COGNITION AND BEHAVIOR IN TWO-PERSON GUESSING GAMES:
AN EXPERIMENTAL STUDY

By Miguel A. Costa-Gomes and Vincent P. Crawford[*]

Final revision: 24 August 2006

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

This paper reports an experiment that elicits subjects' initial responses to 16 dominance-solvable two-person guessing games. The structure is publicly announced except for varying payoff parameters, to which subjects are given free access. Varying the parameters allows very strong separation of the behavior implied by leading decision rules. Subjects' decisions and searches show that most understood the games and sought to maximize payoffs, but that many had simplified models of others' decisions that led to systematic deviations from equilibrium. The predictable component of their deviations is well explained by a structural non-equilibrium model of initial responses based on level-*k* thinking.

Keywords: noncooperative games, experimental economics, guessing games, bounded rationality, cognition, information search (*JEL* C72, C92, C51)

Most applications of game theory assume equilibrium even in predicting initial responses to games played without clear precedents. However, there is substantial experimental evidence that initial responses often deviate systematically from equilibrium, especially when the reasoning that leads to it is not straightforward. This evidence also suggests that a structural model in which some players follow certain kinds of boundedly rational decision rules, in lieu of equilibrium, can out-predict equilibrium in applications involving initial responses.

Modeling initial responses more accurately promises several benefits. It can establish the robustness of the conclusions of equilibrium analyses in games where boundedly rational rules mimic equilibrium. It can challenge the conclusions of applications to games where equilibrium is implausible without learning, and resolve empirical puzzles by explaining the systematic deviations from equilibrium such games often evoke. More generally, it can yield insights into cognition that elucidate many other aspects of strategic behavior. A leading example is learning, where assumptions about cognition determine which analogies between current and previous games players recognize and also sharply distinguish reinforcement from beliefs-based and more sophisticated rules, thereby influencing implications for convergence and equilibrium selection.

The potential for improving on equilibrium models of initial responses is vividly illustrated by Rosemarie Nagel's (1995) and Teck-Hua Ho, Colin Camerer, and Keith Weigelt's (1998; "HCW") "guessing" or "beauty contest" experiments, inspired by Keynes' analogy quoted in our epigraph. In their games, n subjects (n = 15-18 in Nagel, n = 3 or 7 in HCW) made simultaneous guesses between lower and upper limits (0 and 100 in Nagel, 0 and 100 or 100 and 200 in HCW). The subject who guessed closest to a target (p = 1/2, 2/3, or 4/3 in Nagel; p = 0.7, 0.9, 1.1, or 1.3 in HCW) times the group average guess won a prize. There were several treatments, each with identical targets and limits for all players and games. The structures were publicly announced, to justify comparing the results with predictions based on complete information.

Although Nagel's and HCW's subjects played a game repeatedly, their first-round guesses can be viewed as initial responses if they treated their own influences on future guesses as negligible, which is plausible for all but HCW's 3-subject groups. With complete information, in all but one treatment the game is dominance-solvable in a finite (limits 100 and 200) or infinite (limits 0 and 100) number of rounds, with a unique equilibrium in which all players guess their lower (upper) limit when $p < 1$ ($p > 1$). As a result, equilibrium predictions depend only on rationality, in the decision-theoretic sense, and beliefs based on iterated knowledge of rationality.

1

Yet Nagel's subjects never made equilibrium guesses initially, and HCW's rarely did so. Most initial guesses respected from 0 to 3 rounds of iterated dominance, in games where 3 to an infinite number are needed to reach equilibrium (Nagel, Figure 1; HCW, Figures 2A-H and 3A-B). Nagel's and HCW's data resemble neither "equilibrium plus noise" nor "equilibrium taking noise into account" as in quantal response equilibrium ("QRE"; Richard McKelvey and Thomas Palfrey (1995)). Their data do suggest that subjects' deviations from equilibrium have a coherent structure. In Nagel's games, for example, the distributions of guesses have spikes that track $50p^k$ for $k = 1, 2, 3$ across the different targets p in her treatments (Nagel, Figure 1). Like the spectrograph peaks that foreshadow the existence of chemical elements, these spikes are evidence of a partly deterministic structure, one that is discrete and individually heterogeneous.

Similarly structured initial responses have been found in matrix games by Dale Stahl and Paul Wilson (1994, 1995; "SW"), Miguel Costa-Gomes, Vincent Crawford, and Bruno Broseta (1998, 2001; "CGCB"), and Costa-Gomes and Georg Weizäsacker (2005); in other normal-form games by Camerer, Ho, and Juin-Kuan Chong (2004; "CHC"); and in extensive-form bargaining games by Camerer, Eric Johnson, Talia Rymon, and Shankar Sen (1993, 2002; "CJ"). As in the guessing games, subjects usually make undominated decisions; but they rely less often on dominance for others (T. Randolph Beard and Richard Beil (1994)), and reliance on iterated dominance seldom goes beyond three rounds.

The main difficulty in analyzing the data from such experiments is identifying subjects' decision rules, or *types*, within the enormous set of possibilities. The above studies assume that each subject's decisions follow one of a small set of a priori plausible types, with error, and estimate which type best fits each subject's decisions econometrically. Leading types include *L1* (*Level 1*), which best responds to a uniform prior over its partner's decisions; *L2* (or *L3*), which best responds to *L1* (*L2*); *D1* (*Dominance 1*), which does one round of deletion of dominated decisions and best responds to a uniform prior over its partner's remaining decisions; *D2*, which does two rounds of iterated deletion and best responds to a uniform prior over the remaining decisions; *Equilibrium*, which makes its equilibrium decision; and *Sophisticated*, which best responds to the distribution of other subjects' responses, and is included to test whether any subject has a prior understanding of others' decisions that transcends the other simple rules.

Like *Equilibrium*, *Lk* and *Dk* types are rational, with perfect models of the game, and general in that they are applicable to any game. They are usually defined, as we shall do here, to

satisfy subsidiary assumptions of self-interest and risk-neutrality. Thus their only essential departure from *Equilibrium* is replacing its perfect model of others' decisions with a simple non-equilibrium model.[2] *Lk*'s and *Dk*-1's guesses both survive *k* rounds of iterated elimination of dominated decisions and so in two-person games are *k*-rationalizable (B. Douglas Bernheim (1984)). These types mimic equilibrium in games that are dominance-solvable in small numbers of rounds, but deviate systematically in some more complex games, where their decisions can differ significantly, especially for the low values of *k* that are empirically plausible. *Dk* types are closer to how theorists analyze games, and Nagel's results are often taken as evidence of explicit finitely iterated dominance. But *Lk* types have larger estimated frequencies and predominate in applications (Crawford (2003); CHC; and Crawford and Nagore Iriberri (2005a, 2005b)).

Successful applications depend on correctly specifying the structure of initial responses, but previous experiments leave considerable room for doubt on this issue. Previous designs—in which each subject repeatedly plays a single game with a large strategy space as in Nagel and HCW, or a series of different games with small strategy spaces as in SW and CGCB—only weakly separate the types included in the specification from each other and nearby types. Nagel's and HCW's games with $p < 1$ and limits 0 and 100 are an extreme example, where *Lk*'s guesses $[(0+100)/2]p^k$ and *Dk*-1's guesses $([0+100p^{k-1}]/2)p$ both track the spikes at $50p^k$. The freedom to specify the possible types also raises doubts about omitted types and overfitting via accidental correlations with included but irrelevant types. Thus, SW's, CHC's, and CGCB's high estimated numbers of *L1* and *L2* subjects might be no more than proxies for altruistic, spiteful, risk-averse, or confused *Dk* or *Equilibrium* subjects; or other, entirely different omitted types.[3]

Our experiment resolves many of these doubts by eliciting subjects' responses to a series of 16 guessing games designed for this purpose, and using a novel specification test to detect omitted types or overfitting. Like previous designs, ours suppresses learning and repeated-game

---

[2]Compare Reinhard Selten (1998): "Basic concepts in game theory are often circular in the sense that they are based on definitions by implicit properties…. Boundedly rational strategic reasoning seems to avoid circular concepts. It directly results in a procedure by which a problem solution is found." *Lk* makes precise predictions without closing the loop as for equilibrium by anchoring its beliefs in a uniform prior and adjusting them by iterating best responses. *Dk* does so by invoking a uniform prior after finitely iterated deletion of dominated decisions. Keynes' wording in our epigraph connotes *Lk*'s finite iteration of best responses, anchored by true preferences rather than uniform priors, as is natural in a beauty contest. The informal literature on deception also features finite iteration of best responses, anchored by truthfulness or credulity (Crawford (2003, p. 139)).

[3]For example, SW (1994) found large numbers of *L1* and *L2* subjects in an econometric analysis that did not include SW's (1995) *Worldly* type, which best responds to an estimated mixture of a noisy *L1* and a noiseless *Equilibrium*; but SW's (1995) data analysis from a closely related experiment almost completely rejected *L2* in favor of *Worldly*. Our specification analysis suggests that SW's rejection of *L2* may have been incorrect (Section II.D).

effects to justify an analysis of subjects' guesses as initial responses, game by game. Unlike Nagel's and HCW's games, ours have only two players, who make simultaneous guesses within limits. Each player has a lower limit (100 or 300), an upper limit (500 or 900), and a target (0.5, 0.7, 1.3, or 1.5). A player's payoff is higher, the closer his guess is to his target times his partner's guess.[4] The resulting games are asymmetric and, with complete information, dominance-solvable in from 3 to 52 rounds, with essentially unique equilibria determined by players' lower (upper) limits when the product of targets is less (greater) than one. Within this structure, which is publicly announced, the targets and limits vary independently across players and games, with targets either both less than one, both greater than one, or mixed. The targets and limits are normally hidden, but subjects can search for them, game by game, through a computer interface. Low search costs then make the structure effectively public knowledge. Varying the targets and limits makes it impossible for subjects to recall them from previous games and so makes monitoring information search a powerful additional tool for studying cognition.

In our design a subject's sequence of guesses yields a strategic "fingerprint" that often reveals his type with great clarity. Of the 88 subjects in our main treatments, 43 made guesses that comply *exactly* (within 0.5) with one of our type's guesses in from 7 to 16 of the games (20 *L1*, 12 *L2*, 3 *L3*, and 8 *Equilibrium*; Figures 1-4). These compliance levels are far higher than could plausibly occur by chance, given how strongly types' guesses are separated (Figure 5) and that guesses could take from 200 to 800 different rounded values in each game. Because our types specify precise, well-separated guess sequences in a very large space, these subjects' guesses allow one intuitively to "accept" the hypothesis that they followed their apparent types, and so rule out alternative interpretations of their behavior. In particular, because the accepted *Lk* and *Equilibrium* 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 *Lk* can be confidently attributed to non-equilibrium beliefs rather than irrationality, risk aversion, altruism,

---

[4]A subject's guess is not required to be between his limits, but guesses outside his limits are automatically adjusted up to the lower or down to the upper limit as necessary; and payoffs are determined by players' adjusted guesses (Section I.B). Two-person guessing games allow us to focus on the central game-theoretic problem of predicting the decisions of others who view themselves as a non-negligible part of one's own environment. Brit Grosskopf and Nagel (2001) report experiments with a different class of two-person guessing games, in which all subjects have the same limits and targets, the targets are less than one, and subjects are rewarded for guessing closer to a target times the pair's average guess. Guessing the lower limit is a weakly dominant strategy in their games, which therefore do not fully address the issue of predicting others' decisions.

spite, or confusion.[5] By contrast, with SW's or CGCB's coarse strategy spaces even a perfect fit does not distinguish a subject's apparent type from nearby omitted types; and in Nagel's and HCW's designs, with each subject playing a single game, the ambiguity is even more severe.

Our other 45 subjects' fingerprints are less clear. But for all but 14 of them, violations of simple dominance were comparatively rare (less than 20%, versus 38% for random guesses), suggesting that their behavior was coherent, even though less well described by our types. We study all 88 subjects' behavior in more detail via a maximum likelihood error-rate analysis, following SW and CGCB. We assume each subject's behavior in the 16 games is determined, with error, by one of the leading types listed above. Estimates based on guesses strongly reaffirm our type identifications for the 43 subjects whose fingerprints are clear, and assign several more subjects each to *L1*, *L2*, and *Equilibrium*, plus a few to *D1* and *Sophisticated* (Tables 1 and 7).

For these 45 subjects, our econometric type estimates suffer from the same ambiguity of interpretation as the estimates in previous analyses. To learn whether any subjects' guesses could be better explained by types omitted from our specification, or whether any estimated types are artifacts of accidental correlations with irrelevant included types, we conduct a new specification test that compares the likelihood of our estimated types, subject by subject, with those of estimates based on 88 *pseudotypes*, each constructed from one of our subject's guesses in the 16 games. This test reaffirms most of our identifications of *L1*, *L2*, or *Equilibrium* subjects, but calls into question all but one each of our identifications of *L3*, *D1*, or *Sophisticated* subjects (Table 1). It also supports our a priori specification of possible types by giving no indication of significant numbers of SW's *Worldly* type or any other type omitted from our specification.

Information search adds another dimension to our econometric analysis.[6] Following CGCB, we link search to guesses by taking a procedural view of decision-making, in which a subject's type determines his search and guess, possibly with error. Each of our types is naturally associated with algorithms that process information about targets and limits into decisions. We use those algorithms as models of subjects' cognition, making conservative assumptions about how it is related to search that allow a tractable characterization of types' search implications.

---

[5] For these subjects our design is thus an antidote to Jörgen Weibull's (2004) argument that rejections of equilibrium in experiments that do not independently measure preferences are "usually premature."
[6] A companion paper, Costa-Gomes and Crawford (2006; "CGC"), will analyze our subjects' search behavior in more detail, studying the relations between cognition, search, and guesses.

The types then provide a kind of basis for the enormous space of possible guesses and searches, imposing enough structure to make it meaningful to ask if they are related in a coherent way.

Our design separates types' search implications much more strongly than previous designs, while making them almost independent of the game. This allows some subjects' types to be read from their searches alone (Appendix E, CGC); but most subjects' searches less clearly identify their types. We therefore generalize our error-rate analysis to re-estimate subjects' types using search as well as guesses. Taking both into account, 55 of 88 subjects are reliably identified as one of our types, 45 of them non-*Equilibrium* (Table 1). This analysis reaffirms the absence of significant numbers of types other than *L1*, *L2*, *Equilibrium*, or hybrids of *L3* and/or *Equilibrium*. These results are consistent with previous analyses, but significantly refine and sharpen them.

Thus, to the extent that our subjects' deviations from equilibrium can be predicted, they appear to be based almost entirely on level-*k* thinking. Given the definitions of level-*k* types, our results strongly affirm subjects' rationality and ability to comprehend games and reason about others' responses to them. Although they challenge the use of equilibrium as a universal model of initial responses to games, the simplicity of the alternative non-equilibrium model they suggest should help to allay the common fear that if equilibrium is not assumed, "anything can happen."

I. Experimental Design

To test theories of strategic behavior, an experimental design must identify clearly the games to which subjects are responding. This is usually done by having a "large" subject population repeatedly play a given stage game, with new partners each period to suppress repeated-game effects, viewing the results as responses to the stage game. Such designs allow subjects to learn the structure from experience, which reduces noise; but they make it difficult to disentangle learning from cognition, because even unsophisticated learning may converge to equilibrium in the stage game. Our design, by contrast, seeks to study cognition in its purest form by eliciting subjects' initial responses to 16 different games, with new partners each period and no feedback to suppress repeated-game effects, experience-based learning, and experimentation. This section describes the overall structure of our design, the games, and how they are presented.

A. Overall structure

Our sessions were run at the University of California, San Diego's (UCSD) Economics Experimental and Computational Laboratory (EEXCL) or the University of York's Centre for Experimental Economics (EXEC). Subjects were recruited from undergraduates and graduate

students, with completely new subjects for each session.[7] Table 2 summarizes the overall structure of our experiment, which included four Baseline sessions, B1-B4, with a total of 71 UCSD subjects; one Open Boxes session, OB1, with 17 UCSD subjects; and 15 Robot/Trained Subjects sessions, R/TS1-R/TS15, with a total of 148 subjects, 37 UCSD and 111 York.

All treatments used the same 16 games, presented in the same randomized order (Table 3). The games consist of eight pairs that are symmetric across player roles, so that subjects can be paired without dividing them into subgroups. One pair consists of two symmetric games.

We first describe the Baseline and then explain how our other treatments differed. After the instructions and an understanding test, groups of 13 to 21 subjects were randomly paired to play the 16 games, with new partners each period.[8] Subjects received no feedback during play and could proceed independently at their own paces, but were not allowed to change their guesses once confirmed. Although these features suppress experience-based learning, introspective learning may still occur. However, tests reveal no significant difference between subjects' pooled guesses in the symmetric game when played third and twelfth in the sequence, suggesting that the effects of introspective learning were limited (Appendices C and D).[9] Accordingly, we analyze subjects' guesses as initial responses to each game, without considering order of play.

