Discussion Paper No. 613

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

Miguel A. Costa-Gomes
and
Vincent P. Crawford

August 2004

The Institute of Social and Economic Research
Osaka University
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

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

". . . 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 experiments that elicit 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 understand the games and seek to maximize their payoffs, but have boundedly rational models of others' decisions, which lead to systematic deviations from equilibrium.

Keywords: noncooperative games, experimental economics, guessing games, bounded rationality, strategic sophistication, 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, and Joel Sobel for helpful discussions; 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.

1. Introduction

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 in games where the reasoning that leads to it is not straightforward. This evidence also suggests that a structural model that allows for the possibility that players follow certain boundedly rational decision rules, in lieu of equilibrium, can out-predict equilibrium in applications involving initial responses. Such a model is likely to establish more robustly the conclusions of applications that now rely on equilibrium in simple games, where boundedly rational rules often mimic equilibrium; but also to challenge the conclusions of applications to games too complex for equilibrium to be plausible without learning, by predicting the deviations from equilibrium often observed in such games.[2]

The potential usefulness of a structural non-equilibrium model 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 (15-18 in Nagel, 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.

Although Nagel's and HCW's subjects played a game repeatedly in the same group, their initial guesses can be viewed as responses to an independent play of the game if they treated their own influences on others' future guesses as negligible, which is plausible for all but perhaps the 3-subject groups. With publicly announced structures, it is also reasonable to assume complete information. With one exception (p = 4/3 with limits [0, 100]), this makes the games dominance-solvable in finite (limits [100, 200]) or infinite (limits [0, 100]) numbers of rounds, with unique equilibria in which all players guess their lower (upper) limit when $p < 1$ ($p > 1$).

[2]Camerer (2003), Camerer, Ho, and Chong (2004; "CHC"), and Kübler and Weizsäcker (2004) give examples of structural non-equilibrium analyses. Crawford (2003) gives a sample application to strategic communication. A common theme is that allowing structured bounded rationality can resolve puzzles that are intractable assuming equilibrium. Although we focus on initial responses, modeling them more accurately will also inform applications where players can learn to play an equilibrium from experience with analogous games. The mental models that drive initial responses also affect the structure of learning rules, distinguishing reinforcement from beliefs-based and more sophisticated rules. This, in turn, influences predictions of convergence and selection among multiple equilibria.

These equilibrium predictions depend only on rationality (in the decision-theoretic sense) and beliefs derived from iterated or common knowledge of rationality. Yet Nagel's subjects never made their equilibrium guesses initially, and HCW's rarely did so. Most initial guesses respected from 0 to 3 rounds of iterated dominance, in games where from 3 to an infinite number are needed to identify an equilibrium (Nagel, Figure 1; HCW, Figures 2A-H and 3A-B).

These deviations from equilibrium are not adequately explained by social preferences, risk aversion, or failures of rationality. And while such deviations are often modeled as "equilibrium plus noise" or "equilibrium taking noise into account" in the sense of quantal response equilibrium ("QRE"; McKelvey and Palfrey (1995)), Nagel's and HCW's data resemble neither equilibrium plus noise nor QRE for any of the standard distributions, even making allowance for boundary effects and social preferences. Their data do suggest that the deviations have a coherent, partly deterministic structure. In Nagel's [0,100] games, for example, subjects' guesses have spikes that track $50p^k$ for $k = 1, 2, 3, \ldots$ across the different targets p in her treatments (Nagel, Figure 1). Like spectrograph peaks that signal the existence of chemical elements, these spikes are evidence of a structure that is discrete and individually heterogeneous.

Similarly structured deviations from equilibrium have been found in initial responses to matrix games by Stahl and Wilson (1994, 1995; "SW") and Costa-Gomes, Crawford, and Broseta (1998, 2001; "CGCB"), extensive-form alternating-offers bargaining games by Camerer, Johnson, Rymon, and Sen (1993, 2002; "CJ"), and other games (Crawford (1997, Section 4)), Camerer (2003, chapter 5), and CHC). As in the guessing games, subjects' responses are often "strategic" and they make undominated decisions with frequencies well above random, but they are less likely to rely on dominance for others (Beard and Beil (1994)), and reliance on iterated dominance seldom goes beyond 3 rounds. And in these experiments, subjects make equilibrium decisions less often in games where identifying them requires more rounds of iterated dominance or the fixed-point logic of equilibrium in non-dominance-solvable games (CGCB, Table II).

The data from these experiments have been analyzed using a variety of boundedly rational strategic decision rules called *types*. Leading examples include *L1* (for *Level 1*), which chooses its best response given a uniform prior over its partner's decisions; *L2*, which best responds to *L1*; *L3*, which best responds to *L2*; *D1* (for *Dominance 1*), which does one round of deletion of dominated decisions and chooses its best response given a uniform prior over its partner's remaining decisions; and *D2*, which does two rounds of iterated deletion of dominated decisions and best responds given a uniform prior over the remaining decisions.

How do these types differ from an *Equilibrium* type that makes its equilibrium decision? *Equilibrium, Lk*, and *Dk* types are all rational and all have accurate models of the game. All are usually defined, as we shall do here, to satisfy subsidiary assumptions of self-interestedness and risk-neutrality, which imply expected-pecuniary-payoff maximization. *Lk* and *Dk* types' essential departures from equilibrium involve replacing its accurate model of others' decisions with simplified, boundedly rational models. *Lk* anchors its beliefs with a uniform prior and adjusts them via thought-experiments involving iterated best responses, without "closing the loop" as for equilibrium. *Dk* avoids closing the loop by starting with finitely iterated knowledge of rationality and invoking a uniform prior at the end. Both procedures yield workable models of others' decisions, while avoiding much of the cognitive complexity of equilibrium analysis.[3]

Although *Lk* and *Dk* types have similar cognitive requirements by standard measures, and *Dk* types are closer to how theorists analyze dominance-solvable games, *Lk* types are usually taken as the natural specification of boundedly rational strategic reasoning.[4] However, both can explain the empirical relationship between equilibrium compliance and strategic complexity noted above: Because *Dk* respects k+1 rounds of dominance by construction, and *Lk+1* respects k+1 rounds in many games, a suitably heterogeneous mixture of either kind of type will mimic equilibrium in games that are dominance-solvable in small numbers of rounds, while deviating systematically in some more complex games. Further, in Nagel's and HCW's games with target $p < 1$ and limits [0, 100], *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, so both track the spikes in their data; but by the same token, both are perfectly confounded in those games. (They are only weakly separated in the other experiments mentioned above.)

---

[3]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. Each step of the procedure is simple, even if many case distinctions by simple criteria may have to be made."

[4]Keynes' wording in our epigraph connotes finite iteration of best responses—albeit before the notions of iterated dominance and equilibrium were current—but anchored, as seems natural in the beauty contest, by true preferences rather than uniform priors. Crawford (2003, p. 139) briefly discusses the informal literature on deception, which also features decision rules based on finite iteration of best responses, in this case anchored by truthfulness or credulity. Nagel (1995) focuses on *Lk* types, citing her questionnaire responses (1993, pp. 14-15; private communication) in support. Her subjects were University of Bonn students, most with no prior knowledge of game theory; but some of her questionnaire data were from 20 subjects in a pilot at the London School of Economics, who were probably more sophisticated. The LSE subjects' first-period rationales give some evidence of *Dk* types: By our reading, they include 3 *L2*, 3 *L1*, 2 possibly *Worldly* (SW (1995)), one *Equilibrium*, one *D2*, one *D1*, and 9 unclassifiable.

For these and other reasons, the structure of initial responses has not been identified as precisely or documented as convincingly as current methods allow. To move closer to that goal, this paper reports experiments that elicit 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 subjects' behavior as initial responses to independent games.

Our guessing games differ from Nagel's and HCW's in several ways. They have only two players, who make simultaneous guesses. Each player has a lower and an upper limit (either [100, 500], [100, 900], [300, 500], or [300, 900]), but players are not required to guess between their limits: guesses outside them are automatically adjusted up to the lower limit or down to the upper limit as necessary. Each player also has a target (0.5, 0.7, 1.3, or 1.5), and his payoff is higher, the closer his adjusted guess is to his target times his partner's adjusted guess.[5]

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 target and limits are hidden, but subjects are allowed to search for them through a MouseLab computer interface, game by game, as often as desired but one at a time.[6] Low search costs make the entire structure effectively public knowledge, to justify comparing subjects' behavior with predictions based on complete information. The games are asymmetric (with one exception) and, with complete information, dominance-solvable in 3 to 52 rounds, with essentially unique equilibria determined (not always directly) by players' lower (upper) limits when the product of targets is less (greater) than one.[7]

Studying two-person games allows us to focus sharply on the central game-theoretic problem of predicting the decisions of others who view themselves as a non-negligible part of one's own environment.[8] Tracking behavior within subjects across 16 different games with large strategy spaces greatly enhances separation of the behavioral implications of *Equilibrium* and

[5]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 guess. 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.
[6]Subjects were not allowed to write, and the search data suggest that there was very little memorization. MouseLab was developed to study individual decisions; see 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" to be divided in each period. CGCB used it to study matrix games in which subjects could look up their own and their partners' payoffs.
[7]The equilibria are only essentially unique because all guesses that lead to the same adjusted guess are equivalent.
[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.

leading alternative types. Varying the targets and limits within an intuitive common structure makes it easier for subjects to understand the rules and focus on predicting others' guesses, which reduces the noisiness that is typical of initial responses. It also makes it impossible for subjects to recall the targets and limits from previous plays, and so makes monitoring subjects' searches for hidden information about them a powerful tool for studying cognition more directly.

