

TESTING THE MINIMAX HYPOTHESIS: A RE-EXAMINATION OF O'NEILL'S GAME EXPERIMENT

BY JAMES N. BROWN AND ROBERT W. ROSENTHAL¹

In this paper, we re-examine the data from O'Neill's experiment involving a repeated, two-person, constant-sum game. We find that there is less evidence in support of the minimax hypothesis than indicated by O'Neill. There is strong evidence of serial correlation in players' choices, with several players displaying statistically significant dependence on the past moves of their opponents. We interpret this finding as evidence that the players themselves rejected minimax play as the appropriate model for their opponents' behavior. We find no evidence that players' behavior approached minimax behavior as players became more experienced.

KEYWORDS: Experimental game, zero-sum, minimax, statistical tests.

1. INTRODUCTION

IN A RECENT PAPER, Barry O'Neill (1987) reported the results of an ingenious experiment designed to test the minimax hypothesis for two-person, constant-sum games. In comparison with earlier experimental tests of this hypothesis, in which the minimax theory fared poorly, O'Neill's experiment required fewer maintained hypotheses regarding players' utility functions. In addition, O'Neill's experiment involved a game that was simple enough so that players might reasonably be expected to approximate the (nontrivial) strategic behavior implied by the minimax hypothesis. O'Neill interpreted his results as providing strong support for the hypothesis of minimax play, at least in an aggregate sense. He also argued that, due to superior experimental design, his results should be given more weight than earlier experimental results in evaluating the experimental success of the theory.

In this paper, we re-examine the results of O'Neill's experiment and find support for the minimax hypothesis to be much weaker than indicated by O'Neill. First, we find that tests based on observed win frequencies have little power to reject the minimax model when a wide class of other models may be true. Nevertheless, win frequencies for individual player pairs clearly reject the common probability model implied by the minimax hypothesis. Observed wins also reject the weaker hypothesis of independent play in O'Neill's experiment, with contemporaneous correlation in players' card choices accounting for virtually all of the difference between observed win frequencies and win frequencies predicted by minimax play.

¹ We are indebted to Barry O'Neill for useful comments and for generously providing us with the raw data from his experiment. We gratefully acknowledge stimulating conversations with Robert Aumann and Robert Porter on this subject, the research assistance of Jordi Masso, the helpful comments of the referees, and the financial support of the National Science Foundation under Grant SES86-03550. None of the above should be presumed to agree with any opinions expressed herein.

We also study the actual moves in O'Neill's data, and find that tests based on these statistics also may lack power for some types of non-minimax play. Nevertheless, although observed mixtures are roughly suggestive of minimax play in the aggregate, they too reject the minimax model quite clearly. This finding is reinforced when the data are studied at the level of the individual player. Not only is the particular multinomial model implied by minimax play rejected by the data, but the more general hypothesis that any common, stationary multinomial process generated the observed plays is also rejected, primarily on the basis of extensive serial correlation in players' card choices.

In the next section, we describe the relevant aspects of O'Neill's experiment and analysis. In Section 3 we describe areas in which the data are inconsistent with minimax play. Section 4 contains additional evidence that we believe refutes the hypothesis of i.i.d. play in the experiment. In Section 5 we find that players' behavior did not approach minimax behavior as the experiment progressed. Section 6 concludes with a brief summary and some suggestions for alternative experimental designs that might be more effective in identifying the precise forms of players' departures from minimax play in constant-sum situations.

2. O'NEILL'S EXPERIMENT AND ANALYSIS

In O'Neill's experiment, 25 pairs of subjects played a card-matching game 105 times in succession, face to face.² The players chose their cards simultaneously, revealed them to each other and to a referee, and paid each other according to the cards drawn. Each subject was given an initial money stock of \$2.50. The winner in each stage game received 5 cents, paid by the loser. The stage game was constant-sum, with payoff-structure shown below.

		Column Player			
		1	2	3	<i>J</i>
Row Player	1	–	+	+	–
	2	+	–	+	–
	3	+	+	–	–
	<i>J</i>	–	–	–	+

Here, the symbols 1, 2, 3, and *J* denote the cards ace, deuce, trey, and joker, respectively. The symbol + denotes a win for the row player, while – denotes a loss. From debriefing questionnaires, it seems that the subjects understood the rules of the game.

The stage game has a unique equilibrium: both players randomize with probabilities .2, .2, .2, .4, respectively. The value of the stage game (the probability of a win for the row player in equilibrium) is .4. Since both players can guarantee themselves their shares of the equilibrium payoff, there is no reason, according to the theory, that either player should deviate from the grand strategy of i.i.d. plays of the equilibrium mixture. Indeed, this is the unique equilibrium strategy combination in the repeated game, assuming, for instance,

² The players were not told in advance the exact number of repetitions.