To control subjects' preferences, they were paid for their game payoffs as follows. After the session each subject returned in private and was shown his own and his partners' guesses and his point earnings in each game. He then drew five game numbers randomly and was paid $0.04 per point for his payoffs in those games.[10] With possible payoffs of 0 to 300 points per game, this yielded payments from $0 to $60, averaging about $33. Including the $8 fee for showing up at least five minutes early (which almost all subjects received) or the $3 fee for showing up on

---

[7]Appendix A gives instructions and Appendix B describes our pilots and how they influenced the design. To reduce noise we sought subjects in quantitative courses; but to avoid subjects with theoretical preconceptions we excluded graduate students in economics, political science, cognitive science, or psychology, and disqualified subjects who revealed that they had participated in game experiments or (except for a few who had been briefly exposed in an undergraduate course) studied game theory. We allowed roughly four non-faculty university community members.

[8]Some pairings among the 13 subjects in session B1 were repeated once, in a game unknown to them. The games took subjects 1-3 minutes each. Adding 1½ to 2 hours for checking in, seating, instructions, and screening yielded sessions of 2¼ to 2¾ hours, near our estimate of the limit of subjects' endurance for a task of this difficulty.

[9]Even so, our analysis of clusters (Appendix F) suggests introspective learning by two of our 88 Baseline and OB subjects, who appeared to have switched from *L1* to *L2* after the first few games.

[10]It is theoretically possible to control risk preferences using the *binary lottery* procedure, in which a subject's payoff is his probability of winning a given monetary prize. We avoid this complication because payment directly in money usually yields similar results, and risk preferences do not affect iterated dominance or pure-strategy equilibrium.

time, this made Baseline (OB) subjects' average total earnings $41.21 ($40.68). Subjects never interacted directly, and their identities were kept confidential.

The structure of the environment, except the games' targets and limits, was publicly announced via instructions on subjects' handouts and computer screens. During the session, subjects had free access, game by game, to their own and their partners' targets and limits via a MouseLab interface (Figure 6, Section I.C). This made the games' structures effectively public knowledge, and so we compare the data with predictions that assume complete information.[11]

Subjects were taught the mechanics of looking up targets and limits and entering guesses, but not information-search strategies. The instructions took care to avoid suggesting particular guesses or decision rules. Subjects were given ample opportunity for questions, and were then required to pass an Understanding Test to continue. Subjects who failed were dismissed, and the remaining subjects were told that all subjects remaining had passed.[12] Before playing the 16 games, subjects were also required to participate in four unpaid practice rounds, after which they were publicly shown the frequencies of subjects' practice-round guesses in their session and told how they could use them to evaluate the consequences of their own practice-round guesses.[13] After playing the 16 games, subjects were asked to fill out a debriefing questionnaire, in which they were asked how they decided what information to search for and which guesses to make.

Our OB treatment addresses the concern that making subjects look up the targets and limits might distort their responses, reducing comparability between our results and those from more conventional designs. The OB treatment is identical to the Baseline except that the games are presented with the targets and limits continually visible, in "open boxes." We find insignificant differences between Baseline and OB subjects' guesses (Appendix C).[14] Accordingly, we pool the data from the Baseline and OB treatments except when search is involved.

---

[11]The possible values of the targets and limits were *not* revealed, to strengthen subjects' incentives to look up the ones they thought relevant to their guesses. Even so, free access still makes the structures public knowledge.
[12]The dismissal rates (including a few voluntary withdrawals) were 20% for Baseline subjects, 11% for OB subjects, and 20% overall for R/TS subjects.
[13]The practice rounds used two player-symmetric pairs of games, in an order that made their symmetries non-salient, so that the guess frequencies could be generated within each session. The variation in frequencies across sessions appears to have had a negligible effect on subjects' behavior in the 16 games. The practice games had a balanced mix of structures, with different targets and limits than in the 16 games to avoid implicitly suggesting guesses.
[14]There are nonetheless hints that OB subjects made high numbers of types' exact guesses less often: OB subjects made up 19% of the subject pool, but only 11% of those who made 14-16 exact guesses and 7% of those who made 10-13. Possibly our design, which makes models of others easy to express as functions of the targets and limits, more strongly encourages Baseline than OB subjects to substitute such models for less structured strategic thinking.

Our R/TS treatments address the concern that we might fail to recognize an empirically important decision rule simply because of subjects' cognitive limitations or unfamiliarity with the setting. For example, a subject who regularly uses equilibrium logic to make strategic decisions in the field, but has trouble applying it in our abstract decision problems, might not show up as an *Equilibrium* subject. Our R/TS treatments are identical to the Baseline except that each subject plays against the computer ("Robot") and the computer plays according to a pre-specified decision rule. The subject is given the standard Baseline instructions, informed of the computer's rule, trained to identify the guesses it yields in our games, and motivated by being paid for the game payoffs determined by his own guesses against the computer's. There are six different kinds of R/TS treatment, in each of which a subject is trained and motivated to follow one of our leading types: *L1*, *L2*, *L3*, *D1*, *D2*, or *Equilibrium*. In an *L2* R/TS treatment, for instance, a subject is informed that the computer makes *L1* guesses (as in *L2*'s beliefs) and trained to identify them. In an *Equilibrium* R/TS treatment, a subject is informed that the computer makes *Equilibrium* guesses and trained to identify them.[15] To the extent that *Equilibrium* R/TS subjects make their own equilibrium guesses, there is reason for confidence that our Baseline subjects' failures to make equilibrium guesses are due not to cognitive limitations or the unfamiliar setting, but to non-equilibrium beliefs or other factors (possibly including lack of training).

CGC's (2006) analysis of our R/TS data confirms that a large majority of subjects can identify the guesses of any of our leading types, including *Equilibrium*, when they are trained and motivated to do so. However, *Lk* types appear to be cognitively far less difficult than *Equilibrium*, and *Equilibrium* appears less difficult than *Dk* types. These differences are probably part of the reason *Lk* types predominate among Baseline subjects' non-equilibrium responses.

B. Two-person guessing games

Our guessing games have two players, *i* and *j* (for "not *i*"), who make simultaneous guesses, x^i and x^j. Each player *i* has a lower limit, a^i, and an upper limit, b^i, but players are not required to guess between their limits; guesses outside the limits are automatically adjusted up to

---

[15] *Equilibrium* subjects were taught each of the three main ways to identify equilibrium guesses: direct checking for pure-strategy equilibrium, best-response dynamics, and iterated dominance. The R/TS treatments also replace the Baseline's practice rounds with a second Understanding Test of how to identify the assigned type's guesses. Subjects were paid an extra $5 or £2.50 for passing this test, and those who failed were dismissed. York R/TS subjects were paid early and on-time show-up fees of £1 and £2, but only £0.02 rather than $0.04 per point, 70% of the UCSD rates. The average total earnings figures were $45.22, $62.03, $51.74, and $50.93 for UCSD R/TS *L1*, *L2*, *D1*, and *Equilibrium* subjects who finished the experiment, and £23.00, £29.76, £28.50, £27.08, £24.12, and £27.65 (with the pound averaging $1.63) for York R/TS *L1*, *L2*, *L3*, *D1*, *D2*, and *Equilibrium* subjects who finished the experiment.

the lower limit or down to the upper limit. Player i's *adjusted guess*, $y^i \equiv R(a^i,b^i;x^i) \equiv x^i$ if $x^i \, \varepsilon$ $[a^i,b^i]$, $y^i \equiv a^i$ if $x^i < a^i$, or $y^i \equiv b^i$ if $x^i > b^i$. Each player i also has a *target*, p^i, and his payoff is higher, the closer his adjusted guess is to his target times his partner's adjusted guess. Writing $e^i \equiv |R(a^i,b^i;x^i) - p^i R(a^j,b^j;x^j)|$ for the distance between player i's adjusted guess and his target times player j's adjusted guess, player i's point payoff, s^i, is given by:

(1)
$$s^i \equiv \max\{0, 200 - e^i\} + \max\{0, 100 - e^i/10\}$$
$$\equiv \max\{0, 200 - |R(a^i,b^i;x^i) - p^i R(a^j,b^j;x^j)|\} + \max\{0, 100 - |R(a^i,b^i;x^i) - p^i R(a^j,b^j;x^j)|/10\}.$$

With or without adjustment, the payoff function in (1) is quasiconcave in player i's guess for any given distribution of player j's guess; and without adjustment the payoff function is symmetric about $e^i = 0$.[16] The relationship between a player's guess and point payoff is not one-to-one because guesses that lead to the same adjusted guess yield the same outcome. We deal with this ambiguity by using a player's adjusted guess as a proxy for all guesses that yield it, and we call a prediction *essentially* unique if it implies a unique adjusted guess.[17]

We vary the targets and limits independently across players and games within this class of games to make the design as informative as possible, given the need for a balanced mix of strategic structures with no obvious patterns. Table 3 summarizes our games, ordered in a way that emphasizes their structural relationships; it also lists the randomized order in which subjects played the games. We identify a player's lower and upper limits by: α for 100 and 500, β for 100 and 900, γ for 300 and 500, and δ for 300 and 900; and a player's target by: 1 for 0.5; 2 for 0.7; 3 for 1.3; and 4 for 1.5. The combination $\beta 1 \gamma 2$, for example, identifies the game in which player i has limits 100 and 900 and target 0.5, and player j has limits 300 and 500 and target 0.7.

The games in our design are dominance-solvable in 3 to 52 rounds. Observation 1 characterizes their equilibria, assuming complete information.[18] If, as in our design, players'

---

[16]Thus, unlike in Nagel's and HCW's games, a player's guess determines a continuous payoff rather than whether he wins an all-or-nothing prize, as a function of his partner's guess rather than a group average. Like Nagel's and HCW's games, ours limit the effects of altruism, spite, and risk aversion. The point payoff function is *not* concave in player i's guess because the weight on e^i in the second term is smaller in absolute value than in the first term; this strengthens payoff incentives near i's best response while keeping them positive elsewhere despite a lower bound of 0 on a game's payoff. In exceptional cases like game $\alpha 4 \beta 1$ (Table 3), it is theoretically possible for a player to guess more than 1000 units from his target times the other's guess, in the flat part of his payoff function.

[17]This ambiguity could be eliminated by requiring players to guess between their limits. We do *not* do so because automatic adjustment enhances the separation of types' search implications. With quasiconcave payoffs a subject can enter the *ideal guess* that would be optimal given his beliefs, ignoring his limits, and know without checking his limits that his adjusted guess will be optimal; our instructions explain this, and most subjects understood it (CGC).

[18]Guesses are in *equilibrium* if each player's guess maximizes his expected payoff, given the other player's. A player's guess *dominates* (*is dominated by*) another of his guesses if it yields a strictly higher (lower) payoff for each

limits and targets are positive and $p^i p^j \neq 1$, their equilibrium adjusted guesses are determined by their lower limits when $p^i p^j < 1$, or their upper limits when $p^i p^j > 1$. (When $p^i p^j = 1$, the equilibrium correspondence is discontinuous and there are multiple equilibria.) In game $\gamma 2 \beta 4$, for instance, the product of targets is $0.7 \times 1.5 = 1.05 > 1$, player i's equilibrium guess is at his upper limit 500, and player j's is at his best response to 500 of 750 (below his upper limit). In game $\delta 2 \beta 3$ the product of targets is $0.7 \times 1.3 = 0.91 < 1$, player i's equilibrium guess is at his lower limit 300, and player j's is at his best response to 300 of 390, above his lower limit.

Observation 1: Unless $p^i p^j = 1$, each guessing game in the above class has an essentially unique equilibrium, in pure strategies, with adjusted guesses as follows:

If $p^i p^j < 1$,

(a) $y^i \equiv R(a^i, b^i; x^i) = a^i$ if $p^i a^j \leq a^i$, and $y^i = \min\{p^i a^j, b^i\}$ if $p^i a^j > a^i$; and

(b) $y^j \equiv R(a^j, b^j; x^j) = a^j$ if $p^j a^i \leq a^j$, and $y^j = \min\{p^j a^i, b^j\}$ if $p^j a^i > a^j$.

Further, although i's ideal guess is $p^i y^j$ and j's is $p^j y^i$, when $p^i p^j < 1$, i can enter $p^i a^j$, or j can enter $p^j a^i$, in lieu of his ideal guess and still be sure that his adjusted guess will be optimal.

If $p^i p^j > 1$,

(c) $y^i \equiv R(a^i, b^i; x^i) = b^i$ if $p^i b^j \geq b^i$, and $y^i = \max\{a^i, p^i b^j\}$ if $p^i b^j < b^i$; and

(d) $y^j \equiv R(a^j, b^j; x^j) = b^j$ if $p^j b^i \geq b^j$, and $y^j = \max\{a^j, p^j b^i\}$ if $p^j b^i < b^j$.

Further, although i's ideal guess is $p^i y^j$ and j's is $p^j y^i$, when $p^i p^j > 1$, i can enter $p^i b^j$, or j can enter $p^j b^i$, in lieu of his ideal guess and still be sure that his adjusted guess will be optimal.

Observation 1 can be verified by direct checking or by noting that if, say, $p^i p^j < 1$, iterating best responses drives adjusted guesses down until one player's hits his lower limit and the other's is at or above his lower limit. We give a formal proof in Appendix H, where the details play an important role in our analysis of *Equilibrium*'s information search implications.

---

of the other player's possible guesses. A player's guess is *iteratively undominated* if it survives iterated elimination of dominated guesses. A *round* of iterated dominance eliminates all dominated guesses for both players. A game is *dominance-solvable* (in k rounds) if each player has a unique iteratively undominated *adjusted* guess (identifiable in k rounds of iterated dominance). Those iteratively undominated adjusted guesses are players' unique equilibrium adjusted guesses. We distinguish the numbers of rounds players need to identify their own iteratively undominated adjusted guesses; the number of rounds in which the game is dominance-solvable is the higher of these numbers.

Table 3 summarizes the games' structural relationships, which add greatly to the power of our design. For instance, the only important difference between the above-mentioned games $\gamma2\beta4$ (game 13 in Figures 1-5) and $\delta2\beta3$ (game 11) is whether the product of targets is greater or less than one. Observation 1 shows that *Equilibrium* responds strongly to this subtle difference, but low-level *Lk* or *Dk* types, whose guesses vary continuously with the targets, respond much less. Further, games with mixed targets (games 9 to 16 in Figures 1-5) are especially well suited to separating types' guesses and help us diagnose the causes of some subjects' deviations from equilibrium (Section II.A). Finally, moving some of *Equilibrium* and other types' guesses away from the limits and the other structural variations in Table 3 stress-test types' predictions, and with our games' large strategy spaces allow us to "reverse-engineer" some deviations and thereby distinguish cognitive errors from "random" behavior (Appendix F). For example, Figure 7 shows the pattern of iterated dominance and how it converges to equilibrium in game $\gamma4\delta3$, where the product of targets is $1.5\times1.3 > 1$, player *i*'s equilibrium guess is at his upper limit 500, and player *j*'s equilibrium guess is at his best response to 500 of 650 (below his upper limit). Here dominance for player *j* occurs initially at both his limits, which stress-tests *Equilibrium* and *Dk* types. None of these features are shared by Nagel's, HCW's, SW's, or CGCB's games.

Observation 2 simplifies the characterization of types' adjusted guesses. It shows that for guessing games in the class, a player's best responses to uniform beliefs on an interval like those in the definitions of types *L1*, *D1*, and *D2*, and indirectly *L2* and *L3*, equal his target times the midpoint of the interval, adjusted if necessary to lie within his limits. This certainty-equivalence result is independent of risk preferences, but it depends on symmetry and uniform beliefs.[19]

Observation 2: Suppose the point payoff function of a guessing game in the class is a symmetric, continuous, almost everywhere differentiable function $s(x\text{-}pz)$ that is weakly decreasing in $|x\text{-}pz|$, where x is a player's guess; p is his target; and z, his partner's guess, is a random variable uniformly distributed on $[a,b]$. Then for any player with a continuous, almost everywhere differentiable von Neumann-Morgenstern utility function $u(\cdot)$ that values only money (risk-neutral, risk-averse, or risk-loving), his expected-utility maximizing choice is $x^* = p\mathrm{E}z = p(a+b)/2$, and his expected-utility maximizing choice s.t. $x \in [c,d]$ is $R(c,d; p(a+b)/2)$.

---

[19]Observation 2 shows that in our games, *Lk* guesses are *k*-point-rationalizable (Bernheim (1984)), but it also shows that our design is not well-suited to distinguishing *k*-point-rationalizable types from those that are *k*-rationalizable in the usual sense, which allows non-deterministic beliefs.

