GAMES AND ECONOMIC BEHAVIOR **16,** 280–302 (1996)
ARTICLE NO. 0087

Abreu–Matsushima Mechanisms: Experimental Evidence*

Martin Sefton

*School of Economic Studies, University of Manchester, Manchester, M*13 9*PL, United Kingdom*

and

Abdullah Yavaş

Smeal College of Business, Pennsylvania State University, University Park, Pennsylvania 16802

Received April 12, 1995

Abreu–Matsushima mechanisms can be applied to a broad class of games to induce any desired outcome as the unique rationalizable outcome. We conduct experiments investigating the performance of such mechanisms in two simple coordination games. In these games one pure-strategy equilibrium is "focal"; we assess the efficacy of Abreu–Matsushima mechanisms for implementing the other pure-strategy equilibrium outcome. Abreu–Matsushima mechanisms induce some choices consistent with the desired outcome, but more choices reflect the focal outcome. Moreover, "strengthening" the mechanism has a perverse effect when the desired outcome is a Pareto-dominated risk-dominated equilibrium. *Journal of Economic Literature* Classification Number: C7. © 1996 Academic Press, Inc.

1. INTRODUCTION

It is well known that a group of individuals independently pursuing their own interests may fail to attain an outcome that promotes the group's interests. In such situations it is natural to seek mechanisms that change individual incentives, thereby leading to a better outcome for the group. Consequently a large theoretical literature has emerged devoted to designing and evaluating such mechanisms.

* Financial support from the Institute for Real Estate Studies at the Pennsylvania State University is gratefully acknowledged. Donna Chaklos, Tom Geurts, Brian Moran, Mary Ann Morris, and Shiawee Yang provided valuable research assistance. The authors thank Gary Bolton, Robert Forsythe, Ann Gillette, Robert Rosethal, two anonymous referees, and seminar participants at Penn State University, University of Manchester, University of Windsor, 1994 European Econometric Society Meetings, and the Seventh World Congress of the Econometric Society for their comments. E-mail address for first author: martin.sefton@man.ac.uk.

0899-8256/96 $18.00

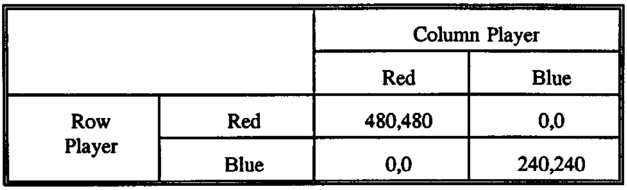
In this paper we present an experimental investigation of the mechanism introduced by Abreu and Matsushima (1992a). This mechanism can be used to implement any outcome of a broad class of games as a unique rationalizable outcome. The mechanism uses two elements to implement the desired outcome of a game. First, it forces the players to play the game in very small pieces; second, it levies fines on the first player(s) whose behavior leads to divergence from the desired outcome. As we shall explain in the next section, this implies that even a very small fine can be effective if the game is broken into sufficiently many pieces. In fact, such a fine achieves its goal through iterative elimination of strictly dominated strategies.

The applicability of the mechanism has, however, been criticized. Glazer and Rosenthal (1992); hereafter, GR) argue that the mechanism will not perform as predicted because it may involve many rounds of iterated dominance. They illustrate their argument by considering whether the mechanism could be used to implement the Pareto-dominated and risk-dominated equilibrium of a coordination game. They suspect that the players "would abandon the logic of iterated dominance in favor of the focal point in this game." In a response to GR, Abreu and Matsushima (1992b; hereafter, AM) disagree. However, this dispute about the performance of the mechanism is entirely speculative: the disagreement concerns how subjects *woud* behave in a carefully controlled experiment.

In order to lend empirical substance to this dispute, we conducted an experiment in which a mechanism was incorporated into the game they discuss: we tested the ability of the mechanism to implement the Pareto-dominated and risk-dominated equilibrium of a coordination game. We also investigated the impact of varying the number of pieces into which the game is broken. The logic of AM suggests that dividing the game into a greater number of pieces amplifies the effect of a given fine, thus enhancing the performance of the mechanism. However, as the number of pieces increases, the number of rounds of iterated dominance required to implement the desired outcome also increases, and this is the basis of the GR critique.

We find significant differences between actual behavior and predicted behavior. A negligible portion of the decisions correspond to the theoretical prediction. Moreover, the effect of varying the number of pieces is interesting. As we increase the number of pieces, the degree of success of the mechanism did not improve, contrary to what one would expect on theoretical grounds. Instead, our results are consistent with the notion that people only carry out iterated dominance arguments for a limited number of iterations.

This experiment represents an extremely special implementation problem: the planner's objective is to implement a *Pareto-inferior* outcome. A more natural implementation problem is suggested by the experimental results of coordination games in which the Pareto-dominant equilibrium is risk-dominated (see Cooper *et al.*, 1992; Straub, 1995; and Sefton and Yavaş, 1995). In these experiments *coordination failures* are prevalent—subjects tend to play the Pareto-dominated



		Column Player	
		Red	Blue
Row Player	Red	480,480	0,0
	Blue	0,0	240,240

FIG. 1. Payoff Matrix for Game I.

equilibrium. A natural question to ask is whether an Abreu-Matsushima mechanism can resolve such coordination failures. We investigated this question in a second experiment where we attempted to implement the Pareto-dominant, but risk-dominated, equilibrium of a coordination game. Neither AM nor GR disputes the ability of the mechanism to implement such an outcome. Nevertheless, given our results from the first experiment, we were uncertain about the effectiveness of the mechanism in such a setting.

Our results for this game are less clear. Although the mechanism was more successful in implementing the desired outcome than that in our first experiment, the theoretical prediction is still a poor predictor of behavior. On the other hand, unlike in our first experiment and consistent with the logic of AM, dividing the game into a larger number of pieces resulted in a higher success rate for the mechanism.

The remainder of the paper is organized as follows. The Abreu–Matsushima mechanism is presented in Section 2. We describe the experimental design in Section 3 and the experimental procedures in Section 4. The experimental results are reported in Section 5, and in Section 6 we provide some concluding remarks.

2. THE ABREU–MATSUSHIMA MECHANISM

Abreu and Matsushima (1992a) introduce a mechanism that implements any desired outcome of a broad class of games as a unique Nash equilibrium. In fact, implementation is achieved via the iterative deletion of strictly dominated strategies so that the desired outcome is also the unique rationalizable outcome. However, a unique Nash equilibrium, and even a unique rationalizable outcome, can be an implausible predictor of actual behavior. This is particularly true when predictions are based on many rounds of iterated dominance (see Kreps, 1990, pp. 393–399; or Basu, 1994). The GR paper argues that when the Abreu–Matsushima mechanism implements an outcome, it does so in a way which is susceptible to this criticism. They illuminate their argument using the two-player coordination game with payoff matrix given in Fig. 1. We will refer to this game as Game I.