TABLE I
RELATIVE FREQUENCIES OF CARD CHOICES IN O'NEILL'S EXPERIMENT^a

Column Player Choice					Marginal Frequencies For Row Player:
Row Player Choice	1	2	3	<i>J</i>	
1	.044 (.040) [.004]	.043 (.040) [.004]	.043 (.040) [.004]	.091 (.080) [.005]	.221 (.200) [.008]
2	.046 (.040) [.004]	.038 (.040) [.004]	.038 (.040) [.004]	.092 (.080) [.005]	.215 (.200) [.008]
3	.049 (.040) [.004]	.032 (.040) [.004]	.037 (.040) [.004]	.085 (.080) [.005]	.203 (.200) [.008]
<i>J</i>	.086 (.080) [.005]	.065 (.080) [.005]	.051 (.080) [.005]	.158 (.160) [.007]	.362 (.400) [.010]
Marginal Frequencies for Column Player:	.226 (.200) [.008]	.179 (.200) [.008]	.169 (.200) [.008]	.426 (.400) [.010]	

^a Numbers in parentheses represent minimax predicted relative frequencies. Numbers in brackets represent standard deviations for observed relative frequencies under the minimax hypothesis.

that each player wishes to maximize the sum of his winnings. In fact, this would constitute the equilibrium set in the repeated game for any of a large class of ways in which the subjects of the experiment could have been evaluating their payoff sequences, but there was no safeguard in the experiment to ensure that the subjects were actually using members of this class.

Of course, finite samples generated from i.i.d. processes can display sample correlations that are nonzero. Also, previous studies suggest that experimental subjects typically find it difficult to produce i.i.d. sequences (see, for example Baddeley (1966)). For these reasons alone, one might expect to see some departure from the predictions of the minimax model. An objective for anyone studying these data, therefore, is to judge whether the departures observed in the data appear more likely to have been caused by intentional departures from the equilibrium strategies, rather than by chance or unintentional departures.

Table I presents aggregate statistics describing the outcomes of O'Neill's experiment.³ In each interior cell of the table is reported the relative frequency with which the pair of moves relevant to that cell occurred. (The entire experiment involved 105 joint moves for each of 25 pairs of players, so there were 2625 joint moves altogether.) In parentheses below these numbers are listed the corresponding frequencies expected under the minimax model. In brackets below these numbers are listed the standard deviations, under the null

³ O'Neill's reported statistics summarizing card choices appear to contain a few minor errors which do not affect any of his or our conclusions.

hypothesis (minimax), for the corresponding cell relative frequencies. At the bottom and to the right of the table are listed the overall relative frequencies with which each player was observed to play a particular card, again accompanied by the corresponding expected frequencies and standard deviations under the minimax model.

Perhaps the most striking aspect of this table is the rough consistency of the observed relative frequencies with those implied by the minimax model. The relative frequencies observed for pairs of plays involving nonjokers are generally in the neighborhood of .04, while the relative frequencies observed for pairs involving one joker and for pairs involving two jokers are generally in the neighborhoods of .08 and .16, respectively.

As a point of reference, a natural alternative to the hypothesis of strategic behavior is the hypothesis that each player behaved purely randomly, choosing each card with probability .25 on each play. With purely random choice, each pair of plays would occur with probability .0625, and so we would expect the observed relative frequencies in each interior cell of Table I to be roughly equal to .0625. We would also expect each of the "overall" relative frequencies listed in the margins of Table I to be roughly equal to .25. The data in Table I clearly reject this hypothesis, however. In comparison with a model of purely random choice, at least, the minimax model seems to be the more appropriate model.

The aggregate win frequencies observed in O'Neill's games also are consistent with the minimax model. For the entire set of 2625 games, the fraction of wins accounted for by subjects in the role of the row player was .410, which, under the null hypothesis (binomial with parameters .4 and 2625), is only one standard deviation away from the expected value of .4.⁴ Moreover, according to O'Neill, the win frequencies observed for individual pairs were similar enough across pairs of players so that the possibility of differential skill in playing the game could be dismissed.

O'Neill did describe some discrepancies between his data and the minimax model (for example, the tendency for players to avoid repeating a winning card, the propensity to play more aces than expected, and the greater-than-expected variance across pairs of players in the frequency with which jokers were played), but he judged these discrepancies to be minor relative to the overall patterns of play shown in Table I. O'Neill's general conclusion was that the data support the minimax model:

"...players were constrained to follow minimax in its gross statistics because these were relatively observable by the opponent. However at each move players felt free to invent patterns, follow hunches or do other things that introduced dependencies and variance into the sequence of plays. They could do this without significant danger because the opponent had a limited ability to calculate all the relevant probabilities especially when only a small sample of moves was available. But a large deviation from the overall minimax proportion was easier to notice so players avoided the risk of loss by sticking close to the minimax proportions" (p. 2109).

⁴ O'Neill reported .401 instead of .410, an apparent typographical error.

3. A REAPPRAISAL

The minimax hypothesis implies that card choices in O'Neill's experiment represent independent drawings from a multinomial probability distribution with parameters .2, .2, .2, .4; and that each player's choices are independent of the past choices of either player in the pair. With minimax play, row player win frequencies are binomially distributed with probability .4. In this section, we re-examine the degree of support each of these predictions finds in the data.

The overall row winning percentage of .41 in O'Neill's data seems to offer strong evidence of minimax play. However, the structure of O'Neill's game is such that a wide range of non-minimax behavior also can lead to a winning percentage for the row player that is close to .40. For example, even in the extreme case in which players choose their cards (independently) with equal probabilities, the row player would win with probability .438.

Alternatively and somewhat more generally, suppose that the two players recognize that jokers and number cards are not equivalent in the game, but that all number cards have symmetric roles relative to each other. Suppose that, in accordance with this recognition, the row player plays a joker with probability x and plays each of the remaining number cards with probability $(1 - x)/3$, while the column player independently plays a joker with probability y and plays the remaining number cards each with probability $(1 - y)/3$. If we evaluate all possible pairs (x, y) in the unit square, 19.1% of those pairs generate a probability of winning for the row player between .38 and .42. If we restrict ourselves to considering values of x and y that lie in the interval (.2, .8), this rises to 38.3%.

