One-play, two-play, five-play, and ten-play runs of Prisoner's Dilemma[1]

L. G. MOREHOUS
Mental Health Research Institute, The University of Michigan

The subjects in this experiment were 200 undergraduate women enrolled in a beginning psychology course at The University of Michigan. The subjects were paired using criteria aimed at minimizing the probability that the subjects were previously acquainted. We tried not to pair subjects who lived in the same residence hall. It was impossible to follow this criterion completely, but in no pair did the subjects live on the same floor, and in only one pair did the subjects appear to have been previously acquainted. This latter pair was dropped from the analysis. One subject in each of two other pairs was familiar with the Prisoner's Dilemma game, and these two pairs were also dropped. Furthermore, the last pair was not used in the analysis in order to limit the total to 96 pairs. The reason will be mentioned presently.

The game matrix, used in a previous study (Morehous, 1964), was as follows:

	C	D
C	5, 5	–10, 10
D	10, –10	–5, –5

Every pair of subjects played a one-trial game, a two-trial game, a five-trial game, and a ten-trial game. In order to compensate for any effect due to the position in which the game was played, the order of the games was varied systematically. Position, as used here, means that a game with a given number of trials was played first, second, third, or fourth in the sequence. The four different games can thus be ordered in 24 ways, with each game being played in each position six times. To equalize the effects that position might have, the number of subject pairs had to equal some multiple of 24. Thus the number of pairs was limited to 96, and each game was played in each position 24 times—enough for reliable analysis.

The payoff to a subject was the algebraic sum of the points she had earned in k trials (k equalling the length of a run), divided by k. Therefore each game was worth a maximum of ten points and a

[1] My special thanks go to Dr. Anatol Rapoport for his advice concerning the development of this study and to Phillip Dale for his work in programming the computer operations.

minimum of minus ten points. The points were then converted into cents on a one-to-one ratio. This procedure served to make each game worth the same amount of money, regardless of the number of trials. A payoff was made at the end of each game. If the subject had a positive score she got money from the experimenter; if she had a negative score she paid the experimenter.

If, after completing these four different games, either subject had actually lost money to the experimenter, the following game matrix was introduced:

	C	D
C	15, 15	0, 25
D	25, 0	10, 10

The new game was played as before except that a payoff occurred after each trial rather than after each game. The pair continued to play this game until both subjects had at least as much money as they had had before the experiment began. (Since the subjects were unpaid and required to participate for course credit, it was felt that they should not lose their own money in the experiment.)

A written feedback explaining the experiment was mailed to each subject after the data from all the subjects had been collected.

Results

If, in a given trial, a subject chose either the top row or the left column of the matrix, her response is indicated by C (cooperation). If she chose either the bottom row or the right column, her response is indicated by D (defection). Thus the outcome of any trial can be represented by CC, CD, DC, or DD. CC and DD may be thought of as bilateral cooperation and bilateral defection, respectively. CD and DC are essentially the same outcome in reverse, unilateral cooperation or unilateral defection. Unilateral cooperation or unilateral defection will be represented here by UNI (= ½ [CD+DC]). Total cooperation, the single letter C, is equal to CC + ½(CD+DC).

The first question raised by the collected data is whether the position in which a particular length of run was played had any effect on the level of cooperation. To test the null hypothesis, a median test for significance was run for each of the four parameters in each of the games. Table 1 presents the chi-squares. All except one are above the .05 level of significance (df = 3). The subjects apparently considered a particular game in the same manner whether it was played first or whether it was played in the light of previous experience. We shall conclude that there was not enough evidence to reject the null hypothesis; the exception will be included in the analysis.

Since we can disregard the position in which a game was played, the means of the four parameters C, CC, DD, and UNI take on greatest interest. In Table 2 we see that, in one- and two-trial games, UNI and CC were about equal; together they formed about one-third of the subjects' responses. Two-thirds of all responses were non-cooperative.

The level of total cooperation (C) in the five- and ten-trial games is slightly above the level of bilateral defection (DD). Again unilateral cooperation and bilateral cooperation are approximately equal, but their values have risen from the shorter games.

The parametric means for five- and ten-trial games are also similar to each other and different from those of the shorter games. The chi-squares in Table 2 make use of this difference, for they were calculated by a median test for significance (df = 1) using only two categories: one- and two-trial games, and five- and ten-trial

TABLE 1
CHI-SQUARES FOR MEDIAN TEST FOR SIGNIFICANCE BETWEEN POSITION AND THE PARAMETERS C, CC, DD, AND UNI

	1-trial	2-trial	5-trial	10-trial
C	4.00	2.04	4.63	0.90
CC	1.48	3.56	5.71	0.16
DD	4.00	1.13	11.10*	5.07
UNI	3.82	5.68	1.86	2.70

* $p < .02$.