Game I has multiple Nash equilibria which cannot be narrowed down using stability and coarser refinement criteria. However, most observers expect players to choose red in this game. The concepts of risk dominance and Pareto dominance (see Harsanyi and Selten, 1988) can be used to support such expectations, and so we identify (red,red) as the focal equilibrium of Game I.[1]

Now, suppose a planner's objective is to implement the (blue, blue) equilibrium. An Abreu–Matsushima mechanism would accomplish this objective by dividing the game into many pieces and introducing a small fine. Specifically, each player submits a sequence of T choices, instead of a single choice, of red or blue. The sequences are then matched, first choice of the row player with first choice of the column player, second choice with second choice, and so on. For each (red,red) combination the players receive a payoff of $480/T$ each, for each (blue, blue) combination the players receive $240/T$ each, and for other combinations the players receive nothing. In addition, a player pays a fine of F if the earliest choice of red in his or her sequence occurs before that of his or her opponent. Both players pay the fine if their earliest choices of red occur at the same time. If $F > 480/T$ the unique rationalizable outcome, determined by iterative elimination of strictly dominated strategies, consists of both players choosing a sequence of T blues. In this sense, even an arbitrarily small fine can implement (blue, blue) as the unique rationalizable outcome, as long as T is large enough. Note that the fine is never actually used in equilibrium.

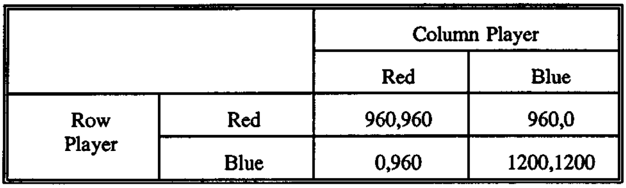
Whether subjects in a carefully controlled experiment will play the assigned equilibrium is an empirical question upon which GR and AM disagree. While GR "would hesitate to give long odds" on successful implementation, AM's "gut instinct is that our mechanism will not fare poorly in terms of the essential feature of their construction." The effect of varying F, given T, is not controversial—it seems reasonable to suppose that increasing the penalty on the earliest choice of red will reduce its incidence. Rather, it is the effect of varying T, given F, that is controversial. According to the logic of AM, increasing T multiplies the effect of a given fine.[2] Thus, a given fine may implement (blue,blue) when T is large but not when T is small. On the other hand, when T increases, the number of rounds of iterated dominance required to implement (blue, blue) increases. Thus the rationality assumption required to implement (blue,blue) becomes stronger.[3] This controversy seems a prime candidate for an experimental test.

While the example nicely illustrates the dispute, it involves an unusual plan-

[1] In fact, in a preliminary experiment in which 12 pairs of subjects played Game I once, Red was chosen 22 of 24 times. We interpret this as supporting our interpretation of (red,red) as the focal equilibrium, although the two rogue decisions should be noted.

[2] For a given F, as T increases, $480/T$ becomes smaller relative to F. This is expected to increase the incentive for each player to play his or her earliest choice of red after that of the other player.

[3] The rationality requirement is that "each player knows the other player knows . . . knows the other player is rational", the length of the sentence growing with the number of iterations of dominance.



		Column Player	
		Red	Blue
Row Player	Red	960,960	960,0
	Blue	0,960	1200,1200

FIG. 2. Payoff Matrix for Game II.

ning problem: left to themselves the players are expected to attain an efficient outcome, and the planner is attempting to undermine this. After conducting experiments with Game I we considered a second game that involves a more interesting implementation problem. This game, Game II, is illustrated in Fig. 2.

Again there are two pure-strategy equilibria, (red,red) and (blue,blue). This time, however, (red,red) is risk-dominant and (blue,blue) is Pareto-dominant, and it is not clear, *a priori*, which strategy will be played. Cooper *et al.* (1992) and Straub (1995) have conducted experiments with this game and find that the risk-dominant equilibrium prevails, and on the basis of this evidence we refer to (red,red) as the focal outcome of Game II.[4]

We then asked whether an Abreu–Matsushima mechanism would be successful in inducing blue (efficient) play. If $F > 960/T$ the unique rationalizable outcome of the modified game is for both players to submit a sequence comprising T choices of blue. While there is no conflict between the planner's objective and the players' payoffs, because of our results from Game I we were uncertain as to whether the mechanism would be successful in Game II.

3. THE EXPERIMENTAL DESIGN

Experiment I involved three sessions in which Game I was modified to incorporate an AM mechanism. Because the modified games are more complicated than Game I, we wanted to give subjects a chance to learn the subtleties of the game and so in each session we had each subject play a single modified game once for practice, then the same game 14 times for cash. Subjects faced a different opponent in each of the 15 rounds.

We attempted to reproduce as closely as possible the essential elements of the game and varied the critical parameter of the game, T, across the sessions while keeping F fixed. According to theory, for sufficiently large values of T we should observe all-blue sequences. We assess the performance of the mechanism

[4] As for Game I, we also confirmed this in a preliminary experiment.

by testing whether the prevalence of blue play increases with T, in line with this comparative static prediction.

Both GR and AM discussed a mechanism with $T = 100$, but given the difficulties associated with making, communicating, and computing the payoffs from a game with such a large number of choices, we considered this infeasible for experimental purposes. Instead, we conducted sessions of games with $T = 4$, 8, and 12 (we denote these sessions by 4T, 8T, and 12T); we see no reason why using shorter sequences should prejudice the experiment against successful implementation.

Using shorter sequences means that larger fines are required to implement (blue,blue). We were concerned that with large fines (relative to the equilibrium payoffs) blue play may be induced for reasons unrelated to the essential "pieced play" feature of the mechanism. For this reason we set $F = 90$ so that (red,red) remains risk-dominant in the $T = 1$ game with a fine. (Straub, 1995, presents compelling evidence that a payoff-dominant outcome will be observed in experiments with this type of game if it is also risk-dominant.) With this fine, the all-blue outcome is uniquely rationalizable for $T \geq 6$.

With these parameters, we can observe whether blue play is more likely when the mechanism implements (blue, blue) than when it does not. Further, we can investigate whether strengthening the mechanism, by increasing T, improves its performance.[5] We do this by comparing the prevalence of blue play across sessions.

For Experiment II we again conducted three sessions, with $T = 4, 8, 12$. The procedures were identical to those used for Experiment I, except for the payoffs and fine. The payoffs were derived from Game II, and the fine was set at $F = 160$, so that, as in Experiment I, the all-blue outcome is uniquely rationalizable for $T \geq 6$. Again, we are able to assess whether there is more blue play when the mechanism implements (blue,blue) than when it does not, as well as assessing whether blue play increases when the mechanism is strengthened.

