COGNITION AND BEHAVIOR IN TWO-PERSON GUESSING GAMES: AN EXPERIMENTAL STUDY[1]

By Miguel A. Costa-Gomes, University of York,

and Vincent P. Crawford, University of California, San Diego

6 April 2004; revised 26 October 2004

". . . 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, game by game, through an interface that records their information searches. Varying the parameters allows strong separation of the behavior implied by leading decision rules and makes monitoring search a powerful tool for studying cognition. Many subjects' decisions and searches show clearly that they understood the games and sought to maximize payoffs, but had boundedly rational models of others' decisions, which led to systematic deviations from equilibrium.

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

---

[1]We thank Bruno Broseta, Colin Camerer, Yan Chen, Graham Elliott, Jerry Hausman, Nagore Iriberri, Eric Johnson, Rosemarie Nagel, Matthew Rabin, Tatsuyoshi Saijo, Jason Shachat, Joel Sobel, and two referees for helpful discussions and comments; Herbert Newhouse, Steven Scroggin, and Yang Li for research assistance; the U.K. Economic & Social Research Council (Costa-Gomes) and the U.S. National Science Foundation (Crawford and Costa-Gomes) for financial support; and the California Institute of Technology and the Institute for Social and Economic Research, Osaka University (Costa-Gomes) and the University of Canterbury, New Zealand (Crawford) for their hospitality. Our experiments were run in the University of California, San Diego's Economics Experimental and Computational Laboratory (EEXCL), with technical assistance from lab administrators Kevin Sheppard and Maximilian Auffhammer; and in the University of York's Centre for Experimental Economics (EXEC). One pilot session was run at Hong Kong University of Science and Technology with the help of Benjamin Hak-Fung Chiao. Appendices A-G and Figures 2A-2P are posted as pdf files at http://weber.ucsd.edu/~vcrawfor/#Guess.

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

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 Nagel's (1995) and Ho, Camerer, and Weigelt's (1998; "HCW") "guessing" or "beauty contest" experiments, inspired by Keynes' famous 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, 100] in Nagel, [0, 100] or [100, 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, in each of which the targets and limits were identical for all players and games. The structures were publicly announced, to justify comparing subjects' behavior 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, 200]) or infinite (limits [0, 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.

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"; McKelvey and Palfrey (1995)). But their data do suggest that subjects' deviations from equilibrium have a coherent structure. In Nagel's [0,100] 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 Stahl and Wilson (1994, 1995; "SW") and Costa-Gomes, Crawford, and Broseta (1998, 2001; "CGCB"); in other kinds of normal-form games (Camerer (2003, Chapter 5); Camerer, Ho, and Chong (2004; "CHC"); Crawford (1997, Section 4)); and in extensive-form bargaining games by Camerer, Johnson, Rymon, and Sen (1993, 2002; "CJ"). As in the guessing games, subjects make undominated decisions with high frequencies; but they rely less often on dominance for others (Beard and Beil (1994)), and reliance on iterated dominance seldom goes beyond three rounds. Subjects also make equilibrium decisions less often in games where identifying them requires more rounds of iterated dominance or the fixed-point logic of equilibrium (CGCB, Table II).

These papers modeled subjects' heterogeneous responses by assuming that each subject's decisions follow one of several boundedly rational strategic decision rules called *types*. 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; and *D2*, which does two rounds of iterated deletion and best responds to a uniform prior over the remaining decisions. Like an *Equilibrium* type that makes its equilibrium decision, *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 selfinterestedness and risk-neutrality. Thus *Lk*'s or *Dk*'s only essential departure from *Equilibrium* is in replacing its perfect model of others' decisions with a simplified, boundedly rational model.[2]

Although *Dk* types are closer to how theorists analyze games, *Lk* types dominate applications and are usually taken as the natural specification of boundedly rational strategic decision rules.[3] But the evidence does not yet justify this degree of confidence, even though *Lk* types have the largest estimated frequencies in most data analyses. In the above experiments *Lk* types are weakly separated from plausible alternatives. Nagel's and HCW's games with $p < 1$ and limits [0, 100] are an extreme example, where *Dk* and *Lk+1* guesses are perfectly confounded—both tracking the spikes at $50p^k$.[4] Further, the data analyses rest on a priori specifications of small numbers of possible types. This may in fact be a necessary evil in this kind of analysis; but it also entails a risk of specification bias that may have gone undetected in the tests that were used. As a result, the analyses cannot confidently rule out the possibility that the high estimated frequencies of *L1* and *L2* are proxies for (say) altruistic, spiteful, risk-averse, or confused *Dk* or *Equilibrium* subjects; or for other, entirely different types inadvertently omitted from the specification.[5]

For these and other reasons, explained below, the structure of initial responses to games has not been identified as precisely or documented as convincingly as current methods allow. To move closer to that goal, this paper reports an experiment that elicits subjects' responses to a series of 16 dominance-solvable two-person guessing games. The design suppresses learning and repeated-game effects to justify an analysis of their behavior as initial responses to each game.

---

[2]*Lk* anchors beliefs in a uniform prior and adjusts them via thought-experiments with iterated best responses, without "closing the loop" as for equilibrium. *Dk* avoids closing the loop by invoking a uniform prior after finitely iterated deletion of dominated decisions. In Selten's (1998) words: "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."

[3]Camerer (2003); CHC; Crawford (2003); Crawford and Iriberri (2004); and Kübler and Weizsäcker (2004) give example applications. Keynes' wording in our epigraph connotes 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)). Nagel (1995) focuses on *Lk* types, citing subjects' questionnaire responses (1993, pp. 14-15). Her subjects were University of Bonn students with little knowledge of game theory, but our reading of her questionnaires from a London School of Economics pilot, whose subjects were probably more sophisticated, provides support for *Dk* as well as *Lk* types.

[4]*Dk*'s guess is $([0+100p^k]/2)p \equiv 50p^{k+1} \equiv [(0+100)/2]p^{k+1} \equiv$ *Lk+1*'s guess. Further, both *Dk* and *Lk+1* can explain the empirical relationship between equilibrium compliance and complexity: *Dk* respects k+1 rounds of dominance, and *Lk+1* respects k+1 rounds in many games, so a suitable mixture of either kind of type mimics equilibrium in games that are dominance-solvable in small numbers of rounds, but deviates systematically in some more complex games.

[5]SW (1994), for example, 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 design differs from Nagel's and HCW's in several ways. Our guessing games have only two players, who make simultaneous guesses. Each player has a lower and an upper limit ([100, 500], [100, 900], [300, 500], or [300, 900]). Each player also has a target (0.5, 0.7, 1.3, or 1.5), and his payoff is higher, the closer his guess is to his target times his partner's guess.[6] Within this common structure, which is publicly announced, the targets and limits vary independently across players and games, with the targets sometimes both less than one, sometimes both greater than one, and sometimes mixed. 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. The targets and limits are hidden, but subjects are allowed to search for them, game by game, through a computer interface.[7] Low search costs then make the structure effectively public knowledge.

Studying two-person games allows us to focus sharply on the central game-theoretic problem of predicting the decisions of other players who view themselves as a non-negligible part of one's own environment.[8] Tracking behavior within subjects across 16 different games with large strategy spaces greatly enhances separation of types' implications (Table III). Varying the targets and limits within an intuitive structure makes it easier for subjects to understand the rules, so that they can focus on predicting others' guesses. It also makes it impossible for subjects to recall the current targets and limits from previous games, and so makes monitoring their searches for hidden information about them a powerful tool for studying cognition more directly.

In our design, a subject's sequence of guesses often yields a clear strategic "fingerprint," so that his type can be read directly from his guesses. 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 7-16 of the games (20 *L1*, 12 *L2*, 3 *L3*, and 8 *Equilibrium*; Table IX). These compliance levels are far higher

---

[6]Thus 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. This eliminates his need to predict how his guess affects an average. Like Nagel's and HCW's games, ours limit the effects of altruism, spite, and risk aversion.
[7]Subjects were not allowed to write, and the search data suggest that there was very little memorization. The interface, 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 it to study matrix games in which subjects could look up their own and their partners' payoffs.
[8]Grosskopf and Nagel (2001) report experiments with two-person guessing games in which subjects were rewarded for guessing closer to a target times the pair's average guess. With targets less than one, guessing the lower limit is a weakly dominant strategy, so their games do not fully address the issue of predicting others' decisions.

than could plausibly occur by chance, given how strongly types' guesses are separated (Tables III-IV) and that guesses could take from 200 to 800 different rounded values in each game.

Because our types specify precise guess sequences in a very large space, these subjects' guesses rule out almost any alternative interpretations. In particular, because the types "build in" risk-neutral, self-interested rationality and perfect models of the game, the guesses of the 35 of the 43 subjects who conform closely to non-*Equilibrium* types can be confidently attributed to non-equilibrium beliefs, rather than irrationality, risk aversion, altruism, spite, or confusion.[9]

Our other 45 subjects' fingerprints are less clear. But for all but 12 of them, violations of simple dominance were comparatively rare (less than 20%, versus 40% for random guesses). This suggests 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, subject by subject. Our econometric framework follows CGCB's in most respects.[10] We assume that each subject's behavior is determined, with error, by a single type, which determines his guesses and searches in all 16 games. Our types include *L1*, *L2*, *L3*, *D1*, *D2*, and *Equilibrium* as defined above. To test whether any subject has a prior understanding of others' decisions that transcends these simple decision rules, we add CGCB's *Sophisticated* type, which represents the ideal of a rational person who can predict the distributions of other subjects' initial responses.

Maximum likelihood type estimates based on guesses reaffirm our type identifications for the 43 subjects whose fingerprints are clear, and assign several additional subjects each to *L1*, *L2*, and *Equilibrium*, plus a few to *D1* and *Sophisticated* (Table IX). To evaluate the reliability of these estimates, we use a new specification test that compares the likelihood of our estimated type, subject by subject, with the likelihoods of estimates based on 88 *pseudotypes*, each constructed from one of our subject's guesses in the 16 games. Such comparisons can detect whether any subjects' guesses would be better explained by an alternative decision rule omitted from our specification, and sometimes help to identify omitted rules. They can also detect when a subject's estimated type is an artifact of accidental correlations with irrelevant included types.