TABLE 2
MEAN C, CC, DD, AND UNI FOR 1-TRIAL, 2-TRIAL, 5-TRIAL, 10-TRIAL, AND 300-TRIAL GAMES

	1-trial	2-trial	5-trial	10-trial	300-trial	χ^2	$p <$
C	.33	.31	.43	.42	.38	9.38	.01
CC	.16	.14	.21	.21	.22	64.60	.001
DD	.50	.52	.35	.38	.46	30.58	.001
UNI	.17	.17	.22	.21	.16	0.00	NS

games. One- and two-trial games demonstrate about the same level of cooperation. Between two-trial games and five-trial games cooperation increases.

On comparing the long run of 300 trials with the short-run games we see that bilateral cooperation is about equal in five-trial, ten-trial, and 300-trial games, and that unilateral cooperation in 300-trial games equals that in one- and two-trial games. In the 300-trial games there is a slightly greater degree of bilateral choices than in the shorter games. We might expect this to be the case, since the first few trials of a game are probably devoted to testing different strategies, and later on the choices become more stable. That is, the "lock-in" sets in later in the game.

Figures 1, 2, 3, and 4 show the time courses for two-trial, five-trial, and ten-trial games and for the first ten trials of 300-trial games on each of four parameters. The time courses show little difference between the five-trial game and the first five trials of the ten-trial game. Because of this similarity we shall describe only the time courses of the ten-trial game. As the trials proceed, overall cooperation begins high and then takes a dip, but rises again on the seventh trial. But for a lower frequency of responses, the curve for bilateral cooperation is much the same as that for total cooperation. Thus the frequency of unilateral cooperation remains fairly constant. Bilateral defection tends to rise and then level off.

Comparing the first ten trials of the 300-trial game with the ten-trial game, we see that both curves for total cooperation have the same tendency. The main exception to this is the fact that the curve for the 300-trial game begins much lower.[2]

Summarizing these results, we can say that five-trial games seem to be played in much the same manner as the first five trials of ten-trial games, and in some respects the ten-trial game is similar to the first ten trials of the 300-trial game.

The data just presented have described the overall levels of cooperation and defection and the time courses generated by the parameters, but the dynamics of this conflict situation can be analyzed even further. To do this, we use four other parameters, x, y, z, and w (Rapoport, 1965a). To x we assign the probability of a player's cooperating after an outcome in which both players cooperated; to y, the probability that a player will cooperate after an outcome in which he cooperated and the other player defected; to z, the probability that a player will cooperate after an outcome in which he defected and the other player cooperated; and to w, the probability that a player will cooperate after an outcome in which both players defected.

[2] See the article by Rapoport and Dale in this issue.

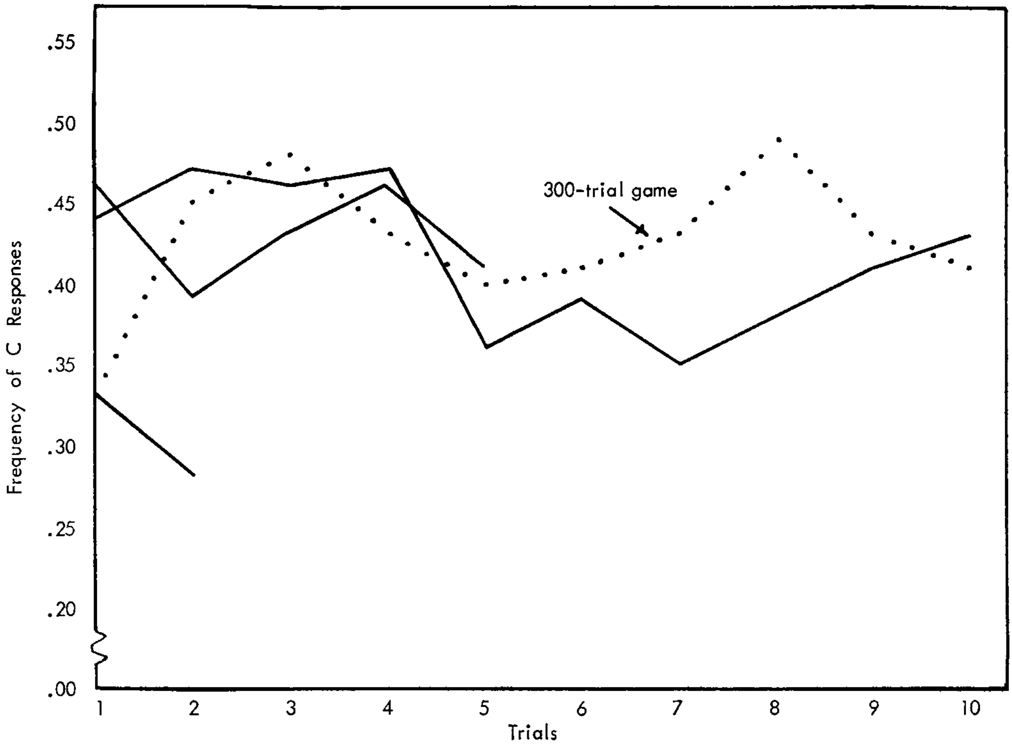

FIG. 1. Time courses for C responses in two-trial, five-trial, ten-trial, and 300-trial games.