In summary, Experiment I consists of three sessions of Game I, spread over 4, 8, and 12 pieces, respectively, with a 90 point fine. With this fine the 8T and 12T games implement (blue,blue), but the 4T game does not.[6] Experiment II consists of three sessions of Game II, spread over 4, 8, and 12 pieces respectively, with a 160 point fine. Again, this fine implements (blue,blue) in the 8T and 12T games, but not in the 4T game.

[5] Another possibility for strengthening the mechanism is to increase the fine, as discussed by AM. We expect that, given T, there is a level of fine sufficiently large to induce all-blue play. However, the question of more interest is whether, given F, there is a level of T sufficiently large to induce all-blue play. Thus, we preferred to manipulate the incentive to delay red choices by varying T.

[6] In the 4T games the best response of a player is to produce a sequence identical to that of his/her opponent. Thus, (blue,blue) is only one of many equilibria in this game.

4. EXPERIMENTAL PROCEDURES[7]

Experiment I was conducted in September 1993 and Experiment II in March 1994 at Pennsylvania State University. Each experiment used 90 subjects, 30 in each of three sessions, who had signed up in response to fliers posted around campus. Each subject participated in one session only. In each session 30 subjects were seated in a large room, read a set of instructions, and given an opportunity to ask questions. We then conducted a practice round in which earnings were hypothetical, and at the end of this practice round we gave subjects another opportunity to ask questions. A partition was then drawn to divide the room into two halves, and this completed the instructional part of the session.

Each session consisted of 15 rounds (including the practice round), with subjects rematched with a new, anonymous opponent after each round. The pairings were also designed to prohibit indirect repeated interactions through common opponents of opponents. Thus, at the beginning of any round, a subject could not have been affected in any way by the previous play of their new opponent. In this sense, we refer to the 15 games played by a subject as a sequence of one-shot games.[8] Throughout each session the only permitted communication between subjects was via their formal decisions.

In each round of a session each subject made a decision consisting of T choices of red and/or blue and recorded it. When all subjects had done this, monitors delivered the decision forms to the appropriate subjects. Subjects then computed their earnings, and monitors checked their calculations. The next round did not begin until the monitors had verified that all subjects had calculated their earnings correctly.

Subjects started the session with an initial balance of 1600 points in Experiment I (2000 points in Experiment II) and accumulated additional points through the outcome of the games they played. At the end of the session they were paid \$0.25 per 100 points in Experiment I (\$0.25 per 250 points in Experiment II). The sessions averaged 85 min, and earnings averaged \$12.60 in Experiment I (\$13.80 in Experiment II).

5. EXPERIMENTAL RESULTS[9]

Figure 3 displays the proportion of entirely blue sequences in Experiment I across rounds. It is clear that the process of iterated dominance does not work

[7] A full set of experimental materials is included in Appendices 1 and 2.

[8] Of course, subjects may have *thought* their current decision could influence future opponents' play. We are skeptical that subjects would modify their play on the basis of such (incorrect) beliefs).

[9] We exclude the data from the practice round from all calculations in this section. A complete set of the data is available from either author upon request.

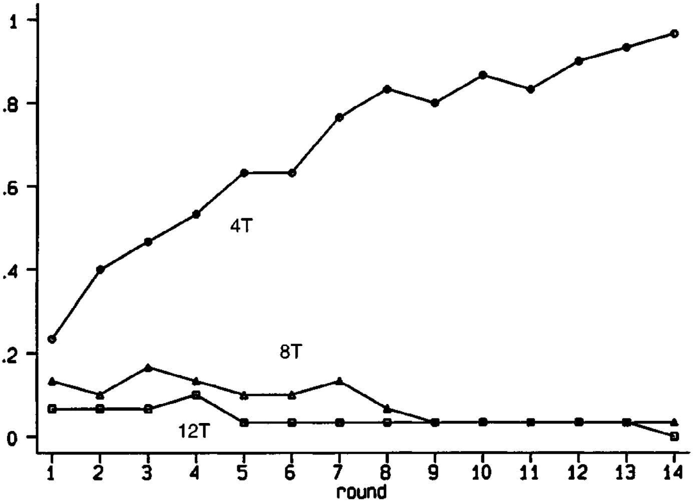


FIG. 3. All-Blue Play in Experiment I.

to the full extent supposed in theory; theoretically, sequences should have been all-blue in the 8T and 12T sessions. In fact, in these sessions, there are few entirely blue sequences. Nor is there any indication that subjects in those two sessions are learning to play blue: in the last 7 rounds of these sessions, only one in thirty decisions corresponded to the precise theoretical prediction. Only in the 4T session, where blue play is not a unique equilibrium, is blue play predominant. In the 4T session blue play increases until, in the last round, only one of thirty sequences included a choice of red.

However, the mechanism did induce some blue play. In the 8T and 12T games it is worth looking at whether *any* blue choices were observed, since one round of iterated dominance would eliminate any all-red choices. In the 8T session 97% and in the 12T session 95% of the sequences involved at least one choice of blue. However, it is not clear whether this was due to "pieced play," rather than the presence of a fine *per se*, since a similar pattern is evident in the 4T session: 97% of sequences in the 4T session include a choice of blue.

Closer inspection reveals an important feature of these data. The vast majority of sequences have the property that once a subject takes the risk of paying the fine by choosing a red at some point in his or her sequence, he or she never plays blue in the remainder of the sequence.[10] We call such sequences "monotonic," and we refer to the position in a sequence of a subject's last choice of blue as that

[10] This appears to be consistent with the observation from our preliminary experiment that in the absence of a fine subjects play red.

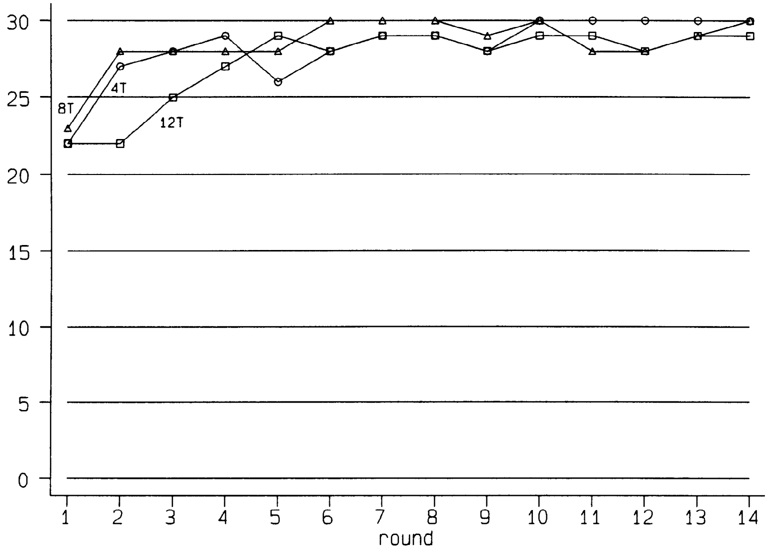


FIG. 4. Monotonic Sequences in Experiment I.

subject's "switchpoint."[11] In Fig. 4 we summarize the incidence of monotonic sequences. Aggregating across the three sessions, around 75% of sequences are monotonic in the first round, but the percentage grows quickly: in the last seven rounds less than 3% of the sequences (less than one subject in a round) are nonmonotonic.