Our main goals are to use subjects' decisions and information searches, in the light of the cognitive implications of alternative theories of behavior, to better identify the decision rules and mental models that underlie their initial responses; and to learn to what extent monitoring search helps to identify subjects' types and predict their deviations from equilibrium decisions.[9] Other goals include learning more about the relationship between cognition, search, and decisions, and comparing the cognitive difficulty of alternative rules. These will be the focus of a companion paper, which will analyze the search behavior of this study's and CGCB's subjects in more detail.

Our theoretical and econometric framework follows CGCB's.[10] We assume each subject's behavior is determined, with error, by one of a finite set of types, 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 variety of subjects' behaviors but small enough to avoid overfitting.[11]

To test whether any of our subjects have a prior understanding of others' decisions that transcends simple, mechanical decision rules, we add a type, CGCB's *Sophisticated*, to represent the ideal of a person who can predict the distributions of people's initial responses to games with

---

[9]The Baseline treatment just described is supplemented by seven subsidiary treatments (Section 2.A). An Open Boxes ("OB") treatment is identical to the Baseline except that each game's target and limits are continually visible; its purpose is to test whether subjects' guesses are affected by the need to look them up. There are also six Robot/Trained Subjects ("R/TS") treatments, identical to the Baseline except that each subject is trained and rewarded as a specific type: *L1*, *L2*, *L3*, *D1*, *D2*, or *Equilibrium*; their purpose is to evaluate subjects' ability to implement our leading types' guesses, and to provide a benchmark against which to judge Baseline subjects' information searches.
[10]CGCB's and our structural approach to modeling strategic behavior builds on the analyses of Holt (1999; circulated in 1990), SW (circulated in 1993), Harless and Camerer (1995), Nagel (1993, 1995), and Stahl (1996).
[11]An ad hoc type could perfectly mimic a subject's decision history, but this would have no explanatory power. It is hard to dispense with a priori specification because there are multiple rationales for any history, 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.

various structures. In theory *Sophisticated* best responds to the probability distributions of its partners' decisions, but those distributions are part of a behavioral theory of games that is not yet fully developed. We therefore operationalize *Sophisticated* using the best predictions of the distributions now available: the population frequencies of our own subjects' guesses.

A subject's guesses in our 16 games with large strategy spaces often yield a clear strategic "fingerprint," so that her/his type can be read with confidence directly from guesses. Of the 88 subjects in our main (Baseline and OB) treatments, 43 made guesses that comply *exactly* (within 0.5) with one of our type's guesses in 7-16 of the games (Table IX). These compliance levels are far higher than could plausibly occur by chance, given how strongly types' guesses are separated (Tables III-IV) and that guesses could take 200 to 800 different rounded values in each game. The remaining 45 subjects' fingerprints are blurred, at least when viewed through the lens of our types. But for all but 12 of them, violations of simple dominance occurred with frequency below 20%, in games where random guesses would yield a frequency of about 40%. This suggests that most of those subjects' behavior was also coherent, though less well described by our types.

We study all 88 subjects' behavior in more detail by conducting a maximum likelihood error-rate analysis of their guesses, subject by subject. Because our subjects made types' exact guesses so frequently, 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 errors that follow a logistic distribution over the rest of the interval between her/his limits.

Our maximum likelihood type estimates based on guesses alone assign 43 subjects to *L1*, 20 to *L2*, 3 to *L3*, 5 to *D1*, 14 to *Equilibrium*, and 3 to *Sophisticated* (Table IX). To evaluate the reliability of these estimates, given our a priori specification of types, we conduct a subject by subject specification test that compares the likelihood of our estimated type with those of estimates based on 88 *pseudotypes*, each constructed from one subject's guesses in the 16 games. Such comparisons help to detect whether any subjects' guesses would be better explained by alternative decision rules omitted from our specification, and sometimes help to identify omitted rules. They also give an indication of whether subjects' estimated types are credible explanations of their guesses or artifacts of overfitting via accidental correlations with irrelevant types.

Our specification test and additional specification analysis generally support our a priori specification of types. After testing, 58 of our 88 subjects appear to be reliably identified from guesses alone: 27 as *L1*, 17 as *L2*, 11 as *Equilibrium*, and one each as *L3*, *D1*, or *Sophisticated*.

Our analysis of search calls 7 of these subjects' type identifications into question (one *L1*, which we ultimately affirm; 4 *L2*s; one *D1*; and one *Equilibrium*). But because our types specify precise guesses in large strategy spaces, 52 subjects' identifications show that they had accurate models of the games and acted as rational, self-interested expected-payoff maximizers. Our analysis also shows that the deviations from equilibrium of the 42 of those subjects whose types are reliably identified as other than *Equilibrium* can be confidently attributed to non-equilibrium beliefs based on simplified models of others, rather than to altruism, spite, confusion, or irrationality.[12]

Information search adds another dimension to our analysis. Following CGCB, we link search to guesses by taking a procedural view of decision-making, in which a subject's type determines her/his search and guess, possibly with error. Each of our types is naturally associated with algorithms that process information about payoff parameters 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 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 her/his searches, without even considering guesses (Table XI). But subjects' searches are noisy and highly heterogeneous, and most of them less clearly identify a type. To extract the information from searches, we generalize our error-rate analysis to re-estimate all 71 Baseline subjects' types, using search compliance as well as guesses.

Type estimates based on search only, or on guesses and information search, generally reaffirm estimates based on guesses alone. In addition to raising doubts about the 7 subjects whose types appeared to be reliably identified from guesses mentioned above, search helps to identify the probable types of 7 subjects for whom the evidence from guesses is unclear. The end result is that 27 subjects are reliably identified as *L1* and another 2 are probable *L1*s; 13 are reliably identified as *L2*; 10 are reliably identified as *Equilibrium*; and one each is reliably identified as *L3* or *Sophisticated*.[13] Overall, 52 of our 88 subjects can be reliably identified based

[12]Compare Weibull's (2004) argument that rejections of equilibrium in game experiments that do not independently measure preferences are "usually premature".

[13]The last 2 subjects are from our searchless OB treatment. Other subjects' low levels of compliance with *Sophisticated*'s search requirements suggest that the very last identification might not survive monitoring search.

on guesses and search, and several more can be probably identified. Of those reliably identified, 42—nearly half our sample—are identified as types other than *Equilibrium*. These results are a powerful affirmation of subjects' rationality and ability to comprehend complex games and reason about others' responses to them, but they are also a challenge to the use of equilibrium as a universal model of initial responses to games. The overwhelming predominance of *Lk* types once they are adequately separated from *Dk* and alternative types is particularly intriguing, given the leading role played by iterated best responses in informal analyses of strategic behavior.

The rest of the paper is organized as follows. Section 2 describes our design. Section 3 derives our types' implications for guesses and information search. Section 4 reports preliminary statistical tests and summarizes the results of R/TS treatments and Baseline subjects' compliance with iterated dominance and equilibrium. It then uses guesses alone to estimate the types that best describe Baseline and OB subjects' behavior, and discusses our specification test. Finally, it uses both guesses and search to re-estimate Baseline subjects' types. Section 5 is the conclusion.

2. 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 the noisiness of their responses; but they also make it difficult to disentangle learning from sophistication, because even unsophisticated learning often converges to equilibrium in the stage game. Our design studies sophistication in its purest form by eliciting subjects' initial responses to 16 different two-person guessing games, with new partners each period and no feedback to suppress repeated-game effects, experimentation, and learning. This section describes our design, first its overall structure, then the games, and finally 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 disqualified other subjects who revealed that they had formally studied game theory or previously participated in game experiments.[14]

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*).[15] 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 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.[16] 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 her/his own and her/his partners' guesses and her/his point earnings in each game. S/he then drew five game numbers randomly and was paid \$0.04 per point for her/his payoffs in those games.[17] With possible payoffs of 0 to 300

---

[14]We also allowed approximately 4 non-faculty university community members, and a few other subjects who had been briefly exposed to game theory in undergraduate courses, Oscar-winning movies, etc.

[15]The instructions are in Appendix A and our pilot experiments and how they influenced the design are described in Appendix B (http://weber.ucsd.edu/~vcrawfor/#Guess). Mixed R/TS treatments are theoretically acceptable because R/TS subjects did not interact with one another. The data exclude one *L1* subject in R/TS1, because her/his guesses revealed clearly that s/he had copied from a nearby *L2* subject. (R/TS subjects were not told whether the treatment was the same for all subjects in a session, but s/he assumed this. Comparing the guesses of all neighboring subjects in all treatments suggests that s/he was the only cheater.)

[16]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.

[17]It is theoretically possible to control subjects' risk preferences using Roth and Malouf's (1979) *binary lottery* procedure, in which a subject's payoff determines his probability of winning a given monetary prize. We avoided the

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. 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 all subjects remaining had passed.[18] 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.[19] 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 4.A), suggesting that subjects' decisions are not seriously distorted 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 her/his assigned

---

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.

[18]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.

[19]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.

type's guesses via programmed instruction on her/his screen and handout.[20] S/he was rewarded for game payoffs as in the Baseline, except that s/he was paired not with other subjects but with a *robot* (framed as "the computer") that followed her/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*.[21] 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.[22] As in the Baseline, all aspects of this structure were publicly announced, except the games' targets and limits.