Our specification analysis reaffirms a large majority of our econometric identifications of *L1*, *L2*, or *Equilibrium* subjects, but calls into question some of each plus all but one each of our identifications of *L3*, *D1*, or *Sophisticated* subjects. It indicates that the questionable *L3* subjects

[9]Compare Weibull's (2004) argument that rejections of equilibrium in experiments that do not independently measure preferences are "usually premature".
[10]CGCB's framework builds on Holt (1999), SW, Harless and Camerer (1995), Nagel (1995), and Stahl (1996).

and some of the questionable *Equilibrium* subjects may instead be complex hybrids of *L3* and/or *Equilibrium*. It also supports our a priori specification of possible types by giving no indication of significant numbers of SW's *Worldly* type or of any other type omitted from our specification.

Information search adds another dimension to our econometric analysis.[11] 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. 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.

Under our assumptions, our design separates types' search implications much more strongly than previous designs, while making them almost independent of the game. This sometimes allows a subject's type to be read directly from his searches, without even considering guesses (Appendix I); but most subjects' searches less clearly identify their type. We therefore generalize our error-rate analysis to re-estimate subjects' types using their searches as well as their guesses.

Taking both search and guesses into account, 54 of our 88 subjects are reliably identified as one of our types, 44 of them non-*Equilibrium* (Table IX). The full analysis reaffirms the absence of significant numbers of subjects of types other than *L1*, *L2*, *Equilibrium*, or hybrids of *L3* and/or *Equilibrium*. Given our definition of *Lk*, these results strongly affirm subjects' rationality and their ability to comprehend complex games and reason about others' responses to them, while challenging the use of equilibrium as the principal model of their initial responses. The surprisingly simple structure of the part of subjects' behavior that our analysis can explain is consistent with previous analyses, but significantly refines and sharpens them. Its simplicity should help to allay the common fear that with bounded rationality, "anything can happen".

The rest of the paper is organized as follows. Section 1 describes our experimental design. Section 2 derives types' implications for guesses and information search. Section 3 reports preliminary statistical tests and results, introduces our econometric model and uses it to estimate subjects' types, and discusses our specification analysis. Section 4 is the conclusion.

---

[11]A companion paper, Costa-Gomes and Crawford (2004), will analyze subjects' search behavior in more detail, studying the relations between cognition, search, and decisions and comparing the cognitive difficulty of types.

1. Experimental Design

To test theories of strategic behavior, an experimental design must clearly identify 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, and using the results to test theories of behavior in the stage game. Such designs allow subjects to learn the structure from experience, which reduces noise; but they also make it difficult to disentangle learning from cognition, because even unsophisticated learning often converges to equilibrium in the stage game. Our design studies 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, experimentation, and learning. This section describes the overall structure of our design, the games, and how they are presented to subjects.

A. Overall structure

All of our sessions were run either 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). In each case subjects were recruited from undergraduates and graduate students (Ph.D. or M.A.), with completely new subjects for each session. 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 we disqualified other subjects who revealed that they had formally studied game theory or previously participated in game experiments.[12]

Table I 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 fifteen Robot/Trained Subjects sessions, R/TS1-R/TS15, with a total of 148 subjects in mixed treatments: 37 UCSD subjects (7 *L1*, 9 *L2*, 11 *D1*, and 10 *Equilibrium*) and 111 York subjects (18 *L1*, 18 *L2*, 18 *L3*, 19 *D1*, 19 *D2*, and 19 *Equilibrium*).[13] All treatments used the same 16 games (Table II), which include eight player-symmetric pairs so that Baseline or OB subjects can be paired with other Baseline or OB subjects without dividing subjects into

[12]We allowed approximately four non-faculty university community members, plus a few students who had been briefly exposed to game theory in undergraduate courses.
[13]Appendix A gives the instructions and Appendix B describes our pilot experiments and how they influenced the design (http://weber.ucsd.edu/~vcrawfor/#Guess). Mixed R/TS treatments are theoretically acceptable because R/TS subjects did not interact. The data exclude one *L1* subject in R/TS1, because his guesses showed that he had copied from a nearby *L2* subject. (R/TS subjects were not told whether all subjects in a session were assigned the same type, but he assumed this.) Comparing all neighboring subjects' guesses suggests that he was the only cheater.

subgroups. The games include one perfectly symmetric pair, so that each subject plays one game twice, allowing a weak test of consistency of responses. All treatments presented the games in the same order, which was randomized ex ante, and which made their symmetries non-salient.

We first describe the Baseline treatment and then explain how other treatments differed. In the Baseline, after an instruction phase and Understanding Test, groups of 13-21 subjects were randomly paired to play the 16 games, with new partners each period.[14] Subjects received no feedback during the games. They could proceed independently at their own paces, but were not allowed to change their guesses once confirmed. These design features suppress learning and repeated-game effects, to justify an analysis of behavior as initial responses to each game.

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.[15] 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 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. The Baseline instructions avoided suggesting guesses or decision rules. During the session, subjects had free access, game by game, to their own and their partners' targets and limits via a MouseLab interface as described below.[16] Subjects were taught the mechanics of looking up targets and limits and entering guesses, but not information-search strategies. They were given ample opportunity for questions, and after the instructions they were required to pass an Understanding Test to continue. Subjects who failed were dismissed, and the remaining subjects were told that

[14]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.

[15]It is theoretically possible to control subjects' risk preferences using the *binary lottery* procedure, in which a subject's payoff determines his probability of winning a given monetary prize. We avoided the complexity of binary lotteries because risk preferences do not influence predictions based on iterated dominance or pure-strategy equilibrium, and results using direct payment are usually close to those using binary lotteries.

[16]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.

all subjects remaining had passed.[17] Before playing the 16 games, Baseline subjects were 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.[18] 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 how they decided which guesses to make.

The OB treatment is identical to the Baseline treatment except that the 16 games are presented with the targets and limits continually visible, in "open boxes." Its purpose is to learn whether subjects' guesses are affected by the need to look up the targets and limits. We find only insignificant differences between Baseline and OB subjects' guesses (Section 3.A), suggesting that subjects' decisions are not strongly affected by the need to look up payoffs.

The R/TS treatments are identical to the Baseline treatment, except each subject is trained and rewarded as a specific type: *L1*, *L2*, *L3*, *D1*, *D2*, or *Equilibrium*. In addition to standard instructions as in the Baseline, each R/TS subject was taught how to identify his assigned type's guesses via programmed instruction on his screen and handout.[19] He was rewarded for game payoffs as in the Baseline, except that he was paired not with other subjects but with a *robot* (framed as "the computer") that followed his type's model of others: guesses uniformly distributed between the partner's limits for *L1*, or on the set of (iteratively) undominated guesses for *D1* (*D2*); *L1* (*L2*) guesses for *L2* (*L3*); equilibrium guesses for *Equilibrium*.[20] The R/TS treatments also replace the Baseline's practice rounds, less relevant when subjects do not interact, with a second Understanding Test of how to identify the assigned type's guesses. Subjects were paid an extra $5 for passing this test, and those who failed were dismissed.[21] As in the Baseline, all aspects of this structure were publicly announced, except the games' targets and limits.

---

[17]The dismissal rates (including a few voluntary withdrawals) were 20% for Baseline subjects, 11% for OB subjects, and 20% for R/TS subjects of all types. Table VII gives dismissal rates for R/TS subjects by assigned type.

[18]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 games had a balanced mix of structures, with different targets and limits than in the 16 games to avoid implicitly suggesting guesses.

[19]*Equilibrium* subjects, for instance, were taught each of the three main ways to identify their equilibrium guesses: direct checking for pure-strategy equilibrium, best-response dynamics, and iterated dominance.

[20]We used realizations of random robot guesses rather than their means to minimize differences from the Baseline.

[21]The average total earnings figures for UCSD R/TS *L1*, *L2*, *D1*, and *Equilibrium* subjects who finished the experiment were $45.22, $62.03, $51.74, and $50.93. York R/TS subjects were paid early and on-time show-up fees of £1 and £2, plus £2.50 for passing the second Understanding Test, but only £0.02 rather than $0.04 per point. With the pound averaging $1.63 during the York sessions, those fees, which seemed adequate, were roughly 70% of the

The R/TS instructions differed from the Baseline in one further way. The predicted behavior of *Lk* or *Dk* depends on best responses to uniform beliefs on intervals. We expected most R/TS *Lk* or *Dk* subjects to treat such beliefs as if concentrated on their means, identifying best responses via certainty-equivalence. To eliminate variation across subjects that is unrelated to our goals, we designed our guessing games to have this certainty-equivalence property, without regard to players' risk preferences (Observation 2, Section 2.B). The R/TS instructions also encouraged *Lk* or *Dk* subjects to use certainty-equivalence to identify best responses.[22]

The main purposes of the R/TS treatments are to learn to what extent Baseline subjects' deviations from equilibrium are due to cognitive limitations; and to learn what the information searches of *Equilibrium* and other types would be like, as a check on the model of cognition and search we use to analyze Baseline subjects' behavior. The R/TS results generally validate our simple model of cognition and information search (Section 2). Most if not all R/TS *Equilibrium* subjects can reliably identify equilibrium guesses, but there are significant, sometimes surprising differences in the apparent cognitive difficulty of our types: *Lk* types appear to be far less difficult than *Equilibrium*, and *Equilibrium* may be less difficult than *Dk* types (Section 3.B).

B. Two-person guessing games

In our guessing games, two players, *i* and *j*, make simultaneous guesses, x^i and x^j. We use *i* for the generic player and *j* for "not *i*". 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; instead guesses outside the limits are automatically adjusted up to the lower limit or down to the upper limit. Thus, player *i*'s *adjusted guess*, $y^i \equiv R(a^i,b^i;x^i) \equiv x^i$ if $x^i \in [a^i,b^i]$, $y^i \equiv a^i$ if $x^i < a^i$, or $y^i \equiv b^i$ if $x^i > b^i$, or equivalently $y^i \equiv R(a^i,b^i;x^i) \equiv \min\{b^i, \max\{a^i,x^i\}\} \equiv \max\{a^i, \min\{b^i,x^i\}\}$. Each player *i* also has a *target*, p^i. Writing $e^i \equiv |R(a^i,b^i;x^i) - p^iR(a^j,b^j;x^j)|$ for the distance between player *i*'s adjusted guess and p^i times player *j*'s adjusted guess, player *i*'s point payoff, s^i, is given by

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

UCSD fees. York R/TS *L1*, *L2*, *L3*, *D1*, *D2*, and *Equilibrium* subjects' average total earnings figures were £23.00, £29.76, £28.50, £27.08, £24.12, and £27.65. The fee for passing the second Understanding Test raises R/TS subjects' average earnings, relative to Baseline and OB subjects, but R/TS *L1*, *D1*, and *D2* subjects' earnings were lower than other R/TS subjects', other things equal, because they faced uncertainty about their simulated partners' guesses.