In short, the surface that relates winning percentages to independent strategy combinations is quite flat around the minimax strategy. As a result, tests of the minimax model that are based on observed winning percentages are unlikely to reject the minimax model, even when any member of a wide class of alternative models may in fact have generated the data. In statistical terminology, tests based on observed winning percentages for this game lack power.⁵

Somewhat similar remarks can be made about tests based on observed plays. Although the frequencies in Table I (and especially the greater relative frequency of joker choices) appear to indicate minimax play, these frequencies also could have been generated by some forms of non-minimax play in O'Neill's game. For example, if players tried to convince their opponents that their choices were purely random and then responded optimally whenever they believed that their opponents expected random play, the resulting choice frequencies could be very similar to those in Table I. This follows from the asymmetric nature of the payoff matrix, which leads each player to choose a joker card whenever the player believes that his opponent is responding optimally to purely random play. In particular, if a player chose cards entirely at

⁵ This lack of power is made worse when players are treated individually, as the analysis to follow suggests they should be.

random 80 percent of the time and believed that his opponent expected him to play randomly the remaining 20 percent, the player's card choice frequencies would (in expectation) be the same as minimax. Thus, although the card choice frequencies in Table I are suggestive of minimax play, they can be equally suggestive of play that is not minimax. Like tests based on observed winning percentages, therefore, tests based on observed mixtures may lack power when certain plausible alternative hypotheses are considered.

Despite the limited power of these tests of the minimax null hypothesis against some alternatives, the minimax null fares poorly in O'Neill's data. Here we list our findings.

(i) Even in Table I (which seems to support minimax), O'Neill's data clearly reject the minimax model. Using the absolute frequencies corresponding to the sixteen interior cells of Table I, a Chi-square test for conformity with minimax play (based on Pearson goodness of fit) produces a test statistic of 60.248 (with nine degrees of freedom). This value allows rejection of the minimax model at virtually any level of significance. Thus, although the data in Table I appear "close" to minimax proportions, they are not close enough to have been plausibly generated by minimax play.

(ii) Data at the level of the individual pair provide further evidence against minimax play. Because the number of plays for any single pair of players is much smaller than the total number of plays for all players combined, tests based on data for individual pairs are less precise and are therefore less likely to reject the null hypothesis when some other hypothesis is true. Nevertheless, for a surprisingly large number of pairs, observed behavior differs in a statistically significant fashion from what the minimax model predicts. As seen in Table II, six row players and nine column players chose frequencies that jointly allow rejection of the minimax multinomial model as the process determining their card choices. Twelve pairs played with frequencies that allow rejection of the minimax model for at least one member of the pair. Taking the two players in each pair jointly, nine of the twenty-five pairs played with frequencies that allow joint rejection of the minimax hypothesis. If the minimax model had generated these data, we would expect only one such rejection in any of these sets of twenty-five tests.⁶

To make the same point somewhat differently, for the two hundred relative frequencies in Table II that represent the card selections of the fifty subjects in O'Neill's experiment, thirty-five are sufficiently different from minimax to allow rejection at the five percent level of significance. For the two-hundred frequen-

⁶ These tests are based on comparisons of the variance across players in the observed number of plays for each card with the variance that is predicted by the minimax model. The ratio of these two numbers is distributed as a Chi-square variate with twenty-five degrees of freedom. Our test result for the joker card confirms O'Neill's statement that the observed variation across players in the number of times jokers were chosen is too large to have been plausibly generated by the minimax model.

TABLE II
RELATIVE FREQUENCIES OF CARD CHOICES AND ROW-PLAYER WINS
IN O'NEILL'S EXPERIMENT BY PLAYER PAIR

Pair #	Winning % for Row Player	Row Player Choice				Column Player Choice				Comments
		1	2	3	J	1	2	3	J	
1	.391	.257	.286*	.276	.181*	.229	.210	.210	.352	a, c
2	.295	.171	.181	.210	.438	.152	.152	.143	.552*	b
3	.390	.162	.114*	.181	.543*	.219	.133	.095*	.552*	a, b, c
4	.419	.219	.267	.181	.333	.067*	.086*	.124	.724*	b, c
5	.343	.171	.171	.190	.467	.324*	.086*	.143	.448	b, c
6	.419	.257	.143	.210	.390	.257	.200	.095*	.448	b
7	.476	.238	.229	.229	.305*	.219	.190	.238	.352	—
8	.467	.171	.286*	.219	.324	.267	.248	.190	.295*	—
9	.362	.181	.257	.267	.295*	.257	.181	.219	.343	—
10	.390	.257	.171	.152	.419	.200	.190	.200	.410	—
11	.390	.276	.248	.171	.305*	.229	.200	.200	.371	—
12	.543	.276	.133	.105*	.486	.210	.200	.162	.429	a, c
13	.410	.219	.267	.248	.267*	.114*	.219	.133	.533*	a, b, c
14	.467	.267	.229	.200	.305*	.267	.248	.257	.229*	b, c
15	.324	.200	.181	.162	.457	.295*	.143	.190	.371	—
16	.343	.152	.248	.162	.438	.219	.238	.162	.381	—
17	.362	.200	.219	.219	.362	.229	.171	.190	.410	—
18	.486	.238	.181	.190	.390	.219	.152	.219	.410	—
19	.390	.286*	.190	.200	.324	.171	.162	.162	.505	—
20	.438	.210	.219	.143	.429	.210	.171	.124	.495	—
21	.476	.190	.229	.210	.371	.276	.219	.181	.324	—
22	.400	.200	.162	.181	.457	.286*	.181	.190	.343	—
23	.448	.229	.286*	.324*	.162*	.295*	.181	.105*	.419	a, b, c
24	.495	.248	.257	.238	.257*	.248	.162	.219	.371	a
25	.333	.238	.229	.200	.333	.181	.152	.076*	.590*	b, c

* Denotes rejection (at .05 level) of minimax binomial model for a given card.