In terms of switchpoints, the predicted switchpoints in the 8T and 12T sessions are 8 and 12 respectively. These predictions can be viewed as the result of T applications of iterated dominance. Further, the observed switchpoint can be regarded as an indication of the number of applications of iterated dominance that subjects actually carry out. The difference between the arguments of AM and GR might then be interpreted as over whether subjects have an unlimited or limited facility for applying iterated dominance. Clearly, it is worth looking more closely at this switching behavior across sessions and rounds.

Figure 5 presents graphs of the switching behavior across rounds for each session discarding nonmonotonic sequences. In Fig. 5a we plot the average switchpoint for each round, where the average is taken over the monotonic sequences observed in that round. Thus, Fig. 5a shows the average *number* of blue choices in a monotonic sequence. Letting μ_T represent the mean switchpoint in a game with T pieces, then theory predicts $\mu_4 < \mu_8 < \mu_{12}$. In contrast, if

[11] Thus, switchpoints range from 0 (an all-red sequence) to T (an all-blue sequence).

subjects only carry out a limited number of iterations of iterated dominance, we would not expect average switchpoints to vary across sessions.[12] Informal inspection of Fig. 5a suggests that in all sessions of Experiment I, sequences generally consist of about 3 or 4 blue choices before switching to red.

In Fig. 5b we plot the average *normalized* switchpoint—the average switchpoint divided by T—across rounds. That is, Fig. 5b shows the average *proportion* of blue choices in a monotonic sequence. Letting $\pi_T = \mu_T/T$, the theoretical prediction is that $\pi_T = 1$ for $T > 240/F$, with no clear prediction otherwise. This prediction clearly fails. In fact, the figure suggests that mean normalized switchpoints are lower in games with larger T.

Table 1 presents statistics, based on final round monotonic sequences[13], for testing the following hypotheses concerning switchpoints: H_{01}: $\mu_4 = \mu_8$ and H_{02}: $\mu_8 = \mu_{12}$. We see that while the average switchpoints are higher in the sessions with larger T, the difference is not significant. Table 1 also features statistics for testing the following hypotheses concerning normalized switchpoints: H_{01}: $\pi_4 = \pi_8$ and H_{02}: $\pi_8 = \pi_{12}$. We see that, contrary to the theoretical prediction, the mean normalized switchpoints are significantly lower in the sessions with larger T.[14]

Taken together, Fig. 5 and Table I support the following description of behavior in Experiment I: in all three sessions, subjects carry out the iterative dominance argument up to around four iterations, thus the fraction of the game up to which the iterated dominance is carried out becomes smaller as T increases.

We also computed chi-square statistics and their associated p-values as summary measures of the stability of the switchpoint distributions (discarding nonmonotonic sequences). The p-values for the 4T, 8T, and 12T sessions are 0.000, 0.870, and 0.001, respectively. Larger p-values indicate more stable distributions of switchpoints, while smaller p-values indicate less stability. The p-values attest to instability in the 4T session, as is evident from casual inspection of Figs. 3 and 5, but the apparent round effect in the 12T session is not so easily interpreted. In the 12T session the switchpoints tend to be higher in the second half of the session relative to the first, but there is little evidence of significant change within the last seven rounds.[15] Thus, we warn against interpreting the round effect as a persistent learning effect, and caution against extrapolating increasing prevalence of blue play beyond the horizon of the 12T session.

[12] Assuming that T exceeds the “limited number” of iterations. Otherwise the mean switchpoint may change with T due to truncation.

[13] The qualitative results are not affected if we include nonmonotonic sequences or if we replace a subject’s final round decision with their average decision over the 14 rounds.

[14] For completeness, Table I also presents statistics for comparing the proportions of all-blue play across sessions.

[15] Based on the last seven rounds, the p-values associated with the χ^2 statistics are 0.797, 1.000, and 0.743 for the 4T, 8T, and 12T sessions.

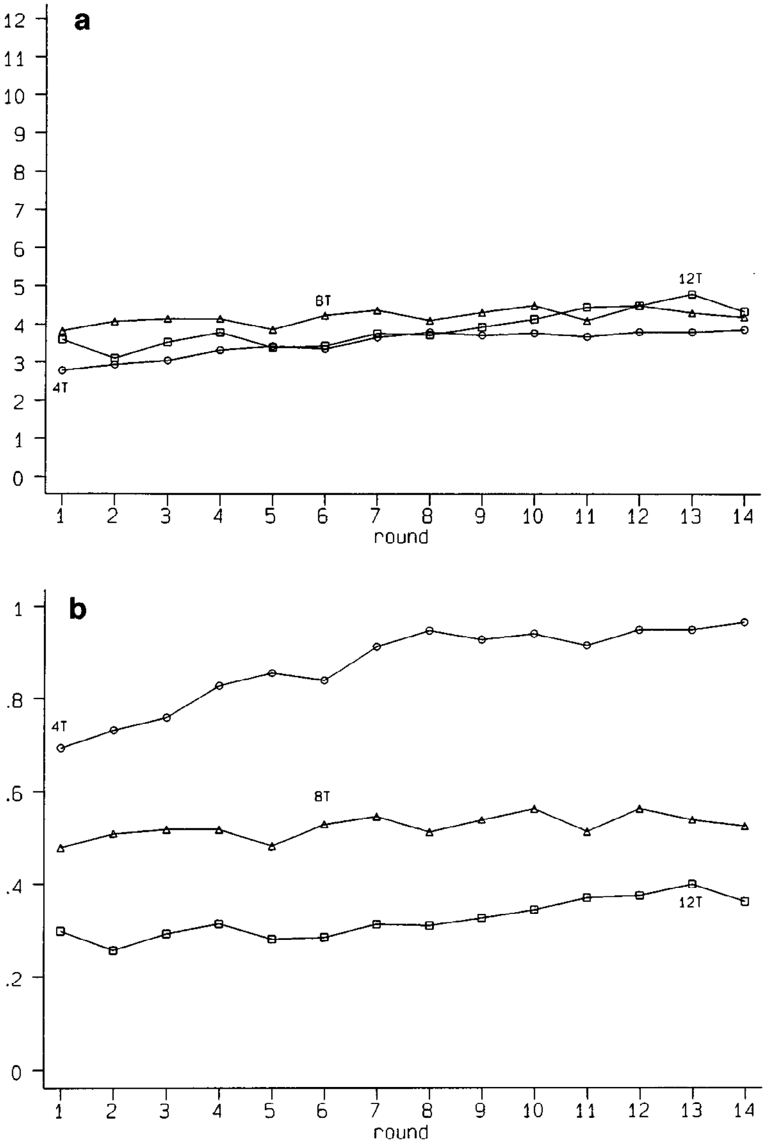


FIG. 5. Switching Behavior in Experiment I: (a) Average Switchpoints, (b) Average Normalized Switchpoints.