[22]The encouragement is implicit in the wording, and does not use the term certainty-equivalence (Appendix A).

With or without adjustment, the point payoff function in (1) is quasiconcave in player *i*'s guess for any given distribution of player *j*'s guess; and without adjustment it is symmetric about $e^i = 0$.[23] The relationship between a player's guesses and his point payoff is not one-to-one, because all 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 that adjusted guess, describing a prediction as *essentially* unique if it implies a unique adjusted guess.

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 his *ideal guess*, the guess that would be optimal given his beliefs, ignoring his limits, and know without checking his own limits that his adjusted guess will be optimal. (Our instructions explain this, and most subjects' behaviors showed that they understood it.) In our design *L1*'s ideal guess depends only on its own target and its partner's limits, while *Equilibrium*'s depends on both players' targets and a combination of its own and its partner's lower or upper limits, and our other types' all depend on both players' targets and limits. Thus, by contrast with CGCB's and other designs, where *L1*'s decisions almost inevitably depend only on own payoff parameters, *L1*'s search implications are sharply separated both from our other types' implications and from those of a solipsistic type that assumes that only its own parameters are relevant. (We find a great deal of evidence of *L1*, but none of solipsism.)

Because our design suppresses learning and repeated-game effects and makes the structures of our guessing games effectively public knowledge, our data analysis will treat them as independent games of complete information. Players' 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 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.

[23]It 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 α4β1 (Table II), 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.

In deriving types' implications, we assume that each player maximizes the expected utility of his total money payment from the 16 games. Because his 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, such a player maximizes his point payoff in any given game. When guesses have uncertain consequences, risk preferences are potentially relevant.[24] But Observation 1 below shows that our games have essentially unique equilibria in pure strategies, so risk preferences do not affect *Equilibrium* guesses. And Observation 2 shows that best responses to uniform beliefs are certainty-equivalent, so risk preferences do not affect *L1*, *D1*, or *D2* guesses, or the best responses that define *L2* or *L3* guesses. This leaves *Sophisticated* guesses, which are normally best responses to non-uniform beliefs, and so are not covered by Observation 2. In characterizing them we assume players are risk-neutral, and thus maximize their expected point payoffs, game by game. Given this, each of our types maximizes its expected point payoff, game by game, given some beliefs; and each implies an essentially unique, pure guess in each game, except *Sophisticated*, for which this is generically true.

We now establish Observations 1 and 2. To avoid trivialities, we assume that all limits and targets are strictly positive, as in our design.

Observation 1: Unless $p^ip^j = 1$, each guessing game in the above class has an essentially unique equilibrium, in pure strategies. If $p^ip^j < 1$, in equilibrium $y^i \equiv R(a^i,b^i;x^i) = a^i$ and $y^j \equiv R(a^j,b^j;x^j) = \min\{p^ja^i,b^j\}$ if and only if ("iff") $p^ja^i \geq a^j$; and $y^i = \min\{p^ia^j,b^i\}$ and $y^j = a^j$ iff $p^ia^j \geq a^i$. If $p^ip^j > 1$, in equilibrium $y^i = b^i$ and $y^j = \max\{a^j,p^jb^i\}$ iff $p^jb^i \leq b^j$; and $y^i = \max\{a^i,p^ib^j\}$ and $y^j = b^j$ iff $p^ib^j \leq b^i$.

Observation 1 shows that unless $p^ip^j = 1$, which is never true in our design, each game in the class from which our guessing games are drawn has an essentially unique equilibrium, in pure strategies, determined (not always directly) by players' lower limits when the product of their targets is less than one, or their upper limits when the product is greater than one.[25] This is true without regard to risk preferences or dominance-solvability, although not all games in this class are dominance-solvable because there is no dominance for extreme parameter values. The

---

[24]Recall that our games do not have the binary lottery, winner-take-all structure of Nagel's and HCW's games.
[25]In game γ2β4 (Table II), for instance, the product of targets is 1.05, so equilibrium is determined by the upper limits. The γ2 player's equilibrium guess is at his upper limit, 500, but the β4 player's equilibrium guess is at 750, below his upper limit, 900. Moving some equilibrium guesses away from the boundaries allows clearer inferences.

proof is straightforward. If $p^ip^j < 1$, say, iterating best responses drives players' adjusted guesses down until one player's hits his lower limit and the other's is at or above his own lower limit.

The discontinuity of the equilibrium correspondence when $p^ip^j = 1$ sharply separates *Equilibrium* guesses from other types': Games such as δ2β3 and γ2β4 (Table II) differ mainly in whether p_ip_j is slightly below or above one; equilibrium responds strongly to this difference but boundedly rational rules, whose guesses vary continuously with the targets, all but ignore it.

Observation 2: Suppose a guessing game's point payoff function 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 an interval $[a,b]$. 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 of x is $x^* = p\mathrm{E}z = p(a+b)/2$, and his expected-utility maximizing choice of x s.t. $x \in [c,d]$ is $R(c,d; p(a+b)/2)$.

Proof: We show that $x^* = p(a+b)/2$ solves $\max_x \int_a^b u(s(x-pz))dz$ (ignoring the positive factor $[1/(b\text{-}a)]$). 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. □

Observation 2 shows that for a class of two-person guessing games including ours, for any player with a continuous, almost everywhere differentiable von Neumann-Morgenstern utility function that is self-interested and values only money, best responses to uniform beliefs on an interval like those in the definitions of types *L1*, *D1*, and *D2*, and, indirectly, of *L2* and *L3*, equal the player's target times the midpoint of the interval, adjusted if necessary to lie within his limits. This result is independent of risk preferences, but it depends on symmetry and uniform beliefs.

We chose our games' limits and targets to make the design as informative as possible, given the need for a balanced mix of parameter values and strategic structures with no obvious correlations across games or players. In each game, each player's lower and upper limits are either [100, 500], [100, 900], [300, 500], or [300, 900], and each player's target is 0.5, 0.7, 1.3, or 1.5. We identify a player's combination of lower and upper limits by a Greek letter: α for [100, 500]; β for [100, 900]; γ for [300, 500]; or δ for [300, 900]. We identify a player's target by a number: 1 for 0.5; 2 for 0.7; 3 for 1.3; or 4 for 1.5. A game is identified by a combination such as β1γ2, in which player *i* has limits β for 100, 900 and target 1 for 0.5, and player *j* has limits γ for 300, 500 and target 2 for 0.7. Recalling that our 16 games include eight player-symmetric pairs, game γ2β1 is the player-symmetric counterpart of β1γ2: β1γ2 from player *j*'s point of view.

Table II summarizes our 16 games, ordered to emphasize structural relationships. It also lists the common, randomized order in which subjects played the games; whether the targets are both < 1 (Low), both > 1 (High), or neither (Mixed); whether the equilibrium is determined by players' upper limits (High) or their lower limits (Low); the number of rounds of iterated dominance player *i* needs to identify his equilibrium guess; whether dominance is alternating (A), simultaneous (S), or simultaneous in the first round but then alternating (S/A); and whether dominance initially occurs at both of a player's limits (Yes) or not (No).[26]

Table III lists the adjusted guesses implied for player *i* by the types *L1*, *L2*, *L3, D1*, *D2*, *Equilibrium*, and *Sophisticated*; and the ranges of guesses that survive 1-4 rounds of iterated dominance. Table IV summarizes the separation of implied guesses across types, measured as the number of guesses that differ by more than 0, or 25. *L2* and *D1* are separated much more strongly than in previous experiments. More generally, separation by more than 0 averages two-thirds of the theoretical maximum for all six types (64/96) and 13/16 of the maximum excluding *D2* and *L3* (52/64), which is hard to improve upon within a simple overall structure like ours.

The games with high numbers of rounds of iterated dominance, which result from a product of targets near one and limits far apart, are particularly well suited to separating types' guesses. The structural variations summarized in Table II stress-test our type identifications by making types' predicted guesses more subtle. They also play a central role in our specification analysis (Section 3.E), where, together with our games' large strategy spaces and the discontinuity of the

[26]Here we distinguish the numbers of rounds a game's players need to identify their own iteratively undominated adjusted guesses; the number of rounds in which the game is dominance-solvable is the maximum of these.

equilibrium correspondence when $p^i p^j = 1$, they sometimes allow us to distinguish intentional behavior from "random" behavior or cognitive errors by "reverse-engineering" subjects' guesses.

We conclude this section by using the observed frequencies of Baseline and OB subjects' pooled guesses, which did not differ significantly (Section 3.A), to estimate the strength of their incentives to make types' guesses. Table V's rows give the expected monetary earnings in dollars over all 16 games of a subject who made a given type's guesses, as a function of a hypothetical type that determines the subject's partners' guesses. The *L0* column refers to a partners' type whose guesses are uniform random between its limits, as in *L1*'s beliefs. The strength of an *L1* subject's incentives to make *L1*'s guesses can be gauged by using the *L0* column to compare the expected earnings of *L1* guesses with those of other leading types. Similarly, the *L1* (*L2*) column reflects *L2*'s (*L3*'s) beliefs; the *R1* (*R2*) column refers to a type whose guesses are uniform random over guesses that survive 1 (2) rounds of iterated dominance, reflecting *D1*'s (*D2*'s) beliefs; the *Equilibrium* column reflects *Equilibrium*'s beliefs; and the B+OB column refers to Baseline and OB subjects' actual frequencies, reflecting *Sophisticated*'s estimated beliefs.

Using Table V to make the suggested comparisons shows that subjects whose beliefs correspond to types *Equilibrium*, *L2*, and *L3* have strong incentives to make their type's guesses. *Equilibrium*, for instance, would earn $46.05 against *Equilibrium*, $12.05 more than the next most profitable type in the table, *L3*, which would earn $34.00. Similar calculations show that *L2*'s and *L3*'s earnings would be $10.25 and $6.90 higher than the next most profitable type's. Our other leading types have comparatively weak incentives by this conservative measure: $1.29 for *D2*, $1.22 for *L1*, $0.85 for *D1*, and $0.46 for *Sophisticated*.[27]

C. Using MouseLab to present guessing games

The games were displayed on subjects' screens via MouseLab. 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 1 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.[28] A

[27]Among our types, only *L1* and *Equilibrium* are *not* fairly close substitutes for *Sophisticated*, given its beliefs.
[28]Subjects were lent calculators to facilitate the arithmetic needed to determine their guesses. It is possible

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.