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.[23]

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 3). 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 4.B).

---

[20] *Equilibrium* subjects, for instance, were taught each of the three main ways to identify their equilibrium guesses: by direct checking for pure-strategy equilibrium, by best-response dynamics, and by iterated dominance.
[21] We used realizations of random robot guesses rather than their means to minimize differences from the Baseline.
[22] 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 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.
[23] The encouragement is implicit in the wording, and does not use the term certainty-equivalence (Appendix A).

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; guesses outside them 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 $x^i\ \varepsilon\ [a^i,b^i]$, $y^i \equiv a^i$ if $x^i < a^i$, or $y^i \equiv b^i$ if $x^i > b^i$, 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) $$s^i \equiv \max\{0,200 - e^i\} + \max\{0,100 - e^i/10\}$$
$$\equiv \max\{0,200 - |R(a^i,b^i;x^i) - p^iR(a^j,b^j;x^j)|\} + \max\{0,100 - |R(a^i,b^i;x^i) - p^iR(a^j,b^j;x^j)|/10\}.$$

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 $e^i = 0$.[24] 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.

The 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' responses 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 most other designs, where *L1*'s decisions almost inevitably depend only on own payoff parameters, *L1*'s search implications are sharply separated both from

[24] It is *not* concave in player *i*'s guess because the weight on e^i in the second term is algebraically larger than in the first; 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.

our other types' and from those of a solipsistic type that assumes 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 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 adjusted guesses are players' unique equilibrium adjusted guesses.[25]

We assume throughout 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 simple 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, however, a player's risk preferences are potentially relevant.[26] We deal with this problem as follows. Observation 1 below shows that our games have essentially unique equilibria in pure strategies, so that risk preferences do not affect *Equilibrium* guesses. Observation 2 shows that best responses to uniform beliefs are certainty-equivalent, so that risk preferences do not affect *L1*, *D1*, or *D2* guesses, or *L2* or *L3* guesses, which are defined as best responses to *L1* or *L2* guesses. *Sophisticated* guesses, however, are normally best responses to non-uniform beliefs, and so are not covered by Observation 2. In characterizing them we assume that players are risk-neutral, and thus maximize their expected point payoffs, game by game. Each of our types then maximizes its expected point payoffs, game by game, given some beliefs; and each implies an essentially unique, pure guess in each of our games, except *Sophisticated*, for which this is generically true.

We now establish two simple observations that are important in interpreting our results. To avoid trivialities, we assume that all limits and targets are strictly positive, as in our design.

---

[25]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.
[26]Recall that our games do not have the binary lottery, winner-take-all structure of Nagel's and HCW's games.

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 (but 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.[27] This is true without regard to risk preferences or dominance-solvability, although not all games in this class are dominance-solvable because for extreme parameter values there is no dominance. The 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 parameters, all but ignore it.

Observation 2 establishes the certainty-equivalence property referred to above:

Observation 2: Suppose that 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 $R(c,d; p(a+b)/2)$.

[27]In game γ2β4 (Table II), for instance, players' targets are 0.7 and 1.5, whose product is 1.05 so the equilibrium is determined by players' upper limits. The γ2 player's equilibrium guess is at his upper limit of 500, but the β4 player's equilibrium guess is at 750, below his upper limit of 900. Moving some equilibrium decisions away from the

Proof: We show that $x^* = p(a+b)/2$ solves $\max_x \int_a^b u(s(x-pz))dz$ (ignoring the positive factor $[1/(b-a)]$). The integral in the maximand is differentiable because $u(s(x - pz))$ is continuous. Its derivative with respect to x, evaluated at x^*, is (ignoring points of nondifferentiability)

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

where the equality holds for $x^* = p(a+b)/2$ by symmetry. Because $u(\cdot)$ is increasing and $s(\cdot)$ is weakly decreasing in $|x - pz|$, raising x above x^* lowers the derivative below 0, and lowering x below x^* raises it above 0; thus, the integral in the maximand is quasiconcave in x. Because $x^* = p(a+b)/2$ satisfies the first-order condition for maximizing the integral, x* is optimal ignoring the constraint $x \varepsilon [c,d]$ 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, and 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, *L2* and *L3*, equal the player in question'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 crucially on symmetry of the payoff function and uniformity of 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.[28] 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.

---

boundaries in this way allows clearer inferences than when equilibrium is always at the boundary.

[28]The possible values of the targets and limits were *not* publicly announced, to strengthen subjects' incentives to look up the ones they thought relevant to their guesses. Free access still makes the structures public knowledge.

Table II summarizes our 16 games, ordered to emphasize structural relationships. It also lists the 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 (always finite in our design); 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).

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); this 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 types by making their predicted guesses more subtle. They also play an important role in our specification test (Section 4.E), where, together with our large strategy spaces, they sometimes allow us to distinguish cognitive errors from intentional behavior 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, to estimate the strength of their incentives to make their 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 then 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*.[29]

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 s/he could open another box or enter her/his guess, s/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.[30] A subject could enter and confirm his guess by moving the cursor into the box labeled "Keyboard Input," clicking, typing the guess, and then moving the cursor into the box at the bottom of the screen and clicking. A subject could move on to the next game only after confirming her/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,

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

and adjacent lower-and-upper-limit pairs that are characteristic of *L1*, *L2*, and other leading theories (Section 3), less likely to occur for reasons unrelated to cognition.

3. 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 (Section 2.B).

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.[31]

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.[32] 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. 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.

An *L1* player *i*, for instance, best responds to the belief that player *j*'s guess is uniformly distributed between his limits. This yields a guess for *j* that is never adjusted, and averages $[a^j+b^j]/2$. By Observation 2, *L1*'s ideal guess is $p^i[a^j+b^j]/2$, which will be automatically adjusted, if necessary, to $R(a^i,b^i; p^i[a^j+b^j]/2) \equiv \min\{b^i, \max\{a^i, p^i[a^j+b^j]/2\}\}$. An *L1* player *i* therefore has relevant look-up sequence: $\{[a^j, b^j]$ (to compute *j*'s average guess), p^i (to identify *i*'s ideal guess)$]$ $\equiv \{[4, 6], 2\}$. Thus, the look-ups $a^j \equiv 4$ and $b^j \equiv 6$ associated with the basic operation $[a^j+b^j]/2$ must appear adjacently at least once, in any order; and the look-up $p^i \equiv 2$ for the operation $p^i[a^j+b^j]/2$ must appear at least once, possibly separated from $[a^j, b^j] \equiv [4, 6]$, and in any order.

An *L2* player *i* best responds to the belief that player *j* is *L1*, taking the adjustment of *j*'s guess into account. An *L1* player *j*'s adjusted guess is $R(a^j,b^j; p^j[a^i+b^i]/2)$, so an *L2* player *i*'s ideal guess is $p^iR(a^j,b^j; p^j[a^i+b^i]/2)$, which will be adjusted to $R(a^i,b^i; p^iR(a^j,b^j; p^j[a^i+b^i]/2))$. An *L2* player *i* therefore has relevant look-up sequence: $\{([a^i, b^i], p^j)$ (to predict *j*'s *L1* ideal guess), a^j, b^j (to predict *j*'s *L1* adjusted guess), p^i (to identify *i*'s ideal guess)$] = \{([1, 3], 5), 4, 6, 2\}$.[33]

[31]These estimates are then rounded to the nearest integer for simplicity.
[32]These assumptions adapt CGCB's Occurrence and Adjacency assumptions to the current design. We stress that their motivation is empirical: In theory a subject could scan the parameters in any order and use memory to perform his type's operations, making the order of look-ups useless in inferring cognition. Real subjects seldom do that.
[33]With automatic adjustment, an *L2* player *i* doesn't need to know his own limits to play the game or to think about the effects of his own guess being adjusted, but he does need to know them to predict *j*'s *L1* guess. By contrast, an *L1* player *i* does not need to know his own limits, only *j*'s. Because the possible values of the limits are not public

An *L3* player i best responds to the belief that player j is *L2*, taking the adjustment of j's guess into account. An *L2* player j's adjusted guess is $R(a^j,b^j; p^jR(a^i,b^i; p^i[a^j+b^j]/2))$, so an *L3* player i's ideal guess is $p^iR(a^j,b^j; p^jR(a^i,b^i; p^i[a^j+b^j]/2))$, which will be adjusted to $R(a^i,b^i; p^iR(a^j,b^j; p^jR(a^i,b^i; p^i[a^j+b^j]/2))$. An *L3* player i therefore has relevant look-up sequence: $\{((([a^j, b^j], p^i), a^i, b^i, p^j)$ (to predict j's *L2* ideal guess), a^j, b^j (to predict j's *L2* adjusted guess), p^i (to identify i's ideal guess)$\}$ = {((([4, 6], 2), 1, 3, 5), 4, 6, 2}. For minimal restrictions, with order within curly brackets unrestricted, this simplifies to $\{(([a^j, b^j], p^i), a^i, b^i, p^j\}$ = {(([4, 6], 2), 1, 3, 5}.