TABLE I
Tests for Session Effects in Experiment[a]

Dependent variable	Statistic[b]	p-Value[c]
All-blue		
4T vs 8T	−19.80	1.000
8T vs 12T	−1.00	0.841
Switchpoint		
4T vs 8T	0.90	0.184
8T vs 12T	0.29	0.386
Normalized switchpoint		
4T vs 8T	−8.07	1.000
8T vs 12T	−3.07	0.999

[a] The null hypothesis is that the means of the dependent variable are the same in each session. (The test is applied to final-round monotonic sequences only.)

[b] The reported statistic is the difference in means, standardized to have an approximate standard normal distribution under the null. A minus sign indicates that the average value is higher in the session with smaller T.

[c] The reported p-value is the probability under the null hypothesis of getting a statistic as large as the one observed.

Finally, a comparison of penultimate and final round behavior suggests that by the end of all sessions average behavior is very stable. This partly reflects the stability of individual subjects' play: the majority of subjects did not change their strategy across these rounds. Of the remaining subjects, as many decreased as increased their switchpoint.[16]

We now turn to the results of Experiment II. Figure 6 shows the proportion of sequences that are all-blue in each session, by round. As in Experiment I, the session with the most blue play, 43% over 14 rounds, is the 4T session, and this is the session in which T is not large enough to implement all-blue sequences as the unique rationalizable outcome. However, there is considerably less all-blue play in this session than was observed in the 4T session of Experiment I. Also, as in Experiment I, the 8T session displays little all-blue play: 11 of 30 first-round sequences were all-blue, but all-blue play dissipates over time, disappearing by the final round. The 12T session in Experiment II, on the other hand, displays many more all-blue sequences than it did in Experiment I. In fact, the amount of all-blue play in the 12T session is almost as much as the 4T session: 38% of sequences are all-blue (with there being little change over time).

As in Experiment I, the vast majority of sequences in Experiment II involves some blue choices. Indeed, as in Experiment I, sequences typically begin with

[16] For the 4T session 1 subject increased, 29 did not change, and none decreased their switchpoint; for the 8T (12T) session 4 (6) increased, 21 (16) did not change, and 4 (7) decreased their switchpoint.

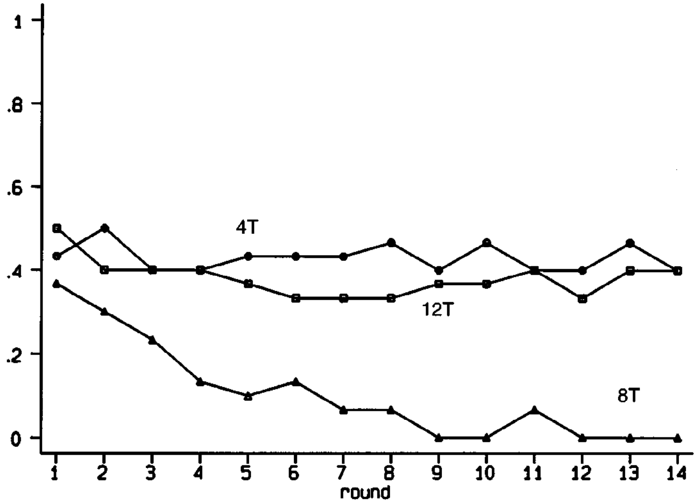


FIG. 6. All-Blue Play in Experiment II.

a blue choice and end with a red choice, switching from blue to red only once. Figure 7 summarizes the incidence of monotonic sequences in Experiment II, and is very similar to the corresponding figure for Experiment I. Around 25% of sequences are nonmonotonic in the first round, but the percentage drops quickly to less than 3% (less than one subject per round) in the last seven rounds.

Figure 8a tracks average switchpoints across rounds for each session and shows that average switchpoints are higher in the sessions with larger T. However, Fig. 8b suggests that when normalized switchpoints are used to summarize sequences, i.e., when blue play is represented by the proportion of blue choices in a sequence, the 4T session is again the session which features the most blue play. Furthermore, in contrast to Experiment I, where the mean switchpoint is invariant to T and the mean normalized switchpoint is *inversely* related to T, in Experiment II both measures of switchpoints are higher in the 12T than in the 8T session. One interpretation of this is that subjects can carry out the iterative dominance argument to a larger length and to a larger fraction of the length of a game as T increases in Experiment II. These results, based on informal inspection of Fig. 8, are confirmed by the statistics in Table II.

Figure 8 suggests that in Experiment II, unlike in Experiment I, average behavior is most stable in the 4T session. A summary measure of the stability of the distributions of switchpoints for Experiment II (discarding nonmonotonic sequences) is given by the p-values associated with χ^2 statistics: for the 4T, 8T, and 12T sessions these are 0.999, 0.000, and 0.001. As in Experiment I, the source of instability in the 8T and 12T sessions appears to lie in the early

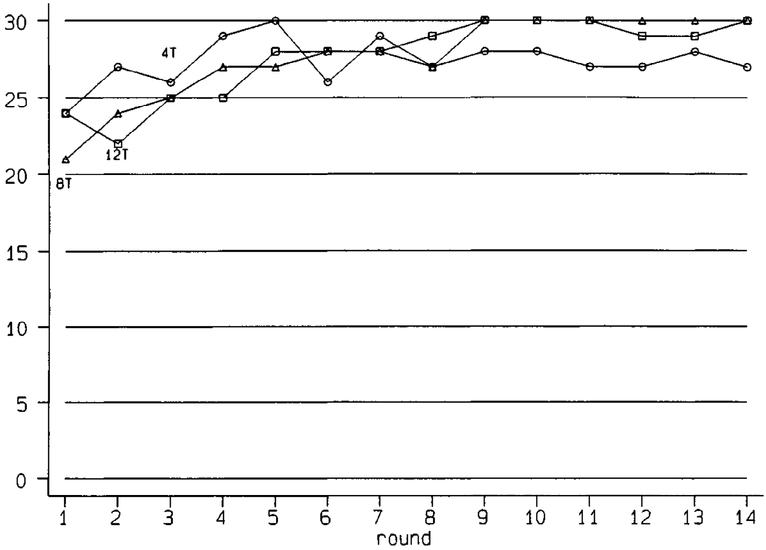


FIG. 7. Monotonic Sequences in Experiment II.

