion in coordination)

Crawford and Iriberri, "Level-*k* Auctions: Can Boundedly Rational Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?," *Econometrica* 2007 (overbidding in common- and independent-private value auctions)

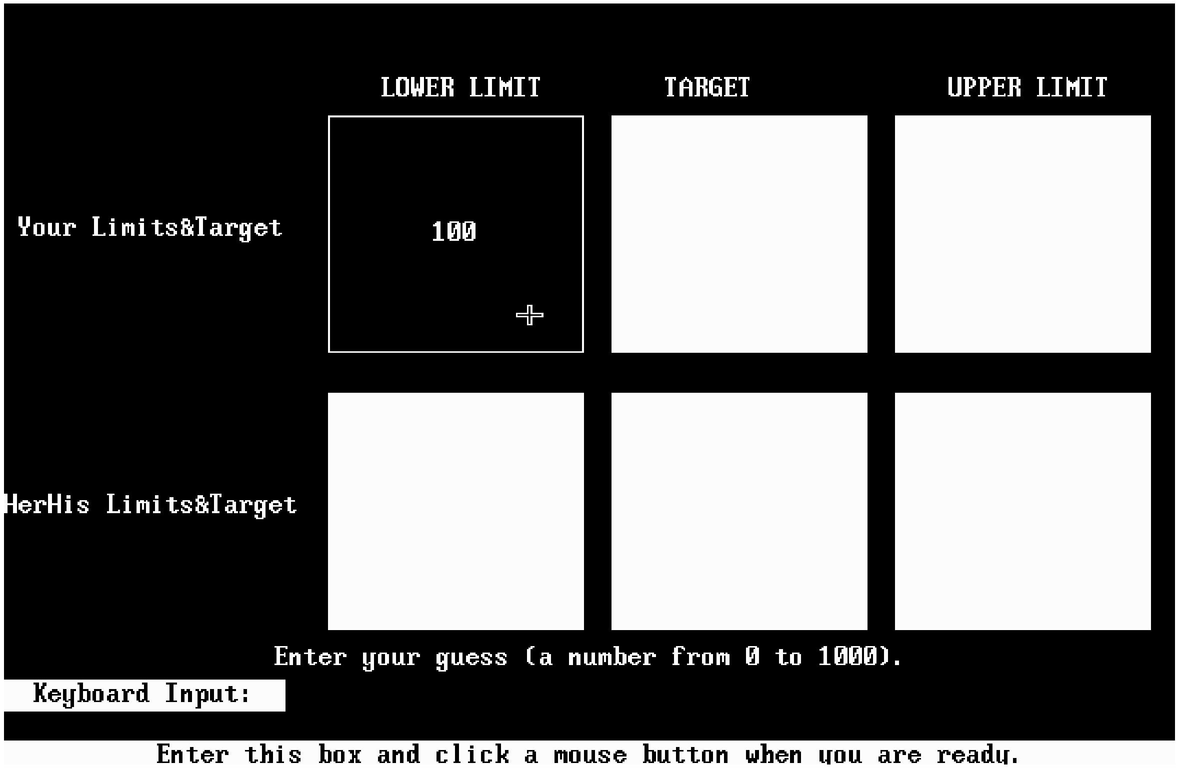
Crawford and Iriberri, "Fatal Attraction: Salience, Naivete, and Sophistication in Experimental Hide-and-Seek Games," *American Economic Review* 2007 (framing-induced deviations from unique mixed-strategy equilibrium)

Crawford, Gneezy, and Rottenstreich, "The Power of Focal Points is Limited: Even Minute Payoff Asymmetry May Yield Large Coordination Failures," *American Economic Review* in press (framing-induced miscoordination in Schelling-style coordination games)

Monitoring Search via MouseLab in Two-Person Guessing Games

Within a publicly announced structure, CGC presented each game to subjects via MouseLab (http://www.cebiz.org/mouselab/mouselab.htm).

The interface normally concealed the targets and limits but allowed subjects to look them up as often as desired, one at a time, by clicking on the boxes. (Subjects were not allowed to write, and the data strongly suggest that they did not memorize the parameters.)



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

With search costs as low as subjects' search patterns make them seem, free access makes the entire structure of the game effectively public knowledge, so the results can be used to test theories of behavior in complete-information versions of the games.

The design also maintains tight control over subjects' motives for search by making information from previous plays irrelevant to current payoffs.

Allowing subjects to search for a small number of hidden payoff parameters within a simple, publicly announced structure allows subjects to focus on the task of predicting others' responses without getting lost in details of the structure.

Independently separating the implications of leading decision rules for search and decisions makes it possible to study the relationship between them, which multiplies the power of the design to identify subjects' decision rules.

Allowing search patterns to vary in a high-dimensional space makes search more informative and allows greater separation of rules via search.

The simple parametric structure also makes leading rules' search implications (almost) independent of the game.

Aside: Where is the Information in Subjects' Searches?