Our design for the display reflects the fact that previous work has revealed a top-left bias in subjects' look-ups and a left-right bias in their transitions (CGCB). The effects of such biases can be transformed by reallocating parameters to boxes, but not eliminated. Our design seeks to minimize the ambiguity of interpretation such biases cause, by putting each player's parameters in a single row, putting Your parameters in the first row, and putting a player's targets between his limits. This makes looking up Her/His parameters, which is a hallmark of strategic thinking, and adjacent lower-and-upper-limit pairs that are characteristic of *L1*, *L2*, and other leading types less likely to occur for reasons unrelated to cognition.

2. Types' Implications for Guesses and Information Search

This section derives our types' implications for guesses and information search, seeking minimal restrictions to avoid imputing irrationality to subjects whose cognition we cannot directly observe. Recall that we take a procedural view of decision-making, in which a subject's type determines his search and guess, both with error. Under our assumptions, each of our types implies an essentially unique, pure adjusted guess in each game, which maximizes its expected payoff given beliefs based on some model of others' decisions. The leading role in the derivations is played by a type's *ideal guesses*, those that would be optimal given the type's beliefs, ignoring its limits. A type's ideal guess completely determines its adjusted guess in a game, and the resulting outcome, via the adjustment function $R(a^i,b^i;x^i) \equiv \min\{b^i, \max\{a^i,x^i\}\}$. A type's ideal guess also determines its *minimal* search implications, because a subject can enter his ideal guess and know that his adjusted guess will be optimal without checking his own limits.

Observation 1 for *Equilibrium* and Observation 2 for *L1*, *L2*, *L3*, *D1*, and *D2* immediately yield expressions for those types' ideal guesses as functions of the game's targets and limits. We estimate *Sophisticated*'s ideal guesses as risk-neutral best responses to the pooled distribution of Baseline and OB subjects' adjusted guesses (which did not differ significantly), game by game,

for subjects to record two parameters at a time in the memory and on the display of their calculators; but that is much less convenient than using the interface, and no subject appeared to use the calculator this way.

rounded to the nearest integer for simplicity. Because we also rounded subjects' guesses to the nearest integer, and few made exact *Sophisticated* guesses, this does not lead to misclassification.

Types' search implications are derived as follows. Under standard assumptions, an expected-payoff maximizing player looks up all costlessly available information that can affect his beliefs. We therefore require that if a type's guess depends on a parameter, that parameter must appear at least once in the type's look-up sequence. This is uncontroversial, but of limited use because most subjects satisfy it by chance for most types in most games. We supplement it by restricting the order of look-ups. Recall that each type is naturally associated with algorithms that process payoff information into guesses. These require series of arithmetic *operations* on parameters; we call operations that logically precede any other operation *basic*.

Subjects' searches in our pilots, our R/TS treatments, and CJ's and CGCB's experiments suggest that most subjects perform operations one at a time via adjacent look-ups, starting with basic operations, remembering their results, and otherwise relying on repeated look-ups rather than memory. We stylize these regularities by requiring that in each game, the basic operations needed to identify a type's ideal guess are represented at least once in the look-up sequence by adjacent look-ups, in any order, and that other operations are represented at least once by the associated look-ups, in any order, but possibly separated by other look-ups. These assumptions adapt CGCB's (Section 3.C) Occurrence and Adjacency assumptions in ways appropriate to current design. We stress that their motivation is empirical: In theory, a subject could scan the parameters in any order and rely on memory to perform his type's operations, making the order of look-ups useless in inferring cognition; but real subjects rarely do that. We call the look-ups that satisfy these search requirements for a given type the type's *relevant* look-ups.

Table VI lists the expressions for our types' ideal guesses and the associated relevant look-ups, in our notation for limits and targets and in terms of the associated box numbers (Figure 1: 1 for a^i, 2 for p^i, 3 for b^i, 4 for a^j, 5 for p^j, 6 for b^j) in which MouseLab records subjects' look-up sequences in our design. Appendix H gives more detailed derivations. 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. The look-ups associated with each type's operations are shown in the order that seems most natural to us, if there is one.

3. Statistical and Econometric Analysis of Subjects' Guesses and Information Searches

This section presents a statistical and econometric analysis of subjects' guesses and information searches.[29] Section 3.A reports preliminary statistical tests. Section 3.B summarizes the aggregate compliance of R/TS subjects' adjusted guesses with the implications of their assigned types, and Section 3.C summarizes the aggregate compliance of Baseline and OB subjects' adjusted guesses with iterated dominance and equilibrium. Section 3.D presents a maximum likelihood error-rate analysis of Baseline and OB subjects' guesses. Section 3.E discusses our specification test and analysis. Section 3.F generalizes Section 3.D's error-rate analysis to use Baseline subjects' searches, along with their guesses, to estimate their types.

A. Preliminary statistical tests

In this section we report tests for differences in subjects' adjusted guesses across the OB treatment and the four sessions of the Baseline treatment. Because the tests compare data from independent samples with no presumption about how they differ, we use exact two-sample Kolmogorov-Smirnoff tests, pairing the five Baseline and OB sessions in all possible ways and, for each pair, conducting the tests separately for each game. This yields 11 *p*-values less than 5% in a total of 160 tests (five sessions taken two at a time, times 16 games per session), a bit more than one would expect by chance (11/160 = 6.9%) but distributed evenly across sessions and games. Similarly, comparing the four Baseline sessions pooled with the OB session yields one *p*-value less than 5% in 16 tests.[30] This suggests that subjects' guesses are not strongly affected by the need to look up payoff parameters, so our results should be representative of those obtained by standard methods. We conclude that differences across Baseline sessions or between Baseline and OB treatments are small enough to justify pooling the data on guesses across sessions.[31]

The tests also reveal no significant difference between Baseline and OB subjects' pooled guesses in the symmetric game, δ3δ3, when played third and twelfth in the sequence (see also

[29]Appendix C gives the complete data on guesses and the order, but not duration, data on look-up sequences. Figures 2A-2P (http://weber.ucsd.edu/~vcrawfor/#Guess) graph game-by-game frequency distributions of adjusted guesses.
[30]Conducting the tests this way would be justified only if subjects' guesses were independent across games and session pairs, which is unlikely in the first case and impossible in the second; but correcting for the dependence is impractical. These tests are presented only as a way to gauge the differences across sessions and treatments. We also found no significant evidence that subjects' guesses in practice rounds differed across the Baseline and OB sessions.
[31]Nonetheless, there are 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.

Figures 2G-2H). This suggests that the effects of introspective learning without feedback are small enough to justify analyzing the data without considering the order of play.

B. R/TS subjects' compliance with assigned types' guesses

Table VII summarizes the aggregate exact compliance (within 0.5) rates of R/TS subjects' adjusted guesses with assigned types' guesses, along with the failure rates in the R/TS treatments' second, type-specific Understanding Test. Overall, compliance is highest for *Lk* types, next highest for *Equilibrium*, and lowest for *Dk* types. Among *Lk* or *Dk* subjects, compliance falls with *k* as one would expect, with the exception that compliance is lower for *L1* than for *L2* and *L3*.[32] These aggregate results mask considerable individual heterogeneity. Many R/TS subjects implement their assigned type's guesses perfectly or almost perfectly, while others do no better than random. (More detailed results will be reported in Costa-Gomes and Crawford (2004).)

The Understanding Test failure rates tell a similar story about the relative cognitive difficulties of our types, except that *Equilibrium* failure rates are much higher than *D1* and *D2* failure rates. This may be due to the greater stringency of our *Equilibrium* Understanding Test, which tests comprehension of the three different ways to identify equilibrium decisions subjects were taught (equilibrium checking, best-response dynamics, and iterated dominance; Appendix A) and may therefore screen out more subjects whose compliance would be low. However, the compliance rates, ranging from 55.6% to 70.3% for *Equilibrium* and *Dk* subjects, which are high for exact compliance, suggest that Baseline subjects' widespread failure to make *Equilibrium* or *Dk* guesses is not directly caused by cognitive limitations. Nonetheless, the striking differences in compliance and failure rates between *Lk*, *Equilibrium*, and *Dk* R/TS subjects are probably an important clue in explaining the predominance of *Lk* and *Equilibrium* over *Dk* Baseline subjects.

C. Baseline and OB subjects' compliance with iterated dominance and equilibrium

We now examine the aggregate compliance of Baseline and OB subjects' adjusted guesses with iterated dominance and equilibrium. Table VIII reports Baseline, OB, and pooled Baseline and OB subjects' compliance with 0-3 rounds of dominance, and with *Equilibrium* adjusted guesses, both overall and in the games ordered as in Table II, with random compliance as a benchmark.[33] Aggregate compliance with 0-3 rounds of dominance is similar for Baseline and

[32]This inversion is due, we suspect, to a curious framing effect, in which some *L1* R/TS subjects try to outguess the computer but *L2* or *L3* subjects do not try to outguess their simulated partners' deterministic responses.

[33]Appendix D gives the analogous results for other types. Almost all zero compliance rates for iterated dominance are due to logical constraints. The rates seldom differ for within 0 or 0.5, but when they do the tables give the latter.

OB subjects game by game, usually far higher than random. In both treatments subjects violate simple dominance at a rate (100 minus compliance with 0 rounds in Table VIII) less than random in each of the 13 games in which it is non-vacuous, by a factor from one-sixth to two-fifths. Overall, subjects respect simple dominance 90% of the time, a typical rate for initial responses to games and much higher than random, which averages about 60% in our games. Compliance varies systematically across games, but there is no clear effect of structure beyond what determines random compliance.[34] Baseline and OB subjects' compliance with *Equilibrium* adjusted guesses are also similar game by game, also with no clear effect of structure per se.

D. Econometric analysis of Baseline and OB subjects' guesses

As explained in the Introduction, a large minority of our Baseline and OB subjects made guesses that conform so closely to one of our types that we can confidently assign the subject to that type by inspection, but most of our subjects' guesses are less conclusive. In this section we conduct a maximum likelihood error-rate analysis of all 88 Baseline and OB subjects' 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.