Proof: We show that $x^* = p(a+b)/2$ solves $\max_x \int_a^b u(s(x - pz))dz$ (ignoring the positive factor

$[1/(b-a)]$). The integral in the maximand is differentiable because $u(s(x - pz))$ is continuous. Its

derivative with respect to x, evaluated at x^*, is (ignoring points of nondifferentiability)

(2) $\qquad \int_a^{(a+b)/2} u'(s(x^*-pz))s'(x^*-pz)dz + \int_{(a+b)/2}^b u'(s(x^*-pz))s'(x^*-pz)dz = 0,$

where the equality holds for $x^* = p(a+b)/2$ by symmetry. Because $u(\cdot)$ is increasing and $s(\cdot)$ is

weakly decreasing in $|x - pz|$, raising x above x^* lowers the derivative below 0, and lowering x

below x^* raises it above 0; thus, the integral in the maximand is quasiconcave in x. Because $x^* =$

$p(a+b)/2$ satisfies the first-order condition for maximizing the integral, x^* is optimal ignoring the

constraint $x \in [c,d]$ and $R(c,d; p(a+b)/2)$ is optimal respecting the constraint.

In deriving our types' implications, we assume that each player maximizes the expected

utility of his total money payment over the 16 games. Each type then implies an essentially

unique adjusted guess in each game, which maximizes its expected payoff given beliefs based on

some model of others' decisions.[20] The key to the derivations is a type's *ideal guess*, the one that

would be optimal given its beliefs, ignoring its limits. A type's ideal guess determines its

adjusted guess in a game via the adjustment function $R(a^i,b^i;x^i) \equiv \min\{b^i, \max\{a^i,x^i\}\}$. We

estimate *Sophisticated*'s ideal guesses as risk-neutral best responses to the pooled distribution of

Baseline and OB subjects' adjusted guesses, game by game, rounded to the nearest integer for

simplicity.[21] *Equilibrium*'s ideal guesses follow immediately from Observation 1, and *L1*'s, *L2*'s,

L3's, *D1*'s, and *D2*'s follow immediately from Observation 2.

The left-hand side of Table 4 lists the general formulas for types' ideal guesses as functions

of the targets and limits. Table 5 lists types' adjusted guesses and the guesses that survive 1 to 4

---

[20]A type's adjusted guesses are all that matters about its choices, and all that our types determine. Because a player's total payment is proportional to his point payoffs in five randomly chosen games, a first-order stochastic dominance argument shows that when guesses have known consequences, the player must maximize his point payoff in any given game. When guesses have uncertain consequences, risk preferences are potentially relevant. But Observation 1 shows that our games have essentially unique equilibria in pure strategies, so risk preferences do not affect *Equilibrium* adjusted guesses. And Observation 2 shows that best responses to uniform beliefs are certainty-equivalent, so risk preferences do not affect *L1*, *D1*, or *D2* adjusted guesses, or the best responses that define *L2* or *L3* adjusted guesses. For *Sophisticated* adjusted guesses, which may best respond to non-uniform beliefs and so not be covered by Observation 2, we must assume that players are risk-neutral to justify the statement in the text. Even so, *Sophisticated* adjusted guesses are only generically unique because their beliefs allows ties in optimal guesses.
[21]Because we also rounded subjects' guesses to the nearest integer, and few subjects made exact *Sophisticated* guesses, this does not lead to misclassification.

rounds of iterated dominance in each game. Figure 5 summarizes the separation of types' adjusted guesses in the 16 games. No two types are separated in less than 8 games; the number of games in which two given types are separated averages 11 2/3 out of 16 (73%), which is hard to improve upon within a simple overall structure like ours; and *L2* and *D1* are separated in 13 games, much more strongly than in any previous experiment (Appendix G).

The right-hand side of Table 4 lists types' minimal implications for information search, which are used in Section II.E's econometric analysis, in general notation and in the box numbers in which searches are recorded. Our derivation of these implications in Appendix H and CGC (2006) is based on a procedural view of decision-making in which the algorithms that can be used to process payoff information into a type's guesses determine its searches as well. Types' minimal search implications are based on their ideal guesses, because (with our quasiconcave payoffs) a subject can enter his ideal guess and know that his adjusted guess will be optimal without checking his own limits. As suggested by our R/TS treatments and CJ's and CGCB's experiments, we assume that subjects perform the most basic operations needed to identify the ideal guess one at a time via *adjacent* (consecutive in the sequence) look-ups, in any order, remembering their results, and otherwise relying on repeated look-ups rather than memory. Basic operations will then be represented by adjacent look-up pairs that can appear in any order, but cannot be separated by other look-ups. Other operations will be represented by the associated look-ups, in any order, possibly separated by other look-ups. We call a minimal look-up sequence that satisfies these requirements for a type the type's *relevant* look-ups.

We close this subsection by discussing the strength of our subjects' incentives to follow particular types. Given the enormous set of possible types in our design, we approach this issue by estimating how costly it would be for a subject of a given type, with its beliefs, to behave according to a different type if he still had the given type's beliefs. By this standard, *Equilibrium*, *L2*, and *L3* subjects all have strong incentives to make their type's guesses (Appendix G). *Equilibrium*'s expected earnings would be $46.05 in our 16 games if its partners made equilibrium guesses, $12.05 more than its earnings would be with *L3* guesses, and even more than its earnings would be with our other types' guesses. The analogous earnings differences for *L2* and *L3* are $10.25 and $6.90. But our other leading types have weaker incentives: analogous differences of $1.29 for *D2*, $1.22 for *L1*, $0.85 for *D1*, and $0.46 for *Sophisticated*.[22]

---

[22]Among our types, only *L1* and *Equilibrium* are *not* fairly close substitutes for *Sophisticated*, given its beliefs.

C. Using MouseLab to present guessing games

The games were displayed on subjects' screens via a computer interface called MouseLab.[23]
To suppress framing effects, a subject was called "You" and his partner was called "S/He," etc. A
subject could look up a payoff parameter by using his mouse to move the cursor into its box and
left-clicking; in Figure 6 the subject has opened the box that gives his own ("Your") lower limit,
100. Before he could open another box or enter his guess, he had to close the box by right-
clicking; a box could be closed after the cursor had been moved out of it. Thus both opening and
closing a box required a conscious choice. Subjects were not allowed to write during the main
part of the experiment. A subject could enter and confirm his guess by moving the cursor into the
box labeled "Keyboard Input," clicking, typing the guess, and then moving the cursor into the
box at the bottom of the screen and clicking. A subject could move on to the next game only
after confirming his guess; after an intermediate screen, the cursor returned to the top-center.
MouseLab automatically records subjects' look-up sequences, look-up durations, and guesses.

II. Analysis of Subjects' Guesses and Information Searches

This section analyzes subjects' guesses and information searches, starting with those of
our subjects whose types are apparent from guesses alone and continuing with all subjects'
compliance with iterated dominance and equilibrium, an econometric analysis of subjects'
guesses and specification test, and an econometric analysis of their guesses and information
search.[24] Table 1 summarizes each phase's conclusions regarding the numbers of subjects of each
estimated type, and Table 7 summarizes the results subject by subject.

A. Subjects whose types are apparent from guesses alone

Figures 1-4 graph the actual against predicted adjusted guesses of the 43 subjects whose
types are apparent from guesses alone, with games ordered to emphasize structural relationships
as in Table 3. Figures 1-3 graph 20 apparent *L1* subjects, 12 apparent *L2* subjects and 3 apparent
L3 subjects, each with *Equilibrium* guesses for comparison; and Figure 4 graphs 8 apparent
Equilibrium subjects, with *L3* guesses for comparison. Only deviations from subjects' apparent

---

[23]MouseLab was developed to study individual decisions (Payne, Bettman, and Johnson (1993, Appendix) and
http://www.cebiz.org/mouselab.htm). CJ 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" in each period. CGCB
used MouseLab to study matrix games in which subjects could look up their own and their partners' payoffs.
[24]Appendix D graphs the aggregate game-by-game frequency distributions of subjects' adjusted guesses. Appendix E
gives the complete data on subjects' guesses and the order (but not duration) of their look-up sequences.

types are shown; the 20 *L1* subjects in Figure 1, for example, who made a total of 320 guesses in 16 games, deviated from *L1* guesses a total of 104 times, each identified by the subject's mark.

Section II.C's econometric analysis confirms that in likelihood-based type inferences, subjects' rates of exact (within 0.5) compliance with types' guesses are far more important than the magnitudes of their deviations. Thus the most important message of Figures 1-4 is that these 43 subjects' rates of exact compliance were very high, despite our large strategy spaces. A second message is that these subjects' guesses usually varied only slightly across the two symmetric games, 7 and 8, suggesting that the effects of introspective learning were limited.

There are systematic differences between subjects' responses to games with (9-16, right sides of Figures 1-4) and without (1-8) mixed targets: Apparent *L2*, *L3*, and *Equilibrium* subjects all deviate from their types' predictions much more often in games with mixed targets. This is surprising, because *L2* and *L3* guesses are determined by simple formulas in which a subject's own and his partner's targets play similar roles (Table 4) and all of the standard methods for identifying *Equilibrium* decisions (direct checking, best-response dynamics, and iterated dominance) work equally well with and without mixed targets. Apparent *L1* subjects, whose ideal guesses do not depend on their partner's target, making the distinction between games with and without mixed targets irrelevant, do not deviate more often in games with mixed targets.[25]

The details of apparent *Equilibrium* and *L3* subjects' guesses deepen the puzzle. Our 8 apparent *Equilibrium* subjects' 44 deviations from *Equilibrium* when it is separated from *L3* (out of a possible 72 = 9 games × 8 subjects) are *all* in the direction of (and sometimes beyond) *L3* guesses. And our apparent *L3* subjects' deviations from *L3* when it is separated from *Equilibrium* are usually in the direction of *Equilibrium* and often coincide with it—even in game 1, our only such game without mixed targets. Thus, many of these subjects appear to be using hybrid rules that blend *Equilibrium* (especially in games without mixed targets) with *L3*. This illustrates the potential empirical importance of the subtlety of identifying equilibrium decisions in our games.

B. Subjects' compliance with iterated dominance and equilibrium

Table 6 reports Baseline and OB subjects' compliance with 1 or more, 2 or more, 3 or more, and 4 or more rounds of iterated dominance—equivalently, levels of *k*-rationalizability—and exact compliance with *Equilibrium*, overall and game by game, with random compliance as

---

[25]There are no clear patterns in the magnitudes of deviations or with respect to other aspects of the games' structures. Only one of our 29 *Equilibrium* R/TS subjects came close to the apparent *Equilibrium* subjects' patterns in the Baseline; the rest made just as many exact guesses with as without mixed targets (Appendix E, CGC (2006)).

a benchmark.[26] Subjects make undominated guesses at a rate well above random in each of the 13 games in which subject *i* has any dominated guesses, with an overall rate of 90%, typical for initial responses and far higher than random (62%). Only 14 subjects make undominated guesses in less than 80% of the games, which suggests that the behavior of a large majority of our subjects was coherent. Compliance with iterated dominance is always higher than random when this is possible, and usually far higher. It varies widely across games, but with no clear effect of structure beyond what determines random compliance. Compliance with *Equilibrium* is lower in games with mixed targets but otherwise shows no clear effect of structure. Because our games with mixed targets coincide with those with many rounds of iterated dominance, in this respect our results correspond to those for CGCB's (2001, Table II) matrix games.

C. Econometric analysis of Baseline and OB subjects' guesses

Although 43 of our 88 Baseline and OB subjects' types are apparent from their guesses, the remaining 45 subjects' guesses are not immediately clear. In this subsection we estimate all 88 subjects' types econometrically, via a maximum likelihood error-rate analysis of their guesses. Our goals are to summarize the implications of the data in a comprehensible way, to assess the strength of the evidence in favor of our types, and to identify those subjects whose guesses are not well explained by our types and guide the search for better explanations of their behavior.

We assume that each subject's behavior is determined, possibly with error, by a single type, which determines his guesses and searches in all games. The types we allow were chosen a priori from general principles of strategic decision-making that have played important roles in the literature, with the goal of specifying a set large and diverse enough to do justice to the heterogeneity of subjects' behaviors but small enough to avoid overfitting. We include *L1*, *L2*, *L3*, *D1*, *D2*, and *Equilibrium* as defined in the Introduction, and add CGCB's *Sophisticated* to test whether any subjects have a prior understanding of others' decisions that transcends these simple rules. In theory, *Sophisticated* best responds to the probability distributions of its partners' decisions; but those distributions are part of a behavioral game theory that is not yet fully developed. We therefore operationalize *Sophisticated* using the best available predictions of those distributions in our setting: the population frequencies of our own subjects' guesses.[27]

---

[26]The differences between Baseline and OB subjects are unimportant here. Appendix G's tables give the analogous results for types other than *Equilibrium*.

[27]An ad hoc type could perfectly mimic a subject's decision history, but would have no explanatory power. It is

Index types $k = 1,\ldots,K$ and games $g = 1,\ldots,G$. In game g, denote subject i's lower and upper limits a_g^i and b_g^i, his *un*adjusted and *adjusted* guess x_g^i and $R_g^i(x_g^i) \equiv \min\{b_g^i, \max\{a_g^i, x_g^i\}\}$, and type k's adjusted guess t_g^k. Write $x^i \equiv (x_1^i,\ldots,x_G^i)$ and $R^i(x^i) \equiv (R_1^i(x_1^i),\ldots,R_G^i(x_G^i))$.

We analyze the data subject by subject.[28] Interpreting a pattern of deviations from types' guesses requires an error structure. We assume that, conditional on a subject's type, his errors are independent across games. Because our subjects so often made types' exact guesses, we use a simple "spike-logit" error structure in which, in each game, a subject has a given probability of making his type's guess exactly and otherwise makes guesses that follow a logistic distribution over the rest of the interval between his limits. Thus in game g a type-k subject makes a guess that leads to type k's adjusted guess t_g^k within 0.5 with probability $1 - \varepsilon$; but with probability $\varepsilon \in [0,1]$, his *error rate*, his adjusted guess has error density $d_g^k(R_g^i(x_g^i), \lambda)$ with *precision λ*.[29]

In describing how payoffs affect the error density, we assume that subjects are risk-neutral for simplicity. Let y be subject i's partner's adjusted guess and $S_g(R_g^i(x_g^i), y)$ be i's own expected monetary payoff in game g, where the expectation is taken only over the random selection of games i is paid for. All of our types best respond to some beliefs; let the density $f_g^k(y)$ represent type k's beliefs. Subject i's expected payoff in game g for type k's beliefs can then be written:

$$(3) \qquad S_g^k(R_g^i(x_g^i)) \equiv \int_0^{1000} S_g(R_g^i(x_g^i), y) f_g^k(y) dy \, .$$

hard to dispense with a priori specification because the space of possible types is enormous and the leading types have no simple, unifying structure. Further, there are multiple rationales for any given history of guesses, but we link guesses and search via a procedural model whose implications depend not only on what guesses a type implies, but why. Our *L1* corresponds to SW's *Level 1* or CGCB's *Naïve*, and is related to *Level 1* or *Step 1* in Nagel, Stahl, HCW, and CHC. Our *L2* (*L3*) corresponds to CGCB's *L2* (*L3*), and is related to *L2* (*L3*) in SW, Nagel, Stahl, HCW, and CHC. Earlier work suggests that types beyond *L3* or *D2* are empirically unimportant, and there is no evidence of them in our data. We also omit three types CGCB found empirically unimportant: *Pessimistic* (maximin), *Optimistic* (maximax), and *Altruistic*. *Pessimistic* and *Optimistic* do not distinguish clearly among guesses in our games; and we judged the effects of own guesses on others' payoffs too weak and non-salient for *Altruistic* to be plausible.
[28]CGCB (2001) used an aggregate mixture model that imposed stronger restrictions on subjects' type distributions, and studied cognition at the individual level by conditioning on individual histories. CGCB (1998) estimated subject by subject using the same dataset, with similar results. Estimating subject by subject seems better suited to studying cognition and more robust to misspecification; but the results are unlikely to differ much from a mixture model.
[29]In our design entered guesses are restricted to the interval [0, 1000], which includes all possible limits. There is no need to allow the error rate and precision to depend on type, because all three are estimated jointly.

18

Let $U_g^{ik} \equiv [t_g^k - 0.5, t_g^k + 0.5] \cap [a_g^i, b_g^i]$, the set of subject i's possible adjusted guesses in

game g that are within 0.5 of type k's adjusted guess t_g^k, and let $V_g^{ik} \equiv [a_g^i, b_g^i]/U_g^{ik}$, the

complement of U_g^{ik} relative to $[a_g^i, b_g^i]$. The density $d_g^k(R_g^i(x_g^i), \lambda^k)$ then satisfies:

$$(4) \qquad d_g^k(R_g^i(x_g^i), \lambda) \equiv \frac{\exp[\lambda S_g^k(R_g^i(x_g^i))]}{\int_{V_g^{ik}} \exp[\lambda S_g^k(z)]dz} \text{ for } R_g^i(x_g^i) \in V_g^{ik}, \text{ and 0 elsewhere.}$$

The precision λ is inversely related to the dispersion of a subject's erroneous guesses:

As $\lambda \to \infty$ they approach a noiseless best response to his type's beliefs, and as $\lambda \to 0$ they

approach uniform randomness between his limits, excluding exact guesses. For a given value of

λ, the dispersion declines with the strength of payoff incentives, evaluated for the type's beliefs.