Studying cognition via search requires a model of how cognition shows up in subjects' look-up sequences. Different papers take different positions on this:

Camerer, Johnson, et al. gave roughly equal weight to look-up durations and to the numbers of look-ups of each pie ("acquisitions") and the transitions between pies.

Rubinstein (*Economic Journal* 2007) considered only durations.

Camerer, Johnson, et al.'s and Rubinstein's analyses were also conducted at a high level of aggregation, both across subjects and over time.

Gabaix, Laibson, Moloche, and Weinberg (*American Economic Review* 2006) focused on numbers of look-ups (as opposed to durations) and considered aspects of their order too.

Gabaix et al. also conducted their analysis mostly at a high level of aggregation.

Studying Cognition via Numbers and Orders of Subjects' Look-ups

CGCB and CGC argued instead that cognition is sufficiently heterogeneous and search sufficiently noisy that they are best studied at the individual level.

CGCB and CGC also assumed that which look-ups subjects make, in which order, reveals at least as much information about cognition as durations or transition frequencies.

This should not be surprising, because simple theories of cognition more readily suggest roles for which look-ups subjects make, in which orders, than durations.

(CGC made no claim that durations are irrelevant, just that they don't deserve the priority they have been given. CGCB did present some results on durations, as "gaze times.")

Types as Models of Cognition, Search, and Decisions

CGC's (and CGCB's) models of cognition, search, and decisions are based on a procedural view of decision-making, in which a subject's type determines his search, and his type and search then determine his decision.

This procedural view is the key to linking cognition, search, and decisions in the analysis.

(Because a type's search implications depend not only on what decisions it specifies, but why, something like a types-based model seems necessary here.)

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

The analysis uses those algorithms as models of cognition, deriving a type's search implications under simple assumptions about how it determines search.

With their derived search implications, the types provide a kind of basis for the enormous space of possible decision and search sequences.

This imposes enough structure to describe subjects' behavior in a comprehensible way, and to make it meaningful to ask how decisions and searches are related.

How Does Cognition Determine Search?

Without further assumptions, nothing precludes a subject's scanning and memorizing the information and "going into his brain" to figure out what to do—in which case his searches will reveal nothing about his cognition.

(Neuroeconomics has an advantage over monitoring search here, because involuntary correlates of such a subject's thinking will still be observable.)

But 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.

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 meaning of the signal.

CGCB's and CGC's assumptions are conservative, in that they rest on types' minimal search implications and they add as little structure beyond these as possible.

The 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.)

Types' Search Implications

Types' search implications are derived as follows.

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.

E.g. $[a^j+b^j]$ (averaging the partner's limits) is the only basic operation for *L1*'s ideal guess, $p^i[a^j+b^j]/2$.

CGCB and CGC 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.

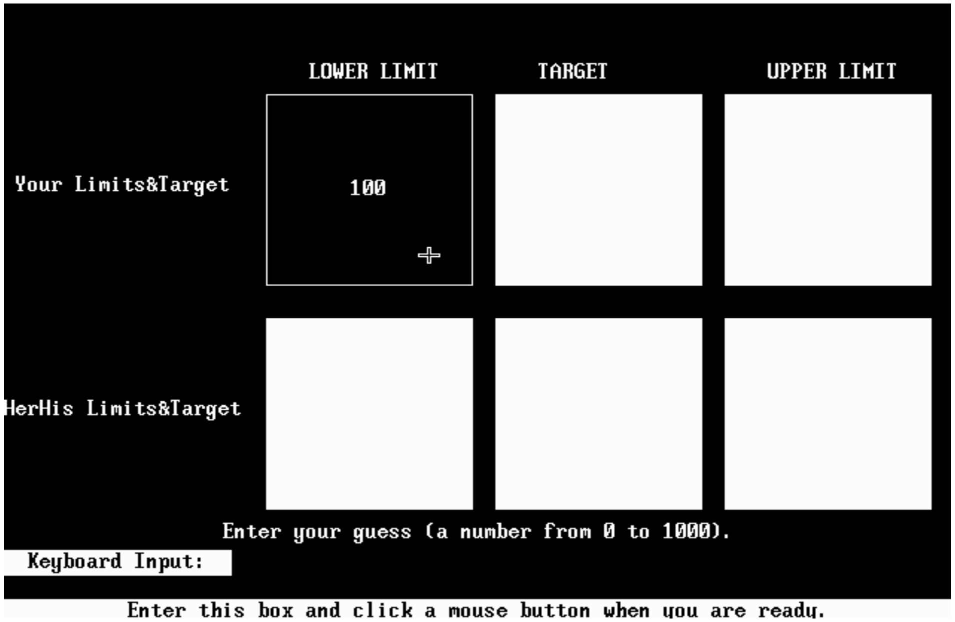
Such pairs are grouped within square brackets, as in $\{[a^j, b^j], p^i\}$ for *L1*.