Recall that in our model, each subject's behavior is determined, possibly with error, by a single type, which determines his guesses and searches in all 16 games. Our types include *L1*, *L2*, *L3*, *D1*, *D2*, and *Equilibrium* as defined above. These types 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.[35] We add CGCB's *Sophisticated* to test whether any of our subjects have a prior understanding of others' decisions that transcends these simple decision rules. In theory, *Sophisticated* best responds to the probability distributions of its

[34]By contrast, the number of rounds of dominance has a strong effect on equilibrium compliance in CGCB's games.

[35]An ad hoc type could perfectly mimic a subject's decision history, but would have no explanatory power. It is hard to dispense with a priori specification because the space of possible decision rules is enormous and the leading rules have no simple, unifying structure; and because there are multiple rationales for any 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. *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. *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 higher-order *Lk* and *Dk* types are empirically unimportant, and there is no evidence of them in our data. We also omit 3 types CGCB allowed but found empirically unimportant: *Pessimistic* (maximin), *Optimistic* (maximax), and *Altruistic*. *Pessimistic* and *Optimistic* do not distinguish 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.

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 the distributions in our setting: the population frequencies of our own subjects' guesses.

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 $x^i \equiv (x_1^i,...,x_G^i)$ and $R^i(x^i) \equiv (R_1^i(x_1^i),...,R_G^i(x_G^i))$.

We analyze the data subject by subject.[36] 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 - ε; but with probability $\varepsilon \in [0,1]$, his *error rate*, his adjusted guess has density $d_g^k(R_g^i(x_g^i),\lambda)$ with *precision* λ.[37]

In describing how payoffs affect the error density $d_g^k(R_g^i(x_g^i),\lambda)$, we assume for simplicity that subjects are risk-neutral. Let y and $S_g(R_g^i(x_g^i), y)$ be subject i's partner's adjusted guess and i's own expected monetary payoff in game g, given y and i's own adjusted guess $R_g^i(x_g^i)$. Let the density $f_g^k(y)$ represent the beliefs about y implicit in type k.[38] Subject i's expected payoff in game g for type k's beliefs can then be written:

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

[39]

[36]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. We believe that estimating subject by subject is better suited to studying cognition, given subjects' heterogeneous behavior, and more robust to misspecification.

[37]Because the error rate, precision, and type are estimated jointly for each subject, there is no need to allow the error rate and precision to depend on type.

[38]The expectation in $S_g(R_g^i(x_g^i), y)$ is taken only over the random selection of games for which subject i is paid. All of our types can be viewed as best responding to some beliefs about their partner's guesses.

[39]In our design entered guesses are restricted to the interval [0, 1000], which includes all possible limits.

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

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

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

---

[40]The conditional density could be allowed to extend to U_g^{ik}, but our specification is simpler, and almost 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 $n^{ik} = 0$ or G, after setting the products with no terms in (5) equal to 1, 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.

The left-hand side of Table IX reports Baseline and OB subjects' numbers of dominated guesses and maximum likelihood estimates based on (7) of their types k, precisions λ, numbers of exact type-k guesses (which equal 16(1 – ε), where ε is the error rate). Subjects are ordered by estimated type, in decreasing order of likelihood within type. These point estimates assign 43 subjects to *L1*, 20 to *L2*, 3 to *L3*, 5 to *D1*, 14 to *Equilibrium*, and 3 to *Sophisticated*. 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.[41] The hypothesis λ = 0 is

[41]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.

rejected at the 1% (5%) level for the 21 (34) subjects whose estimates are superscripted ** (*) in Table IX, so the logit model's payoff-sensitive errors significantly improve the fit over a spike-uniform model like CGCB's for about a third of our subjects. The joint restriction $\varepsilon \approx 1$ and $\lambda = 0$, which approximates a completely random model of guesses, is rejected at the 5% (and 1%) level for all but the 10 subjects whose type indicators are superscripted † in Table IX.

E. Specification test and analysis

For reasons explained in the Introduction, subjects' point type estimates cannot all be taken at face value. As in previous analyses, they might be artifacts of our a priori specification of possible types, which could err by either by omitting relevant types, by including empirically irrelevant ones, or both. We now describe a specification test that addresses these issues.

Subject by subject, the test compares the likelihoods of our type estimate and analogous estimates based on 88 *pseudotypes*, each constructed from one of our subject's guesses over the 16 games.[42] Such comparisons help to detect whether any subject's guesses are better explained by an alternative decision rule, omitted from our specification; or whether a subject's estimated type is an artifact of overfitting, via accidental correlations with irrelevant included types.

First imagine that we had omitted an empirically important type, say *L2*. Then the pseudotypes of subjects now estimated to be *L2* would tend to outperform the non-*L2* types we estimated for them, and would also make approximately the same (*L2*) guesses. Define a *cluster* as a group of two or more subjects such that: (i) each subject's pseudotype has higher likelihood than the estimated type for each other subject in the group; and (ii) subjects' pseudotypes make "sufficiently similar" guesses.[43] Finding a cluster should lead us to diagnose an omitted type, and studying the common 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.[44]

---

[42]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.

[43]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. Finally, although the logic of our definition allows overlapping but non-nested clusters, that problem does not arise here.

[44]Because pseudotypes incorporate decision errors, they only approximate the omitted types we seek to identify. 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

Appendix E summarizes the results of comparing the likelihoods of our estimated types with the likelihoods of the 88 pseudotypes. Subjects are associated with rows, ordered by type and likelihood as in Table IX but with types ordered alphabetically; pseudotypes are associated with columns; and the entries give likelihoods. Appendix F summarizes the results of the likelihood comparisons in part (i) of the definition of a cluster and lists the 25 subsets of pseudotypes and subjects who satisfy part (i).[45] There are 5 (non-overlapping) subsets in which subjects' guesses appear close enough to warrant checking part (ii) of the definition. The subjects in those subsets that we judge to be part of a cluster are identified in the left-hand side of Table IX by superscripts on their type identifiers corresponding to the cluster labels in Appendix F and below. Appendix F also collects the guesses for subjects in those five subsets from Appendix C, and presents them with the games' parameters and types' guesses to facilitate the analysis.

We now discuss the similarities in subjects' guesses in each of these subsets, diagnosing misspecification by omitted decision rules and identifying the omitted rules when possible:

A. Subjects 202, 310, and 417, all estimated to be *Equilibrium*: All made *Equilibrium* guesses in our 8 games without mixed targets, and 310 also did so in 3 of our games with mixed targets; there was no apparent pattern with respect to other aspects of the structures (Table II). 202's and 417's deviations are always in the same direction, but to different guesses; all but one of 310's deviations in games without mixed targets was in the same direction, also to different guesses. This pattern of deviations is intriguing because the standard methods for identifying equilibrium guesses all work equally well in games with and without mixed targets.[46] We judge

---

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 a priori restrictions and 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.

[45] None of the likelihood comparisons are very close, except for 210's estimated type versus 302's pseudotype. We also made two exceptions to part (i) of the requirement: Subject 310 is included as a potential member of cluster A because his guesses are close to those of others in cluster A, and subject 204 is included as a potential member of cluster E because its likelihood is very close to the standard and its guesses are similar to other members' guesses.

[46] Only one of our 29 *Equilibrium* R/TS subjects came at all close to these subjects' patterns (1203 with 11 exact guesses, 4 of them with mixed targets), and the rest made as many exact guesses with as without mixed targets. In our debriefing questionnaire, subject 417 explicitly distinguishes games with mixed targets, in which, he says, "I usually assumed my partner chose from fairly near the center of his range, assuming it would deviate from this appropriately based on the difference of our multipliers (i.e., that the average of our guesses would be near the median of the overlapping part of our ranges)." We take this to mean that he adjusted his beliefs upward (downward) when his own target was lower (higher), but only half of 417's deviant guesses are consistent with this. For games without mixed targets, 417 gives a clear definition of equilibrium: "I made a greedy choice, always assuming my partner also made a greedy choice…."; there is no clue why he did not also follow this rule with mixed targets. Subject 202's responses are too vague to be helpful. Subject 310 says (without distinguishing games with mixed

202's and 417's guesses similar enough to meet the definition of a cluster, but we are unable to tell how they were determined; we suspect that they were using "homemade" rules that happen to mimic *Equilibrium* in games without mixed targets. However, we exclude 310 and so provisionally accept his identification as *Equilibrium*, which fits his guesses significantly better than 202's and 417's pseudotypes do, despite the similarities. This cluster illustrates the potential empirical importance of the subtlety of the arguments needed to identify equilibrium decisions.

B. Subjects 210 and 302, both estimated to be *L3* (with *Equilibrium* a fairly close second for both): Both deviate from *L3* guesses in 7 games, 6 of which have mixed targets; and 302 also has minor deviations in games 11 (also with mixed targets) and 14. There is no apparent pattern with respect to other aspects of the structures. Of the 7 common deviations, 6 are in the same direction, all to similar guesses. Both subjects make exactly the equilibrium guess in game 6, our only game without mixed targets in which *Equilibrium* is separated from *L3*. We are unable to tell how those subjects' guesses were determined, but we judge them similar enough to meet the definition of a cluster. Their decision rules appear to be hybrids of *L3* and *Equilibrium*, perhaps switching from one to the other according to some cue in the structure that we cannot discern.

C. Subjects 407, estimated to be *L2*; and 516, estimated to be *L1*: Both make *L1* guesses in most (5 and 7, respectively) of the first 9 games played and *L2* guesses in most (6 and 4) of the last 7. (*L1* and *L2* guesses are separated in all but game 9, in which both make the *L1* and *L2* guess.) There is no apparent pattern in their deviations from *L1* or *L2* with respect to the structures. We judge their guesses similar enough to meet the definition of a cluster, but we do not believe these subjects followed an omitted hybrid type. The time pattern of deviations and the fact that most of their later guesses followed a more sophisticated rule suggest introspective learning during play, of a kind ruled out by assumption in our analysis.[47]

---

targets), "Used what would be best for me and what was best for them" and then gives the formula for the equilibrium adjusted guess *without* mixed targets. These responses amply illustrate the pitfalls of using questionnaire statements as a substitute for data. From now on we refer to questionnaire responses only when they are helpful.