A *D1* player i deletes one round of dominated guesses for player j and then best responds to uniform beliefs over j's remaining guesses. The first round of dominance reduces j's guesses to the interval $[\max\{a^j, p^ja^i\}, \min\{p^jb^i, b^j\}]$. Thus, a *D1* player i's ideal guess is $p^i(\max\{a^j, p^ja^i\} + \min\{p^jb^i, b^j\})/2$, which will be adjusted to $R(a^i,b^i; p^i(\max\{a^j, p^ja^i\} + \min\{p^jb^i, b^j\})/2))$. A *D1* player i therefore has relevant look-up sequence: $\{(a^j, [p^j, a^i]), (b^j, [p^j, b^i])$ (to delete j's dominated guesses), p^i (to identify i's ideal guess)] = {(4, [5, 1]), (6, [5, 3]), 2}.[34] *D1*'s look-up implications differ somewhat from *L2*'s, although both respond similarly to iterated dominance.

A *D2* player i deletes two rounds of dominated guesses for player j and best responds to uniform beliefs over j's remaining guesses. The first round reduces i's guesses to the interval $[\max\{a^i, p^ia^j\}, \min\{p^ib^j, b^i\}]$ and j's guesses to $[\max\{a^j, p^ja^i\}, \min\{p^jb^i, b^j\}]$. The second round further reduces j's guesses to $[\max\{\max\{a^j, p^ja^i\}, p^j\max\{a^i, p^ia^j\}\}, \min\{p^j\min\{p^ib^j, b^i\}, \min\{p^jb^i, b^j\}\}]$. A *D2* player i's ideal guess is therefore $p^i[\max\{\max\{a^j, p^ja^i\}, p^j\max\{a^i, p^ia^j\}\} + \min\{p^j\min\{p^ib^j, b^i\}, \min\{p^jb^i, b^j\}\}]/2$, which will be adjusted to $R[a^i,b^i; p^i[\max\{\max\{a^j, p^ja^i\}, p^j\max\{a^i, p^ia^j\}\} + \min\{p^j\min\{p^ib^j, b^i\}, \min\{p^jb^i, b^j\}\}]/2]$. A *D2* player i therefore has relevant look-up sequence: $\{(a^i, [p^i, a^j]), (b^i, [p^i, b^j])$ (to delete i's first-round dominated guesses), $(a^j, [p^j, a^i]), (b^j, [p^j, b^i]), p^j$ (to delete j's first- and second-round dominated guesses), p^i (to identify i's ideal guess)$\}$ ≡ {(1, [2, 4]), (3, [2, 6]), (4, [5, 1]), (6, [5, 3]), 5, 2}.

An *Equilibrium* player i's ideal guess is p^ia^j if $p^ip^j < 1$ or p^ib^j if $p^ip^j > 1$. This guess can be identified by evaluating this formula (assuming that logical implications of things that are public knowledge are also public knowledge), or by equilibrium-checking, best-response dynamics, or

knowledge, an *L2* player i cannot infer that adjustment of player j's ideal guess can occur only at his upper (lower) limit when $p^j > 1$ ($p^j < 1$). An *L2* subject who incorrectly infers this may omit $a^j = 4$ ($b^j = 6$) when $p^j > 1$ ($p^j < 1$).

[34]With automatic adjustment, a *D1* player i needs to know his limits only to delete player j's dominated guesses, and need not otherwise consider the adjustment of j's guess. Also, when $p^j > 1$ (< 1), dominance for j usually occurs only near his lower (upper) limit (Table II). A *D1* subject who incorrectly assumes that this is true in all of our games may omit $(b_j, [p^j, b^i])$ = (6, [5, 3]) when $p^j > 1$ or $(a_j, [p^j, a^i])$ = (4, [5, 1]) when $p^j < 1$.

iterated dominance. Evaluating the formula yields the look-up sequence: $\{[p^i,\ p^j]$ (to check if the product of the targets is greater or less than one), a^j (to identify i's ideal guess when $p^ip^j < 1)\}$ = {[2, 5], 4} if $p^ip^j < 1$; or $\{[p^i, p^j], b^j$ (to identify i's ideal guess when $p^ip^j > 1)]$ = {[2, 5], 6} if $p^ip^j > 1$. In principle, minimal restrictions for equilibrium-checking should allow for the possibility that a subject correctly conjectures whether his equilibrium guess is determined by players' lower or upper limits. If so, he need only identify his ideal guess and verify that it is consistent with equilibrium. This requires the look-up sequence: $\{p^i, a^j$ (to identify i's ideal guess, given his conjecture), p^j (to verify that it is consistent with equilibrium for player j)} if $p^ip^j < 1$ or $\{p^i, b^j$ (to identify i's ideal guess, given his conjecture), p^j (to verify that it is consistent with equilibrium for player j)} if $p^ip^j > 1$. These restrictions, $[p^i, a^j, p^j]$ = [2, 5, 4] if $p^ip^j < 1$ or $[p^i, b^j, p^j]$ = [2, 5, 6] if $p^ip^j > 1$, are less stringent than those for any other method of identifying equilibria. However, a subject cannot be sure of her/his conjecture without checking whether $p^ip^j > 1$, and it seems more realistic to add the [2, 5] order requirement, as needed for evaluating the formula.[35]

Finally, although *we* define *Sophisticated*'s ideal guess as its best response to the distribution of its potential partners' guess, as estimated from our subjects' guesses, *Sophisticated* must deduce its beliefs from the structure of the game. We assume that this requires identifying both the game's equilibrium and the other player's guesses that survive two rounds of dominance for the other player and one for the subject. Because the search requirements for these are a subset of *D2*'s, we take *Sophisticated*'s relevant look-up sequence to be the same as *D2*'s.[36]

[35]Even two rounds of iterated dominance implies more stringent restrictions; and best-response dynamics requires still more, $\{[a^i, b^i], [a^j, b^j], p^i, p^j\} \equiv$ {[1, 3], [4, 6], 2, 5} just to identify a starting profile of guesses within the limits. Without the [2, 5] order requirement, *Equilibrium*'s search implications are the only ones with no order restrictions, which makes them much easier to satisfy than other types' and obscures the implications of the search data, leading our econometric model to spuriously identify some subjects' searches as *Equilibrium* even though by inspection they are obviously more consistent with other types. Even with the order requirement *Equilibrium*'s search implications are as simple as *L1*'s and simpler than other boundedly rational types', unlike in CGCB's and CJ's designs. Our 17 (of 29) highly successful R/TS *Equilibrium* subjects (those with 15-16 exact equilibrium guesses) violated *Equilibrium*'s search implications with the order requirement 6% of the time, only 3 of them in more than 1 game.
[36]We stop at two rounds of dominance for the other player and one for the player himself because in previous work few subjects have responded to dominance beyond these levels. Because requiring more rounds, for either player, would make the search requirements for *Sophisticated* more stringent, and few subjects comply with them even as defined here, this assumption strengthens our ultimate conclusion that none of our subjects are *Sophisticated*.

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

This section presents a statistical and econometric analysis of subjects' guesses and information searches.[37] Section 4.A reports preliminary statistical tests. Section 4.B summarizes the aggregate compliance of R/TS subjects' adjusted guesses with the implications of their assigned types, and Section 4.C summarizes the aggregate compliance of Baseline and OB subjects' adjusted guesses with iterated dominance and equilibrium. Section 4.D presents a maximum likelihood error-rate analysis of Baseline and OB subjects' guesses, estimating the types that best describe their guesses in the 16 games and the associated error distributions. Section 4.E discusses our specification test and analysis, and Section 4.F introduces our analysis of information search by describing R/TS and Baseline subjects' compliance with types' search implications. Section 4.F generalizes Section 4.D's analysis to use Baseline subjects' search compliance, with their guesses, to estimate their types and the associated error distributions.

A. Preliminary statistical tests

In this section we report tests for differences in subjects' adjusted guesses across the sessions of the Baseline treatment and the OB 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 pooled Baseline sessions with the OB session yields one *p*-value less than 5% in 16 tests.[38] 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.[39]

[37] Appendix C gives the complete data on guesses and the order, but not duration, data on look-up sequences. Figures 2A-2P (also at http://weber.ucsd.edu/~vcrawfor/#Guess) graph the frequency distributions of adjusted guesses, game by game. Tables X-XI give selected R/TS and Baseline subjects' look-up sequences in the first 2-3 games.

[38] 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.

[39] Despite our failure to find significant aggregate differences, there are hints that OB subjects made high numbers of exact guesses for our types less often: OB subjects made up 19% of this subject pool, but only 11% of those who made 14-16 exact guesses and 7% of those who made 10-13; they were 30% of those who made 7-9 exact guesses.

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

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

Table VII summarizes the exact compliance (within 0.5) of R/TS subjects' adjusted guesses with their 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*.[40]

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' near-universal failure to make *Equilibrium* or *Dk* guesses is not due mainly to 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.

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. These variations will be studied in more detail in our companion paper.

C. Baseline and OB subjects' aggregate 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

---

Perhaps the fact that our design 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.

[40]This inversion is due, we suspect, to a curious framing effect, in which some *L1* R/TS subjects try to outguess the

guesses, both overall and in the games ordered as in Table II, with random compliance as a benchmark.[41] Aggregate compliance with 0-3 rounds of dominance is similar for Baseline and OB subjects game by game, and 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.[42] 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 compelling. 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.