^a Denotes joint rejection (at .05 level) of minimax multinomial model for all cards chosen by the row player, based on Pearson statistic and $\chi^2(3)$.

^b Denotes joint rejection (at .05 level) of minimax multinomial model for all cards chosen by the column player, based on Pearson statistic and $\chi^2(3)$.

^c Denotes joint rejection (at .05 level) of minimax multinomial model for all cards chosen by both players, based on Pearson statistic and $\chi^2(6)$.

cies taken as a whole, a Chi-square test for conformity with the minimax model produces a test statistic of 348.24 (with 150 degrees of freedom). If the minimax model had generated the data in Table II, the probability of observing such an extreme value for this statistic would be zero to at least fifteen decimal places. Clearly, the minimax model fails to fit the data at the level of the individual pair.

(iii) Compared with Table I, the data in Table II also show weaker support for minimax when tested against the specific alternative of purely random play. In choosing between these two alternatives, the relevant likelihood-ratio test can be stated in terms of joker probabilities alone. If model choice is based simply on maximum likelihood, the critical value for the test between minimax and purely random play is 34 jokers in 105 plays, or a relative frequency of .324. As seen in Table II, there are 10 row players and 3 column players for whom

purely random play yields a higher likelihood than minimax. Thus, although minimax outperforms purely random play for a majority of players, the margin of preference appears less than Table I might suggest, especially for row players.⁷

(iv) More generally, the data in Table II indicate that no single multinomial model applied to all pairs in common can explain the card choices observed in O'Neill's experiment. This conclusion follows from the excessive variation across players in the frequencies with which each card was played. On the basis of Chi-square tests for the variance in the total number of plays for each card, we reject the hypothesis of homogeneity for the cards deuce and trey at the .05 level of significance. We reject homogeneity for the ace card at the .005 level of significance, and we reject homogeneity for the joker card at virtually any significance level. In the joint test for all card choices by all players, we again reject homogeneity at virtually any level of significance. On the basis of these tests, we conclude that it is inappropriate to view O'Neill's data as having been generated by a single multinomial process.

(v) The observed variance across player pairs in the number of row-player wins is too large to have been plausibly generated from the binomial model implied by minimax play. The minimax model predicts that the number of row-player wins in a set of 105 games is distributed with a variance of 25.2. In O'Neill's data, however, the sample variance for the number of row-player wins is equal to 42.7. On the basis of the appropriate Chi-square test (25 degrees of freedom and a Chi-square value of 40.67), the difference between these two numbers allows rejection of the minimax null hypothesis at the .025 level of significance.⁸

(vi) Table II shows no general tendency for a player to benefit from playing minimax frequencies. Of course, if the opponent were playing minimax, the player's own choice would have no effect on his payoff. Here, however, the data seem to indicate that deviation from minimax frequencies is profitable! To use an extreme example, the pair with the highest winning percentage for the row player (pair 12) is a pair for which the row player deviated significantly from the minimax frequency, while the column player did not. At the other end of the spectrum, the pair with the lowest winning percentage for the row player

⁷ The asymmetry between row and column in these comparisons may itself reveal something about players' departures from minimax. To illustrate, suppose that the row player chooses cards purely at random. Suppose the column player does likewise, except on $X\%$ of the plays, for which the column player can tell whether the row player has chosen a joker card. In this case, the column player will choose the joker card with overall probability $(.75)(X) + (.25)(1 - X)$, and expected choice frequencies for the column player will be exactly minimax if X is equal to .3. If row and column players are reversed in this scenario, however, all cards will still be chosen with equal probability. This asymmetry results from the game's payoff structure, which encourages row players to match, but column players to complement, the joker choices of their opponents.

⁸ This rejection appears to contradict O'Neill's statement that, based on win-frequency statistics, differential ability (in the sense of heterogeneous winning probabilities) is unlikely to be present in these data.

(pair 2) is a pair for which the column player deviated significantly from the minimax frequency, while the row player did not. These findings are inconsistent with the hypothesis of i.i.d. play, and they suggest to us that this hypothesis should be questioned more carefully.

4. THE I.I.D. HYPOTHESIS

In this section, we ask whether players' observed card choices can be modeled as i.i.d. drawings from pair-specific stationary multinomial distributions. We begin by comparing observed winning percentages with those winning percentages implied by temporally and contemporaneously independent play, using the observed card frequencies. The difference between these two numbers measures the contribution to payoffs of contemporaneous correlation in the card choices made by players in each player pair.