Because unadjusted guesses that lead to the same adjusted guess yield the same payoffs, the

error structure treats them as equivalent, and the likelihood can be expressed entirely in terms of

a subject's adjusted guesses. For subject i, let N^{ik} be the set of games g for which $R_g^i(x_g^i) \in V_g^{ik}$,

and n^{ik} be the number of games in N^{ik}, so that the number of games for which $R_g^i(x_g^i) \in U_g^{ik}$ is G

- n^{ik}. For a type-k subject i in game g, the probability of observing an adjusted guess

$R_g^i(x_g^i) \in U_g^{ik}$ is $(1 - \varepsilon)$, the probability of observing an adjusted guess $R_g^i(x_g^i) \in V_g^{ik}$ is ε, and

the conditional density of an adjusted guess in V_g^{ik} is then $d_g^k(R_g^i(x_g^i), \lambda)$ as in (4).[30] Because

errors are independent across games, the density of a sample with adjusted guesses

$R^i(x^i) \equiv (R_1^i(x_1^i), ..., R_G^i(x_G^i))$ for a type-k subject i is:

$$(5) \qquad d^k(R^i(x^i), \varepsilon, \lambda)) \equiv (1 - \varepsilon)^{(G-n^{ik})} \varepsilon^{n^{ik}} \prod_{g \in N^{ik}} d_g^k(R_g^i(x_g^i), \lambda),$$

where products with no terms (if $n^{ik} = 0$ or G) are taken to equal 1. Letting $p \equiv (p^1, ..., p^K)$

denote the vector of prior type probabilities, weighting by p^k, summing over k, and taking

logarithms yields subject i's log-likelihood:

$$(6) \qquad L^i(p, \varepsilon, \lambda | R^i(x^i)) \equiv \ln\left[\sum_{k=1}^{K} p^k d^k(R^i(x^i), \varepsilon, \lambda)\right]$$

---

[30] The conditional density could be allowed to extend to U_g^{ik} as well as V_g^{ik}, but our specification is simpler, and

approximately equivalent given the near-constancy of payoffs within the narrow interval of exact guesses U_g^{ik}.

It is clear from (6) that the maximum likelihood estimate of p sets $p^k = 1$ for the (generically unique) k that yields the highest $d^k(R^i(x^i), \varepsilon, \lambda))$, given the estimated ε and λ. The maximum likelihood estimate of ε can be shown from (5) to be n^{ik}/G, the sample frequency with which subject i's adjusted guesses fall in V_g^{ik}. The maximum likelihood estimate of λ is the standard logit precision, restricted to guesses in V_g^{ik}.

The maximum likelihood estimate of subject i's type maximizes the logarithm of (5) over k, given the estimated ε and λ. When n^{ik} is between 0 and G, the maximand is:

(7) $\ln d^k(R^i(x^i), \varepsilon, \lambda)) \equiv (G - n^{ik})\ln(G - n^{ik}) + n^{ik}\ln(n^{ik}) + \sum_{g \in N^{ik}} \ln d_g^k(R_g^i(x_g^i), \lambda) - G\ln G$.

When $n^{ik} = 0$, the maximand reduces to 0; and when $n^{ik} = G$, the maximand reduces to the sum over g on the right-hand side of (7).

The likelihood takes the separation of types' guesses across games into account, favoring a type only to the extent that it explains a subject's guesses better than other types. It treats a guess as stronger evidence for a type the closer it is to the type's guess, because the payoff function is quasiconcave and the logit term increases with payoff; and it treats a guess that exactly matches a type's guess as the strongest possible evidence for the type, discontinuously stronger than one that is close but not within 0.5. If n^{ik} is near 0 for only one k, that k is usually the estimated type. If n^{ik} is nearly the same for all k, the estimated type is mainly determined by the logit term; and if n^{ik} is near G for all k, the type estimate is close to the estimate from a standard logit model.

Table 1 (column 3) reports the aggregate frequencies of subjects' type estimates based on (7). Table 7 (left-hand side) reports each Baseline or OB subject's number of dominated guesses and the estimates of his type k, precision λ, number of exact type-k guesses (= $16(1 - \varepsilon)$, where ε is the error rate), with subjects ordered by type, in decreasing order of likelihood within type. The joint restriction $\varepsilon \approx 1$ and $\lambda = 0$, which approximates a completely random model of guesses, cannot be rejected at the 5% (and 1%) level for 10 subjects (6 estimated *L1*, 2 *D1*, 1 *Equilibrium* and 1 *Sophisticated*, with type indicators superscripted † in Table 7); Table 1 (column 4) reports the aggregate type frequencies excluding these subjects as "unclassified."

Likelihood ratio tests reject the hypothesis $\varepsilon \approx 1$, which approximates a standard logit model, at the 5% (1%) level for all but 7 (2) of our 88 subjects (110 and 213 at the 1% level, plus

109, 113, 212, 421, and 515 at the 5% level), so the spike in our specification is necessary.[31] The hypothesis $\lambda = 0$ is rejected at the 1% (5%) level for the 21 (34) subjects whose estimates are superscripted ** (*) in Table 7, so the logit model's payoff-sensitive errors significantly improve the fit over a spike-uniform model such as CGCB's for only about a third of our subjects. This suggests that many of our subjects' deviations are due to cognitive errors rather than insufficient motivation or lack of effort; and this view is reinforced by Appendix F's analysis of clusters.

D. Specification test and analysis

As explained in the Introduction, our a priori specification might omit empirically relevant types and/or include irrelevant ones, and this leaves some room for doubt regarding our 45 subjects whose types are not apparent from their guesses. To learn whether any of their guesses could be better explained by omitted types, or whether any of their type estimates are due to accidental correlations with included irrelevant types, we conduct a new specification test.

To understand the logic of the test, imagine that we had estimated subjects' types as before, but using a specification that omitted an empirically relevant type, say *L2*. Then the patterns of guesses across games of subjects whose behavior is best described by *L2* would tend to resemble one another more than any included type. We search for such clusters of subjects by reëstimating Section II.C's model, subject by subject, with the original list of possible types augmented by 88 *pseudotypes*, one constructed from each of our subject's guesses in the 16 games.[32] We then compare the likelihood of each subject's original type estimate with the likelihoods of the 87 other subjects' pseudotypes. Define a *cluster* as a group of two or more subjects such that: (i) each subject's original estimated type has smaller likelihood than the pseudotypes of all other subjects in the group; and (ii) all subjects in the group make "sufficiently similar" guesses.[33] Finding such a cluster should lead us to diagnose an omitted type, and studying the common

---

[31]We report these tests only as a simple way to gauge the strength of the evidence provided by our data. Their standard justifications are unavailable, here and below, because the null hypotheses involve boundary parameter values. We approximated the test for $\varepsilon = 1$ using a non-boundary value of ε just below one.

[32]We are grateful to Jerry Hausman for suggesting the idea of this test. We allow spike-logit errors for pseudotypes to avoid biasing the tests against them. The logit term's dependence on expected payoffs means that to define a pseudotype's error density we must infer beliefs, because pseudotypes do not come with built-in models of others. We do this as simply as possible, by assuming that the pseudotypes' guesses are best responses and inferring point beliefs, game by game, from their subjects' guesses. For a dominated guess, or a guess at a limit that is a best response to multiple beliefs, we infer the beliefs that bring the pseudotype's guess closest to maximizing payoff.

[33]Not requiring significantly higher likelihood in (i) avoids ruling out cluster candidates because their pseudotypes offer only slight improvements in fit; few of the comparisons are very close. The "sufficiently similar" in (ii) could be made more precise, but it is more informative to consider possible clusters on a case by case basis (Appendix F). Although the logic of our definition allows overlapping but non-nested clusters, that problem does not arise here.

elements of its subjects' guesses may help to reveal its decision rule. Conversely, not finding a cluster suggests that there are no empirically important omitted types.[34]

Similarly, we can diagnose overfitting via accidental correlations with included irrelevant types if a subject's estimated type performs no better than random against the pseudotypes other than his own. For a pseudotype to have higher likelihood than the subject's estimated type it must come first among our 7 possible types plus itself. If the likelihoods of the subject's estimated type and of our types and pseudotypes are approximately independent and identically distributed ("i.i.d."), this has probability approximately 1/8. We therefore diagnose overfitting if a subject's estimated type does not have higher likelihood than all but at most $87/8 \approx 11$ of the pseudotypes.

Appendix F summarizes the results of our search for clusters, identifying five, involving 11 subjects, indicated in Table 7 by superscript cluster labels A, B, C, D, and E on their type identifiers. Table 7 also identifies 15 subjects whose estimated types have lower likelihoods than 12 or more pseudotypes, indicated in Table 7 by superscripts + on their type identifiers. Table 1 (column 5) reports the aggregate type frequencies excluding these 26 subjects, and 4 others previously excluded, as "unclassified". Each of the remaining 58 subjects' estimated types, in bold with no superscripts in Table 7: (i) does better at the 1% level than a random model of guesses within our specification; (ii) has a higher likelihood than all but at most a random number of pseudotypes; and (iii) is not a member of any cluster.

Despite the differences between our games and those in previous studies, our type classification is close to Nagel's, HCW's, CGCB's, and SW's. There are two main differences. We find more *Equilibrium* subjects than all previous studies except SW's. And we find no significant numbers of types other than *L1*, *L2*, *Equilibrium*, and *L3*/*Equilibrium* hybrids, in contrast to SW's (1995) classification of many subjects as *Worldly*, almost to the exclusion of *L2*.

Our analysis also sheds light on the specification of level-*k* models. We stress that the issue here is how best to describe subjects' decision rules empirically, not how an analyst would model other subjects' behavior. However tempting it may be to assume that subjects use sophisticated econometric specifications, it might not be the best way to describe their behavior.

---

[34]The qualification "empirically important" is necessary because there may be subjects who follow rules that differ from our types but are unique in our dataset. Such subjects are unlikely to repay the cost of constructing theories of their behavior, and it seems difficult to test for them. Our test makes the search for omitted types manageable within the enormous space of possible types, while avoiding judgment calls about possible types by focusing on patterns of guesses like those subjects actually made. Our notion of cluster is similar in spirit to notions that have been proposed elsewhere, but it imposes much more structure, in a way that seems appropriate here.

Our *Lk* best responds to a noiseless *Lk-1*, while SW define it as best responding to an *Lk-1* with decision noise—as in QRE, but with non-equilibrium beliefs. (The issue here is not whether subjects' own decisions are noisy, but whether the deterministic part of their guesses responds to others' decision noise.) SW's and CHC's definition of *L1* as a best response to uniform beliefs is identical to ours. SW's *Worldly* best responds to an estimated mixture of *L1* and a noiseless *Equilibrium*; and CHC's *Lk* best responds to an estimated mixture of noiseless *Lk-1* and lower-level *Lk* types. Our results favor our *Lk* definition over SW's *Lk* with regard to decision noise, and suggest that SW's *Worldly* is misspecified. SW's *L2* best responds to a noisy *L1*, which depending on the noise parameter ranges from *L0* (uniform random) to our noiseless *L1*. By a kind of "median-voter" result, our not-everywhere-differentiable payoff function (Section I.B) makes it optimal to best respond to the median type in the population as if it were the only type.[35] Thus in our games *Worldly* ignores *Equilibrium* when its frequency is less than 0.5—as in all published estimates—and is then equivalent to SW's noisy *L2*. Because our payoff function is quasiconcave, SW's *L2* and *Worldly* guesses between our *L1* and *L2*—strictly except for extreme parameter values. Yet only one of our 88 subjects made guesses between our *L1* and *L2* in as many as 10 games, one in 9, and 2 in 8, while 43 made *exact* guesses for our *L1*, *L2*, *L3*, or *Equilibrium* in 7 or more games (Appendix E).[36]

Our results are inconclusive with regard to our *Lk* definition versus CHC's definition of *Lk* types as best responses to estimated Poisson mixtures of noiseless lower-level *Lk* types. Because CHC's mixture parameter depends on others' behavior, which subjects do not observe, their definition implicitly assumes that subjects have prior understandings of it; similar assumptions are implicit in SW's definitions of *L2* and *Worldly*. CGCB (2001, Section 3.A) argued that the *Sophisticated* type tests for such prior understandings more cleanly, without imposing structural restrictions, than types that depend on estimated parameters like CHC's *Lk* or SW's *Worldly*. More evidence on this would be useful, but in our games CHC's *L2* and *L3* both make the same guesses as our *L2*.[37] Thus, our results do not discriminate between CHC's and our *Lk* definitions.

---

[35]The derivative of our payoff function to the left (respectively, right) of its peak is positive (negative), and the two are equal in magnitude. Thus the sign of the expected derivative is determined by the median type in the distribution.
[36]On average, random guesses would fall in the range in 4.14 games. The 3 subjects with 8 or 9 guesses (115, 501, and 506) gave no useful information in their questionnaires, but the subject with 10 (517) stated a homemade rule inconsistent with *Worldly*: "I took the midpt of my bound times his/her target, avg'd that with his/her midpt, then mult'd that number by my target, and finally avg'd that result with my midpt." The prevalence of OB subjects in this group may seem significant, but there were no OB subjects among the 5 subjects with 7 guesses in the range.
[37]CHC's *L2* best responds to a mixture of *L0* and *L1* in the proportions 1:τ, which for τ > (<) 1 puts more weight on

E. Econometric analysis of Baseline subjects' guesses and information searches

In this section we generalize Section II.C's model of guesses to obtain an error-rate model of guesses and information searches, and use it to reëstimate Baseline subjects' types. The model follows Section II.C's model, avoiding unnecessary differences in the treatment of guesses and search. Our main goals are to summarize the implications of the search data and to assess the extent to which monitoring search modifies the view of behavior suggested by subjects' guesses.

The assumptions about how cognition drives information search and decisions that underlie our econometric analysis are supported by the search behavior of our Baseline subjects whose types are apparent from their guesses and of our R/TS subjects (Appendix E). The main new issue is measuring compliance with types' search implications, which we propose to define as the density of the type's relevant look-ups (Table 4) in the look-up sequence.

Two aspects of the look-up data (Appendix E) are important here. First, many subjects (e.g. 202 and 210) usually start with "123456" or some variation, and many end with an optional "13," checking their own limits even if their type does not require it (e.g. 101 and 206). We do *not* filter out these patterns because subjects may use the information they yield, and the choice of how to filter would involve hidden degrees of freedom.

Second, subjects' look-up patterns are heterogeneous in timing: Many Baseline subjects whose types are apparent from their guesses usually look first at their type's relevant sequence and then either make irrelevant look-ups or stop (e.g. 108, 118, and 206). A smaller number consistently make irrelevant look-ups first, and look at the relevant sequence only near the end (e.g. 413). Others repeat the relevant sequence over and over (e.g. 101). Thus one can identify three styles, "early," "late," and "often"; but "often" subjects are almost always well described as either "early" or "late". We filter out this heterogeneity using a binary nuisance parameter called *style*, which is assumed constant across games and modifies type in a way that affects only search implications. We take each subject to have style $s = e$ for "early" or $s = l$ for "late". For a given game, subject, type, and style, we define search compliance as the density of relevant look-ups early or late in the sequence. If $s = e$, we start at the beginning and continue until we obtain a complete relevant sequence. If we never obtain such a sequence, compliance is 0. Otherwise compliance is the ratio of the length of the relevant sequence to the number of look-ups that first

---

L1 (*L0*). By the above "median-voter" result, CHC's *L2* best responds to *L1* alone if $\tau > 1$, or *L0* alone if $\tau < 1$. They argue that $\tau \approx 1.5$ in most applications, in which case their *L2* is confounded with our *L2*. Their *L3*, which best responds to a mixture of *L0*, *L1*, and *L2* in proportions $1{:}\tau{:}\tau^2/2$, is also confounded with our *L2* when $\tau \approx 1.5$.

yields a complete sequence. If, for instance, the relevant sequence has length six, and the first complete sequence is obtained after eight look-ups, then compliance is 0.75. The definition of search compliance is identical if $s = l$, but starting from the end of the sequence. Compliance for a given type is thus a number from 0 to 1, comparable across styles, games, and subjects.[38]

To reduce the need for structural restrictions, we discretize search compliance as follows.[39] For each game, subject, type, and style, we sort compliance into three categories: $C_H \equiv [0.667, 1.00]$, $C_M \equiv [0.333, 0.667]$, and $C_L \equiv [0, 0.333]$, indexed by $c = H, M, L$. We call compliance c for type k and style s *type-k style-s compliance c*, or just *compliance c* when the type and style are clear from the context. All products over c are taken over the values H, M, and L.