We analyze the data subject by subject.[43] Recall that in our model, each subject's behavior in all 16 games is determined, possibly with error, by one of the seven types. Index the types $k = 1,\ldots,K$ and the games $g = 1,\ldots,G$. In game g, denote subject i's lower and upper limits a_g^i and b_g^i, his *un*adjusted and *adjusted* guess x_g^i and $R_g^i(x_g^i) \equiv \min\{b_g^i, \max\{a_g^i, x_g^i\}\}$, and type k's adjusted guess t_g^k. Write $x^i \equiv (x_1^i,\ldots,x_G^i)$ and $R^i(x^i) \equiv (R_1^i(x_1^i),\ldots,R_G^i(x_G^i))$.

computer but *L2* or *L3* subjects do not try to outguess their simulated partners' attempts to outguess the computer.

[41]Appendix D gives the analogous results for our other types. Almost all of the zero compliance rates for iterated dominance are due to logical constraints rather than empirical tendencies. The rates seldom differ for within 0 and within 0.5, but when they do the tables give the latter.

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

[43]CGCB (2001) used an aggregate mixture model that imposed stronger restrictions on subjects' type distributions, and studied cognition at the individual level by using an uninformative prior over the parameters to condition on individual histories. Estimating subject by subject without cross-subject restrictions is better suited to subjects' heterogeneous behavior, more robust to misspecification, and seems more appropriate because we believe cognition is best studied at the individual level. Comparing CGCB's (1998) subject by subject estimates with CGCB's (2001) estimates, however, suggests that the methods yield similar results.

Interpreting patterns of deviations from types' guesses requires an error structure. We assume that, conditional on a subject's type, his errors are independent across games. We use a spike-logit specification in which, in each game, a subject has a given probability of making his type's guess exactly and otherwise makes errors 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 conditional density $d_g^k(R_g^i(x_g^i),\lambda)$ with *precision* λ.[44]

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*.[45] Subject *i*'s expected payoff in game *g* for type *k*'s beliefs can then be written:

(3) $$S_g^k(R_g^i(x_g^i)) \equiv \int_0^{1000} S_g(R_g^i(x_g^i), y) f_g^k(y)dy .^{46}$$

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

[44]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.
[45]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.

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).[47] Because errors are independent across games, the density of a sample with adjusted guesses $R^i(x^i) \equiv (R_1^i(x_1^i), \ldots, R_G^i(x_G^i))$ for a type-*k* subject *i* is:

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

where products with no terms (if n^{ik} = 0 or *G*) are taken to equal 1. Weighting by p^k, summing over *k*, and taking logarithms yields subject *i*'s log-likelihood:

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

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:

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

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

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

[47] The conditional density could be allowed to extend to U_g^{ik}, but our specification is simpler, and almost

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 $16(1-\varepsilon)$, where ε is the error rate). Subjects are ordered by estimated type, in decreasing order of likelihood within type.[48] These 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 (5) 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.[49] The hypothesis $\lambda = 0$ is 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

These maximum likelihood type estimates cannot all be taken at face value. Some of them could be sensitive to our a priori specification of possible types, which might err by omitting relevant types and/or overfitting by including empirically irrelevant ones. We now conduct a specification test that addresses these issues, which we hope will prove useful in other settings.

Subject by subject, the test compares the likelihood of our type estimate with the likelihoods of analogous estimates based on 88 *pseudotypes*, each constructed from one of our

---

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

[48]Table IX's right-hand side reports estimates based on guesses and information search, discussed in Section 4.G.

[49]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.

subject's guesses over the 16 games.[50] Such comparisons help to detect whether any of our subjects' guesses are better explained by an alternative decision rule, omitted from our specification, or whether a subject's estimated type is a credible explanation of his guesses or an artifact of overfitting via accidental correlations with irrelevant 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.[51] 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.[52]

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).[53] There are 5 (non-overlapping) subsets in which

---

[50]We are grateful to Jerry Hausman for suggesting the idea of this test. We allow spike-logit errors for the pseudotypes, as for the estimated types, 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 the pseudotypes do not come with "built-in" models of other players. We do this in the simplest possible way, by assuming that the pseudotypes' guesses are best responses and inferring point beliefs, game by game, from their subjects' guesses. For a dominated guess, which is not a best response to any beliefs, or for a guess at a limit that is a best response to multiple beliefs, we extend this definition by inferring the beliefs that bring the guess closest to maximizing payoff.
[51]Requiring only higher (rather than significantly higher) likelihood in (i) prevents us from ruling out cluster candidates because their pseudotypes offer only slight improvements in fit; we find below that few of the comparisons are close. The "sufficiently similar" in (ii) could be made more precise, but we have found it more informative to consider possible clusters on a case by case basis. Finally, although logic of our definition allows overlapping but non-nested clusters, in our analysis this problem does not arise.
[52]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 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.
[53]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

subjects' guesses also appear close enough to warrant checking part (ii) of the definition. The subjects in these subsets are identified in the left-hand side of Table IX by superscript letters 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 along with the games' parameters and types' guesses in a way that facilitates the analysis below.

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 (Section 3) work equally well in games with and without mixed targets.[54] We judge 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 provisionally accept 310's identification as *Equilibrium*, which fits her/his guesses significantly better than 202's and 417's pseudotypes do, despite their 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

---

because her/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.

[54]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, s/he says, "I usually assumed my partner chose from fairly near the center of his/her 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 s/he adjusted her/his beliefs upward (downward) when her/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 s/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 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, illustrating the risks of taking subjects' questionnaire statements at face value. From now on we refer to questionnaire responses only when they are helpful.

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. 6 of the 7 common deviations 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 econometric analysis.[55]

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.[56] The common deviations have no apparent pattern with respect to timing or structures. 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.

[55]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 her/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.
[56]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 her/his questionnaire.

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 3 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.[57] Their guesses are superficially quite similar, but it is plain that they were 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.

With regard to overfitting, we take the position that for a subject's estimated type to be a credible explanation of his behavior it should perform at least as well against the pseudotypes as it would, on average, at random.[58] Suppose that a subject's behavior is random relative to our types and all pseudotypes other than his own, in the sense 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 a credible explanation of her/his guesses it should have higher likelihood than all but at most 87/8 ≈ 11 of the pseudotypes. Those subjects whose estimated

---

[57]Subject 204 says s/he first found the person whose "spread" (defined as own target times the difference between the partner's limits) was smaller. If her/his spread was smaller, s/he guessed the average of the range between her/his target times the partner's lower and upper limits; and if the partner's spread was smaller, s/he guessed the average of the analogous partner's range, thus without taking her/his own target into account, which makes no sense decision-theoretically. In fact s/he adjusted the ranges according to the limits; with this adjustment the stated rule explains her/his guesses in 11/16 games. Subject 313 says that s/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/her 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 s/he separately adjusted each term in the above formula to her/his own limits before averaging them (see her/his game 14 guess), which makes no sense decision-theoretically. With adjustment, the stated rule explains her/his guesses in 14/16 games. Subject 409 says that s/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 her/his guesses in 13/16 games.
[58]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.

types have lower likelihoods than 12 or more pseudotypes have type identifiers superscripted + in Table IX; they include most subjects of each estimated type with the lowest likelihoods: 10 of those estimated *L1*, 2 estimated *L2*, and one each estimated *D1*, *Equilibrium*, and *Sophisticated*.

Our type estimates, as modified by the test results, suggest that of the 43 subjects whose type estimate is *L1*, 27 are reliably identified as *L1*. The remaining 16 subjects (5 in clusters C, D, or E and 11 others whose estimated types do not do significantly better than randomness within our specification and/or better than enough pseudotypes) probably have spurious type estimates. Of the 20 subjects whose type estimate is *L2*, 17 are reliably identified. The remaining 3 subjects (one in cluster C and 2 whose estimated types do not do better than enough pseudotypes) probably have spurious estimates. Of the 3 subjects whose type estimate is *L3*, only one appears reliably identified. The other 2 (both in cluster B) probably have spurious estimates. Of the 5 subjects whose type estimate is *D1*, only one appears reliably identified. The other 4 (2 in cluster E and 2 whose estimated types do not do significantly better than randomness and/or better than enough pseudotypes) probably have spurious estimates. Of the 14 subjects whose type estimate is *Equilibrium*, 11 appear to be reliably identified. The remaining 3 subjects (2 in cluster A, omitting potential cluster member 310, and one whose estimated type does neither significantly better than randomness nor better than enough pseudotypes) probably have spurious estimates. Of the 3 subjects whose type estimate is *Sophisticated*, only one appears to be reliably identified. The other 2 subjects (one who does not do significantly better than randomness and one who does not do better than enough pseudotypes) probably have spurious estimates.

Overall, 58 of our 88 subjects appear to be reliably identified from guesses alone: 27 as *L1*, 17 as *L2*, 11 as *Equilibrium*, and one each as *L3*, *D1*, or *Sophisticated*. Our search analysis calls 7 of these subjects' type identifications into question (one *L1*, which we ultimately affirm; 4 *L2*s; one *D1*; and one *Equilibrium*). But because our types specify precise guesses in large strategy spaces, 52 subjects' identifications show clearly that they had accurate models of the games and acted as rational, self-interested expected-payoff maximizers. Our analysis also shows that for at least the 42 of those subjects whose types are reliably identified as other than *Equilibrium*, their deviations from equilibrium can be confidently attributed to non-equilibrium beliefs based on simplified models of others, rather than confusion, altruism, spite, or irrationality.