TABLE II
Tests for Session Effects in Experiment II[a]

Dependent variable	Statistic[b]	p-Value[c]
All-Blue		
4T vs 8T	−4.56	1.000
8T vs 12T	4.40	0.000
Switchpoint		
4T vs 8T	2.53	0.006
8T vs 12T	6.97	0.000
Normalized switchpoint		
4T vs 8T	−2.72	0.997
8T vs 12T	4.05	0.000

[a] The null hypothesis is that the means of the dependent variable are the same in each session. (The test is applied to final round monotonic sequences only.)

[b] The reported statistic is the difference in means, standardized to have an approximate standard normal distribution. A minus sign indicates that the average value is higher in the session with smaller T.

[c] The reported p-value is the probability under the null hypothesis of getting a statistic as large as the one observed.

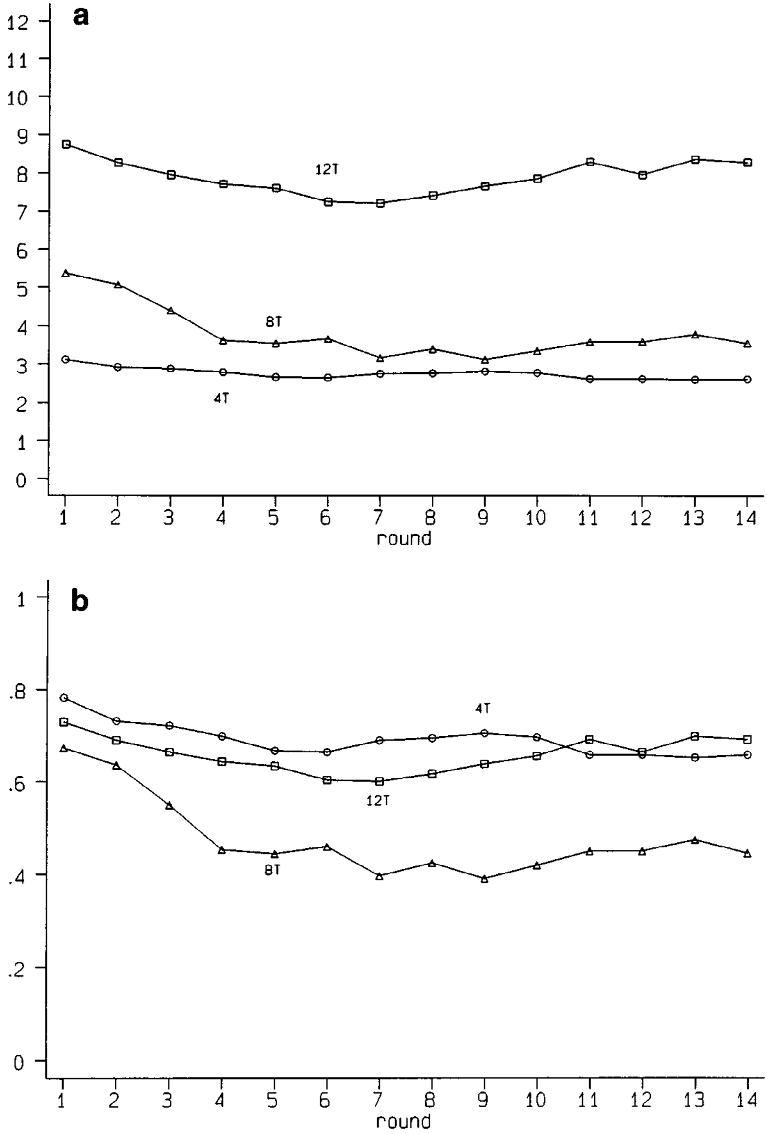


FIG. 8. Switching Behavior in Experiment II: (a) Average Switchpoints, (b) Average Normalized Switchpoints.

rounds of the session.[17] Also, and as in Experiment I, by the end of each session of Experiment II average play is very stable. Most subjects do not change their behavior between the penultimate and final rounds, and among those subjects who do change their behavior, the numbers who decrease or increase their switchpoints are similar.[18]

6. CONCLUSION

We have presented the results of two experiments designed to investigate the empirical performance of a recently proposed implementation mechanism. The mechanism requires players to submit their decisions as a sequence of T choices. The way in which actual play will depend on T is controversial. According to theory, if T is sufficiently large a desired outcome can be implemented via iterated dominance. On the other hand, many economists feel uncomfortable with this prediction because it is based on many rounds of iterated dominance, and the number of iterations grows with T. Our main concern was to investigate the role of this crucial parameter.

In each experiment we conducted two sessions where T was sufficiently large for the mechanism to implement a unique rationalizable outcome of a coordination game. In neither experiment were the data consistent with this predicted outcome, nor was there any evidence of movement toward the predicted outcome in repeated play. Thus, the use of rationalizability as a solution concept is perhaps uncontroversial, but it lacks predictive power in our experimental games.

Theory also delivers comparative static predictions—in fact, these are often of most interest. However, in Experiment I, not only were the experimental outcomes significantly different from the predicted outcome, but also the effect of increasing T was perverse. With a fixed fine, a larger T should strengthen the mechanism, yet we observed less blue play with larger T. This finding provoked our second experiment. Because the mechanism failed to overturn the focal outcome in Game I, it was not obvious whether it would overturn the focal outcome of Game II. Again, we observed more blue play in the 4T games, where T is not large enough to implement (blue,blue), than in the 8T and 12T games, where T is large enough to implement (blue,blue). However, we did find that strengthening the mechanism increased blue play in Experiment II, in the sense that there is more blue play in the 12T game than in the 8T game.

These results suggest several alternative directions for future research. One would be to search for conditions under which the theoretical predictions work well in the laboratory. This might be accomplished by modifying the experimental procedures in various ways. Here we discuss three such modifications.

[17] Focusing on the last seven rounds, the p-values are 0.99, 0.191, and 0.363.

[18] The numbers are: 5 increased, 17 did not change, 3 decreased (4T); 3 increased, 23 did not change, 4 decreased (8T); 6 increased, 20 did not change, 3 decreased (12T).

First, AM argue that in any application the mechanism could be used together with an explanation of how it works. This may help players appreciate the logic behind the mechanism, and hence induce them to play as predicted. Our experiment does not include such an explanation, partly because we are concerned that this might lead to changes in subject behavior that reflect their perceptions about what the experimenter wants them to do, regardless of any enhanced understanding of the incentive structure of the game. Also, any explanation would have to be incorporated into the experiment in a formal way, so that in principle the same experimental procedures could be used by other researchers.