Listed in Table III are the observed winning percentages for each player pair, accompanied by the winning percentages that would have resulted if players' card choices had been statistically independent of opponents' card choices. Also listed are the separate contributions of chosen mixtures (i.e., marginal frequencies) and correlated play to observed row-player winning percentages. The former is measured by the difference between the row winning percentage implied by independent play based on observed mixtures and the minimax value of .40. The latter contribution is measured as the residual difference between the observed row winning percentage and the percentage predicted by independent play.⁹

It is clear from Table III that correlated play is the dominant determinant of the difference between observed row-player winning percentages and the minimax predicted value of .40. The average absolute contribution of mixtures to this difference is .011. In contrast, the average absolute contribution of correlated play is .052. A Chi-square test for the joint significance of the correlation effects listed in Table III yields a value of 40.428, which is statistically significant at the .026 level. The corresponding Chi-square statistic for the mixture effects listed in Table III is 2.655, the probability value for which is virtually unity.¹⁰ Clearly, to understand deviations from minimax winning frequencies in O'Neill's experiment, we must focus on the determinants of contemporaneous correlation in players' card choices.

We therefore turn to the question of how the correlated play evident in Table III might have arisen. In particular, we investigate the possibility that past plays (both own and opponent's) were used to condition current plays. For simplicity

⁹ The fact that the residual difference represents the effect of correlated play is easily seen by noting that for any particular pair of card choices, differences between observed and expected winning frequencies arise solely from differences between observed joint probabilities and the corresponding product of two marginal probabilities. These latter differences, however, are simply the covariances between the two players' choices of the cards in question.

¹⁰ This finding derives from the flatness (with respect to mixtures) of the payoff surface in O'Neill's game when cards are chosen independently.

TABLE III
OBSERVED AND EXPECTED WINNING PERCENTAGES
UNDER THE INDEPENDENCE HYPOTHESIS^a

Pair #	Observed Row Win %	Expected Row Win %	Mixture Effect	Correlated Play Effect
1	.391	.418	.018	-.027
2	.295	.410	.010	-.115
3	.390	.437	.037	-.046
4	.419	.366	-.034	.054
5	.343	.406	.006	-.063
6	.419	.397	-.003	.022
7	.476	.408	.008	.069
8	.467	.414	.014	.053
9	.362	.413	.013	-.051
10	.390	.400	.000	-.010
11	.390	.403	.003	-.013
12	.543	.401	.001	.142
13	.410	.368	-.032	.042
14	.467	.427	.027	.040
15	.324	.395	-.005	-.071
16	.323	.396	-.004	-.053
17	.362	.400	.000	-.038
18	.486	.398	-.002	.087
19	.390	.386	-.014	.004
20	.438	.402	.002	.037
21	.476	.405	.005	.072
22	.400	.393	-.007	.007
23	.448	.402	.002	.046
24	.495	.407	.007	.088
25	.333	.377	-.023	-.043
Mean Absolute Value:	—	—	.011	.052
Chi-Square Statistic:	—	—	2.655	40.428
Chi-Square Prob. Value:	—	—	.999	.026

^a Expected winning percentages assume independent play with mixtures observed over all 105 games for each pair. Mixture effects are defined as the difference between values in column (2) and .40. Correlation effects are defined as the difference between values in columns (1) and (2).

and because of limited sample sizes, we focus on that aspect of O'Neill's game that involves the choice of a joker card.¹¹ Under the minimax hypothesis, each player's choice of a joker card should be independent of past joker choices made by that player and by his opponent. Our work here tests whether this condition holds, and when it does not hold, identifies the source(s) of the failure.

¹¹ A multinomial model would provide a more general and more complete framework within which to study player behavior. For O'Neill's game, however, the data can be viewed as having been generated by a sequence of binary decisions which jointly result in an equivalent multinomial model. Because the joker card has a unique role in O'Neill's game and because the joker card is played roughly twice as frequently as any other card, a focus on the choice of a joker seems to us the most informative starting point. Further analysis would then focus on the choice of other cards, conditional on not having chosen a joker. Unfortunately, for O'Neill's experiment, the data are too sparse to allow precise results at these subsequent stages.

TABLE IV
RESULTS OF SIGNIFICANCE TESTS FROM LOGIT EQUATIONS
FOR THE CHOICE OF A JOKER CARD

Estimating Equation ^a : $J = G[a_0 + a_1 \text{lag}(J) + a_2 \text{lag}2(J) + b_0 J^* + b_1 \text{lag}(J^*) + b_2 \text{lag}2(J^*) + c_1 \text{lag}(J)\text{lag}(J^*) + c_2 \text{lag}2(J)\text{lag}2(J^*)]$		
Null Hypothesis	Player Pairs Whose Behavior Allows Rejection of the Null Hypothesis at the .05 Level	
(1) $a_1, a_2, b_0, b_1, b_2, c_1, c_2$ all = 0	Row: 2, 5, 7, 8, 10, 11, 12, 14, 16, 17, 20, 21, 22 Column: 2, 4, 6, 7, 8, 9, 10, 11, 12, 14, 17, 18, 19, 20, 21, 23, 24, 25	
(2) $a_1, a_2 = 0$	Row: 6, 7, 8, 10, 12, 17, 21, 22 Column: 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17, 19, 20, 23, 24, 25	
(3) b_1, b_2, c_1, c_2 all = 0	Row: 4, 5, 7, 8, 10, 11, 12, 14, 16, 17, 21, 22, 23, 25 Column: 1, 2, 17, 18, 19, 21, 24, 25	
(4) $c_1, c_2 = 0$	Row: 8, 9, 10, 11, 12, 14, 25 Column: 2, 6, 9, 17, 21, 24, 25	
(5) $b_1, b_2 = 0$	Row: 4, 5, 7, 10, 12, 14, 16, 17, 21, 23, 25 Column: 1, 2, 17, 19, 21, 25	
(6) $b_0 = 0$	Row: 2, 4 Column: 2, 4	

^a The symbols J and J^* denote the choice of a joker card by a player and by his opponent, respectively. The function $G[x]$ denotes the function $\exp(x)/[1 + \exp(x)]$. Rejections are based on likelihood-ratio tests.