In our model, in each game a subject's type and style determine his information search and guess, each with error. We assume that, given type and style, errors in search and guesses are independent of each other and across games. We describe the joint probability distribution of guesses and search by specifying compliance probabilities and guess error rates and precisions, given type and style.[40] Let I be an indicator variable for style, with $I_s = 1$ when the subject has style s ($= e$ or l) and 0 otherwise. Given a subject's type and style, let ζ_c be the probability that he has type-k style-s compliance c in any given game, where $\sum_c \zeta_c = 1$, and let $\zeta \equiv (\zeta_H, \zeta_M, \zeta_L)$. As in Section II.C, in each game g, a subject i of type k and style s makes an adjusted guess in U_g^{ik} with probability $1 - \varepsilon$; but with probability $\varepsilon \in [0,1]$, his adjusted guess in V_g^{ik} has conditional density $d_g^k(R_g^i(x_g^i), \lambda)$ with precision λ defined as in (4). Let M_c^{isk} be the set of games g for which subject i has type-k style-s compliance c, let $M^{isk} \equiv (M_H^{isk}, M_M^{isk}, M_L^{isk})$, and let m_c^{isk} be the number of games in M_c^{isk}, so $\sum_c m_c^{isk} = G$. Let N_c^{isk} be the set of games g for which subject i has both type-k style-s compliance c and $R_g^i(x_g^i) \in V_g^{ik}$, let $N^{isk} \equiv (N_H^{isk}, N_M^{isk}, N_L^{isk})$, let n_c^{isk} be the number of games in N_c^{isk}, and let $n^{ik} = \sum_c n_c^{isk}$ (for $s = e$ or l) be the number of games g for which

---

[38]The compliance data are in Appendix H. For *D1*, *D2*, and *Sophisticated* we take the relevant sequence to have length 6, the minimum with which one could satisfy their requirements, e.g. via "153426" for *D1* with requirements {(4,[5,1]),(6,[5,3]),2}, or for *D2* or *Sophisticated* with requirements {(1,[2,4]),(3,[2,6]),(4,[5,1]),(6,[5,3]),5,2}.
[39]Compliance is inherently discrete, but our discretization is coarser than necessary.
[40]A natural generalization would allow search and guess errors to be correlated by allowing compliance-contingent error rates and precisions as in CGCB. We dispense with this for simplicity. This is a convenient place to correct a typographical error in CGCB's equation (4.3), where the summation ($\sum$) should be a product ($\prod$).

subject i has $R_g^i(x_g^i) \in V_g^{ik}$. With i.i.d. errors, the density of a sample with compliance M^{isk} and N^{isk} and adjusted guesses $R^i(x^i) \equiv (R_1^i(x_1^i),...,R_G^i(x_G^i))$ for a subject i of type k and style s is:

$$(8) \qquad d^{sk}(M^{isk}, N^{isk}, R^i(x^i); \varepsilon, \lambda, \zeta) \equiv \prod_c \left[(\zeta_c)^{m_c^{isk}} (1-\varepsilon)^{m_c^{isk}-n_c^{isk}} (\varepsilon)^{n_c^{isk}} \prod_{g \in N_c^{isk}} d_g^k(R_g^i(x_g^i), \lambda) \right],$$

where products with no terms are taken to equal 1. Weighting by I_s and p_k, summing over s and k, and taking logarithms yields subject i's log-likelihood:

$$(9) \qquad L^i(p, s, \varepsilon, \lambda, \zeta \mid M^{isk}, N^{isk}, R^i(x^i)) \equiv \ln \left[\sum_{k=1}^K p^k \sum_{s=e,l} I_s d^{sk}(M^{isk}, N^{isk}, R^i(x^i); \varepsilon, \lambda, \zeta) \right].$$

It is clear from (8) and (9) that the maximum likelihood estimate of p sets $p^k = 1$ and $I_s = 1$ for the (generically unique) type k and style s with the highest $d^{sk}(M^{isk}, N^{isk}, R^i(x^i); \varepsilon, \lambda, \zeta)$, given the estimated ε, λ, and ζ. The maximum likelihood estimates of ε and ζ_c, conditional on type k and style s, can be shown from (8) to be n^{ik}/G and m_c^{isk}/G, the sample frequencies with which subject i's adjusted guesses fall in V_g^{ik} for that k and he has compliance c for that k and s. The maximum likelihood estimate of λ is again the logit precision, restricted to guesses in V_g^{ik}.

The maximum likelihood estimate of subject i's type k maximizes the logarithm of (8) over k and s, given the estimated ε and λ. When n^{ik} is between 0 and G, substituting the estimated ζ_c, ε, and λ into (8), taking logarithms, using $\sum_c m_c^{isk} = G$, $\sum_c n_c^{isk} = n^{ik}$, and $\bigcup_c N_c^{isk} = N^{ik}$ (all for $s = e$ or l), simplifying and collecting terms, yields the maximand:

$$\ln d^{sk}(M^{isk}, N^{isk}, R^i(x^i); \varepsilon, \lambda, \zeta) \equiv$$

$$(10) \quad \sum_c \left[m_c^{isk} \ln(\zeta_c) + (m_c^{isk} - n_c^{isk}) \ln(1-\varepsilon) + n^{isk} \ln(\varepsilon) + \sum_{g \in N_c^{isk}} \ln d_g^k(R_g^i(x_g^i), \lambda) \right] \equiv$$

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

$$\ln d^k(R^i(x^i), \varepsilon, \lambda)) + \sum_c \left[m_c^{isk} \ln m_c^{isk} \right] - G \ln G,$$

where $\ln d^k(R^i(x^i), \varepsilon, \lambda))$ is the log-likelihood of the guesses-only model defined in (7). Thus search adds an additively separable term in search compliance, minus an additional term $G \ln G$. As in Section II.C's model, when $n^{ik} = 0$ or G, $\ln d^k(R^i(x^i), \varepsilon, \lambda))$ reduces to the sum over g in

the second-to-last line of (10). When n_c^{isk} or both m_c^{isk} and $n_c^{isk} = 0$ for some c ($m_c^{isk} \geq n_c^{isk}$ by definition), the corresponding terms drop out of (8) and their analogs are eliminated from (10).

The model now has six independent parameters per subject: error rate ε, precision λ, type k, style s, and two independent compliance probabilities ζ_c. The maximum likelihood estimates of ε, ζ_c, and λ, given k and s, are n^{ik}/G, m_c^{isk}/G, and the standard logit precision. The estimates of k and s maximize the expression in (10), given the other estimates.

Guesses influence these estimates exactly as in Section II.C's model, and unless the estimated k changes the estimates of ε and λ are the same; but now the estimated k is influenced by information search as well as guesses. The search term in the last line of (10) is a convex function of the m_c^{isk}. This favors k-s combinations for which the m_c^{isk} (or the estimated ζ_c) are more concentrated on particular levels of c, because their search implications explain more of the variation in search patterns. Note that such combinations are favored without regard to whether the levels of c on which the m_c^{isk} are concentrated are high or low. We avoid such restrictions because levels of search compliance are not meaningfully comparable across types and it would be arbitrary to favor a type just because its compliance requirements are easier to satisfy. Without them, however, the likelihood may favor a type simply because compliance is 0 in many or all games (0 compliance is independent of style). We deal with this as simply as possible, by ruling out a priori types for which a subject has 0 (not just L) compliance in 8 or more games.[41]

The center and right-hand side of Table 7 report estimates of Baseline subjects' types and styles, error rates, precisions, and rates of search compliance, based first on search only and then on guesses and search combined. For the latter we report separate as well as total log-likelihoods, to give a better indication of what drives them.[42] Most subjects' type estimates based on guesses and search reaffirm the guesses-only estimates.[43] For some subjects, however, the guesses-and-

---

[41]The cutoff of 8 is a conservative response to the difficulty of specifying a precise model of search compliance. A more standard but more complex approach, in the spirit of CGCB's use of their Occurrence assumption in defining search compliance, would add a separate category for 0 compliance; estimate a subject's probability, given type and style, of having positive compliance; and require it to be sufficiently greater than 0. This would have a similar effect.
[42]Ties in the search-only or guesses-and-search type-style estimates are *not* rare, due to our coarse categorization. When they occur we report the tied estimate closest to the guesses-only estimate, indicating the others in the notes. Most subjects' style estimates are early but there is a sizeable minority of late estimates, suggesting that without the style parameter, our characterization of search compliance would distort the implications of some subjects' searches.
[43]This happens in part because the guess part of the log-likelihood is nearly 6 times larger than the search part, and so has much more weight in determining the estimates based on guesses and search. The difference in weights arises

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.[44]

Table 1 (column 6) reports the aggregate type frequencies based on (10). When the guesses-and-search type estimate differs from the guesses-only estimate, we favor the former but require it to pass the analogs of the guesses-only criteria. Under the updated criteria, we can classify 43 of our 71 Baseline subjects: 22 *L1*, 13 *L2*, and 8 *Equilibrium*.[45] The search analysis allows us to classify 2 subjects as *L1* who were previously classified as *L2*; and to classify 1 previous *L1* and 1 previous *L2* subject whose guesses-only estimates were inconclusive. It also calls into question the classification of 4 subjects: 1 each *L1* (subject 415), *L2*, *D1*, and *Equilibrium*. Each of the 43 Baseline subjects now classified, with guesses-and-search type identifiers in bold and with no superscripts (though some have subscripts) in Table 7: (i) does better at the 1% level than a random model of guesses and search within our specification; (ii) the guesses-only part of its likelihood is higher than the guesses-only likelihood for all but at most a random number of pseudotypes; and (iii) is not a member of any cluster.[46]

Adding to these Baseline subjects the 12 of 17 OB subjects (7 *L1*, 1 *L2*, 1 *L3*, 2 *Equilibrium*, and 1 *Sophisticated*) previously classified, with guesses-only type identifiers in bold in Table 7, we have a total of 55 of 88 subjects who can be classified: 29 *L1*, 14 *L2*, 1 *L3*, 10 *Equilibrium*, and 1 *Sophisticated*.[47] Going beyond our criteria, one might add subject 415 as a probable *L1* and the 4 subjects in clusters A and B (Appendix F) as likely hybrids of *L3* and/or

---

because our theory makes sharper predictions about guesses than about search, which are far less likely to be satisfied by chance. If we tried to put search on a more equal footing by making sharper predictions, e.g. requiring more precise levels of compliance within a finer categorization, our subjects' searches would rarely satisfy types' search implications, and the stronger restrictions would cause severe specification bias.

[44]This group includes subject 415, estimated *L1* on guesses (with 9 exact) but (noisy) *D1* on guesses and search. Subject 415 has 9 games with 0 *L1* search compliance due to no adjacent [4,6]'s or [6,4]'s (Table 4), but his sequences are rich in [4,2,6]'s and [6,2,4]'s and *L1* search compliance across games is weakly correlated with *L1* guesses. We therefore believe that this subject simply violated our assumption that basic operations are represented by adjacent look-ups (Section I.B).This group also includes several subjects whose guesses-only type estimates we believe were rightly excluded: 115, 204, and 401, estimated *D1* based on guesses but *Equilibrium* or *L1* on guesses and search; 112, estimated *Equilibrium* based on guesses but *L2* on guesses and search; and 304 and 421, estimated *Sophisticated* based on guesses but *Equilibrium* or *L1* on guesses and search.

[45]A guesses-and-search type estimate can satisfy the classification criteria even if it did not satisfy the guesses-only criteria if it does sufficiently better than random in explaining search. But a guesses-and-search estimate may fail the new criteria because it must have the same or lower likelihood for guesses than the guesses-only type estimate.

[46]In (ii) we include OB subjects' pseudotypes for comparability with guesses-only results, so random still means 11.

[47]Other subjects' low levels of compliance with *Sophisticated*'s search requirements suggest that the identification of the one *Sophisticated* subject, who was a noisy OB subject, might not have survived monitoring search.

Equilibrium. Either way, the search analysis refines and sharpens our conclusions, and confirms the absence of significant numbers of subjects of types other than *L1*, *L2*, *Equilibrium*, or possibly hybrids of *L3* and/or *Equilibrium*. For the 28 to 33 unclassified subjects, our specification analysis suggests that it will be difficult to improve upon a random model.

III. Conclusion

This paper has reported an experiment that elicits subjects' initial responses to a series of 16 two-person guessing games, monitoring their searches for hidden payoff information along with their guesses. Our design yields strong separation of the guesses and searches implied by leading decision rules in a very large space of possible behaviors. Many subjects' guesses yield clear strategic fingerprints, so that their types can be read directly from their guesses. Other subjects' types can be identified via an econometric and specification analysis. The full analysis reveals large numbers of *L1*, *L2*, *Equilibrium*, and *L3* and/or *Equilibrium* hybrid subjects, and indicates the absence of significant numbers of other types. Thus, to the extent that our subjects' deviations from equilibrium can be predicted, they appear to be based almost entirely on level-*k* thinking.

Because our level-*k* types build in risk-neutral, self-interested rationality and perfect models of the game, many subjects' systematic deviations from equilibrium can be confidently attributed to non-equilibrium beliefs rather than irrationality, risk aversion, altruism, spite, or confusion. Thus our results affirm subjects' rationality and ability to comprehend complex games and reason about others' responses to them, while challenging the use of equilibrium as the principal model of initial responses. They are consistent with the results of previous analyses, but significantly refine and sharpen them. The surprisingly simple structure of the alternative non-equilibrium model they suggest should help to allay the common fear that if equilibrium is not assumed, "anything can happen." Moreover, such models have already been used in several applications, including Crawford (2003); CHC; and Crawford and Iriberri (2005a, 2005b).

We close by noting that the cognitive implications of our results suggest conclusions about the structure of learning rules. Our subjects' comprehension of the games and tendencies toward exact best responses to the beliefs implied by simplified models of others point clearly away from reinforcement learning and toward beliefs-based models like weighted fictitious play or hybrids like Camerer and Ho's (1999) experience-weighted attraction learning. We plan in future experiments to use information search to discriminate among alternative theories of learning,

whose search implications are often more sharply separated than their implications for decisions.

References

Beard, T. Randolph and Beil, Richard. "Do People Rely on the Self-interested Maximization of Others?: An Experimental Test," *Management Science*, 1994, 40 (2), pp. 252-262.

Bernheim, B. Douglas. "Rationalizable Strategic Behavior," *Econometrica*, 1984, 52 (4), pp. 1007-1028.

Camerer, Colin and Ho, Teck-Hua. "Experience-weighted Attraction Learning in Normal Form Games," *Econometrica*, 1999, 67 (4), pp. 827-874.

Camerer, Colin, Ho, Teck-Hua and Chong, Juin Kuan: "A Cognitive Hierarchy Model of Games," *Quarterly Journal of Economics*, 2004, 119 (3), pp. 861-898.

Camerer, Colin, Johnson, Eric, Rymon, Talia and Sen, Sankar "Cognition and Framing in Sequential Bargaining for Gains and Losses," in Kenneth Binmore, Alan Kirman, and Piero Tani, eds., *Frontiers of Game Theory*. Cambridge: MIT Press, 1993, pp. 27-47.

Costa-Gomes, Miguel and Crawford, Vincent. "Studying Cognition via Information Search in Experimental Two-Person Guessing Games," Unpublished Paper, 2006.

Costa-Gomes, Miguel, Crawford, Vincent and Broseta, Bruno. "Cognition and Behavior in Normal-Form Games: An Experimental Study." University of California, San Diego, Discussion Paper Series: DP 98-22, 1998.

Costa-Gomes, Miguel, Crawford, Vincent and Broseta, Bruno. "Cognition and Behavior in Normal-Form Games: An Experimental Study," *Econometrica*, 2001, 69 (5), pp. 1193-1235.

Costa-Gomes, Miguel and Weizsäcker, Georg. "Stated Beliefs and Play in Normal-Form Games," Unpublished Paper, 2005.

Crawford, Vincent and Iriberri, Nagore. "Fatal Attraction: Focality, Naivete, and Sophistication in Experimental Hide-and-Seek Games." University of California, San Diego, Discussion Paper 2004-12R, revised February 2005. (a)

Crawford, Vincent and Iriberri, Nagore. "Level-*k* Auctions: Can a Non-Equilibrium Model of Strategic Thinking Explain the Winner's Curse and Overbidding in Private-Value Auctions?" University of California, San Diego, Discussion Paper 2005-13, November 2005. (b)

Grosskopf, Brit and Nagel, Rosemarie. "Rational Reasoning or Adaptive Behavior? Evidence from Two-Person Beauty Contest Games." Harvard Business School Negotiation, Organizations and Markets Unit, Research Paper No. 01-09, October 2001.

Ho, Teck Hua, Camerer, Colin and Weigelt, Keith. "Iterated Dominance and Iterated Best Response in Experimental '*p*-Beauty Contests'." *American Economic Review*, 1998, 88 (4), pp. 947-969.

Johnson, Eric, Camerer, Colin, Sen, Sankar and Rymon, Talia. "Detecting Failures of Backward Induction: Monitoring Information Search in Sequential Bargaining." *Journal of Economic Theory*, 2002, 104 (1), pp. 16-47.