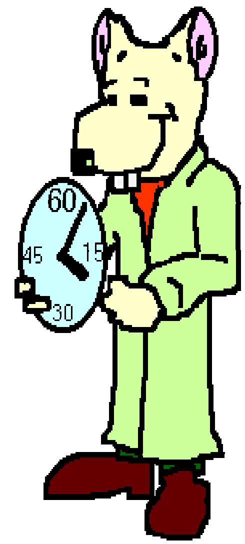
Other operations can appear in any order and their look-ups can be separated.

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!



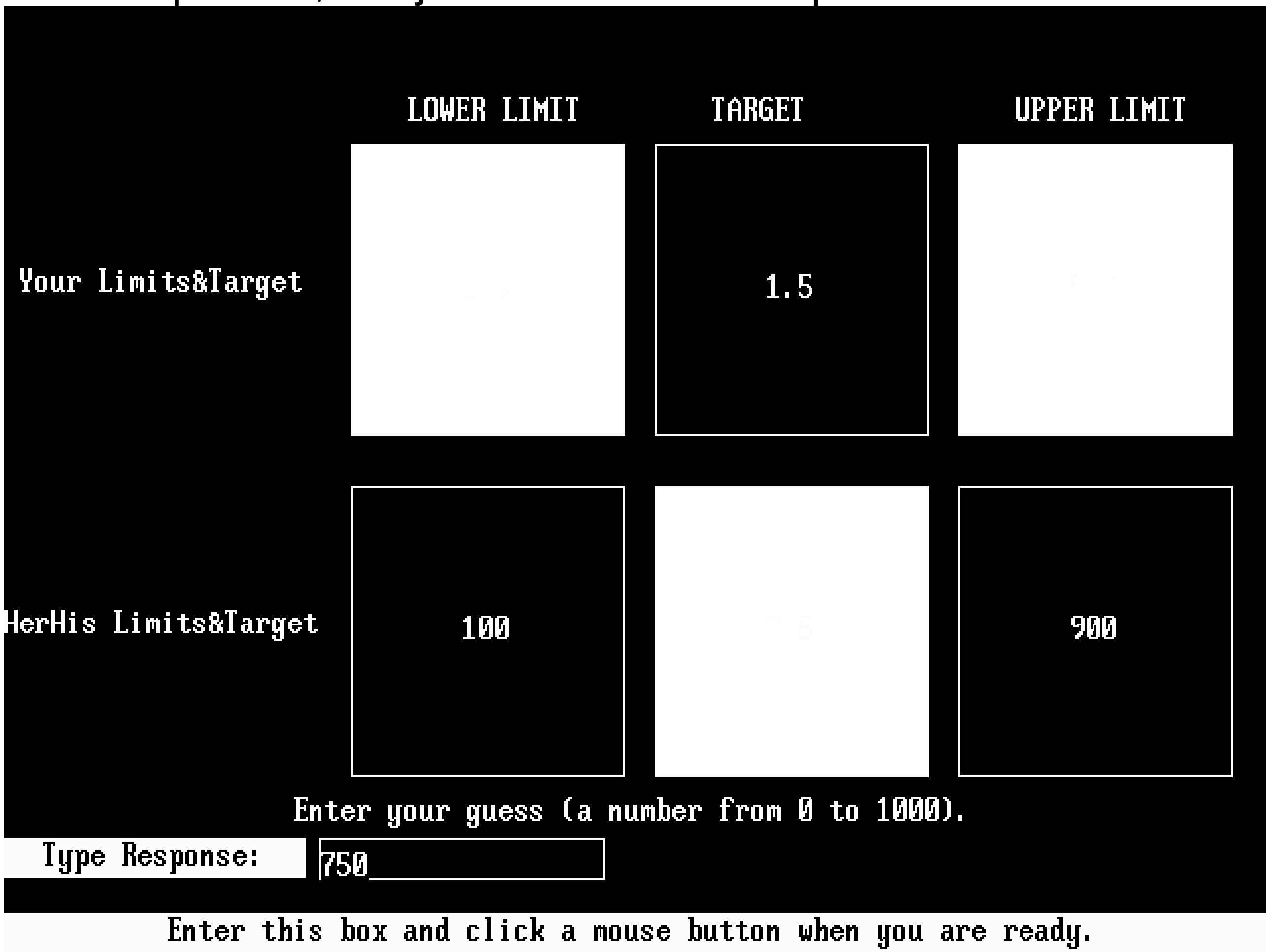


	a	***p***	***b***
You (*i*)	1	2	3
S/he (*j*)	4	5	6

MouseLab Box Numbers