To assess the role of lagged choices in determining current choices, we first estimated for each player a logit equation in which the dependent variable was a dichotomous indicator of the choice of a joker card. The independent variables in these equations were: first and second lagged indicators for both players' past choices; an indicator for the opponent's current choice (included to allow for the possibility that a player might be able to "read the face" of his opponent); and first and second lags for the product of the two players' choice indicators (included to allow for the possibility that successful past plays might affect current plays differently than would unsuccessful past plays).^{12,13} We then performed tests of significance for the joint influence of lagged own choices, lagged opponent's choices, contemporaneous opponent's choices, and lagged interaction terms in these equations. The results of these tests are summarized in Table IV.

Table IV contains the results of six tests. The first test measures the joint significance of all explanatory variables in accounting for a player's choice of the joker card. According to the minimax model, all of the explanatory variables

¹² In his discussion, O'Neill mentions that such differential effects appear to have been present in his data.

¹³ We have experimented with the inclusion of additional (third through sixth) lags in the equations underlying Table IV, and we have found these lags to be statistically unimportant in explaining current choices for all but four of the fifty players in O'Neill's experiment. If Table IV were based on equations that included these extra lagged terms, the row player in pair 6 and the column player in pair 3 would be added to the list of rejections for test #4. The column player for pair 3 would also be added to the list of rejections for test #2.

should be extraneous, and so we should be unlikely to find many cases in which these variables appear to be important in explaining players' choices. It is clear from test #1, however, that the statistical significance of these variables in explaining players' choices is the rule rather than the exception. For nineteen of the twenty five pairs, at least one player's behavior is significantly determined (at the .05 level) by the set of explanatory variables included in our equations.¹⁴

The rejections reported for test #1 result solely from the presence of play that is correlated either with past choices or with the current choice of the opponent. The intercept term is not constrained in this test, so the rejections reported for test #1 are actually rejections for any binomial model as the source of the observed data. If, instead, we maintain as our null hypothesis the particular binomial model implied by minimax play, an additional five row players and two column players are added to our list of rejections for test #1, leaving only one pair (pair 15) for which both players pass our first test and at the same time play joker cards with a probability that is statistically not distinguishable from .40.

Since the minimax model fails to fit the data on joker card choices, we next try to uncover the source of this failure. We distinguish between the influence of a player's own past choices and the influence of all other terms involving the play of the opponent, focusing first on the explanatory importance of a player's own past two choices in determining his current choice. The results (test #2) indicate that a substantial number of players made card choices that were significantly related to their own previous choices. For eight row players and sixteen column players, one can reject (at the .05 level) the null hypothesis that own past choices are uncorrelated with current choices.

Our findings here are consistent either with the notion that players deliberately played in an autocorrelated manner, or that, although wishing to do so, players were unable to achieve independence in their choices. We are unable to choose between these alternatives with the data at hand, and so we hesitate to take the results of our second test as any clear failure of the minimax model as a model for players' intended behavior. Note, however, that the observable correlations found in players' choices could have been exploited by their opponents. We therefore look for evidence that opponents' choices as well as own past choices influenced current play.

The third test summarized in Table IV measures the significance of terms involving the opponent's lagged joker choices in determining a player's current joker choice. If players intended to play minimax and if players believed their opponents to be playing minimax, then, although a player's own lagged play might influence his current play through an inability to achieve randomness, the various terms involving opponent's lagged plays should not influence that

¹⁴ For a large number of these pairs, the minimax null hypothesis can be rejected at much smaller levels of significance. At the .01 level, for instance, the minimax null is rejected for the row player in ten pairs and is rejected for the column player in twelve pairs.

player's card choices. Only if the player believed that his opponent was not playing minimax should these terms influence the player's current card choices.¹⁵

The results of the third test are that for fourteen row players and eight column players, the set of terms involving opponents' lagged plays are either jointly significant or include individual terms that are statistically significant at the .05 level in determining current joker choices. For nineteen pairs, moreover, at least one player's choices are significantly determined by the opponent's choices. We take this finding as strong evidence against the minimax model. If we cannot reject the importance of opponents' plays in determining players' card choices, then we must reject the notion that the players themselves believed their opponents to be playing minimax.

The remaining tests reported in Table IV focus more closely on the specific means by which opponents' plays influenced players' current choices. Test #4 concerns the significance of the lagged interaction terms included in our estimating equations. In his discussion of the data, O'Neill indicated that players seemed less likely to repeat a card choice if that choice had been successful on the previous play. Our tests confirm this statement. For seven row players and for seven column players, the effect of own past choices on current choices is statistically dependent on the corresponding past choices made by the opponent, and for all but a few of these players, the sign of the effect is as suggested by O'Neill. For eleven player pairs, lagged interactions are significant for at least one member of the pair. In O'Neill's discussion, this aspect of player behavior was not emphasized as an important departure from minimax predictions. If this were the sole source of dependence on opponents' past choices, we might agree with this perspective. Our next test, however, shows that the results of test #4 are only symptomatic of a more general dependency on opponents' past plays.