Despite the differences between our games and those in previous studies, our classification of subjects by type is quite close to Nagel's, HCW's, CGCB's, and SW's. There are two main differences between previous classifications and ours. First, we find more *Equilibrium* subjects (12.5% of all Baseline and OB subjects, focusing on identifications we have argued are reliable) than most previous studies, including Nagel's and HCW's studies of guessing games; the main exceptions are SW's studies of matrix games. Second, SW (1995) used versions of *L2*, *L3*, etc., that depend on others' decision noise, represented by an estimated population parameter; and also included a *Worldly* type that best responds to a mixture of a noisy *L1* and a noiseless *Equilibrium* with estimated weights.[59] Including *Worldly* led SW (1995) to identify many subjects as *Worldly* and correspondingly few as *L2*, by contrast with SW (1994), CGCB, CHC, and this paper.

CGCB (Section 3.A) argued in favor of noiseless definitions of types like those of *L2* and *L3* used here and more generally against types that depend on parameters estimated from the population's behavior, like SW's *Worldly* or CHC's versions of *Lk* types that best respond to mixtures of lower-level *Lk* types. Because subjects do not observe others' behavior, such types implicitly assume that they have prior understandings of it. CGCB argued that the issue of prior understandings is more cleanly addressed, with less risk of overfitting, by including a type like *Sophisticated*, which represents the idea of worldliness without introducing additional parameters, imposing structural restrictions, or raising delicate specification issues.

Our analysis adds evidence to this debate, in that our subjects' large numbers of exact *L2* (or in some cases *L3*) guesses by our noiseless definitions suggest that more complex definitions would not add to the model's ability to explain individual subjects' behavior. Our results are actually inconclusive with respect to CHC's versions of *L2* and *L3*, which in our games, under risk-neutrality, are both behaviorally equivalent to our *L2*.[60] Our results are conclusive with respect to *Worldly*. By an argument like that sketched in footnote 60, in our games a risk-neutral best response to the mixture of *L1* and *Equilibrium* by which SW define *Worldly* will completely ignore *Equilibrium* as long as its estimated frequency is less than 0.5, which is true of all such estimates that have been published. Given this, *Worldly* reduces to a best response to SW's noisy *L1*, which ranges from *L0* to a noiseless *L1* depending on estimated parameters. With our

---

[59]The issue is not whether subjects' own decisions should be 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 ours.
[60]CHC's *L2* best responds to a mixture of *L0* (uniform randomness) and *L1* in the proportions $1{:}\tau$, which for $\tau > (<) 1$ puts more weight on *L1* (*L0*). By a kind of "median-voter" result, our (not-everywhere-differentiable) payoff function makes it optimal to best respond to *L1* alone if $\tau > 1$, or to *L0* alone if $\tau < 1$. CHC argue that $\tau \approx 1.5$ in most applications, in which case their *L2* is confounded with our *L2*. A similar argument shows that CHC's *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 $\tau \approx 1.5$ (because *L1* is still the median type). These behavioral equivalences extend to search implications if one assumes (as

quasiconcave payoff function, such a best response lies between a noiseless *L1* and *L2*—strictly between them except for extreme parameter values that make *Worldly* equivalent to one or the other. 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.[61] By contrast, 45 of our subjects made *exact* guesses for *L1*, *L2*, *L3*, or *Equilibrium* in 7 or more games and they and our other subjects' guesses appear random, relative to *Worldly*'s.

F. R/TS and Baseline subjects' compliance with types' search implications

Table X gives the first two games' information search data for a sample of R/TS subjects chosen from those who played their assigned types' guesses with very high frequencies, but representative with regard to such subjects' search compliance (Appendix C gives the complete data). The table also gives Table VI's search implications and MouseLab box numbers for reference. The subjects' frequencies of making their assigned types' (and when relevant, alternate types') exact guesses are in parentheses after the assigned type.

These R/TS subjects' look-up patterns conform closely to predictions based on our theory of cognition and search, with 2 exceptions: Our *Equilibrium* subjects search longer and in much more complex patterns than our theory suggests. And a number of *D1* subjects made more (sometimes many more) *L2* than *D1* guesses, even after passing our *D1* Understanding Test, in which *L2* answers were incorrect. The table includes one of 7 (out of 30) *D1* subjects who fit this pattern, 804; the others were 802, 809, 1213, 1401, 1509, and 1511). There was also one (out of 19) *D2* subject (1913) who made slightly more *L3* than *D2* guesses, but there was never any such "morphing" from *Lk+1* to *Dk*. These results strongly suggest that *Dk* is less natural than *Lk+1*.[62]

For comparison, Table XI gives the information search data from the first three games for a sample of Baseline subjects chosen, type by type, from those whose guesses most closely fit each type, but representative with regard to such subjects' search compliance (Table IX). The subjects' frequencies of making their apparent types' (and when they exist, alternate types') guesses are in

---

CHC's approach requires) that subjects have prior understandings of the population mixture of lower-level types.

[61]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 like those in clusters A and E: "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." However the ambiguities are resolved, this mechanical rule is inconsistent with *Worldly*. 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 (202, 219, 312, 401, and 416).

[62]Some other R/TS subjects made more correct guesses within 0.5 and/or 25 for a type other than their assigned type (613, who made more *L2* than *L1* guesses; 801, who made more *L1* than *D1* guesses; and 1504, who made more *L1* than *Equilibrium* guesses), but transitions from *D1* to *L2* are by far the most common. This phenomenon is probably an important clue regarding the predominance of *Lk* types in the Baseline treatment, which will be examined in more detail in our companion paper.

parentheses after their types. It is clear that these "naturally occurring" Baseline *L1*, *L2*, and possibly *L3* or *Equilibrium* subjects have look-up sequences very close to those of their R/TS counterparts, suggesting that R/TS subjects' training did not have a large effect on their look-up sequences. Baseline look-up sequences also tend to validate our theory of cognition and search, except that estimated *Equilibrium* Baseline subjects had look-up patterns that, like our *Equilibrium* R/TS subjects', were longer and much more complex than our theory suggests.[63]

Two aspects of the search data, evident in Tables X and XI, are important in our econometric specification. First, many subjects (e.g. Baseline subjects 202 and 210 and R/TS subjects 704 and 904) consistently start with "123456" or some variation of this, and many of these and other subjects end with an optional "13," checking their own limits even when this is not required for their type (e.g. Baseline subjects 101 and 206 and R/TS subjects 1412 and 1607). We make no attempt to filter out these patterns because subjects may use the information they yield, and the choice of how to filter them would involve hidden degrees of freedom.

Second, individual look-up patterns differ widely in look-up style: Many successful R/TS subjects, and many Baseline subjects whose types are evident from their guesses, consistently look first at the relevant sequence needed to identify their type's guess, and then either continue looking at irrelevant sequences or stop and confirm a guess (e.g. Baseline subjects 108, 118, and 206, and R/TS subjects 805, 1807, and 1811; Tables X-XI). A smaller number of such subjects consistently look at irrelevant sequences at first and then at the relevant sequence only near the end (e.g. Baseline subject 413 and R/TS subject 904). Still others return to the relevant sequence repeatedly throughout the sequence (e.g. Baseline subject 101 and R/TS subjects 1607 and 1716). Thus one can identify three distinct styles, "early," "late," and "often"; but the data suggest that "often" subjects are almost always either well described as "early" or "late".[64]

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

In this section we generalize Section 4.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 4.D's model, avoiding unnecessary differences in the treatment of guesses

[63]Note however that Baseline subject 309, who made exact *L2* guesses in 16 games, made enough look-ups to be sure of identifying them only in games 6-16; he made search errors in games 1-5, including the 3 shown here. R/TS *L2* subject 2008, who also made exact *L2* guesses in 16 games, and several other R/TS subjects whose compliance with their type's assigned guesses was very high, made similar search errors.

[64]Table X gives subjective style estimates for the R/TS subjects mentioned in the text, and Table XI gives style estimates for Baseline subjects from Section 4.G's econometric analysis of search (Table IX).

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 main issue that arises in extending our analysis to search is measuring the extent to which subjects' look-up sequences comply with types' search implications (Section 3). We define compliance with a type's search implications as the density of the type's relevant look-ups in the look-up sequence (Section 3, Table VI). But because our subjects vary widely in where the relevant look-ups tend to be located in their sequences, we filter out some idiosyncratic noise using a binary nuisance parameter called *style*. Style is assumed constant across games, like type, and modifies type in a way that affects only its search implications.

Specifically, we assume that each subject has either style $s = e$ for "early" or $s = l$ for "late" (Section 4.F). For a given game g, subject i, type k, and style s, we define search compliance as the density of relevant look-ups in the part of the sequence identified by the style. If $s = e$, we start at the beginning of the sequence for the game and continue until we obtain a complete relevant sequence for the type as characterized in Table VI. If we never obtain such a sequence, compliance is 0. Otherwise compliance is the ratio of the length of the type's relevant sequence to the number of look-ups that yields the first complete relevant sequence. If, for instance, the type's relevant sequence has length six, and the first complete sequence is obtained in eight look-ups (with two optional or redundant look-ups interspersed), 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.[65]

To reduce the need for structural restrictions, we discretize search compliance as follows.[66] 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 below are taken over the values H, M, and L.

[65]The compliance data are in Appendix G. For *D1*, *D2*, and *Sophisticated* we take the length of the relevant sequence to be 6, the minimum length with which it is possible to satisfy their requirements, for example 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}. Due to programming constraints we treat multiple sequential lookups as a single lookup; this makes little difference. Given Table VI's characterization of types' search requirements, our definition of compliance refines CGCB's Occurrence and Adjacency in a way that is appropriate for our search data.