L1’s search implications

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

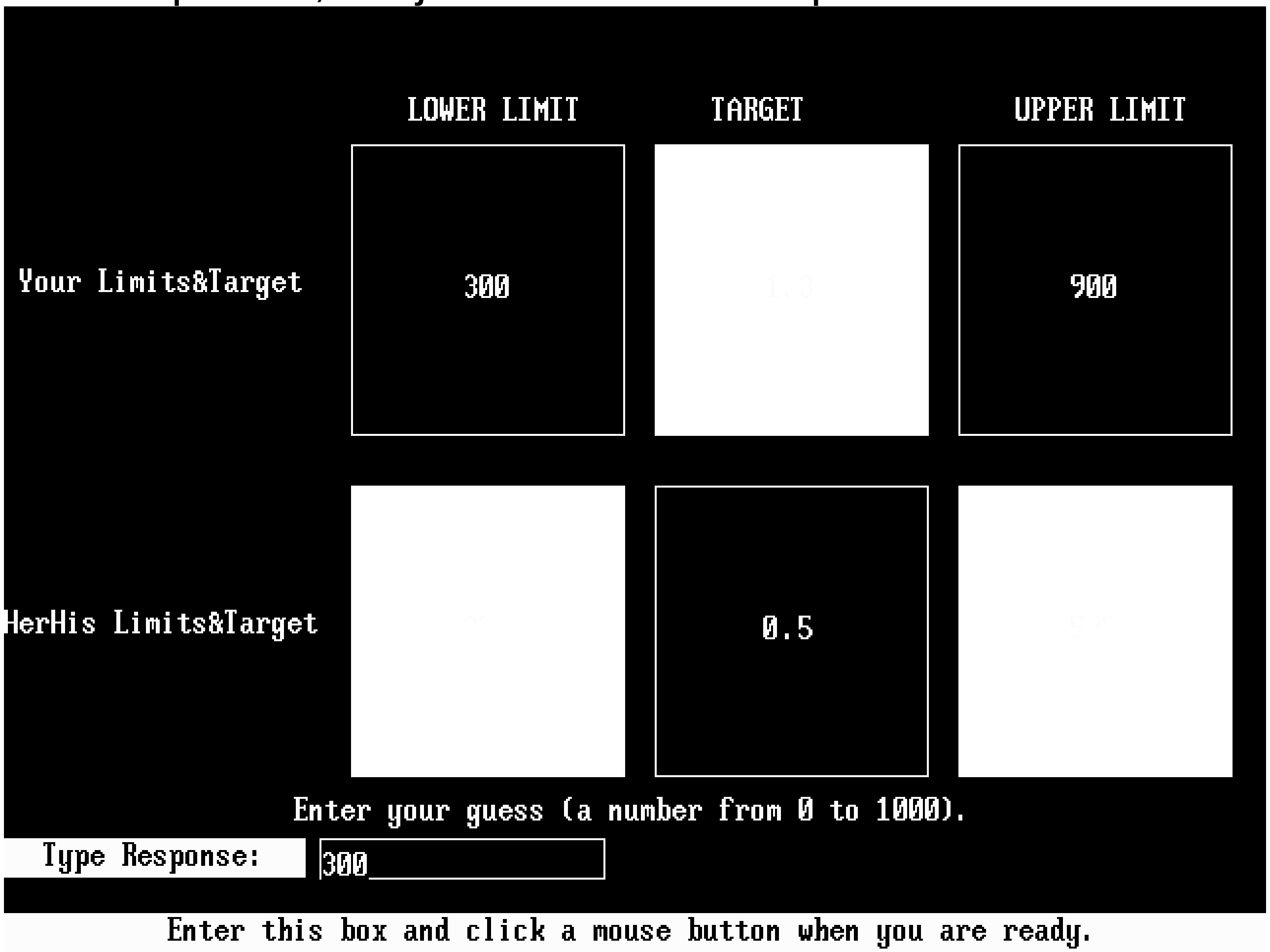


L1’s ideal guess: $p^j[a^j+b^j]/2 = 750$. *L1*’s search implications: $\{[a^j, b^j], p^j\} \equiv \{[4, 6], 2\}$.

(*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.)