The fifth test reported in Table IV concerns the significance of the linear terms involving opponents' lagged plays. If players attempted to predict their opponents' current choices partly on the basis of the opponents' past choices, one would expect these terms to influence observed card choices. As indicated in Table IV, the data are consistent with this notion for many players. For eleven row players and for six column players, the linear terms involving the opponent's past two plays are either jointly significant or include one statistically significant determinant of the current card choice. For fourteen pairs, at least one player's behavior is significantly influenced by these terms.

The final test reported in Table IV concerns the ability of one player to discern the current choice of his opponent, even after controlling for the

¹⁵ One might argue that these cross terms also represent purely inadvertent effects of the sort described in the psychological literature. If one takes this view, however, then one must recognize that O'Neill's experiment can be of little value in testing the minimax hypothesis, for it provides no means by which the observer can distinguish strategic effects from "psychological" effects. We note also that the signs of the effects measured in our fifth test are frequently consistent with the inference that players' choices were guided by attempts to predict their opponents' current choices from their opponents' past choices.

influence of past choices. Specifically, the sixth test measures the contemporaneous correlation between row and column players' choices of the joker card, conditional on the particular configuration of past choices that is observed. Our results indicate that for two player pairs, "face reading" by one player may have occurred for the joker/nonjoker dimension of the experiment. We hesitate, however, to make much of this finding, for the probability values associated with these effects are not terribly small (.023 for pair 2 and .034 for pair 4), and in a group of twenty five pairs, one would expect to see one or two such rejections purely by chance. With reference to this last test, therefore, our conclusion is that "face reading" was not widespread in the joker/nonjoker dimension of O'Neill's game.¹⁶

After evaluating the six tests reported in Table IV, we are left with seven row players and four column players whose choice of joker cards appears consistent with some binomial model. If we further restrict ourselves to that binomial model implied by minimax play, we are left with only three row players and the same four column players whose behavior is fully consistent with the null model. For only one pair of players is the behavior of both players consistent with a binomial model, and for no pair is the behavior of both players fully consistent with the minimax binomial model.

To this point, we have found little support for the minimax model, but our work thus far has assumed stationarity in the process generating players' choices. In the following section, we relax this assumption and look for evidence that the process generating players' choices approached the minimax process as the experiment progressed.

5. LEARNING MINIMAX?

Despite the evidence that players in O'Neill's experiment did not play minimax and apparently did not think their opponents played minimax, if we observed that such evidence was concentrated in the early stage games, we might retain some confidence in the minimax model as a useful source of predictions for player behavior. To address this question, we have repeated much of our earlier analysis, but this time separating out the final fifty stage games played by each pair. Since there had been fifteen practice games allotted to each pair before money was at stake, each pair had already played seventy times before these observations of their behavior took place. If there were a tendency for player behavior to approach that implied by the minimax model, it seems to us that with this much experience some signs of it should begin to show.

Table V presents a decomposition of winning percentages similar to that presented in Table III, but this time differentiating between the first fifty-five and last fifty games played by each pair.

¹⁶ Of course, "face reading" may have occurred in other dimensions of the game. We believe that some such face reading did occur for pair #12. For this pair, we observe a sequence in which the row player won seventeen out of eighteen games, and we observe contemporaneous correlations in the choices of aces and number cards that are significant at the .0001 level.

TABLE V
WINNING PERCENTAGES, EFFECTS OF MIXTURES, AND EFFECTS
OF CORRELATED PLAY BY PLAYER PAIR AND STAGE OF PLAY^a

Pair #	Games 1-55			Games 56-105		
	Row Win %	Mixture Effect	Correlated Play Effect	Row Win %	Mixture Effect	Correlated Play Effect
1	.473	.018	.055	.300	.020	-.120
2	.327	-.036	-.036	.260	.020	-.160
3	.382	.036	-.055	.400	.040	-.040
4	.382	-.036	.018	.460	.000	.060
5	.364	.018	-.055	.320	.000	-.080
6	.382	-.018	.000	.460	.000	.060
7	.527	-.018	.145	.420	.020	.000
8	.400	.036	-.036	.540	.020	.120
9	.309	.000	-.091	.420	.000	.020
10	.291	-.018	-.091	.500	.000	.100
11	.327	.000	-.073	.460	.020	.040
12	.600	.018	.182	.480	.000	.080
13	.364	-.036	.000	.460	-.040	.100
14	.436	.018	.018	.500	.060	.040
15	.400	.000	.000	.240	.000	-.160
16	.309	-.036	-.055	.380	.000	-.020
17	.418	.000	.018	.300	.020	-.080
18	.491	.018	.073	.480	-.020	.100
19	.473	.018	.055	.300	.000	-.100
20	.491	.000	.091	.380	.020	.000
21	.436	.036	.000	.520	.000	.120
22	.473	-.018	.091	.320	-.020	-.060
23	.455	-.018	.073	.440	.020	.060
24	.527	.000	.127	.460	.000	.060
25	.273	-.055	-.073	.400	-.020	.020
Mean Absolute Value:	—	.020	.060	—	.014	.072
Chi-Square Statistic:	—	1.594	35.346	—	1.711	39.545
Chi-Square Prob. Value:	—	.999	.082	—	.999	.032

^a "Mixture" and "Correlated Play" effects are defined as in Table III.

From Table V we can see that there was no manifest tendency for O'Neill's subjects to approach minimax winning percentages as the experiment progressed. For thirteen pairs, winning percentages over the final fifty games are farther in absolute value from .40 than the corresponding winning percentages for the first fifty-five games. For eleven pairs, the measured effects of choosing a non-minimax mixture are larger in the final fifty games than in the first fifty-five, and for fourteen pairs the measured effects of correlated play are larger in the final fifty games than in the first fifty-five games (although these differences are not generally statistically significant when pairs are studied individually).