It should be noted that even if a player is enlightened by an explanation of the logic of the mechanism, the explanation would only be effective in inducing the desired outcome in so far as the player believes other players are likewise enlightened. Further, the explanation may be more or less persuasive, depending on context. One can imagine a very persuasive explanation for Game II that begins with the sentence, “We will now explain how the fine can help you attain maximum earnings, even though you will never have to pay the fine,” whereas the same sentence is not available for Game I.

A second possible modification of the experiment might be to change the nature of repetition. We allowed subjects to learn about the game by having them play repeatedly, but with changing partners. The performance of the mechanism might be enhanced if partners were not rematched: if players best-respond to past play of their opponent, the unraveling that is supposed to take place at the level of introspection may then take place over time, leading, ultimately, to all-blue play. With changing partners, players get less information from opponents’ past choices, and this may inhibit unraveling. Of course, whether outcomes would change in this way can be tested.[19] The problem with unchanged pairings is that it may introduce complicated repeated game effects, making interpretation of the outcomes difficult.

A third approach might be for the experimenter to increase the fine until the mechanism works. The difference between the fine which is theoretically large enough to induce the desired outcome and the effective fine could then be a measure of “psychological barriers.” The outcome of this approach might be the determination of an excessively large effective fine: it may have to be so large as to swamp other incentives. Also, our results from Game I suggest that the effective fine for given T may cease to be effective for larger values of T.

An alternative direction for future research would be to explain the observed discrepancies between theory and evidence. We hope that our experiments can provide some insights and suggestions for alternative theoretical models of the observed behavior under the mechanism.

[19] Andreoni (1988) finds that rematching has a small but statistically significant effect in a public good experiment.

APPENDIX: EXPERIMENTAL MATERIALS

This appendix contains materials for the Game I, $T = 4$ session. The other sessions included the obvious changes. Subject folders contained copies of the instructions, a record sheet and a decision form.

Protocol

Upon entering the room, subjects are randomly seated. The door is closed after 30 subjects have been seated. They complete and return a consent form and are then given folders and the following oral directions.

> May I have your attention please. We are ready to begin. Thank you for coming. Each of you will be paid in cash at the end of the session. What you do during the session will determine how much.
>
> With the exception of the pencil and folder, please remove all materials from your desk. Open your folder, check that you have a set of instructions, a decision form, and a record sheet inside. I will now read the instructions aloud. After the instructions have been read you will have an opportunity to ask questions.

The experimenter reads the instructions.

> Are there any questions?

At this stage there may be questions: The experimenter answers them by referring back to the instructions if possible. If the question will be answered through the practice round, the experimenter says, “I think you’ll figure out the answer to that question when we go through the practice round, if not ask it again.”

> If there are no questions, let’s begin the practice round. Circle your choices on the decision form. Circle all four choices. And when you’ve made your choices, record them on your record sheet. If anyone needs help raise your hand.
>
> You record your choices on line zero-a. Just write R for red and B for blue. You can’t fill out the other lines yet.
>
> When you’ve made your decision and recorded it, raise your hand so that a monitor can come and get your decision form.

Monitors collect forms.

> We won’t deliver the forms until everyone in both rooms has completed their decision.

Monitors deliver forms.

> Now you should be receiving a form from the person in the other room that you’re paired with. Check you have the right form. Their number is at the top of the form and should match the number on your record sheet. Write their choices on row zero-b. Just use R for red and B for blue.

Monitors start circulating rooms, checking that subjects are doing it correctly.

> This part is important. Record your earnings on line zero-c. Each time your choices match you get some points. Check the instructions if you don’t remember. If they don’t match record a zero.
>
> Then figure out if you must subtract 90 points. If the first time you chose red was before or at the same time as the other person, subtract 90. If you need help raise your hand.
>
> To get your earnings for the round, add across line zero-c. Enter the total on the far right. The next round will count toward your earnings so it’s important to be straight on this now.

Monitors check calculations.

> If there are no questions, we’ll close the screen and begin round one. The monitor will bring you a decision form. Just put the used decision forms to one side.

Instructions

General Rules. This is an experiment in the economics of decision making. If you follow the instructions carefully and make good decisions, you can earn a considerable amount of money. You will be paid in private and in cash at the end of the experiment.

The experiment will consist of fifteen rounds, the first of which will be a practice round. The purpose of the practice round is to familiarize you with the experimental procedures. Nothing that you do in the practice round will affect your earnings.

Notice that there are two rooms of subjects in this experiment. In each round you will be paired with another person who is in the other room. You will never be paired with a person in your own room. You will be paired with a different person in each round. You will not know who is paired with you in any round. Similarly, the other people in this experiment will not know who they are paired with in any round. To accomplish this, the partition between the two rooms will be closed before the beginning of round one. It is important that you do not look at other peoples’ work, and that you do not talk, laugh or exclaim out loud during the experiment. If you violate this rule you will be warned once. If you violate this rule a second time you will be asked to leave and you will not be paid.

Description of Each Round. Each round consists of the following steps: (a) you will make a decision and record it; (b) you will record the decision of the person you are paired with; and (c) you will compute your earnings for the round. We will describe each step in turn.

(a) At the beginning of each round we will give you a “Decision Form.” The Decision Form for the practice round (round zero) is in your folder. Look at it now. You make four choices in each round. Each choice is between “RED” and “BLUE”. You will enter your choices by circling the appropriate colors on the decision form. You will then copy your choices onto your “Record Sheet”. You will find this record sheet in your folder. Look at it now. You will record your choices in round one on row 1(a). (You will record your choices in round two on row 2(a), and so on.) You will have three minutes to make your choices and record them. Then the monitor will collect your decision form. After the monitors have collected all of the decision forms from both rooms, they will deliver your decision form to the person in the other room with whom you are paired. They will also deliver to you the decision form completed by the person with whom you are paired.

(b) When you receive the decision form of the person you are paired with, you will record their choices on your record sheet. In round one, you will record their choices on row 1(b). (In round two you will record their choices on row 2(b), and so on.)

(c) You will then compute your earnings for the round according to rules we will discuss below and record them. In round one, you will record your earnings on row 1(c). (In round two you will record your earnings on row 2(c), and so on.) At the end of each round, you will record your earnings for that round in the far right column of your record sheet. A monitor will record your earnings for that round in the far right column of your record sheet. A monitor will check your calculations and, after everyone has entered their earnings for the round correctly, the next round will begin.