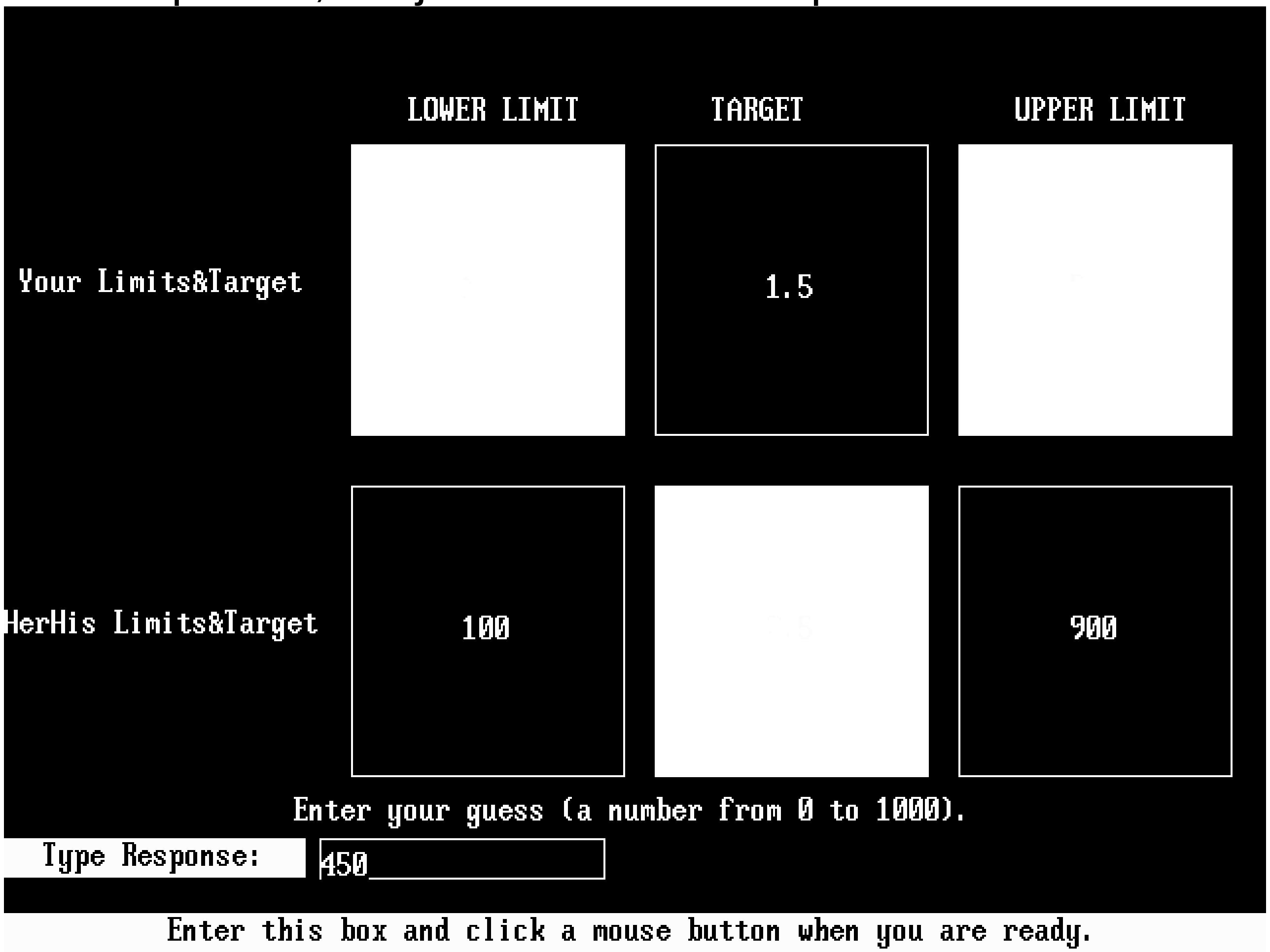


L2's model of its partner's *L1* guess: $p^i[a^j+b^j]/2 = 300$.
Search implications: $\{[a^i,b^i],p^j\} \equiv \{[1,3],5\}$.

(*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: $p^i R(a^j,b^j;\, p^j[a^i+b^i]/2) = 450$.

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.)

The left side of Table 4 lists formulas for types' ideal guesses in CGC's games. The right side lists types' search implications, first in terms of our notation, then in terms of the box numbers in which MouseLab records the data. A type's operations are listed in the order that seems most natural, if there is one; but this is not a requirement of the theory.

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	$p^i R(a^j,b^j; p^j[a^i+b^i]/2)$	$\{([a^i,b^i],p^j),a^j,b^j,p^i\} \equiv \{([1, 3], 5), 4, 6, 2\}$
L3	$p^i R(a^j,b^j; p^j R(a^i,b^i; p^i[a^j+b^j]/2))$	$\{([a^j,b^j],p^i),a^i,b^i,p^j\} \equiv \{([4, 6], 2), 1, 3, 5\}$
D1	$p^i(\max\{a^j,p^j a^i\} + \min\{p^j b^i,b^j\})/2$	$\{(a^j,[p^j,a^i]),(b^j,[p^j,b^i]),p^i\} \equiv$ $\{(4,[5,1]),(6,[5,3]),2\}$
D2	$p^i[\max\{\max\{a^j,p^j a^i\},p^j \max\{a^i,p^i a^j\}\}$ $+\min\{p^j \min\{p^i b^j,b^i\},\min\{p^j b^i,b^j\}\}]/2$	$\{(a^i,[p^i,a^j]),(b^i,[p^i,$ $b^j]),(a^j,[p^j,a^i]),(b^j,[p^j,b^i]),p^j,p^i\}$ $\equiv \{(1,[2,4]),(3,[2,6]),(4,[5,1]),(6,[5,3]),5,2\}$
Eq.	$p^i a^j$ if $p^i p^j < 1$ or $p^i b^j$ if $p^i p^j > 1$	$\{[p^i,p^j],a^j\} \equiv \{[2, 5], 4\}$ if $p^i p^j < 1$ or $\{[p^i,p^j],b^j\} \equiv \{[2, 5], 6\}$ if $p^i p^j > 1$
Soph.	[no closed-form expression, but we take its search implications to be the same as *D2*'s]	$\{(a^i,[p^i,a^j]),(b^i,[p^i,$ $b^j]),(a^j,[p^j,a^i]),(b^j,[p^j,b^i]),p^j,p^i\}$ $\equiv \{(1,[2,4]),(3,[2,6]),(4,[5,1]),(6,[5,3]),5,2\}$