McKelvey, Richard and Palfrey, Thomas. "Quantal Response Equilibria for Normal-Form Games." *Games and Economic Behavior*, 1995, 10 (1), pp. 6-38.

Nagel, Rosemarie. "Unraveling in Guessing Games: An Experimental Study." *American Economic Review*, 1995, 85 (5), pp. 1313-1326.

Selten, Reinhard. "Features of Experimentally Observed Bounded Rationality," *European Economic Review*, 1998, 42 (3-5), pp. 413-436.

Stahl, Dale and Wilson, Paul. "Experimental Evidence on Players' Models of Other Players." *Journal of Economic Behavior and Organization*, 1994, 25 (3), pp. 309-327.

Stahl, Dale and Wilson, Paul. "On Players' Models of Other Players: Theory and Experimental Evidence," *Games and Economic Behavior*, 1995, 10 (1), pp. 218-254.

Weibull, Jörgen. "Testing Game Theory." in Steffen Huck (ed.), *Advances in Understanding Strategic Behaviour: Game Theory, Experiments, and Bounded Rationality: Essays in Honour of Werner Guth*. New York: Palgrave MacMillan, 2004, pp. 85-104.
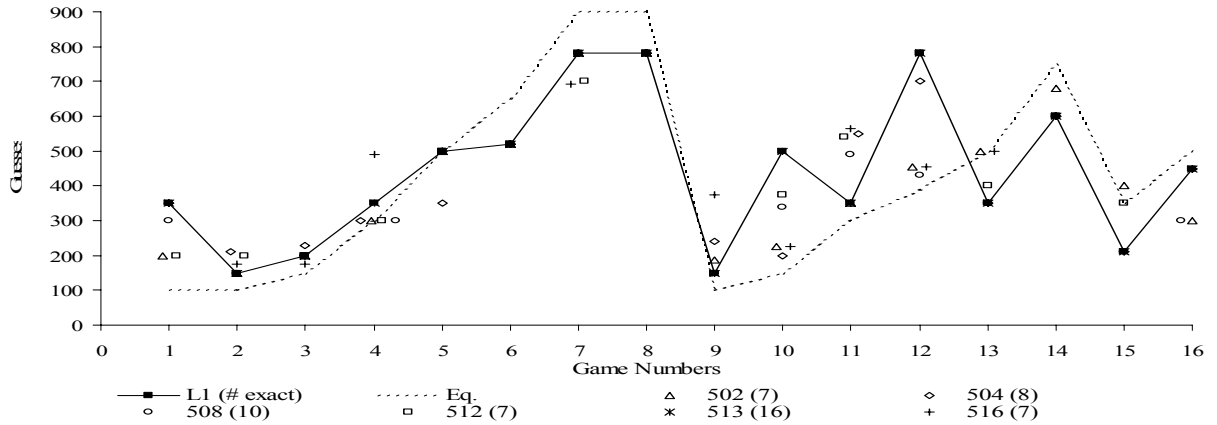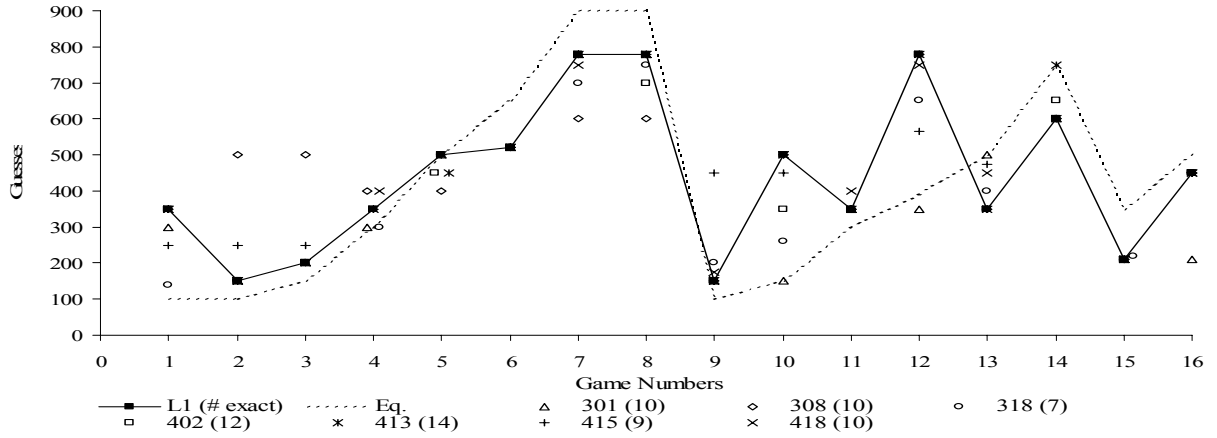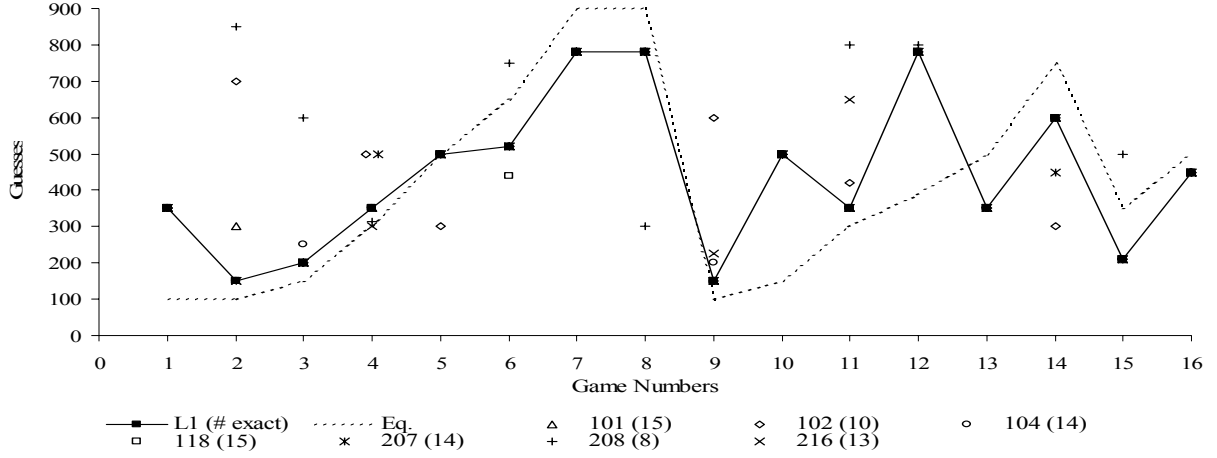
Figure 1. "Fingerprints" of 20 Apparent *L1* Subjects

Note: Only deviations from *L1*'s guesses are shown.

216 (68%) of these subjects' 320 guesses were exact *L1* guesses.
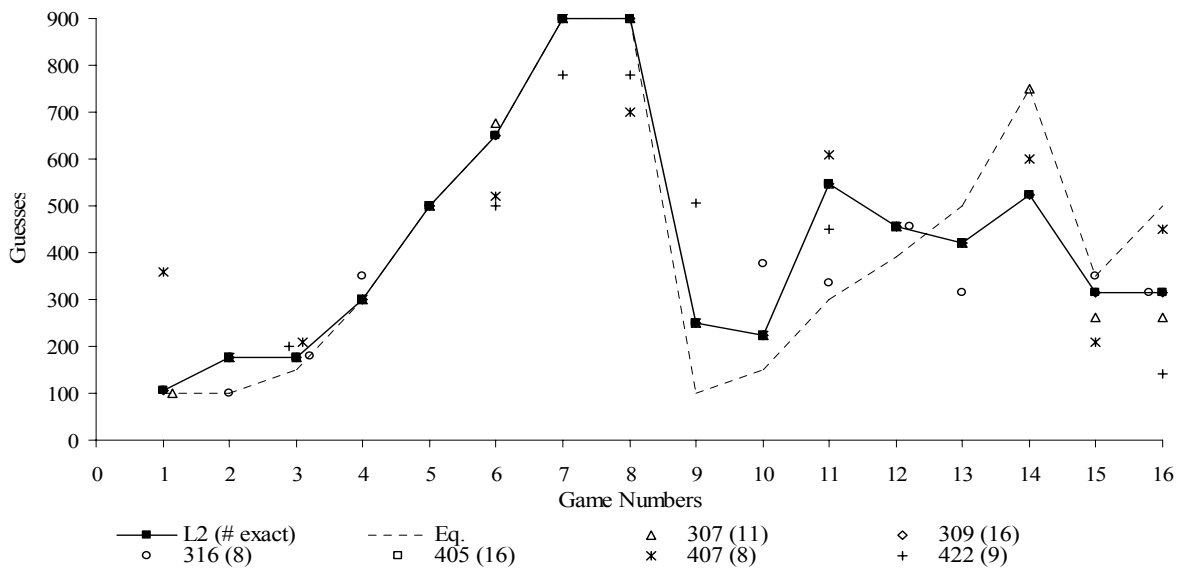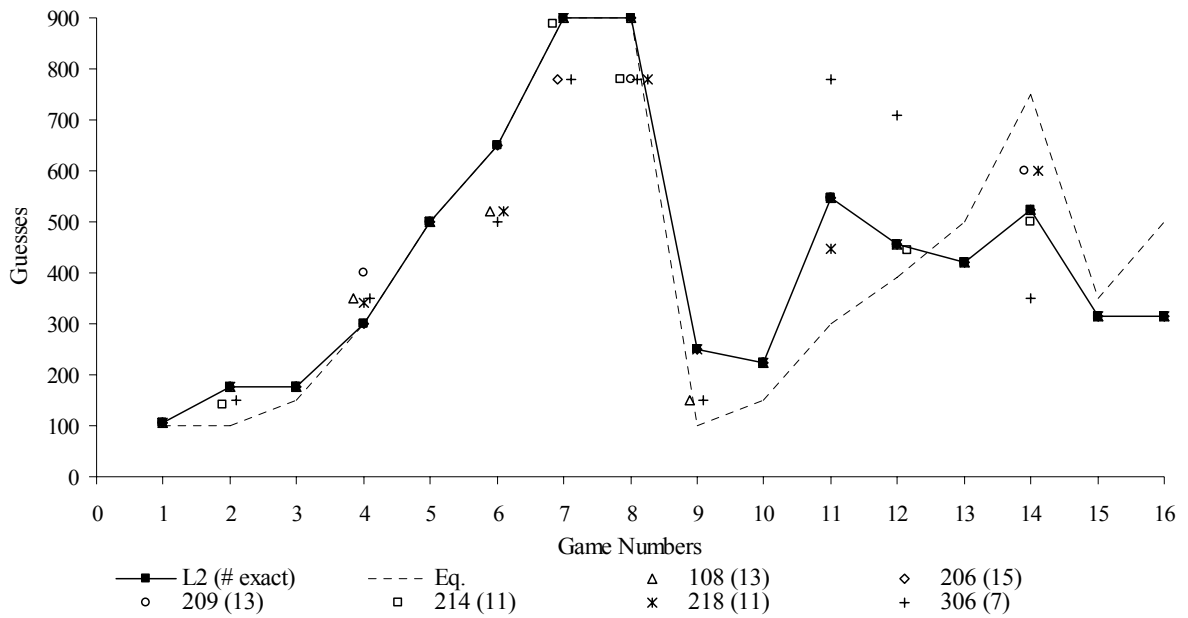
Figure 2. "Fingerprints" of 12 Apparent *L2* Subjects

Note: Only deviations from *L2*'s guesses are shown.

138 (72%) of these subjects' 192 guesses were exact *L2* guesses.
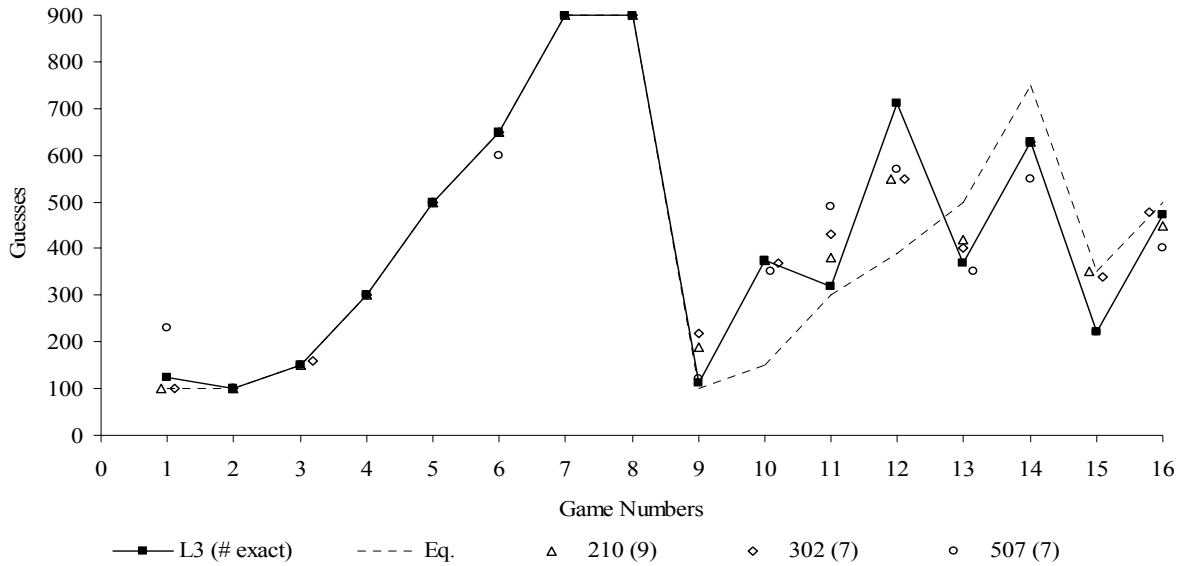
Figure 3. "Fingerprints" of 3 Apparent *L3* Subjects

Note: Only deviations from *L3*'s guesses are shown.

23 (48%) of these subjects' 48 guesses were exact *L3* guesses.
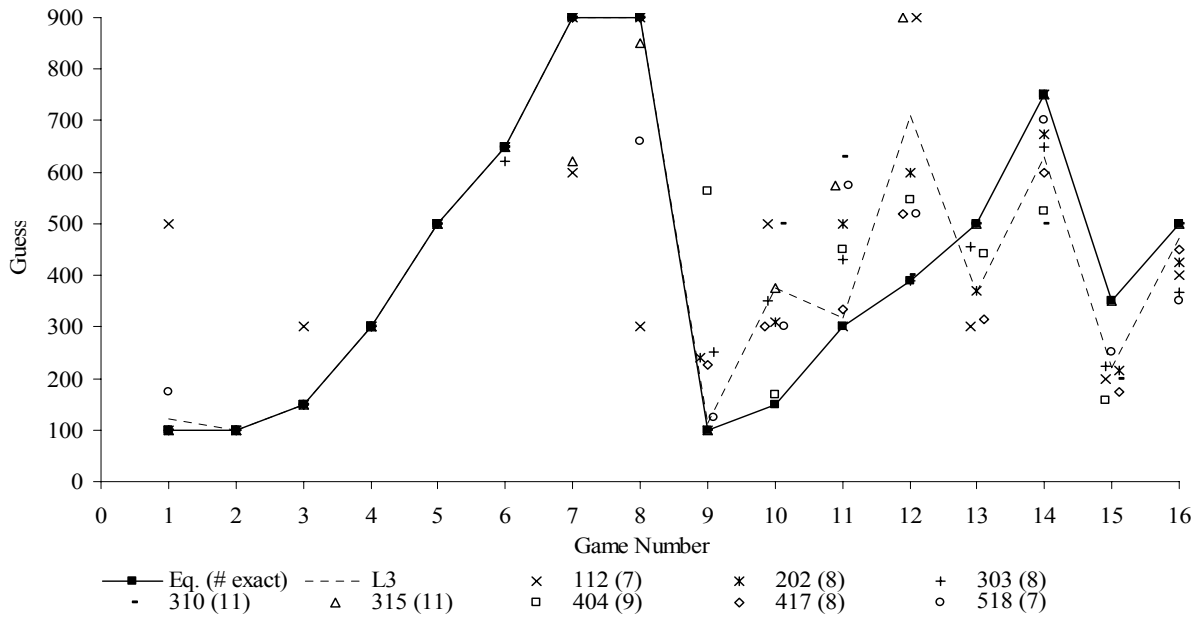
Figure 4. "Fingerprints" of 8 Apparent *Equilibrium* Subjects

Note: Only deviations from *Equilibrium*'s guesses are shown.

