A. Preliminary Statistical Tests

In this section we test for aggregate differences in subjects' decisions across the two runs of the Baseline treatment, B1 and B2; across the Baseline, OB, and TS treatments; and across player roles in isomorphic games. These tests confirm simplifying restrictions suggested by theory and answer questions that are helpful in evaluating our methods. Because the tests compare categorical data from independent samples with no presumption about how they differ, we use Fisher's exact probability test, conducting the tests separately for each game, pooling the data for all subjects in each player role, and for some purposes pooling the data for subjects with isomorphic player roles in different games.[48] Details can be found in CGC&B, Section 4.A.[49]

The tests reveal no differences in subjects' decisions in the B1 and B2 runs that are significant at the 5% level except in game 4C for Column subjects, well within the limits of chance for 36 comparisons. Accordingly, from now on we pool the data from the Baseline runs. The tests also reveal no differences between Baseline and OB subjects' decisions that are significant at the 5% level except in game 6A for Column subjects, again well within the limits of chance. We therefore pool Baseline and OB data when necessary to obtain adequate sample sizes. As expected, there are noticeable differences between Baseline and TS subjects' decisions in 16/18 games, which are significant at any reasonable level in 4/6 games where the subject had three decisions and at the 5% level in 9 games in total. There are no differences between Row and Column subjects' decisions in isomorphic games that are significant at the 5% level except in games 4B and 4D in the Baseline and 9A and 9B in OB, about what would be expected by chance. We therefore pool the data across isomorphic games when necessary to obtain adequate sample sizes. Because these tests include several pairs of isomorphic games that were widely separated in the sequence (5B and 6B, by 12 games; 5A and 6A, by 7; and 7B and 8B, by 5), and we did not control for decision order and labeling across isomorphic games, they provide some assurance that learning and decision labeling and order had little effect on subjects' decisions.

B. Aggregate Compliance with Dominance, Iterated Dominance, and Equilibrium

We now examine subjects' decisions in the aggregate for compliance with dominance, iterated dominance, and equilibrium in the different kinds of games we study.

[48] These tests have low power because of our small sample sizes. Conducting tests separately for each game is fully justified only if subjects' decisions are statistically independent across games, which is unlikely because some games are related. However, the correct test without independence (comparing decision histories) is impractical.

[49] In the TS treatment, we exclude the 3 out of 15 TS subjects who revealed by their comments or exit questionnaires that they did not try to identify equilibria. CGC&B gives the results for the full TS sample, which are similar.

TABLE II

PERCENTAGES OF DECISIONS THAT COMPLY WITH EQUILIBRIUM BY TYPE OF GAME

Type of Game (rounds of dominance for player to identify own equilibrium decision)	Baseline	OB	B + OB	TS
2 × 2 with dominant decision (1)	85.6%	92.6%	88.2%	100.0%
(2A, 2B for Rows; 3A, 3B for Cols.)	(77/90)	(50/54)	(127/144)	(24/24)
2 × 3 with dominant decision (1)	82.2%	100.0%	88.9%	100.0%
(4D for Rows; 4B for Cols.)	(37/45)	(27/27)	(64/72)	(12/12)
3 × 2 with dominant decision (1)	86.7%	92.6%	88.9%	100.0%
(4C for Rows; 4A for Cols.)	(39/45)	(25/27)	(64/72)	(12/12)
4 × 2 with dominant decision (1)	88.9%	96.3%	91.7%	100.0%
(9A for Rows; 9B for Cols.)	(40/45)	(26/27)	([illegible]6/72)	(12/12)
2 × 2, partner has dominant decision (2)	61.1%	79.6%	68.1%	95.8%
(3A, 3B for Rows; 2A, 2B for Cols.)	(55/90)	(43/54)	(98/144)	(23/24)
2 × 3, partner has dominant decision (2)	62.2%	63.0%	62.5%	100.0%
(4A for Rows; 4C for Cols.)	(28/45)	(17/27)	(45/72)	(12/12)
3 × 2, partner has dominant decision (2)	60.0%	55.6%	58.3%	83.3%
(4B for Rows; 4D for Cols.)	(27/45)	(15/27)	(42/72)	(10/12)
2 × 4, partner has dominant decision (2)	73.3%	70.4%	72.2%	100.0%
(9B for Rows; 9A for Cols.)	([illegible]3/45)	(19/27)	([illegible]2/72)	(12/12)
2 × 3 with 2 rounds of dominance (2)	62.2%	68.5%	64.6%	100.0%
(6A, 6B for Rows; 5A, 5B for Cols.)	(56/90)	(37/54)	(93/144)	(24/24)
3 × 2 with 3 rounds of dominance (3)	11.1%	22.2%	15.3%	87.5%
(5A, 5B for Rows; 6A, 6B for Cols.)	(10/90)	(12/54)	(22/144)	(21/24)
2 × 3, unique equilibrium, no dominance	50.0%	51.9%	50.7%	91.7%
(7A, 7B for Rows; 8A, 8B for Cols.)	(45/90)	(28/54)	(73/144)	(22/24)
3 × 2, unique equilibrium, no dominance	17.8%	27.8%	21.5%	91.7%
(8A, 8B for Rows; 7A, 7B for Cols.)	(16/90)	(15/54)	(31/144)	(22/24)