[47]Both subjects' questionnaires give fairly clear statements of *L2*, but no indication that they did not always follow it. It is interesting to compare their guesses with subject 108's, which mostly follow *L2*'s guesses but deviate to *L1'*s in games 2, 10, and 16. 108's *L1* guesses are mostly late, and *L2* fits his guesses significantly better than any pseudotype. 108's questionnaire also gives a clear statement of *L2*, but a vague discussion of the switches to *L1*. A few subjects give weaker evidence of introspective learning, also in the form of early-late *L1* to *L2* switches: 209 makes *L1* guesses in games 1 and 3 and *L2* guesses in all other games but 10; 218 makes *L1* guesses in games 1-3 and *L2* guesses in all other games but 4 and 10; and subjects 301, 504, 508, and 516 have similar, noisier patterns. It is particularly telling that 209 and 218 make *L1* and then *L2* guesses early and late in the symmetric games 3 and 12.

D. Subjects 301 and 508, both estimated to be *L1*: These subject's pseudotypes are the only ones with higher likelihood than each other's estimated type. They have five common deviations from *L1*, always downward, though almost always to different guesses; and each subject also has one lone (upward) deviation.[48] The common deviations have no apparent pattern with respect to timing or structure. Both lone deviations seem due to forgetting to multiply by own target and some common deviations also seem due to forgetting or interchanging targets or limits. We judge these subjects' guesses to be similar enough to meet the definition of a cluster, but we are not fully convinced that they followed an omitted type. There is a chance that they are just sloppy *L1* subjects whose cognitive errors for some reason occurred mostly in the same games.

E. Subjects 204 and 313, both estimated to be *D1*, and 409, estimated to be *L1*: These subjects all made similar guesses, including 645s inexplicable by our types in the symmetric games 3 and 12 and, for 204 and 409, in asymmetric game 13. They are among the minority of subjects who explained their guesses clearly in their questionnaires: All stated homemade rules that depart from standard decision theory (and so from our types) in different ways, but which, properly reinterpreted, explain most of their guesses.[49] We treat them as a cluster because their guesses are similar, but they were plainly not following *L1*, *D1*, or any single omitted type.

The subjects in cluster E illustrate what seems to be a widespread tendency to invent rules by which to process the data of our games into decisions. We find it unremarkable that these 3 subjects' rules deviate from standard decision theory. What is remarkable is the high frequency with which our other subjects' rules (mostly *L1*, *L2*, or *Equilibrium*) *do* conform to standard decision theory, even though most of them are best responses to non-equilibrium beliefs.

[48]Curiously, 3 of subject 301's 6 deviations from *L1* guesses are to equilibrium guesses (twice when they are separated from all other types' guesses), though there is no hint of *Equilibrium* in his questionnaire.

[49]Subject 204 says that he first found the person whose "spread" (defined as own target times the difference between the partner's limits) was smaller. If his spread was smaller, he guessed the average of the range between his target times the partner's lower and upper limits; and if the partner's spread was smaller, he guessed the average of the analogous partner's range, thus without taking his own target into account, which makes no sense decision-theoretically. In fact he adjusted the ranges according to the limits; with this adjustment the stated rule explains his guesses in 11/16 games. Subject 313 says that he guessed $[\max\{a_ip_j, a_jp_i\} + \min\{b_ip_j, b_jp_i\}]/2$ ("I multiplied my upper and lower limits w/ partner's target, then multiplied his upper and lower limits w/ my target. Then I chose the largest of the lowers and smallest of the uppers to find my new more refined range. Then I guessed the average of this range."). In fact he separately adjusted each term in the above formula to his own limits before averaging them (see especially his game 14 guess), which makes no sense decision-theoretically. With adjustment, the stated rule explains his guesses in 14/16 games. Subject 409 says that he guessed $[\max\{a_i, a_jp_i\} + \min\{b_i, b_jp_i\}]/2$ ("Basically, I took his/her lower limit and multiplied it by my target. If the resulting number was between my upper and lower limits, I kept that in mind. Otherwise I picked my lower limit. Then I took his/her upper limit and multiplied it by my target. Again, if the resulting number was within my range, I took it. Otherwise I picked the upper limit. Then I found the average of the two numbers.") The stated rule explains his guesses in 13/16 games.

With regard to the overfitting part of our specification test, we assume that for a subject's estimated type to be credible it should perform at least as well against the pseudotypes as it would, on average, at random.[50] Suppose that a subject's behavior is random relative to our types and pseudotypes other than his own, in that their likelihoods are independent and identically distributed ("i.i.d."). Then for a pseudotype to have higher likelihood than our estimated type it must come first among our 7 types plus itself, which has probability 1/8. Thus, for a subject's estimated type to be credible, it should have higher likelihood than all but at most 87/8 ≈ 11 of the pseudotypes. Those subjects whose estimated types have lower likelihoods than 12 or more pseudotypes have type identifiers superscripted + in Table IX; they include 10 subjects estimated to be *L1*, 2 *L2*, and one *Sophisticated*, all with likelihoods among the lowest for their types.

We now combine our guesses-only type estimates with the results of the specification tests to give a preliminary assessment of the reliability of subjects' type estimates. We say that a guesses-only type estimate *appears reliable* if: (i) it does significantly better at the 5% (or 1%, which yields the same result here) level than a random model of guesses within our specification; (ii) it has higher likelihood than all but at most a random number of pseudotypes; and (iii) it is not a member of any cluster.[51] By these criteria, 58 of our 88 subjects' guesses-only type estimates appear reliable: 27 *L1*, 17 *L2*, 11 *Equilibrium*, and one each *L3*, *D1*, or *Sophisticated*. These subjects' guesses-only type identifiers have no superscripts, and are in bold in Table IX.

Despite the differences between our games and those in previous studies, this type classification is reasonably close to Nagel's, HCW's, CGCB's, and SW's. There are two main differences. First, we find more *Equilibrium* subjects (12.5%, focusing on identifications that appear reliable) than previous studies, except SW's. Second, we find no indication of significant numbers of types other than *L1*, *L2*, *Equilibrium*, and hybrids of *L3* and/or *Equilibrium*, in contrast to SW's (1995) classification of many subjects as *Worldly*, almost to the exclusion of *L2*.

Our results allow us to go beyond our specification test, which looks for unspecified omitted types, to reach more definite conclusions on *Worldly* and SW's noisy version of *L2*.[52] SW's *L2* is defined as a risk-neutral best response to a noisy *L1*, which depending on the noise

[50]This should hold even for pseudotypes associated with subjects of the same estimated type, because under the null hypothesis, another subject's deviations from that type should not help explain the subject's own deviations. This is plainly a weak test, which can be counted on to detect only the most obvious artifacts of overfitting.

[51]"Appears" because this assessment is preliminary, pending our analysis of search.

[52]The issue is not whether subjects' own decisions are noisy, but whether they are assumed to respond to others' decision noise. SW's and CHC's definition of *L1* as a best response to uniform beliefs is identical to our definition.

parameter ranges from *L0* (uniform randomness) to our noiseless *L1*. SW's *Worldly* is defined as a risk-neutral best response to an estimated mixture of a noisy *L1* and a noiseless *Equilibrium*. By a kind of "median-voter" result, our not-everywhere-differentiable payoff function (Section 1.B) makes it optimal to best respond to the median type in the population as if it were the only type.[53] It follows that *Worldly* does not respond to the frequency of *Equilibrium* when it is less than 0.5—as it is in all estimates that have been published—and hence that *Worldly* is equivalent to SW's noisy *L2*. Because our payoff function is quasiconcave, the best response that defines them lies between those of our noiseless *L1* and *L2*—strictly between except for extreme values of the noise parameter that make *Worldly* and SW's *L2* coincide with our *L1* or *L2*. Yet only one of our 88 subjects made guesses in that range in as many as 10 games, one in 9, and 2 in 8.[54] By contrast, 43 made *exact* guesses for our noiseless *L1*, *L2*, *L3*, or *Equilibrium* in 7 or more games; and both their and our other subjects' guesses appear random relative to *Worldly*'s and SW's *L2*'s.

A related specification issue concerns CHC's definition of *Lk* types as best responses to estimated mixtures of noiseless lower-level *Lk* types. This captures SW's idea of worldliness without making *Lk*'s behavior depend on the noisiness of lower-level *Lk* types or mixing in *Equilibrium* as SW do. But because CHC's mixture parameter depends on others' behavior, which subjects do not observe, their definition assumes that subjects have prior understandings of it. CGCB (2001, Section 3.A) argued that the *Sophisticated* type tests for prior understandings more cleanly, without imposing structural restrictions or raising delicate specification issues, than types that depend on estimated parameters like *Worldly* or CHC's *Lk*. More evidence on this issue would be useful, but in our games CHC's noiseless *L2* and *L3* both make exactly the same guesses as our *L2*.[55] Thus, while our results argue against making *Lk* respond to the noisiness of lower-level types, they do not discriminate between CHC's mixture definition of *Lk* and ours.

[53]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.
[54]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.
[55]CHC's *L2* best responds to a mixture of *L0* and *L1* in the proportions $1{:}\tau$, which for $\tau > (<)\ 1$ puts more weight on *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 $\tau \approx 1.5$.

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

In this section we generalize Section 3.D's model of guesses to obtain an error-rate model of guesses and information searches, and use it to re-estimate Baseline subjects' types. The model follows Section 3.D'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 searches of our Baseline subjects whose guesses clearly identify a type, and of our R/TS subjects (Appendix C) generally support our theory of cognition and search (Section 2). The main issue in extending our econometric model of guesses to take search into account is measuring compliance with types' search implications. Two aspects of the data are important here. First, many subjects (e.g. 202 and 210) consistently start with "123456" or some variation, and many end with an optional "13," checking their own limits even when 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 guess fingerprints are clear consistently look first at their type's relevant sequence (Table VI) 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). Still others repeat the relevant sequence over and over (e.g. 101). Thus one can identify three styles, "early," "late," and "often"; but the data also suggest that "often" subjects are almost always well described as either "early" or "late". Accordingly, we define compliance with a type's search implications as the density of the type's relevant look-ups in the look-up sequence, filtering out some idiosyncratic noise using a binary nuisance parameter called *style*.