69 (54%) of these subjects' 128 guesses were exact *Equilibrium* guesses.
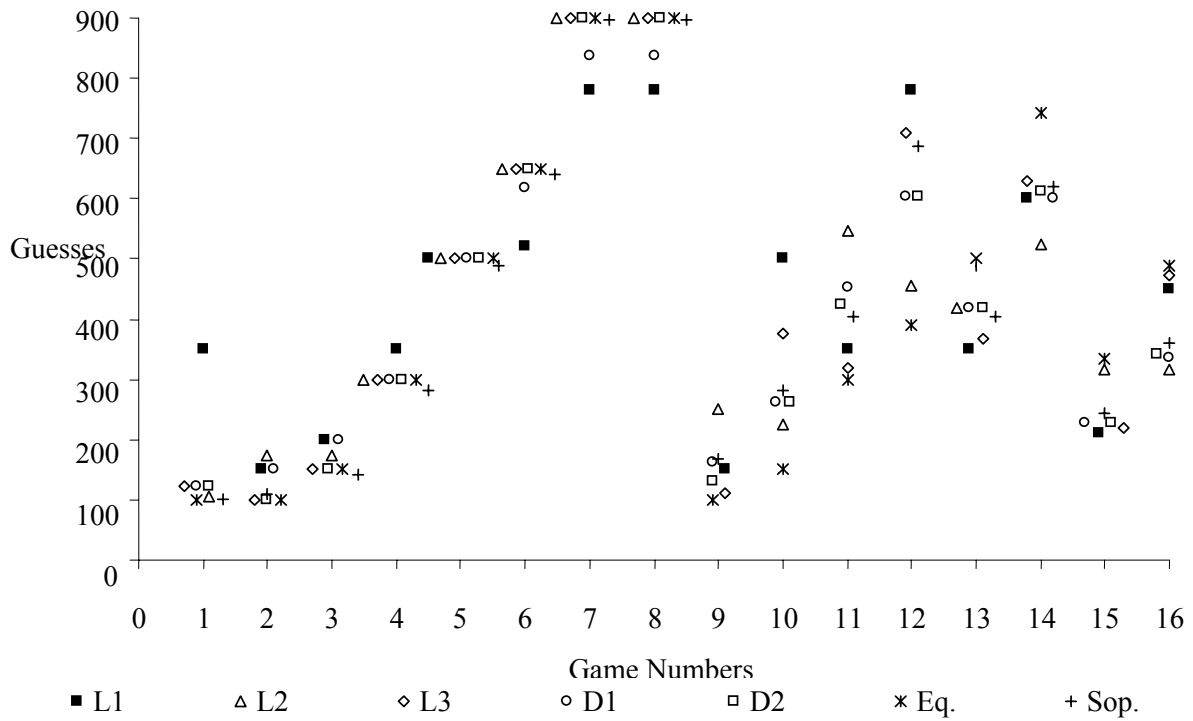
Figure 5. Separation of Types' Predicted Guesses Across Games
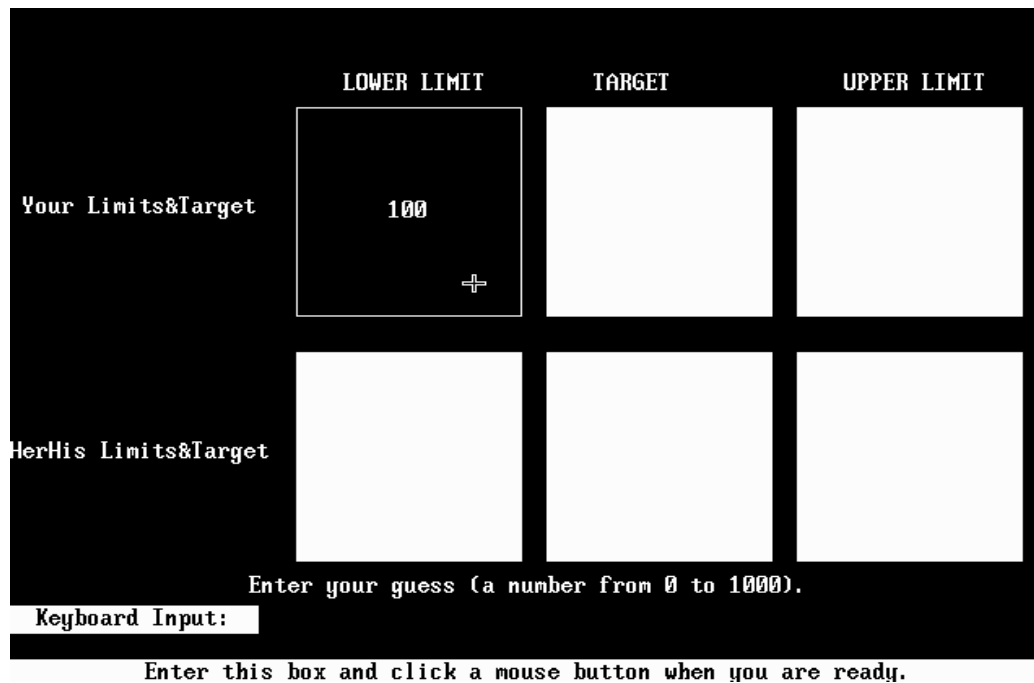
Figure 6. Screen Shot of the MouseLab Display

36

Player *i* $p^i = 1.5$ Eq. Guess
                                        ↓

          $R(1)$ $R(2)$

    |——————————————|———|

  300   450   500

Player *j* $p^j = 1.3$ Eq. Guess
                                                      ↓

        $R(1)$ $R(2)$ $R(3)$ $R(1)$

    |————|————————————|———|———————————————|

  300   390   585   650   900
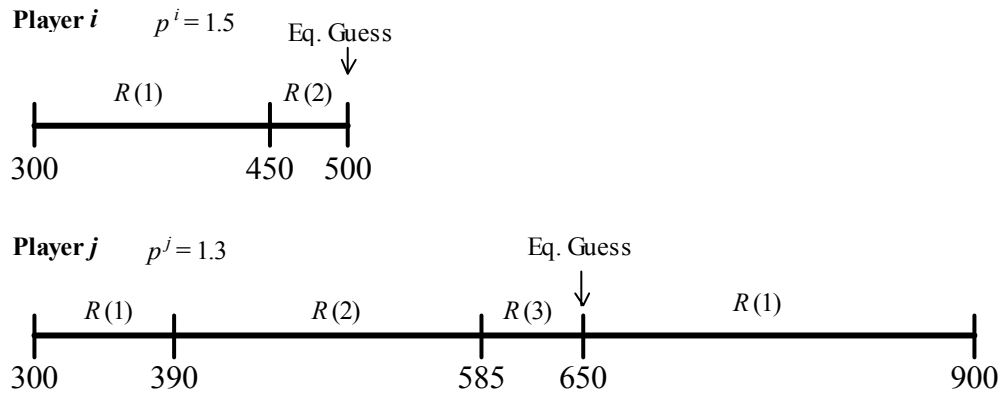
Table 1. Summary of Baseline and OB Subjects' Estimated Type Distributions

Type	Apparent from Guesses	Econometric from Guesses	Econometric from Guesses, Excluding Random	Econometric from Guesses, with Specification Test	Econometric from Guesses and Search, with Specification Test
L1	20	43	37	27	29
L2	12	20	20	17	14
L3	3	3	3	1	1
D1	0	5	3	1	0
D2	0	0	0	0	0
Eq.	8	14	13	11	10
Soph.	0	3	2	1	1
Unclassified	45	0	10	30	33

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

Table 2. Overall Structure of the Experimental Design

Session	Date	Location	Subjects
B1	1/31/2002	UCSD	13
B2	4/19/2002 (a.m.)	UCSD	20
B3	4/19 2002 (p.m.)	UCSD	17
B4	5/24/2002 (a.m.)	UCSD	21
OB1	5/24/2002 (p.m.)	UCSD	17
R/TS1	2/1/2002	UCSD	13: 4 *L1*, 5 *L2*, 4 *Equilibrium*
R/TS2	5/20/2002 (a.m.)	UCSD	5 *Equilibrium*
R/TS3	5/20/2002 (p.m.)	UCSD	8 *D1*
R/TS4	5/23/2002	UCSD	11: 3 *L1*, 4 *L2*, 3 *D1*, 1 *Equilibrium*
R/TS5	4/25/2003	York	10 *L3*
R/TS6	4/30/2003	York	11: 2 *L3*, 9 *D2*
R/TS7	5/1/2003	York	11: 3 *L2*, 2 *L3*, 1 *D1*, 2 *D2*, 3 *Equilibrium*
R/TS8	5/6/2003	York	8: 3 *D1*, 2 *D2*, 3 *Equilibrium*
R/TS9	5/9/2003	York	12: 1 *L2*, 1 L3, 3 *D1*, 1 *D2*, 6 *Equilibrium*
R/TS10	5/14/2003	York	12: 2 *L2*, 5 *D1*, 1 *D2*, 4 *Equilibrium*
R/TS11	5/21/2003	York	10: 3 *L1*, 4 *L2*, 3 *D1*
R/TS12	5/23/2003	York	5 *L1*
R/TS13	5/28/2003	York	8: 4 *L1*, 4 *L2*
R/TS14	5/30/2003	York	12: 3 *L1*, 2 *L2*, 2 *L3*, 2 *D1*, 3 *D2*
R/TS15	6/10/2003	York	12: 3 *L1*, 2 *L2*, 1 *L3*, 2 *D1*, 1 *D2*, 3 *Equilibrium*

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

Note: Game identifiers: Limits α for 100 and 500, β for 100 and 900, γ for 300 and 500, or δ for 300 and 900; targets 1 for 0.5, 2 for 0.7, 3 for 1.3, 4 for 1.5. Low targets are < 1; High targets are > 1; Mixed targets are one < 1, one > 1. High equilibrium is determined by players' upper limits; Low equilibrium is determined by players' lower limits. Rounds of dominance refers to the number player i needs to identify his equilibrium guess. Alternating dominance (A) occurs first for one player, then the other, then the first, etc.; simultaneous dominance (S) occurs for both players at once; and simultaneous then alternating dominance (S/A) is simultaneous in the first round and then alternating. Dominance at both ends refers to whether guesses are eliminated near both of a player's limits.

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

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	$p^i[\max\{\max\{a^i,p^ja^i\},p^j\max\{a^i,p^ja^j\}\}$ $+\min\{p^j\min\{p^jb^j,b^i\},\min\{p^jb^i,b^j\}\}]/2$	$\{(a^i,[p^j,a^i]),(b^i,[p^i, b^j]),(a^j,[p^j,a^i]),(b^j,[p^j,b^i]),p^j,p^i\}$ $\equiv \{(1,[2,4]),(3,[2,6]),(4,[5,1]),(6,[5,3]),5,2\}$
Eq.	$\{a^i$ if $p^ia^j \le a^i$ or $\min\{p^ia^j,b^i\}$ if $p^ia^j > a^i\}$ if $p^ip^j < 1$ or $\{b^i$ if $p^ib^j \ge b^i$ or $\max\{a^i,p^ib^j\}$ if $p^ib^j < b^i\}$ 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; *Sophisticated*'s search implications are the same as *D2*'s]	$\{(a^i,[p^j,a^i]),(b^i,[p^i, b^j]),(a^j,[p^j,a^i]),(b^j,[p^j,b^i]),p^j,p^i\}$ $\equiv \{(1,[2,4]),(3,[2,6]),(4,[5,1]),(6,[5,3]),5,2\}$

Note: The most basic operations are represented by the innermost look-ups, grouped within square brackets; these can appear in any order, but *may not* be separated by other look-ups. Other operations are represented by look-ups grouped within parentheses or curly brackets; these can appear in any order, and *may* be separated by other look-ups. *Equilibrium*'s minimal search implications are derived not directly from *Equilibrium*'s ideal guesses, but from p^ia^j when $p^ip^j < 1$ and p^ib^j when $p^ip^j > 1$ via Observation 1 (see Appendix H).

Table 5. Types' Adjusted Guesses and Guesses that Survive Iterated Dominance

Game	*L1*	*L2*	*L3*	*D1*	*D2*	*Eq.*	*Soph.*	1 round	2 rounds	3 rounds	4 rounds
					Player *i*'s guess for type				Range of iteratively undominated guesses		
1. α2β1	350	105	122.5	122.5	122.5	100	122	100, 500	100, 175	100, 175	100, 100
2. β1α2	150	175	100	150	100	100	132	100, 250	100, 250	100, 100	100, 100
3. β1γ2	200	175	150	200	150	150	162	150, 250	150, 250	150, 150	150, 150
4. γ2β1	350	300	300	300	300	300	300	300, 500	300, 300	300, 300	300, 300
5. γ4δ3	500	500	500	500	500	500	500	450, 500	500, 500	500, 500	500, 500
6. δ3γ4	520	650	650	617.5	650	650	650	390, 650	585, 650	650, 650	650, 650
7. δ3δ3	780	900	900	838.5	900	900	900	390, 900	507, 900	659.1, 900	856.8, 900
8. δ3δ3	780	900	900	838.5	900	900	900	390, 900	507, 900	659.1, 900	856.8, 900
9. β1α4	150	250	112.5	162.5	131.25	100	187	100, 250	100, 250	100, 187.5	100, 187.5
10. α4β1	500	225	375	262.5	262.5	150	300	150, 500	150, 375	150, 375	150, 281.27
11. δ2β3	350	546	318.5	451.5	423.15	300	420	300, 630	300, 630	300, 573.3	300, 573.3
12. β3δ2	780	455	709.8	604.5	604.5	390	695	390, 900	390, 819	390, 819	390, 745.29
13. γ2β4	350	420	367.5	420	420	500	420	300, 500	315, 500	315, 500	330.75, 500
14. β4γ2	600	525	630	600	611.25	750	630	450, 750	450, 750	472.5, 750	472.5, 750
15. α2α4	210	315	220.5	227.5	227.5	350	262	100, 350	105, 350	105, 350	110.25, 350
16. α4α2	450	315	472.5	337.5	341.25	500	375	150, 500	150, 500	157.5, 500	157.5, 500

Table 6. B and OB Subjects' Aggregate Compliance with Iterated Dominance and *Equilibrium*

Game (# rounds of dominance)	Respects 1 or more rounds of dominance	Respects 2 or more rounds of dominance	Respects 3 or more rounds of dominance	Respects 4 or more rounds of dominance	*Equilibrium* exact compliance
All games	90 (62)	75 (41)	53 (34)	40 (26)	18 (0)
1. α2β1 (4)	100 (100)	34 (19)	34 (19)	13 (0)	13 (0)
2. β1α2 (3)	78 (19)	78 (19)	16 (0)	- (-)	16 (0)
3. β1γ2 (3)	73 (13)	73 (13)	10 (0)	- (-)	10 (0)
4. γ2β1 (2)	100 (100)	44 (0)	- (-)	- (-)	44 (0)
5. γ4δ3 (2)	81 (25)	69 (0)	- (-)	- (-)	69 (0)
6. δ3γ4 (3)	88 (43)	35 (11)	26 (0)	- (-)	26 (0)
7. δ3δ3 (5)	97 (85)	91 (66)	70 (40)	26 (7)	24 (0)
8. δ3δ3 (5)	95 (85)	94 (66)	68 (40)	20 (7)	20 (0)
9. β1α4 (9)	70 (19)	70 (19)	34 (11)	34 (11)	5 (0)
10. α4β1 (10)	100 (88)	56 (56)	56 (56)	31 (33)	3 (0)
11. δ2β3 (17)	86 (55)	86 (55)	81 (46)	81 (46)	5 (0)
12. β3δ2 (18)	97 (64)	92 (54)	92 (54)	69 (44)	1 (0)
13. γ2β4 (22)	100 (100)	97 (93)	97 (93)	94 (85)	20 (0)
14. β4γ2 (23)	88 (38)	88 (38)	84 (35)	84 (35)	8 (0)
15. α2α4 (52)	90 (63)	90 (61)	90 (61)	90 (60)	11 (0)
16. α4α2 (51)	98 (88)	98 (88)	94 (86)	94 (86)	6 (0)

Note: Compliance percentages are rounded to the nearest integer, with random compliance percentages in parentheses. Guesses that respect *k* or more rounds of dominance are *k*-rationalizable.