In the first half of the experiment, the overall deviations of row winning percentages from .40 are jointly significant at a level of .038, with the contributions of correlated play jointly significant at a level of .082. In the latter half of the experiment, the overall deviations of row winning percentages from .40 are jointly significant at the slightly higher level of .062, but the contributions of correlated play are now jointly significant at the level of .032. In neither half of

the experiment are the contributions of mixtures to “excess” winnings anything but negligible. The average absolute deviation from .4 attributable to mixtures is .020 for the first fifty-five games and .014 for the last fifty games. In contrast, the average absolute deviations attributable to correlated play in these two sets of games are .060 and .072, respectively.

The conclusion that players’ behavior showed no general tendency to approach minimax is also supported by a direct look at the card choices made by the players, but we shall spare the reader the details. In short, the story told by the data for the experiment taken as a whole is also told by the data for the final fifty games of the experiment. We can find no evidence that players came any closer to minimax play as the experiment progressed.

6. SUMMARY AND CONCLUSIONS

If we are to take the minimax model seriously as a source of empirically testable hypotheses, we must assume that the data have been generated from the probability model implied by minimax play. We need not expect the data observed in any single sample (such as O’Neill’s experiment) to conform exactly with minimax predictions, but we should expect observed discrepancies to be consistent with random deviations implied by the model itself. Applying this standard, we find little support for the minimax hypothesis in O’Neill’s data.

We have studied O’Neill’s data along a number of dimensions, and we have found almost uniform rejection of the minimax hypothesis. We emphasize that these rejections are in no way marginal. Using probability values as a measure of coherence with the theory, the data reject minimax very clearly. These rejections are not simply the product of large samples. Indeed, these rejections generally occur despite the low power associated with pair-by-pair tests and despite the fact that some forms of non-minimax behavior can generate mixtures and win frequencies consistent with minimax play in O’Neill’s game.

Of course, one might argue that some weaker interpretation of the theory is more appropriate and that the data reflect play that is “close” to minimax, if not “exactly” minimax. However, our reading of the data suggests that any definition of “closeness” that would make the data support minimax also would make the data support a number of other models having nothing to do with minimax play. Such a definition would have to admit large departures from minimax mixtures for roughly half the player pairs. It would have to admit correlated play for almost all pairs, with more than half the players apparently rejecting temporally independent play as the model for their opponents’ behavior. It would also have to allow for the peculiar asymmetry in the pattern of departures from minimax seen in Table II. With such a broad definition, we think little would be learned from the data. For example, if row choices in Table II can be seen as “close” to minimax, they may also be seen as “close” to random. We see no value in making the theory so imprecise that the data at hand can no longer discriminate effectively between alternative models. To us, the empirical value of the minimax hypothesis lies precisely in its ability to be rejected by the data. It is from these rejections (both their source and strength) that we learn how theory and future experiments should be amended.

In O'Neill's data, we think the most informative rejections come from the interdependent choices shown by so many player pairs, although the effect of this behavior on payoffs is relatively small. Since such behavior must be counterproductive for one of the players, we would be interested to learn to what extent such behavior is sensitive to variations in the experimental design. The following changes in the experiment appear worth pursuing.

First, we would like to see similar experiments repeated with higher stakes. Although O'Neill reported that the subjects felt involved, it is not difficult to imagine that exploring patterns of behavior in play could have been a nonnegligible factor in their involvement, instead of there being a singleminded focus on achieving highest possible payoffs. Second, we tend to agree with O'Neill concerning the advantages of his face-to-face format, but we would like to know if more anonymous play through computer terminals would alter the form and severity of the serial correlations. If so, it might cause one to rethink the usefulness of the computer-terminal format in many ongoing game experiments which aim to simulate market conditions that do not involve anonymity. (We suspect that both higher stakes and more anonymity would tend to eliminate the contemporaneous correlations exhibited by pairs 2, 4, and 12 in the experiment.) Third, we suggest that some kind of rotation of opponents during the experiment might produce useful benchmarks for studying serially correlated play. With rotation, players ought to feel freer to condition on their own past choices as an aid in achieving any desired move frequencies. On the other hand, they would have, in principle, nothing to learn about the current opponent by analyzing past moves of other opponents. Finally, although O'Neill's stage game possesses the advantageous features he described, it has the disadvantage of a relatively flat payoff surface near the equilibrium combination. Also, it can generate choice frequencies indistinguishable from minimax frequencies with plausible non-minimax strategies. These two features limit the power of tests deriving from O'Neill's experiment and indicate to us a need for some more explicit attention to the operating characteristics of tests based on alternative payoff structures.

Department of Economics, State University of New York at Stony Brook, Stony Brook, NY 11794, U.S.A.

and

Department of Economics, Boston University, 270 Bay State Rd., Boston, MA 02215, U.S.A.

Manuscript received July, 1988; final revision received December, 1989.

REFERENCES

- BADDELEY, A. D. (1966): "The Capacity for Generating Information by Randomization," *Quarterly Journal of Experimental Psychology*, 18, 119–128.
 O'NEILL, B. (1987): "Nonmetric Test of the Minimax Theory of Two-person Zerosum Games," *Proceedings of the National Academy of Sciences, U.S.A.*, 84, 2106–2109.