[66]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 (∏).

Recall that 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 in each game, and that each is independent 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.[67] 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 (assumed independent of type and style) 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 4.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 $R^i(x^i) \equiv (R_1^i(x_1^i),\ldots,R_G^i(x_G^i))$ for a subject i of type k and style s is:

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

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

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

[67]A natural generalization would allow search and guess errors to be correlated for a given game and subject, while remaining i.i.d. across games and subjects, as in CGCB. In our specification, this would amount to allowing compliance-contingent error rates and precisions. We dispense with this refinement for simplicity.

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 s/he has compliance c for that k and s. The maximum likelihood estimate of λ is again the logit precision, restricted to guesses in V_g^{ik}.

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

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

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

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

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

where $\ln d^k(R^i(x^i), \varepsilon, \lambda))$ is the log-likelihood of the guesses-only model defined in (7). Thus search adds an additively separable term in search compliance, minus an additional $G \ln G$ term. As in Section 4.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 2 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 4.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 types for which a subject has 0 (not just *L*) compliance in 8 or more games a priori.[68]

The right-hand side of Table IX reports maximum likelihood estimates of each subject's type and style, error rate, precision, and rates of search compliance, first based on search only and then based 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 estimates.

Most subjects' style estimates are early but there is a sizeable minority of late estimates, suggesting that our characterization of search compliance would distort the implications of some subjects' searches without the style parameter. Most subjects' type estimates based on search only, or on guesses and information search, reaffirm the estimates based on guesses, including 51 of the 58 subjects whose types we described as reliably identified from their guesses.[69] Other subjects' type estimates change when search is taken into account, for one of two reasons.

For some subjects there is a tension between guesses-only and search-only type estimates that is resolved in favor of a type other than the guesses-only estimate.[70] These include 105, 113,

[68]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.
[69]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.
[70]This happens despite the fact that with random behavior the guess part of the log-likelihood is nearly 6 times larger than the search part, and so imperfect compliance has considerably more weight in determining the estimates based on guesses and search combined. The difference in weights may seem counterintuitive, because our model avoids unnecessary differences in the treatment of guesses and search. It arises because our theory of behavior makes much sharper predictions about guesses than about search, which as implemented in the estimation are less likely to be satisfied by chance. We could try to put search on a more equal footing by making its predictions sharper, for example by requiring more precise levels of search compliance within a finer categorization, except for errors. But with the heterogeneity and noisiness of our subjects' searches, they would then satisfy types' search implications

and 420, estimated as noisy *L1* based on guesses but as *Equilibrium* or *L3* based on search or guesses and search; 205, 306, 403, and 414, estimated as *L2* based on guesses but as *L1* or *Equilibrium* based on search or guesses and search (though for 414 the guesses and search estimate differs from the search-only estimate, *L1*); 302, estimated as *L3* based on guesses but as *Equilibrium* based on search or guesses and search; and 312 and 313, estimated as *D1* (312 noisy) based on guesses but as *L1* or *L2* based on search or guesses and search.

For other subjects the type estimate based on guesses has 0 search compliance in 8 or more games, and is therefore ruled out by our a priori constraint. These include 115, 204, and 401, estimated as *D1* (401 very noisy) based on guesses but as *Equilibrium* or *L1* based on search or guesses and search; 112, estimated as *Equilibrium* based on guesses but as *L2* based on search or guesses and search; and 304 and 421, estimated as *Sophisticated* based on guesses but as *Equilibrium* or *L1* based on search or guesses and search. This category also includes subject 415, estimated as *L1* based on 9 exact *L1* guesses, but as *D1* based on search or guesses and search. 415 has 9 games with 0 *L1* search compliance due to no adjacent [4,6]'s; but her/his look-up sequences are rich in 4,2,6's and 6,2,4's and across games, 0 search compliance is very weakly correlated with exact *L1* guesses. We therefore count 415 as a reliably identified *L1* subject who violated our assumption that basic operations are represented by adjacent look-ups (Section 3).

Overall, taking search into account, 52 subjects are reliably identified: 27 as *L1*, 13 as *L2*, 10 as *Equilibrium*, and one each as *L3* or *Sophisticated*. In addition, several more subjects can now be probably identified, as *L1*, *L2*, *L3*, or *Equilibrium*. Thus, the search analysis reinforces our conclusion concerning the predominance of *Lk* types over all other types but *Equilibrium*.

5. Conclusion

This paper has reported experiments that elicit subjects' initial responses to 16 dominance-solvable two-person guessing games, monitoring their searches for hidden but freely accessible payoff information along with their guesses. The design yields strong separation of the guesses and information searches implied by leading decision rules. Of our 88 Baseline and OB subjects, 52 can be reliably identified as one of our types based on guesses and search. Because our types specify precise guesses in large strategy spaces, the identifications show that those subjects had

only rarely, and the stronger restrictions would cause severe specification bias. We have chosen to rely on weaker assumptions about how type determines search than about how it determines guesses, at the cost of extracting less information from subjects' searches than might be possible if we could confidently impose stronger restrictions.

accurate models of the games and acted as rational, self-interested expected-payoff maximizers. 42 of those subjects—nearly half our sample—are reliably identified as types other than *Equilibrium*, so that given the specifications of our types, their systematic deviations from equilibrium can be attributed to non-equilibrium beliefs based on simplified mental models of others, rather than altruism, spite, confusion, or irrationality.

Among our non-*Equilibrium* subjects, *Lk* types are overwhelmingly predominant. This, and the evidence from our R/TS treatments that *Lk* types are more natural than *Dk* and other types, lends support to the leading role given iterated best responses in informal analyses of strategic behavior. Together with the evidence presented in previous work on this topic, which suggests that the underlying behavioral principles will be descriptive of initial responses to a wide range of games, our results suggest that a structural model of initial responses that combines low-level *Lk* types with *Equilibrium* in the right proportions will reliably out-predict equilibrium.

We close by noting that although our results directly concern only initial responses, they also suggest conclusions about the structure of learning rules. In particular, our subjects' comprehension of the games and their strong tendency to choose exact best responses to the beliefs implied by simplified mental models of others point clearly away from reinforcement learning and toward beliefs-based models such as weighted fictitious play or hybrids like Camerer and Ho's (1999) experience-weighted attraction learning. In future experiments we plan to use information search to discriminate among alternative theories of learning, whose search implications are often far 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, ???-???.

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, Vincent Crawford, and Bruno Broseta (1998): "Cognition and Behavior in Normal-Form Games: an Experimental Study," Discussion Paper 98-22, University of California, San Diego (http://weber.ucsd.edu/~vcrawfor/#CogBeh).

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.

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.

Roth, Alvin and Michael Malouf (1979): "Game-Theoretic Models and the Role of Information in Bargaining," *Psychological Review*, 86, 574-594.

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," Stockholm School of Economics Working Paper 382, revised 2004.