Style 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 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.[56]

To reduce the need for structural restrictions, we discretize search compliance as follows.[57] 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.[58] 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 3.D, in each game g, a subject i of type k and style s makes an adjusted guess in U_g^{ik} with probability 1 - ε; 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 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:

---

[56]The compliance data are in Appendix G. 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}.
[57]Compliance is inherently discrete, but our discretization is coarser than necessary. This is a convenient place to correct a typographical error in CGCB's equation (4.3), where the summation (∑) should be a product (∏).
[58]A natural generalization would allow search and guess errors to be correlated for a given game and subject by allowing compliance-contingent error rates and precisions as in CGCB. We dispense with this for simplicity.

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

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

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

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

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

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 3.D'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 3.D'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.[59]

The right-hand side of Table IX reports maximum likelihood estimates of each Baseline subject's type and style, error rate, precision, and rates of search compliance, based first on search only and then on guesses and search combined. For the latter estimates we report separate as well as total log-likelihoods, to give a better indication of what drives the type estimates.[60]

Most subjects' type estimates based on guesses and search reaffirm the guesses-only estimates, including those of 39 of the 46 Baseline subjects whose estimates appeared reliable and 51 out of all 71 Baseline subjects.[61] For some subjects, however, the guesses-and-search type

[59]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.

[60]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.

[61]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 because our theory makes sharper predictions about guesses than about search, which are far less likely to be

estimate resolves a tension between guesses-only and search-only estimates in favor of a type different than the guesses-only estimate.[62] In more extreme cases, a subject's guesses-only type estimate is excluded because it has 0 search compliance in 8 or more games. 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 VI), 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 2).[63]

We update Section 3.E's reliability criteria to incorporate search as follows. 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. We say that a guesses-and-search type estimate is *reliable* if: (i) it does significantly better at the 5% (or 1%, which yields the same result here) 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) it is not a member of any cluster.[64]

A guesses-and-search type estimate that does sufficiently better than random in explaining search can satisfy the updated criterion (i) even if it does not satisfy the guesses-only criterion (i). In fact the only Baseline subject who does *not* satisfy the updated criterion (i) is 109, who had 0 search compliance in at least 8 games for every type. But the updated criterion (ii) is more stringent than the guesses-only criterion (ii), because the guesses-and-search type estimate can only have the same or lower likelihood for guesses than for the guesses-only type estimate. Criterion (ii) calls into question the type identifications of 7 subjects who satisfied the guesses-only criterion (ii), in addition to those of the 15 subjects who didn't satisfy that criterion.

---

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.

[62]These subjects include 105, 113, 213, and 420, estimated as *L1* based on guesses but *Equilibrium*, *L2*, or *L3* based on guesses and search; 110, 205, 306, 403, and 414, estimated as *L2* based on guesses but *L1* or *Equilibrium* based on guesses and search; 302, estimated as *L3* based on guesses but *Equilibrium* based on guesses and search; and 312 and 313, estimated as *D1* (312 noisy) based on guesses but *L1* or *L2* based on guesses and search.

[63]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.

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

Under the updated criteria, we classify 43 of our 71 Baseline subjects' guesses-and-search type estimates as reliable: 22 *L1*, 13 *L2*, and 8 *Equilibrium*. These subjects' guesses-and-search type identifiers have no superscripts (though some have subscripts) and are in bold in Table IX. The search analysis confirms the reliability of 39 of our 46 Baseline subjects whose guesses-only type estimates appeared reliable. It also reliably identifies 2 subjects as *L1* who had appeared reliably identified as *L2*; and confirms the reliability of 1 *L1* and 1 *L2* subject whose guesses-only estimates were inconclusive. Finally, it calls into question the type estimates of 4 subjects who had appeared reliably identified: 1 each *L1* **(**subject 415), *L2*, *D1*, and *Equilibrium*.

Adding to these Baseline subjects the 11 of 17 OB subjects whose guesses-only type estimates (in bold in Table IX) appeared reliable (7 *L1*, 1 *L2*, 1 *L3*, 2 *Equilibrium*, and 1 *Sophisticated*), we have a total of 54 of 88 subjects whose types can be identified with confidence: 29 *L1*, 14 *L2*, 1 *L3*, 10 *Equilibrium*, and 1 *Sophisticated*.[65] Going beyond our criteria, one might add subject 415 as a probable *L1* and the 4 subjects in clusters A and B as likely hybrids of *L3* and/or *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*. This suggests that it will be difficult to improve upon a random model of the behavior of the 29-34 unclassified subjects.

4. 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 (types) 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 reliably be identified via an econometric and specification analysis.

Our subject population includes significant numbers of reliably identified *L1*, *L2*, and *Equilibrium* subjects, and possibly some hybrids of *L3* and/or *Equilibrium*. A large majority of these subjects follow types other than *Equilibrium*; and because their types build in risk-neutral, self-interested rationality and perfect models of the game, their systematic deviations from equilibrium can be confidently attributed to non-equilibrium beliefs rather than irrationality, risk

[65]Other subjects' low levels of compliance with *Sophisticated*'s search requirements suggest that the identification of the 1 *Sophisticated* subject, who was a noisy OB subject, might not have survived monitoring search.

aversion, altruism, spite, or confusion. Our results strongly 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 their initial responses. The surprisingly simple structure of their behavior is consistent with previous analyses but refines and sharpens them, supporting the leading role given *Lk* types in applications and informal analyses. Its simplicity should help to allay the common fear that "anything can happen" with bounded rationality.

A companion paper, Costa-Gomes and Crawford (2004), will analyze our R/TS and Baseline subjects' search behavior in more detail, studying the relations between cognition, search, and guesses. Preliminary analysis provides support for our model of cognition and search and confirms that subjects find *Lk* types easier or more natural to implement than *Equilibrium* or *Dk*. This may help to explain the prevalence of *Lk* types in the Baseline and OB treatments.

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

Camerer, Colin (2003): *Behavioral Game Theory: Experiments on Strategic Interaction*, Princeton University Press, Princeton, NJ.

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

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

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

Costa-Gomes, Miguel, and Vincent Crawford (2004): "Studying Cognition via Information Search in Experimental Two-Person Guessing Games," manuscript in preparation.

Costa-Gomes, Miguel, Vincent Crawford, and Bruno Broseta (1998): "Cognition and Behavior in Normal-Form Games: an Experimental Study," UCSD Discussion Paper 98-22.

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

Crawford, Vincent (1997): "Theory and Experiment in the Analysis of Strategic Interaction," *Advances in Economics and Econometrics, Seventh World Congress: Theory and Applications, Vol. I*, Econometric Society Monographs No. 27, ed. by David Kreps and Kenneth Wallis. New York: Cambridge University Press, pp. 206-242; reprinted in *Advances in Behavioral Economics*, ed. by Colin Camerer, George Loewenstein, and Matthew Rabin. Princeton, NJ: Princeton University Press, 2004.

Crawford, Vincent (2003): "Lying for Strategic Advantage: Rational and Boundedly Rational Misrepresentation of Intentions," *American Economic Review* 93, 133-149.

Crawford, Vincent, and Nagore Iriberri (2004): "Fatal Attraction: Focality, Naivete, and Sophistication in Experimental Hide-and-Seek Games," UCSD Discussion Paper 2004-12.

Grosskopf, Brit, and Rosemarie Nagel (2001): "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.

Harless, David, and Colin Camerer (1995): "An Error Rate Analysis of Experimental Data Testing Nash Refinements," *European Economic Review*, 39, 649-660.

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

Holt, Debra (1999): "An Empirical Model of Strategic Choice with an Application to Coordination Games," *Games and Economic Behavior*, 27, 86-105.

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

Kübler, Dorothea, and Georg Weizsäcker (2004): "Limited Depth of Reasoning and Failure of Cascade Formation in the Laboratory," *Review of Economic Studies*, 71, 425-441.

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

Nagel, Rosemarie (1993): "Experimental Results on Interactive Competitive Guessing," Discussion Paper B-236, University of Bonn.

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

Payne, John, James Bettman, and Eric Johnson (1993): *The Adaptive Decision Maker.* New York: Cambridge University Press.

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

Stahl, Dale (1996): "Boundedly Rational Rule Learning in a Guessing Game," *Games and Economic Behavior*, 16, 303-330.

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

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

Weibull, Jörgen (2004): "Testing Game Theory," in *Advances in Understanding Strategic Behaviour: Game Theory, Experiments, and Bounded Rationality: Essays in Honour of Werner Guth,* Palgrave.

Figure 1. Screen Shot of the MouseLab Display

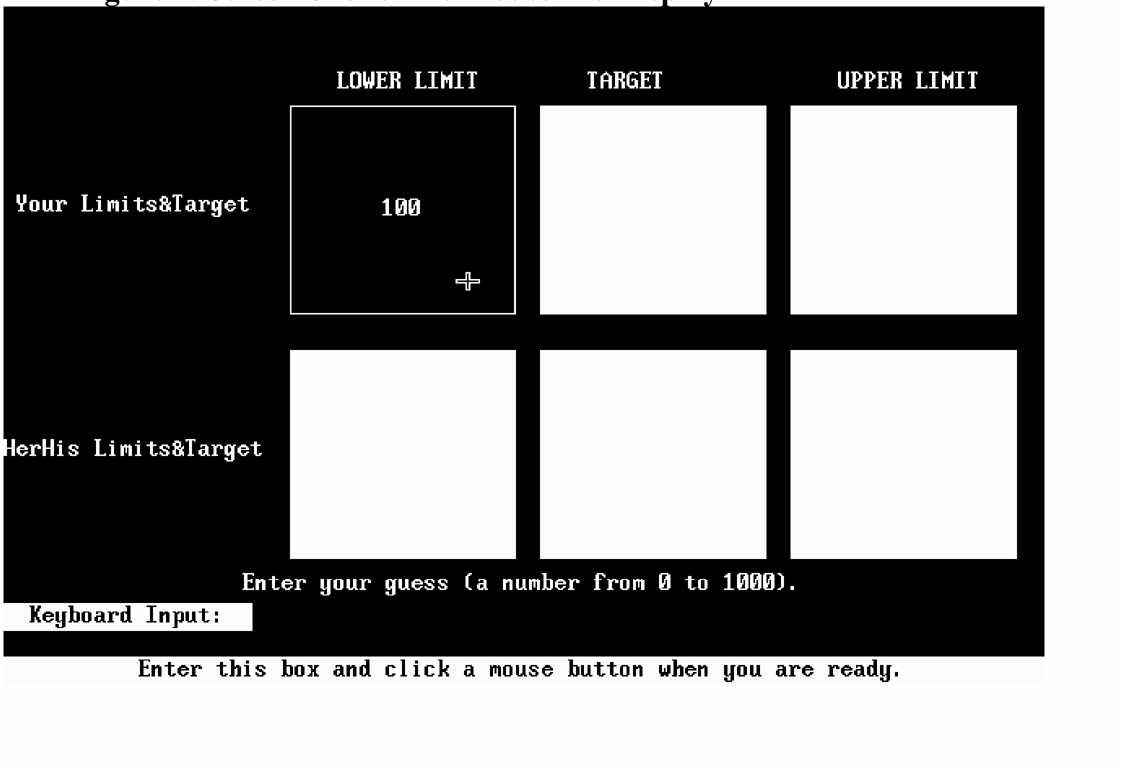

Table I. Overall Structure

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 II. Strategic Structures of the Games

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

Limits: α [100, 500], β [100, 900], γ [300, 500], δ [300, 900]. Targets: 1 for 0.5, 2 for 0.7, 3 for 1.3, 4 for 1.5. Patterns of dominance: A for Alternating; S for Simultaneous; and S/A for Simultaneous in first round, then Alternating.