CGC's Table 4. Types' Ideal Guesses and Relevant Look-ups

Aside: Background evidence on cognition and search

CGC's (and CGCB's assumptions how cognition determine search are based on several sources of evidence:

(i) Camerer, Johnson, et al.'s Robot/Trained Subjects' ("R/TS") searches, which led them to characterize subgame-perfect equilibrium via backward induction search in terms of transitions between the second- and third-round pies

(ii) CGCB's Trained Subjects' searches, which suggest a similar view of *Equilibrium* search in matrix games

(iii) CGC's R/TS subjects with high compliance with their assigned type's guesses, and CGC's Baseline subjects with high compliance with their apparent type's guesses, whose searches suggest a similar view of *L1* and *L2* search

(CGC's six R/TS treatments were identical to their Baseline treatment except that each subject was trained and rewarded as a type: *L1*, *L2*, *L3*, *D1*, *D2*, or *Equilibrium*.)

Search Data for Representative R/TS and Baseline Subjects

Now consider 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.

The data will suggest the following conclusions:

(i) 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)

(ii) The sequences the theory identifies as relevant for a type (Table 4) are unusually dense in the sequences of subjects of that type, at least for the simpler types (CGC's econometric analysis measures search compliance for a type as the density of its relevant sequences in the subject's look-up sequence)

(iii) Those who can speak rodent can quickly learn to read the algorithms many subjects are using directly from the data (if we can do it, the right econometrics can do it too: many of CGC's subjects' types can be reliably identified from search alone (CGC, Table 7))

(iv) 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 Camerer, Johnson, et al.'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)

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	46246254623154623 1
805	D1 (16)	154356423213264 2	51453561536423
1601	D1 (16)	25451436231	5146536213
804	D1 (3)/L2 (16)	1543465213	5151353654623
1110	D2 (14)	1354642646*313	13513464216345146321113 6 414262135362*146546
1202	D2 (15)	2464661354646413213424 62 4226461246255*1224654646	1236451324624262622413 56 462*1352424246613564 62
704	DEq (16)	1234563632565653656263 65 6526514522626526	1234565251236526256352 56 262365456
1205	Eq (16)	12345642465256252563 52*465	1234562445655652632125 54 1466626542514452 6*31
1408	Eq (15)	1231234564456321321 1	1234564561236435241
2002	Eq (16)	1425361253652536163614 54 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)	123456545612346З*	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)	144652313312546232 12512	1324562531564565 4546312315656262	3124565231*123654 55233**513
210	L3 (9) Eq (9) D2(8)	1234561234562134 56 254213654	1234564655622316 54456*2	123455645612 3
302	L3 (7) Eq (7)	2211354656452132 13 45456*541	2135465662135454 6321*26654123	265413232145563214 563214523*654123
318	L1 (7) D1 (5)	13245646525213242* 1462	132465132*462	1346521323*4
417	Eq (8) L3 (7) L2 (5)	252531464656446531 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	123456445613255 4 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

CGC's Econometric Analysis of Guesses and Search

In CGC's econometric analysis of guesses and search, most subjects' type estimates reaffirm the guesses-only estimates.

For some subjects the guesses-and-search type estimate resolves a tension between guesses-only and search-only estimates in favor of a type other than the guesses-only estimate.

In more extreme cases, a subject's guesses-only type estimate is excluded because it has 0 search compliance in 8 or more games.

(For example, Baseline subject 415, whose apparent type was *L1* with 9 exact guesses, had 0 *L1* search compliance in 9 of the 16 games because s/he had no adjacent $[a^j,b^j]$ pairs as we required for *L1*. Her/his look-up sequences were unusually rich in (a^j,p^i,b^j) and (b^j,p^i,a^j) triples, in those orders. Because the sequences were not rich in such triples with other superscripts, this is clear evidence that 415 was an *L1* who happened to be more comfortable with several numbers in working memory than our characterization of search assumes, or than our other subjects were comfortable with. But because this violated our assumptions on search, this subject was "officially" estimated to be *D1*.)

Overall, the incorporating search into the econometric analysis refines and sharpens our conclusions, and confirms the absence of significant numbers of subjects of types other than *L1*, *L2*, *Equilibrium*, or hybrids of *L3* or *Equilibrium*.

What Else is Search Good for?

To better illustrate the possibilities for search analysis, I now discuss two puzzles raised by CGC's analysis, which will be addressed in CGC's sequel, "Studying Cognition via Information Search in Two-Person Guessing Game Experiments."

Puzzle a. What are those Baseline *Equilibrium* subjects really doing?