Figure 1. Screen Shot of the MouseLab Display

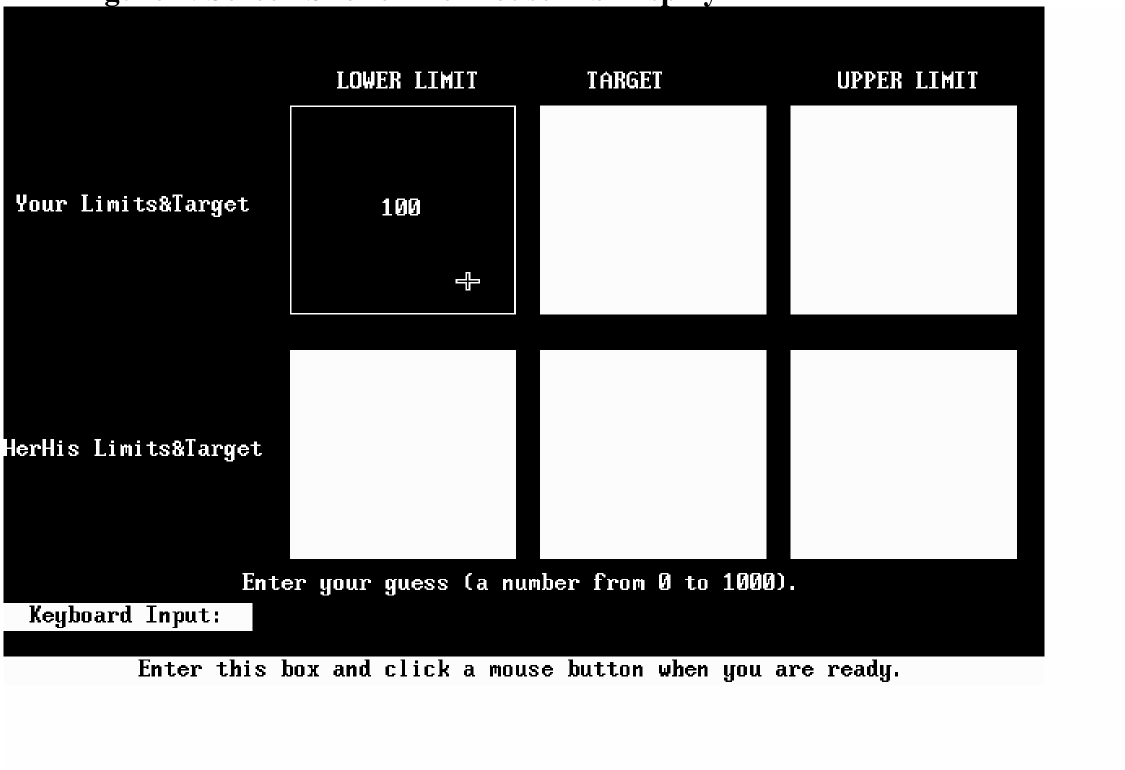

Table I. Overall Structure

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

Notes: Limits are denoted α for [100, 500], β for [100, 900], γ for [300, 500], δ for [300, 900]. Targets are denoted 1 for 0.5, 2 for 0.7, 3 for 1.3, 4 for 1.5. Patterns of dominance are denoted 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 (85)
L2	31.20 (89)	51.81 (100)	31.34 (65)	39.30 (94)	38.68 (90)	31.37 (68)	38.73 (96)
L3	32.99 (94)	35.01 (68)	48.14 (100)	38.70 (93)	41.14 (95)	34.00 (74)	39.34 (97)
D1	33.73 (97)	41.13 (79)	37.56 (78)	41.64 (100)	41.11 (95)	29.42 (64)	39.50 (97)
D2	32.86 (94)	41.56 (80)	40.57 (84)	40.79 (98)	43.13 (100)	32.43 (70)	40.07 (99)
Eq.	30.14 (86)	36.67 (71)	36.09 (75)	35.87 (86)	38.30 (89)	46.05 (100)	35.98 (89)
Soph.	33.04 (95)	41.38 (80)	41.24 (86)	40.77 (98)	41.84 (97)	31.67 (69)	40.53 (100)

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**	***k***	**exact**	**λ**	**ln L**	**k_s**	**ζ_H**	**ζ_M**	**ln L_t**	**ln L_g**	**ln L_s**	**k_s**	**exact**	**λ**	**ζ_H**	**ζ_M**
513	0	0.00	*L1*	16	-	-	-	-	-	-	-	-	-	-	-	-	-
118	0	-9.62	*L1*	15	1.85	-7.41	$L1_e$	0.88	0.06	-17.03	-9.62	-7.41	$L1_e$	15	1.85	0.88	0.06
101	1	-10.27	*L1*	15	0.55	-9.94	$L1_e^{\ddagger}$	0.69	0.31	-20.21	-10.27	-9.94	$L1_e^{\ddagger\ddagger}$	15	0.55	0.69	0.31
104	0	-16.63	*L1*	14	2.20^{*}	-3.74	$L1_e$	0.00	0.94	-20.37	-16.63	-3.74	$L1_e$	14	2.20	0.00	0.94
413	0	-17.81	*L1*	14	0.88	-6.03	$L1_l$	0.13	0.88	-23.84	-17.81	-6.03	$L1_l$	14	0.88	0.13	0.88
207	0	-17.96	*L1*	14	0.42	0.00	$L1_e$	1.00	0.00	-17.96	-17.96	0.00	$L1_e$	14	0.42	1.00	0.00
216	1	-25.41	*L1*	13	1.06	-11.25	$L3_e$	0.75	0.19	-38.69	-25.41	-13.29	$L1_e$	13	1.06	0.31	0.63
402	0	-30.93	*L1*	12	5.65^{*}	-9.00	$L1_e$	0.00	0.75	-39.93	-30.93	-9.00	$L1_e$	12	5.65	0.00	0.75
418	0	-42.23	*L1*	10	21.22^{**}	-7.41	$L2_e$	0.88	0.06	-52.16	-42.23	-9.94	$L1_e$	10	21.22	0.00	0.69
301	1	-45.84	$L1^{D}$	10	0.00	-3.74	$L1_e$	0.06	0.94	-49.58	-45.84	-3.74	$L1_e$	10	0.00	0.06	0.94
508	0	-46.19	$L1^{D}$	10	2.05	-	-	-	-	-	-	-	-	-	-	-	-
308	3	-47.34	*L1*	10	0.00	-9.63	$L3_e$	0.81	0.13	-60.65	-47.34	-13.30	$L1_{el}$	10	0.00	0.19	0.69
102	4	-47.63	*L1*	10	0.00	-9.63	$L2_e$	0.81	0.06	-57.57	-47.63	-9.94	$L1_e$	10	0.00	0.00	0.69
415	1	-53.64	*L1*	9	0.88	-16.38	$D1_e$	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	$L1_l$	8	0.00	0.06	0.94
318	0	-62.61	*L1*	7	3.18^{*}	-3.74	$L1_e^{\ddagger}$	0.00	0.94	-66.36	-62.61	-3.74	$L1_e$	7	3.18	0.00	0.94
512	0	-63.33	*L1*	7	1.56	-	-	-	-	-	-	-	-	-	-	-	-
502	1	-64.55	*L1*	7	1.01	-	-	-	-	-	-	-	-	-	-	-	-
516	1	-64.93	$L1^{C}$	7	1.10^{*}	-	-	-	-	-	-	-	-	-	-	-	-
409	0	-73.59	$L1^{E}$	4	9.90^{**}	-10.59	$L1_l$	0.00	0.38	-84.18	-73.59	-10.59	$L1_l$	4	9.90	0.00	0.38
106	0	-75.82	*L1*	5	1.19	-7.72	Eq_e	0.00	0.19	-85.75	-75.82	-9.94	$L1_l$	5	1.19	0.00	0.31
305	3	-79.89	*L1*	5	0.37	-6.03	$L1_e$	0.88	0.13	-85.92	-79.89	-6.03	$L1_e$	5	0.37	0.88	0.13
411	1	-80.58	*L1*	4	1.45^{**}	0.00	$L3_e$	1.00	0.00	-86.61	-80.58	-6.03	$L1_e$	4	1.45	0.13	0.88
509	1	-81.81	*L1*	4	0.86	-	-	-	-	-	-	-	-	-	-	-	-
203	4	-83.90	*L1*	4	0.00	-9.94	Eq_e	0.00	0.31	-94.49	-83.90	-10.59	$L1_e$	4	0.00	0.00	0.63
505	4	-84.13	*L1*	4	0.43	-	-	-	-	-	-	-	-	-	-	-	-
317	3	-86.58	*L1*	3	0.92^{*}	-3.74	$L1_e$	0.94	0.06	-90.32	-86.58	-3.74	$L1_e$	3	0.92	0.94	0.06
416	1	-86.74	$L1^{\dagger}$	1	4.48^{**}	-3.74	$L1_e^{\ddagger}$	0.00	0.94	-90.48	-86.74	-3.74	$L1_e$	1	4.48	0.00	0.94
217	3	-87.12	*L1*	3	0.68	-10.59	$L1_e$	0.00	0.38	-97.71	-87.12	-10.59	$L1_e$	3	0.68	0.00	0.38

219	3	-87.32	$L1^{+}$	3	0.89^{*}	-7.72	$L1_{e}$	0.00	0.81	-95.04	-87.32	-7.72	$L1_{e}$	3	0.89	0.00	0.81
501	1	-87.93	$L1^{\dagger}$	0	4.38^{**}	-	-	-	-	-	-	-	-	-	-	-	-
410	3	-89.18	$L1$	2	1.53^{**}	-7.72	$L1_{el}^{\ddagger}$	0.00	0.19	-96.90	-89.18	-7.72	$L1_{el}$	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	$L2_{el}$	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	$L2_{e}$	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	$L2_{e}$	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	$L2_{l}$	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	$L2_{e}$	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	$L2_{e}$	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	$L2_{l}$	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	$L2_{l}$	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	$L2_{e}$	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	$L2_{l}$	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	$L2_{e}$	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	$L1_{l}$	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	$L2_{e}$	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	$L1_{e}$	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	$L2_{e}$	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	$L3_e$	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	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	$L1_e^{\ddagger\ddagger}$	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	$L1_e$	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^{A}	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	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	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 type identifier superscripted † means that the subject's estimated type was not significantly better than a random model of guesses ($\lambda = 0$ and $\varepsilon \approx 1$) at the 5% (or 1%) level. A type identifier superscripted + means that the estimated type had lower likelihood than 12 or more pseudotypes in our specification test, more than expected at random. A type identifier superscripted A, B, C, D, or E indicates at least potential membership in a cluster identified in the specification test. An estimated λ superscripted ** (*) means $\lambda = 0$ is rejected at the 1% (5%) level. ln L_t, ln L_g, and ln L_s refer to total, guesses-only, and search-only likelihoods. A type-style identifier subscripted *el* indicates that both styles have equal likelihoods and ζ_c. A search-only type-style identifier superscripted ‡ indicates 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 superscripted ‡‡ indicates alternatives with different ζ_c: $L1_l$ for subjects 101 and 313. No search estimates are reported for subject 109 because s/he had 0 search compliance in 8 or more games for every type.

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

MouseLab box numbers

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

Types' Search Implications

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

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

Notes: The subjects' frequencies of making their assigned types' (and when relevant, alternate types') exact guesses are in parentheses after the assigned type. A * in a subject's look-up sequence means that the subject entered a guess there without immediately confirming it.

Table XI. Selected Baseline Subjects' Information Searches and Estimated Types' Search Implications

MouseLab box numbers

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

Types' Search Implications

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

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

Notes: The subjects' frequencies of making their apparent types' (and when relevant, alternate types') exact guesses are in parentheses after theassigned type. A * in a subject's look-up sequence means that the subject entered a guess there without immediately confirming it.