Table III. Types' Guesses and Guesses that Survive Iterated Dominance

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

Table IV. Numbers of Games in which Types' Guesses are Separated by More than 0, 25

	L1	*L2*	*L3*	*D1*	*D2*	*Eq.*	*Soph.*
L1	-	15, 13	15, 12	12, 10	15, 12	15, 15	15, 14
L2	15, 13	-	11, 9	13, 9	10, 8	11, 9	10, 8
L3	15, 12	11, 9	-	13, 12	8, 5	9, 6	9, 8
D1	12, 10	13, 9	13, 12	-	9, 7	14, 13	12, 10
D2	15, 12	10, 8	8, 5	9, 7	-	9, 8	9, 6
Eq.	15, 15	11, 9	9, 6	14, 13	9, 8	-	11, 9
Soph.	15, 14	10, 8	9, 8	12, 10	9, 6	11, 9	-

Table V. Strength of Baseline and OB Subjects' Incentives to Make Types' Guesses

	L0	*L1*	*L2*	*R1*	*R2*	*Eq.*	B+OB
L1	34.95 (100)	28.41 (55)	36.81 (76)	34.38 (83)	33.61 (78)	25.98 (56)	34.63 (8
L2	31.20 (89)	51.81 (100)	31.34 (65)	39.30 (94)	38.68 (90)	31.37 (68)	38.73 (9
L3	32.99 (94)	35.01 (68)	48.14 (100)	38.70 (93)	41.14 (95)	34.00 (74)	39.34 (9
D1	33.73 (97)	41.13 (79)	37.56 (78)	41.64 (100)	41.11 (95)	29.42 (64)	39.50 (9
D2	32.86 (94)	41.56 (80)	40.57 (84)	40.79 (98)	43.13 (100)	32.43 (70)	40.07 (9
Eq.	30.14 (86)	36.67 (71)	36.09 (75)	35.87 (86)	38.30 (89)	46.05 (100)	35.98 (8
Soph.	33.04 (95)	41.38 (80)	41.24 (86)	40.77 (98)	41.84 (97)	31.67 (69)	40.53 (1

Note: The entries are in US dollars, expressed as percentages of the column maximum in parentheses.

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

Table VII. R/TS subjects' compliance with assigned type's guesses

	L1	*L2*	*L3*	*D1*	*D2*	*Eq.*
UCSD subjects	7	9	-	11	-	10
% Compliance	77.7	81.3	-	55.1	-	58.1
% Failed UT2	0.0	0.0	-	8.3	-	28.6
York subjects	18	18	18	19	19	19
% Compliance	80.9	95.8	84.4	66.1	55.6	76.6
% Failed UT2	0.0	0.0	0.0	0.0	5.0	13.6
UCSD + York subjects	25	27	18	30	19	29
% Compliance	80.0	91.0	84.4	62.1	55.6	70.3
% Failed UT2	0.0	0.0	0.0	3.2	5.0	19.4

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

Game (#rounds)	Respects 0 rounds	Respects 1 round	Respects 2 rounds	Respects 3 rounds	*Equilibrium* within 0 or 0.5	*Equilibrium* within 25
	B, OB, B+OB	**B, OB, B+OB**	**B, OB, B+OB**	**B, OB, B+OB**	**B, OB, B+OB**	**B,OB, B+OB**
All games	10,11,10 (39)	15,16,15 (20)	22,21,21 (7)	13,14,14 (8)	18,15,18 (0,0)	23,15,22 (3)
α2β1 (4)	0,0,0 (0)	62,82,66 (81)	0,0,0 (0)	23,18,22 (19)	15,0,12 (0,0)	31,0,25 (0)
β1α2 (3)	21,24,22 (81)	0,0,0 (0)	62,65,63(19)	17,12,16 (0)	17,12,16 (0,0)	20,12,18 (2)
β1γ2 (3)	27,29,27 (88)	0,0,0 (0)	63,59,63(12)	10,11,10 (0)	10,12,10 (0,0)	28,24,27 (6)
γ2β1 (2)	0,0,0 (0)	55,59,56 (100)	45,41,44(0)	0,0,0 (0)	45,41,44 (0,0)	48,59,50 (0)
γ4δ3 (2)	18,24,19 (75)	14,0,11 (25)	68,77,69 (0)	0,0,0 (0)	68,76,69 (0,0)	72,76,73 (0)
δ3γ4 (3)	11,18,13 (57)	51,59,52 (32)	10,6,9 (11)	28,18,26 (0)	28,18,26 (0,0)	31,18,28 (8)
δ3δ3 (5)	4,0,3 (15)	4,12,6 (19)	23,12,21 (26)	42,53,44 (33)	25,18,24 (0,0)	27,24,26 (0)
δ3δ3 (5)	6,0,5 (15)	0,6,1 (19)	28,18,26 (26)	44,65,48 (33)	23,12,20 (0,0)	23,12,20 (0)
β1α4 (9)	31,24,30 (81)	0,0,0 (0)	37,35,36 (8)	0,0,0 (0)	6,0,5 (0,0)	6,12,7 (0)
α4β1 (10)	0,0,0 (12)	47,35,44 (32)	0,0,0 (0)	23,35,25 (23)	3,6,3 (0,0)	4,6,5 (13)
δ2β3 (17)	14,12,14 (45)	0,0,0 (0)	4,12,6 (9)	0,0,0 (0)	6,0,5 (0,0)	6,0,5 (0)
β3δ2 (18)	6,6,3 (36)	0,6,5 (10)	28,0,0 (0)	44,18,23 (10)	1,0,1 (0,0)	7,0,6 (6)
γ2β4 (22)	0,0,0 (0)	4,0,3 (7)	0,0,0 (0)	3,0,2 (8)	18,29,20 (0,0)	23,29,24 (0)
β4γ2 (23)	11,18,13 (62)	0,0,0 (0)	4,0,3 (3)	0,0,0 (0)	8,6,8 (0,0)	10,6,9 (6)
α2α4 (52)	9,18,10 (38)	0,0,0 (1)	0,0,0 (0)	0,0,0 (1)	13,6,11 (0,0)	20,6,17 (13)
α4α2 (51)	3,0,2 (12)	0,0,0 (0)	3,6,3 (2)	0,0,0 (0)	7,0,6 (0,0)	8,0,7 (0)

Note: The table gives compliance percentages rounded to the nearest integer, with random compliance percentages in parentheses.

Table IX. Type Estimates Based on Guesses Only, Search Only, and Guesses and Search

		Guesses only				Search only				Guesses and search							
ID	**dom.**	**ln L**	$\boldsymbol{k}$	**exac**	$\boldsymbol{\lambda}$	**ln L**	$\boldsymbol{k_s}$	ζ_H	ζ_M	**ln** $\mathbf{L_t}$	**ln** $\mathbf{L_g}$	**ln** $\mathbf{L_s}$	$\boldsymbol{k_s}$	**exa**	$\boldsymbol{\lambda}$	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	$\boldsymbol{L1_e}$	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	$\boldsymbol{L1_{e‡‡}}$	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	$\boldsymbol{L1_e}$	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	$\boldsymbol{L1_l}$	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	$\boldsymbol{L1_e}$	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	$\boldsymbol{L1_e}$	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	$\boldsymbol{L1_e}$	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	$\boldsymbol{L1_e}$	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	$\boldsymbol{L1_{el}}$	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	$\boldsymbol{L1_e}$	10	0.00	0.00	0.69
415	1	-53.64	***L1***	9	0.88	-16.38	$D1_e$	0.31	0.50	-107.28	-90.90	-16.38	$D1_e^{+}$	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	$\boldsymbol{L1_l}$	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	$\boldsymbol{L1_e}$	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	$\boldsymbol{L1_l}$	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	$\boldsymbol{L1_e}$	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	$\boldsymbol{L1_e}$	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	$\boldsymbol{L1_e}$	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	$\boldsymbol{L1_e}$	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	$\boldsymbol{L1_e}$	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	$\boldsymbol{L1_e}$	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	**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	**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	$L3_{el}^{\ddagger}$	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	**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	**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	**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	**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	**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	**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	**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	**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	**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	**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	$L2_{e}^{C}$	8	0.44	0.88	0.13
306	2	-68.48	**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	**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	**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	**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	**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	**L2**	5	0.98**	-	-	-	-	-	-	-	-	-	-	-	-
503	3	-88.21	$L2^{+}$	3	0.00	-	-	-	-	-	-	-	-	-	-	-	-
414	4	-89.00	**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	$D1^{E}$	2	2.68^{**}	-6.03	$L1_e^{\ddagger}$	0.00	0.88	-90.93	-84.90	-6.03	$L1_{e\ddagger\ddagger}^{E}$	2	3.28	0.00	0.88
312	0	-80.45	$D1^{\dagger}$	3	5.85^{**}	-3.74	$L2_e^{\ddagger}$	0.94	0.06	-84.74	-81.00	-3.74	**$L2_e$**	3	1.37	0.94	0.06
204	2	-84.86	$D1^{E}$	2	1.22^{**}	0.00	$L1_e^{\ddagger}$	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	$D1^{\dagger+}$	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	$L3_e^{\ddagger}$	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	$Eq_e^{\ddagger}$	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	$Eq_e^{\ddagger}$	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***	2	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	$Eq^{\dagger+}$	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	$So^{\dagger}$	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 equal 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; $L1_e$ 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.