How Your Earnings Are Determined. You will start the experiment with an initial balance of 1600 points. This amount has already been entered in your record sheet. Your additional earnings in a round will depend on your choices and the choices of the person you are paired with. Your earnings from a round will be determined as follows. If your first choice and the first choice of the person you are paired with are both “Red” you will each earn 120 points, if the first choices are both “Blue” you will each earn 60 points, and if the first choices do not match (that is, if one of you chooses “Red” and the other chooses “Blue”) you will each earn 0 points. Your earnings from the remaining three choices will be determined in exactly the same way. In each round you may also have to subtract 90 points from your earnings. Whether or not you subtract this amount in a round will depend on the *earliest* choices of "Red" by you and the person you are paired with. Specifically, 90 points will be subtracted from your earnings in a round if your earliest choice of "Red' in that round comes *before* that of the person you are paired with. If you both make your earliest choice of "Red" at the same time, you will both lose 90 points. No points will be subtracted from your earnings in a round if your earliest choice of "Red" in that round comes *after* the person you are paired with, or if you do not choose "Red" in that round at all.

Your earnings from a round will be the sum of your earnings from your four choices in that round less any points you may have to subtract. Note that although you will make all of your four choices in a round at once, the order of your choices is important since it may determine whether or not you subtract 90 points.

At the end of round fourteen, you will add your earnings from rounds 1 through 14 to your initial balance and enter the total on the bottom line of your record sheet. This will determine your total point earnings. At the end of the experiment you will receive $0.25 for every 100 points you earned.

Are there any questions?

DECISION FORM FOR SUBJECT ___

ROUND ____

PAIRED WITH ____

MY FIRST CHOICE IS RED / BLUE (circle one)

MY SECOND CHOICE IS RED / BLUE (circle one)

MY THIRD CHOICE IS RED / BLUE (circle one)

MY FOURTH CHOICE IS RED / BLUE (circle one)

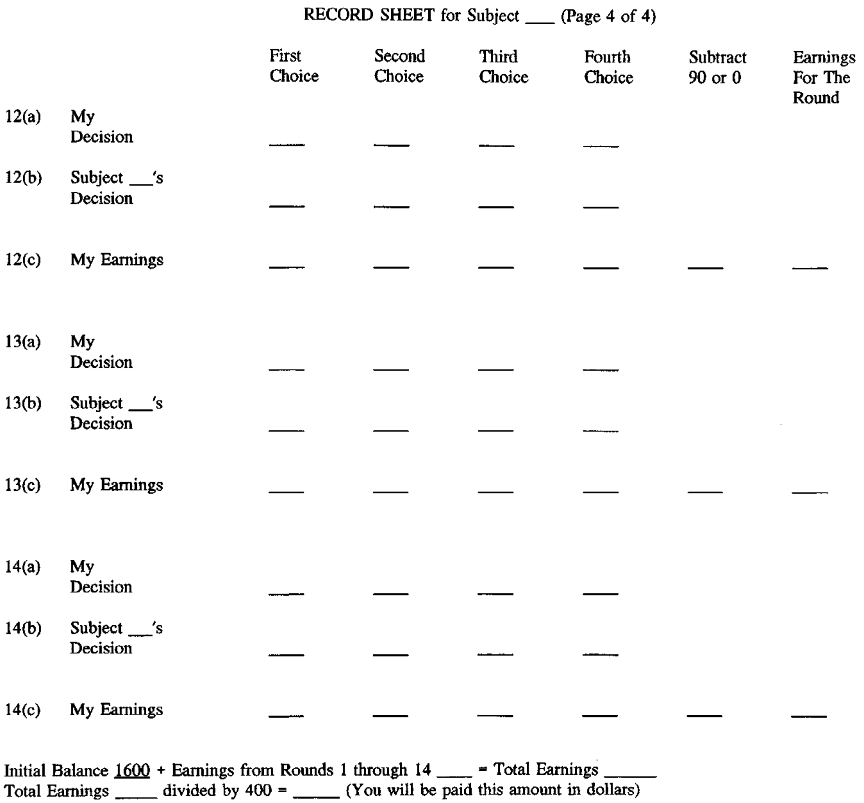
RECORD SHEET for Subject ___ (Page 1 of 4)

		First Choice	Second Choice	Third Choice	Fourth Choice	Subtract 90 or 0	Earnings For The Round
0(a)	My Decision	____	____	____	____		
0(b)	Subject ___'s Decision	____	____	____	____		
0(c)	My Earnings	____	____	____	____	____	____
1(a)	My Decision	____	____	____	____		
1(b)	Subject ___'s Decision	____	____	____	____		
1(c)	My Earnings	____	____	____	____	____	____
2(a)	My Decision	____	____	____	____		
2(b)	Subject ___'s Decision	____	____	____	____		
2(c)	My Earnings	____	____	____	____	____	____

•

•

•



RECORD SHEET for Subject ___ (Page 4 of 4)

		First Choice	Second Choice	Third Choice	Fourth Choice	Subtract 90 or 0	Earnings For The Round
12(a)	My Decision	___	___	___	___		
12(b)	Subject ___'s Decision	___	___	___	___		
12(c)	My Earnings	___	___	___	___	___	___
13(a)	My Decision	___	___	___	___		
13(b)	Subject ___'s Decision	___	___	___	___		
13(c)	My Earnings	___	___	___	___	___	___
14(a)	My Decision	___	___	___	___		
14(b)	Subject ___'s Decision	___	___	___	___		
14(c)	My Earnings	___	___	___	___	___	___

Initial Balance 1600 + Earnings from Rounds 1 through 14 ____ = Total Earnings ______
Total Earnings _____ divided by 400 = _____ (You will be paid this amount in dollars)

REFERENCES

Abreu, D., and Matsushima, H. (1992a). "Virtual Implementation in Iteratively Undominated Strategies: Complete Information," *Econometrica* **60**, 993–1008.

Abreu, D., and Matsushima, H. (1992b). "A Response to Glazer and Rosenthal," *Econometrica* **60**, 1439–1442.

Andreoni, J. (1988). "Why Free Ride? Strategies and Learning in Public Goods Experiments," *J. Public Econ.* **35**, 291–304.

Basu, K. (1994). "The Traveler's Dilemma: Paradoxes of Rationality in Game Theory," *Amer. Econ. Rev.* **84**, 391–395.

Cooper, R., DeJong, D. V., Forsythe, R., and Ross, T. W. (1992). "Communication in Coordination Games," *Quart. J. Econ.* **107**, 739–771.

Glazer, J., and Rosenthal, R. W. (1992). "A Note on Abreu-Matsushima Mechanisms," *Econometrica* **60**, 1435–1438.

Harsanyi, J. C., and Selten, R. (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge, MA: MIT Press.

Kreps, D. M. (1990). *A Course in Microeconomic Theory*. Princeton, NJ: Princeton Univ. Press.

Sefton, M., and Yavaş, A. (1995). "Risk and Coordination: Experimental Evidence," Center for Negotiation and Conflict Resolution working paper. Penn State University.

Straub, P. G. (1995). "Risk Dominance and Coordination Failures in Static Games," *Quart. Rev. Econ. Finance* **35**, 339–364.