Table 7. Subject-by-Subject Econometric Type Estimates and Results of Specification Test

ID	dom.	Guesses	Only			Search	only			Guesses	and	search					
		$\ln L$	k	exact	λ	$\ln L$	k	ζ_H	ζ_M	$\ln L_t$	$\ln L_g$	$\ln L_s$	k	exact	λ	ζ_H	ζ_M
513	0	0.00	$L1$	16	-	-	-	-	-	-	-	-	-	-	-	-	-
118	0	-9.62	$L1$	15	1.85	-7.41	$L1_e$	0.88	0.06	-17.03	-9.62	-7.41	$L1_e$	15	1.85	0.88	0.06
101	1	-10.27	$L1$	15	0.55	-9.94	$L1_e^‡$	0.69	0.31	-20.21	-10.27	-9.94	$L1_{e‡‡}$	15	0.55	0.69	0.31
104	0	-16.63	$L1$	14	2.20*	-3.74	$L1_e$	0.00	0.94	-20.37	-16.63	-3.74	$L1_e$	14	2.20	0.00	0.94
413	0	-17.81	$L1$	14	0.88	-6.03	$L1_l$	0.13	0.88	-23.84	-17.81	-6.03	$L1_l$	14	0.88	0.13	0.88
207	0	-17.96	$L1$	14	0.42	0.00	$L1_e$	1.00	0.00	-17.96	-17.96	0.00	$L1_e$	14	0.42	1.00	0.00
216	1	-25.41	$L1$	13	1.06	-11.25	$L3_e$	0.75	0.19	-38.69	-25.41	-13.29	$L1_e$	13	1.06	0.31	0.63
402	0	-30.93	$L1$	12	5.65*	-9.00	$L1_e$	0.00	0.75	-39.93	-30.93	-9.00	$L1_e$	12	5.65	0.00	0.75
418	0	-42.23	$L1$	10	21.22**	-7.41	$L2_e$	0.88	0.06	-52.16	-42.23	-9.94	$L1_e$	10	21.22	0.00	0.69
301	1	-45.84	$L1^D$	10	0.00	-3.74	$L1_e$	0.06	0.94	-49.58	-45.84	-3.74	$L1_e^D$	10	0.00	0.06	0.94
508	0	-46.19	$L1^D$	10	2.05	-	-	-	-	-	-	-	-	-	-	-	-
308	3	-47.34	$L1$	10	0.00	-9.63	$L3_e$	0.81	0.13	-60.65	-47.34	-13.30	$L1_{el}$	10	0.00	0.19	0.69
102	4	-47.63	$L1$	10	0.00	-9.63	$L2_e$	0.81	0.06	-57.57	-47.63	-9.94	$L1_e$	10	0.00	0.00	0.69
415	1	-53.64	$L1$	9	0.88	-16.38	DI_e	0.31	0.50	-107.28	-90.90	-16.38	DI_e^+	2	0.76	0.31	0.50
504	1	-56.97	$L1$	8	1.68**	-	-	-	-	-	-	-	-	-	-	-	-
208	6	-61.62	$L1$	8	0.00	-3.74	$L1_l$	0.06	0.94	-65.37	-61.62	-3.74	$L1_l$	8	0.00	0.06	0.94
318	0	-62.61	$L1$	7	3.18*	-3.74	$L1_e^‡$	0.00	0.94	-66.36	-62.61	-3.74	$L1_e$	7	3.18	0.00	0.94
512	0	-63.33	$L1$	7	1.56	-	-	-	-	-	-	-	-	-	-	-	-
502	1	-64.55	$L1$	7	1.01	-	-	-	-	-	-	-	-	-	-	-	-
516	1	-64.93	$L1^C$	7	1.10*	-	-	-	-	-	-	-	-	-	-	-	-
409	0	-73.59	$L1^E$	4	9.90**	-10.59	$L1_l$	0.00	0.38	-84.18	-73.59	-10.59	$L1_l^E$	4	9.90	0.00	0.38
106	0	-75.82	$L1$	5	1.19*	-7.72	Eq_e	0.00	0.19	-85.75	-75.82	-9.94	$L1_l$	5	1.19	0.00	0.31
305	3	-79.89	$L1$	5	0.37	-6.03	$L1_e$	0.88	0.13	-85.92	-79.89	-6.03	$L1_e$	5	0.37	0.88	0.13
411	1	-80.58	$L1$	4	1.45**	0.00	$L3_e$	1.00	0.00	-86.61	-80.58	-6.03	$L1_e$	4	1.45	0.13	0.88
509	1	-81.81	$L1$	4	0.86	-	-	-	-	-	-	-	-	-	-	-	-
203	4	-83.90	$L1$	4	0.00	-9.94	Eq_e	0.00	0.31	-94.49	-83.90	-10.59	$L1_e$	4	0.00	0.00	0.63
505	4	-84.13	$L1$	4	0.43	-	-	-	-	-	-	-	-	-	-	-	-
317	3	-86.58	$L1$	3	0.92*	-3.74	$L1_e$	0.94	0.06	-90.32	-86.58	-3.74	$L1_e$	3	0.92	0.94	0.06
416	1	-86.74	$L1^†$	1	4.48**	-3.74	$L1_e^‡$	0.00	0.94	-90.48	-86.74	-3.74	$L1_e$	1	4.48	0.00	0.94
217	3	-87.12	$L1$	3	0.68	-10.59	$L1_e$	0.00	0.38	-97.71	-87.12	-10.59	$L1_e$	3	0.68	0.00	0.38

219	3	-87.32	$L1^+$	3	0.89*	-7.72	$L1_e$	0.00	0.81	-95.04	-87.32	-7.72	$L1_e^+$	3	0.89	0.00	0.81
501	1	-87.93	$L1^\dagger$	0	4.38**	-	-	-	-	-	-	-	-	-	-	-	-
410	3	-89.18	$\mathbf{L1}$	2	1.53**	-7.72	$L1_{el}^\ddagger$	0.00	0.19	-96.90	-89.18	-7.72	$\mathbf{L1_{el}}$	2	1.53	0.00	0.19
510	5	-89.60	$\mathbf{L1}$	3	0.00	-	-	-	-	-	-	-	-	-	-	-	-
420	2	-89.68	$L1^+$	2	1.25**	-3.74	Eq_l	0.00	0.06	-94.26	-90.52	-3.74	Eq_l^+	3	0.19	0.00	0.06
408	2	-89.71	$L1^+$	2	1.09*	-6.03	$L1_e$	0.00	0.88	-95.74	-89.71	-6.03	$L1_e^+$	2	1.09	0.00	0.88
201	3	-90.26	$L1^+$	2	1.21**	-3.74	$L1_e^\ddagger$	0.00	0.94	-94.00	-90.26	-3.74	$L1_e^+$	2	1.21	0.00	0.94
105	2	-90.58	$L1^+$	2	1.29**	-9.00	Eq_e	0.25	0.75	-102.56	-93.56	-9.00	Eq_e^+	2	0.11	0.25	0.75
103	3	-90.61	$L1^+$	2	1.12*	-6.03	$L1_e$	0.00	0.13	-96.63	-90.61	-6.03	$L1_e^+$	2	1.12	0.00	0.13
213	2	-95.57	$L1^{\dagger+}$	0	1.19*	-3.74	$L2_e$	0.94	0.00	-100.34	-96.60	-3.74	$L2_e^+$	0	0.62	0.94	0.00
515	4	-95.68	$L1^{\dagger+}$	1	0.60	-	-	-	-	-	-	-	-	-	-	-	-
113	5	-96.61	$L1^{\dagger+}$	1	0.07	-9.63	$L3_e^\ddagger$	0.81	0.06	-108.49	-98.86	-9.63	$L3_{el}^+$	4	0	0.81	0.06
109	8	-97.31	$L1^{\dagger+}$	1	0.00	-	-	-	-	-	-	-	-	-	-	-	-
309	0	0.00	$\mathbf{L2}$	16	-	-9.94	$L2_{el}^\ddagger$	0.69	0.00	-9.94	0.00	-9.94	$\mathbf{L2_{el}}$	16	0.00	0.69	0.00
405	0	0.00	$\mathbf{L2}$	16	-	-13.30	$L3_e$	0.69	0.13	-14.40	0.00	-14.40	$\mathbf{L2_e}$	16	0.00	0.63	0.25
206	0	-10.07	$\mathbf{L2}$	15	0.79	-7.41	$L2_e$	0.88	0.06	-17.49	-10.07	-7.41	$\mathbf{L2_e}$	15	0.79	0.88	0.06
209	0	-25.51	$\mathbf{L2}$	13	0.96	-9.00	$L1_e$	0.00	0.75	-35.45	-25.51	-9.94	$\mathbf{L2_l}$	13	0.96	0.69	0.31
108	0	-25.88	$\mathbf{L2}$	13	0.45	0.00	$L2_e^\ddagger$	1.00	0.00	-25.88	-25.88	0.00	$\mathbf{L2_e}$	13	0.45	1.00	0.00
214	2	-35.30	$\mathbf{L2}$	11	2.73**	-3.74	$L1_e$	0.00	0.94	-41.33	-35.30	-6.03	$\mathbf{L2_e}$	11	2.73	0.88	0.13
307	1	-38.88	$\mathbf{L2}$	11	1.04*	-7.72	Eq_e	0.00	0.19	-48.51	-38.88	-9.63	$\mathbf{L2_l}$	11	1.04	0.81	0.13
218	0	-40.54	$\mathbf{L2}$	11	0.60	-7.72	$L1_e$	0.00	0.81	-53.84	-40.54	-13.30	$\mathbf{L2_l}$	11	0.60	0.69	0.19
422	2	-55.79	$\mathbf{L2}$	9	0.22	0.00	$L1_e$	0.00	1.00	-61.82	-55.79	-6.03	$\mathbf{L2_e}$	9	0.22	0.88	0.13
316	1	-58.43	$\mathbf{L2}$	8	0.73	-10.97	$Eq_e^\ddagger$	0.00	0.44	-72.26	-58.43	-13.84	$\mathbf{L2_l}$	8	0.73	0.06	0.38
407	0	-60.98	$L2^C$	8	0.44	-6.03	$L2_e^\ddagger$	0.88	0.13	-67.00	-60.98	-6.03	$L2^C$	8	0.44	0.88	0.13
306	2	-68.48	$\mathbf{L2}$	7	0.18	-3.74	$L1_l$	0.00	0.06	-75.68	-71.94	-3.74	$\mathbf{L1_l}$	6	0.71	0.00	0.06
412	0	-69.43	$\mathbf{L2}$	6	1.05**	0.00	$L2_e^\ddagger$	1.00	0.00	-69.43	-69.43	0.00	$\mathbf{L2_e}$	6	1.05	1.00	0.00
205	0	-72.81	$\mathbf{L2}$	6	0.01	0.00	$L1_e$	0.00	1.00	-75.80	-75.80	0.00	$\mathbf{L1_e}$	4	3.27	0.00	1.00
220	1	-72.96	$\mathbf{L2}$	6	0.32	0.00	$L1_e$	0.00	1.00	-76.70	-72.96	-3.74	$\mathbf{L2_e}$	6	0.32	0.94	0.06
403	0	-73.60	$\mathbf{L2}$	6	0.50	-6.03	$Eq_l^\ddagger$	0.00	0.13	-86.91	-80.88	-6.03	Eq_l^+	4	0.84	0.00	0.13
517	0	-73.70	$\mathbf{L2}$	5	0.98**	-	-	-	-	-	-	-	-	-	-	-	-
503	3	-88.21	$L2^+$	3	0.00	-	-	-	-	-	-	-	-	-	-	-	-
414	4	-89.00	$\mathbf{L2}$	2	0.78*	-7.72	$L1_e$	0.00	0.19	-102.56	-92.62	-9.94	Eq_e^+	2	0.36	0.00	0.31
110	3	-92.51	$L2^+$	2	0.00	-9.00	$L1_l$	0.00	0.75	-107.03	-98.03	-9.00	$L1_l^+$	0	0.56	0.00	0.75

210	0	-51.13	$L3^B$	9	0.92*	-10.59	$L1_e$	0.00	0.38	-68.44	-51.13	-17.32	$L3_e^B$	9	0.92	0.38	0.25
302	0	-61.46	$L3^B$	7	1.11**	-6.03	Eq_e	0.00	0.13	-71.14	-65.12	-6.03	Eq_e^B	7	1.11	0.00	0.13
507	0	-63.23	***L3***	7	0.94**	-	-	-	-	-	-	-	-	-	-	-	-
313	0	-79.12	$D1^E$	3	2.68**	-6.03	$L1_{e‡}$	0.00	0.88	-90.93	-84.90	-6.03	$L1_{e‡‡}^E$	2	3.28	0.00	0.88
312	0	-80.45	$D1^†$	1	5.85**	-3.74	$L2_{e‡}$	0.94	0.06	-84.74	-81.00	-3.74	***$L2_e$***	3	1.37	0.94	0.06
204	2	-84.86	$D1^E$	3	1.22**	0.00	$L1_{e‡}$	0.00	1.00	-88.47	-88.47	0.00	$L1_e^{+E}$	2	1.59	0.00	1.00
115	1	-86.10	***D1***	2	1.74**	-9.94	Eq_e	0.00	0.31	-107.99	-98.05	-9.94	Eq_e^+	0	0.39	0.00	0.31
401	2	-91.99	$D1^{†+}$	0	1.58**	-6.03	Eq_l	0.00	0.13	-104.35	-98.32	-6.03	Eq_l^+	0	0.32	0.00	0.13
310	0	-41.69	***Eq***	11	0.00	-9.94	$L1_l$	0.00	0.31	-56.84	-41.69	-15.15	***Eq_{el}***	11	0.00	0.13	0.31
315	0	-41.80	***Eq***	11	0.00	0.00	$L3_{e‡}$	1.00	0.00	-50.80	-41.80	-9.00	***Eq_e***	11	0.00	0.00	0.75
404	1	-54.69	***Eq***	9	0.03	-9.00	$Eq_{e‡}$	0.00	0.75	-63.69	-54.69	-9.00	***Eq_e***	9	0.03	0.00	0.75
303	0	-59.93	***Eq***	8	0.41	-3.74	$Eq_{e‡}$	0.00	0.06	-63.68	-59.93	-3.74	***Eq_e***	8	0.41	0.00	0.06
417	0	-60.52	Eq^A	8	0.30	-10.97	$L1_e$	0.00	0.44	-73.80	-60.52	-13.29	Eq_e^A	8	0.30	0.31	0.63
202	0	-60.78	Eq^A	8	0.10	-9.94	Eq_e	0.00	0.31	-70.72	-60.78	-9.94	Eq_e^A	8	0.10	0.00	0.31
518	0	-66.38	***Eq***	7	0.61	-	-	-	-	-	-	-	-	-	-	-	-
112	2	-66.39	***Eq***	7	0.00	-16.64	$L2_e$	0.25	0.25	-106.23	-89.60	-16.64	$L2_e^+$	3	0	0.25	0.25
215	0	-73.85	***Eq***	6	0.55	-3.74	$L1_e$	0.00	0.06	-81.57	-73.85	-7.72	***Eq_e***	6	0.55	0.00	0.19
314	5	-78.06	***Eq***	5	0.52	-9.94	Eq_e	0.00	0.69	-87.99	-78.06	-9.94	***Eq_e***	5	0.52	0.00	0.69
211	3	-79.14	***Eq***	5	0.00	-7.72	Eq_e	0.00	0.19	-86.86	-79.14	-7.72	***Eq_e***	5	0.00	0.00	0.19
514	8	-85.98	***Eq***	4	0.00	-	-	-	-	-	-	-	-	-	-	-	-
406	2	-86.73	***Eq***	3	0.59	-6.03	$L1_l$	0.00	0.13	-99.17	-86.73	-12.44	***Eq_l***	3	0.59	0.06	0.25
212	5	-96.62	$Eq^{†+}$	1	0.00	-6.03	$L1_e$	0.00	0.88	-104.34	-96.62	-7.72	Eq_e^+	1	0.00	0.00	0.81
506	0	-82.10	***So***	3	1.26**	-	-	-	-	-	-	-	-	-	-	-	-
304	5	-93.29	So^+	2	0.25	0.00	Eq_e	0.00	1.00	-97.31	-97.31	0.00	Eq_e^+	1	0	0.00	1.00
421	4	-96.78	$So^†$	1	0.31	-10.59	Eq_e	0.00	0.38	-109.34	-98.38	-10.97	$L1_e^+$	0	0.43	0.00	0.56

Notes: A guesses-only or guesses-and-search type identifier superscripted † means the subject's estimated type was not significantly better than a random model of guesses ($\lambda = 0$, $\varepsilon \approx 1$) at the 5% (or 1%) level. A guesses-only or guesses-and-search type identifier superscripted + means the estimated type had lower likelihood than 12 or more pseudotypes, more than expected at random. A guesses-only or guesses-and-search type identifier superscripted A, B, C, D, or E indicates membership in a cluster. A guesses-only or guesses-and-search type identifier in bold appears or is reliable by the criteria stated in the text. An estimated λ superscripted ** (*) means that $\lambda = 0$ is rejected at the 1% (5%) level. ln L_t, ln L_g, and ln L_s refer to total, guesses-only, and search-only likelihoods. A type-style identifier subscripted *el* indicates that both styles have equal likelihoods and ζ_c. A search-only type-style identifier subscripted ‡ indicates that there are alternatives with different types and/or ζ_c: $L1_l$ for subjects 101 and 404; $L2_e$ and $L3_e$ for 318 and 204, $L3_e$ for 416 and 201; $L2_l$ for 113; $L1_e$ and $L3_{el}$ for 309; $L1_e$ and $L3_e$ for 108; $L1_e$ for 316, 407, 403, and 315; $L1_e$, $L3_e$, and Eq_e for 412 and 312; $L1_l$, $D2_e$, and So_e for 313; and $D1_e$ for 303. A guesses-and-search type-style identifier subscripted ‡‡ indicates that there are alternatives with different ζ_c: $L1_l$ for subjects 101 and 313. No search estimates are reported for subject 109, who had 0 search compliance in 8 or more games for every type.