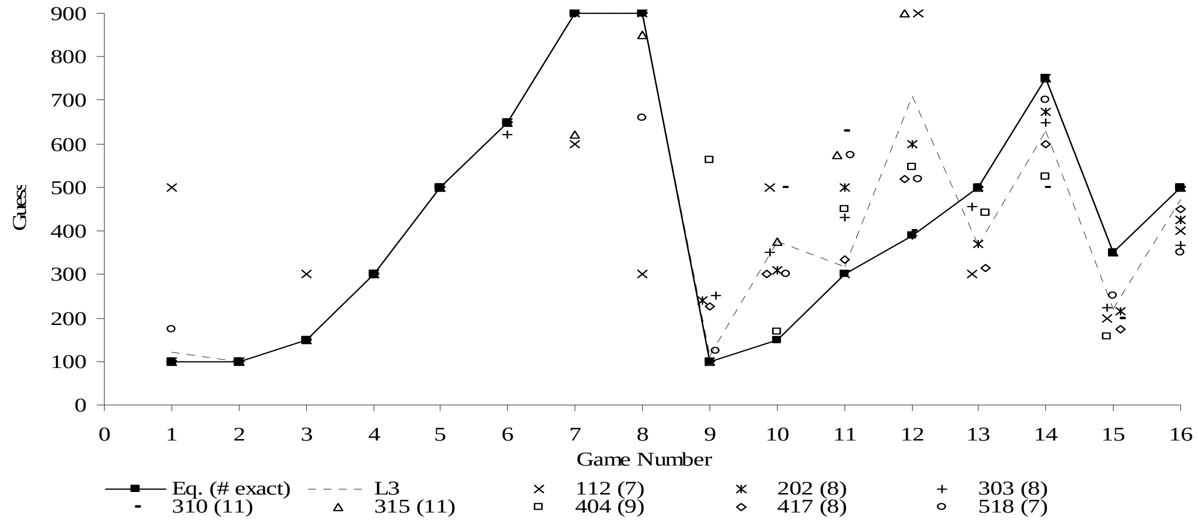
Consider CGC's 8 Baseline subjects with near-*Equilibrium* fingerprints.

Ordering the games by strategic structure, with CGC's eight games with mixed targets (one > 1, one < 1) on the right, shows that their deviations from equilibrium almost always occur with mixed targets (CGC's Figure 4).

Thus it is (nonparametrically) clear that these subjects, whose compliance with *Equilibrium* guesses 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, iterated dominance) work just as well with mixed targets.

Thus whatever these Baseline *Equilibrium* subjects are doing, it's something we haven't thought of yet. (And their debriefing questionnaires don't tell us what it is either.)



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.)

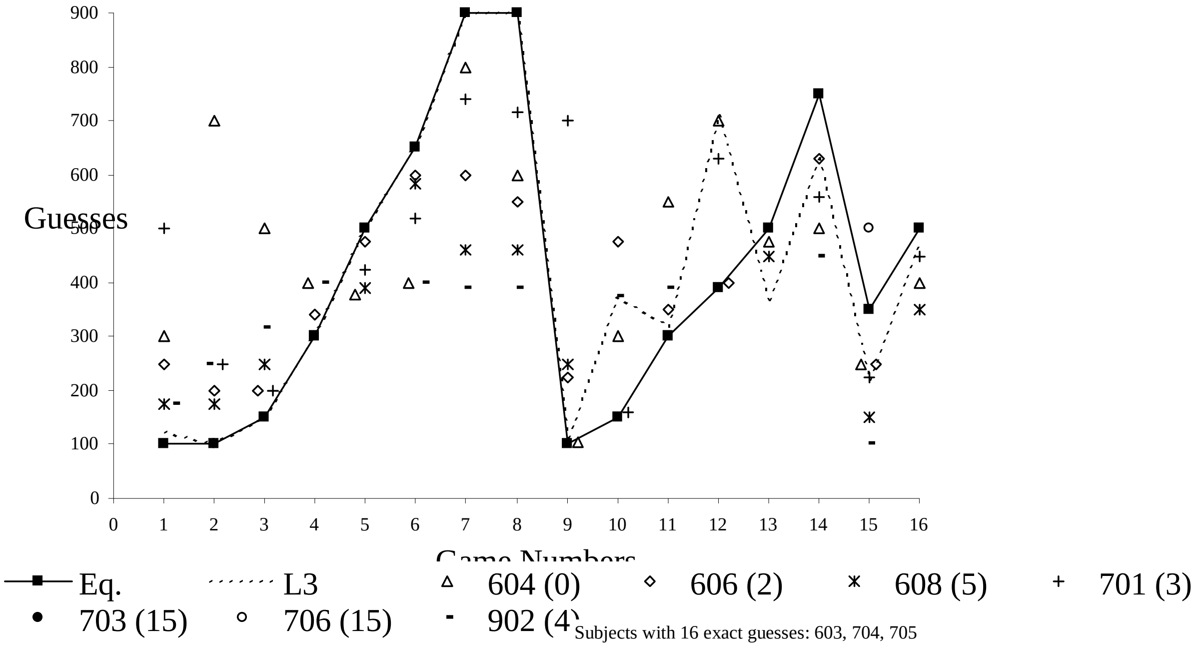
But whatever these Baseline *Equilibrium* subjects are doing, it has a structure:

All 44 of these subjects' deviations from Equilibrium (solid line) when it is separated from *L3* (dotted line) are in the direction of (and sometimes beyond) *L3* guesses

However, this could reflect no more than the fact that in CGC's games, *Equilibrium* guesses are always more extreme than other types' guesses.