Table II reports subjects' rates of equilibrium compliance in the B, OB, and TS treatments, pooling the data from isomorphic games, with population fractions in parentheses. The games are grouped by the complexity of the strategic reasoning they require, measured by the number of rounds of iterated pure-strategy dominance needed to identify the subject's equilibrium decision. Baseline and OB subjects' compliance rates are similar across games of similar complexity; and holding complexity constant, the number of own or other's decisions has little effect. Compliance with equilibrium is quite high for initial responses to abstractly framed games, in most cases well above random. As in previous experiments, compliance is highest in games that can be solved by one or two rounds of iterated dominance, and subjects played dominant decisions with frequencies near 90%.[50] But compliance falls steadily as complexity increases, dropping below random in our 3 × 2 games that are dominance-solvable

[50] Interestingly, in 3 × 2 games with unique pure-strategy equilibria and dominance only via mixed strategies, subjects played dominated decisions with frequencies (10% Baseline, 4% OB) similar to those for pure-strategy dominance.

in three rounds or our 3×2 games with unique equilibria but no pure-strategy dominance.[51]

These results are consistent with subjects' initial responses to games in other experiments, where subjects typically comply with 1–3 rounds of iterated dominance. However, S&W found much higher equilibrium compliance for symmetric 3×3 games solvable by three rounds of iterated pure-strategy dominance or with unique pure-strategy equilibria but no (pure- or mixed-strategy) dominance (68% and 57%, respectively) than we found for 3×2 games of comparable complexity (11–22% and 18–28% in the Baseline and OB, respectively).[52] This difference may stem from S&W's use of symmetric player roles and payoff displays and round-number payoffs, or from our attempt to separate strategic from nonstrategic decision rules as sharply as possible.

TS subjects identified their dominant decisions with frequencies well above 90%. In striking contrast to Baseline and OB subjects, their equilibrium compliance rates fell only slightly in more complex games, averaging about 90% even in games in which Baseline and OB compliance fell below random. This suggests that Baseline and OB subjects' low compliance in complex games is unlikely to be due to the difficulty of looking up payoffs via MouseLab or cognitive limitations. This leaves several possible explanations for the difference: TS subjects' training in identifying equilibria or their higher dismissal rate (see footnote 18); bounded rationality, in the form of decision rules that do not fully analyze others' incentives; a widespread prior understanding of others' decisions like that reflected in our *Sophisticated* type, coupled with a failure of common knowledge that most subjects are *Sophisticated*; or a combination of these. We now turn to a more detailed econometric investigation of the latter possibilities.

C. *Econometric Analysis of Decisions*

In this section we conduct a maximum likelihood error-rate analysis of Baseline and OB subjects' decisions. Recall that our econometric model is a mixture model in which each subject's type is drawn from a common prior distribution over nine types and remains constant for all 18 games.[53] Combining evidence from different patterns of deviation from types' decisions requires an error structure, which we specify as neutrally as possible, in the spirit of H&C's and EG&G's error-rate analyses. We combine *Naïve* and *Optimistic* in this section because their decisions are not separated in our games. We include both

[51] In most cases compliance is slightly higher in OB than in the Baseline. Although this is unlikely to be due entirely to chance, the difference is too small to be significant in our samples.

[52] Crawford (1997, Section 4) surveys other experimental evidence for dominance-solvable games. Our results for games with unique equilibria but no pure-strategy dominance are consistent with the evidence from other settings summarized by Selten (1998, Section 5), which tends to favor decision rules that employ step-by-step reasoning (such as iterated dominance) over what Selten calls "circular concepts" (such as our *Equilibrium* type in non-dominance-solvable games, our *Sophisticated* type, and, as explained in Section 3.A, all of S&W's strategic types).

[53] We are grateful to Glenn Ellison for suggesting this approach.