By contrast, CGC's *Equilibrium* R/TS subjects' (those subjects were taught to identify equilibria, and rewarded as if their partners always chose equilibrium decisions) compliance is equally high with and without mixed targets.

Thus, training eliminates whatever the Baseline *Equilibrium* subjects are actually doing.



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

Possible sources of answers to puzzle a

(i) Can we tell how Baseline *Equilibrium* subjects find equilibrium in games without mixed targets: best-response dynamics, equilibrium checking, iterated dominance, or something else that doesn't "work" with mixed targets?

(The absence of Baseline *Dk* subjects suggests that they are not using iterated dominance. Best-response dynamics, perhaps truncated after 1-2 rounds, seems more likely to us. We can check by refining CGC's characterization of *Equilibrium* search and redoing the econometrics, separately with and without mixed targets.)

(ii) Is there any identifiable difference in Baseline *Equilibrium* subjects' search patterns in games with and without mixed targets? If so, how do the differences compare to those for *L1*, *L2*, or *L3* subjects?

(Our 20 apparent Baseline *L1* subjects' compliance with *L1* guesses (CGC, Figure 1) is almost the same with and without mixed targets: unsurprisingly because the distinction is irrelevant to *L1*.

But our 12 apparent *L2* (Figure 2) and 3 apparent *L3* subjects' compliance with apparent types' guesses is noticeably lower with mixed targets. This is curious, because for *L2* and *L3*, unlike for *Equilibrium*, games with mixed targets require no deeper understanding.)

(iii) Can we tell how R/TS *Equilibrium* subjects with high compliance manage to find their *Equilibrium* guesses even with mixed targets? How does their search in those games differ from Baseline *Equilibrium* subjects' search?

(CGC strove to make the R/TS *Equilibrium* training as neutral as possible, but something must come first. CGC taught them equilibrium checking first, then best-response dynamics, then iterated dominance. To the extent that subjects used one of those methods, it explains why they have equal compliance with mixed targets. If subjects used something else, and it deviates from equilibrium in games with mixed targets, it might provide a clue to what CGC's Baseline *Equilibrium* subjects did.)

(Note that CGC's Baseline subjects with high compliance for some type are, to the extent that we are confident in inferring their beliefs, like robot *un*trained subjects. These don't usually exist because you can't tell robot subjects how they will be paid without teaching them how the robot works, and so training them.

Thus CGC's design provides an unusual opportunity to separate the effects of training and strategic uncertainty, by comparing Baseline and R/TS subjects:

Either *Equilibrium* is natural with mixed targets, but subjects don't see it without training; or *Equilibrium* is unnatural, and/or subjects don't believe that others, even with training, will make *Equilibrium* guesses with mixed targets.)

Puzzle b. Why are *Lk* the only types other than *Equilibrium* with nonnegligible population frequencies?

CGC's analysis of decisions and search revealed significant numbers of subjects of types *L1*, *L2*, *Equilibrium* (with the qualifications expressed above), or hybrids of *L3* and/or *Equilibrium*, and nothing else.

(More precisely, a careful analysis of the data reveals no other types that do better than a random model of guesses for more than one subject.)

Why do these decision rules predominate, out of the enormous number of possible non-equilibrium rules?

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

Possible sources of answers to puzzle b

Most R/TS subjects could reliably identify their type's guesses, even *Equilibrium* or *D2*. (These average rates are for exact compliance, so 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

	L1	***L2***	***L3***	***D1***	***D2***	***Eq.***
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

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

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

(ii) 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.

(iii) Among *Lk* and *Dk* types, compliance was higher for lower *k* as one would expect, except that *L1* compliance was lower than *L2* or *L3* compliance.

(We suspect that this is just 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 simulated *L1* or *L2* robot, which doesn't invite "gambling" behavior.)

(iv) Remarkably, 7 of our 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. (Recall that it is *L2* that is *D1*'s cousin; these subjects seem to have intuited this.)

E.g. R/TS *D1* subject 804 made 16 exact *L2* (and so only 3 exact *D1*) guesses; her/his search also suggests *L2* not *D1* thinking.

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

Subject	Type/Alt[a]	Game 1[b]	Game 2[b]
804	D1 (3)/L2 (16)	1543465213	5151353654623

L2	{([1,3],5),4,6,2}
D1	{(4,[5,1], (6,[5,3]),2}

This kind of morphing, in this direction, is the only kind that occurred.

We view this as pretty compelling evidence that *Dk* types are simply unnatural.

However, a comparison of *Lk*'s and *Dk-1*'s search and storage requirements may have something to add. (E.g. *Dk-1* requires more memory than *Lk*.)