These parameters are particularly interesting in light of what they represent. Since x demonstrates the willingness to continue bilateral cooperation, it may be thought of as a measure of "trustworthiness." "Forgiveness" or "teaching by example" may be suggested by y. The next parameter, z, suggests "responsiveness" to the other player's action. Finally, w indicates a player's attempt to break the bilateral defection and his hope of moving to a situation like z, "trust."

Table 3 presents these parameters for two-trial, five-trial, ten-trial, and 300-trial games. The meanings of the parameters may be more fully understood if we go through the two-trial game step by step. If, on the first trial, a player cooperates and his partner does too (CC), the player is likely to cooperate on the next trial ($x = .77$). If, however, he cooperates and his partner defects on the first trial (CD), he is likely to defect on the second trial ($1 - y = .68$). In other words, if a player cooperates on the first trial, on the second trial he is likely to do what his partner did on the first trial ($p = .73$). The correlation (phi coefficient) between the partner's response on the first trial and the player's response on the second trial is .81. In the case where a player defects on the first trial, he is likely to defect on the second trial too ($1 - z = .84$, $1 - w = .83$), regardless of his partner's response on the first trial. The correlation between the partner's response on the first trial and the player's response on the second trial is

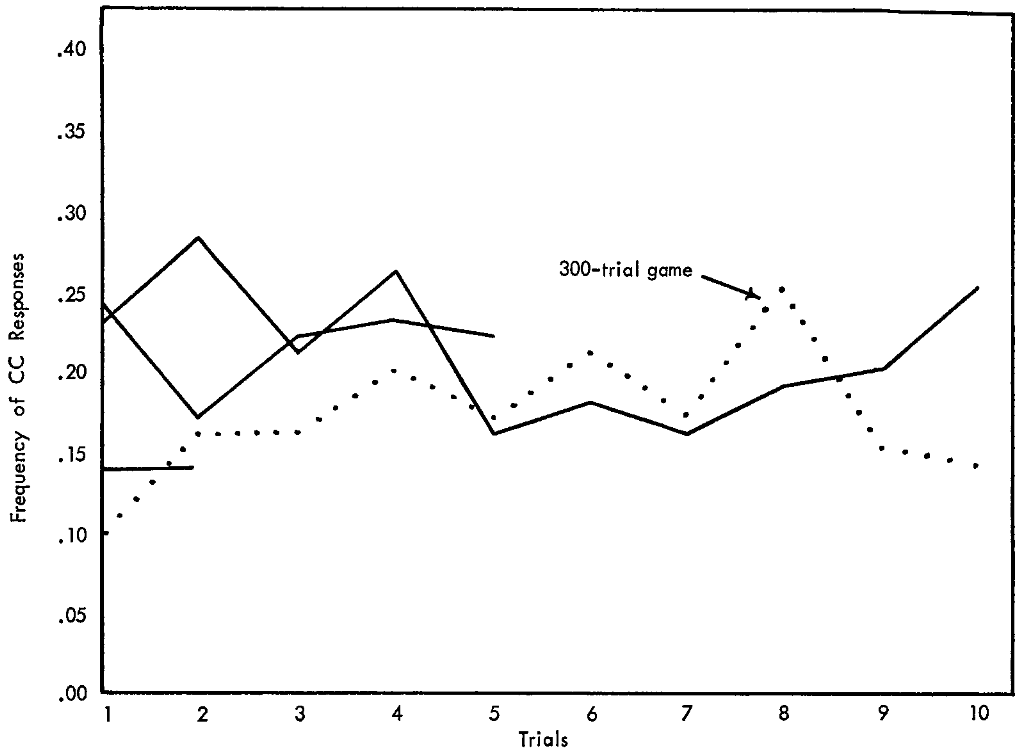

FIG. 2. Time courses for CC responses in two-, five-, ten-, and 300-trial games.

–.024. Once a player defects on the first trial, the probability that he will defect on the second trial appears to be a relatively stable characteristic. Only if a player cooperates on the first trial does his partner's behavior influence his response on the second trial.

This last statement holds true for five-, ten-, and 300-trial games, but the intensity changes. In the two shorter-run games, when a player cooperates on one trial there is not quite so strong a relationship between the partner's response on that trial and the player's response on the next trial. There is also a much greater propensity, in the longer games, to cooperate on a trial when the player has defected on the previous trial. This greater degree of "responsiveness" and "trust" is probably the main reason why five- and ten-trial games have a higher level of cooperation than the two-trial game. If a player defects on the first trial of a two-trial game, the fact that the second trial is the last one apparently leads to a defensive strategy. His reasoning may go like this: "Since I defected on the first trial, my partner will not trust me to cooperate on the second trial, and he will de-

TABLE 3
MEAN x, y, z, AND w FOR 2-TRIAL, 5-TRIAL, 10-TRIAL, AND 300-TRIAL GAMES

	2-trial	5-trial	10-trial	300-trial
x	.77	.61	.60	.73
y	.32	.36	.37	.50
z	.16	.41	.33	.24
w	.18	.36	.31	.20

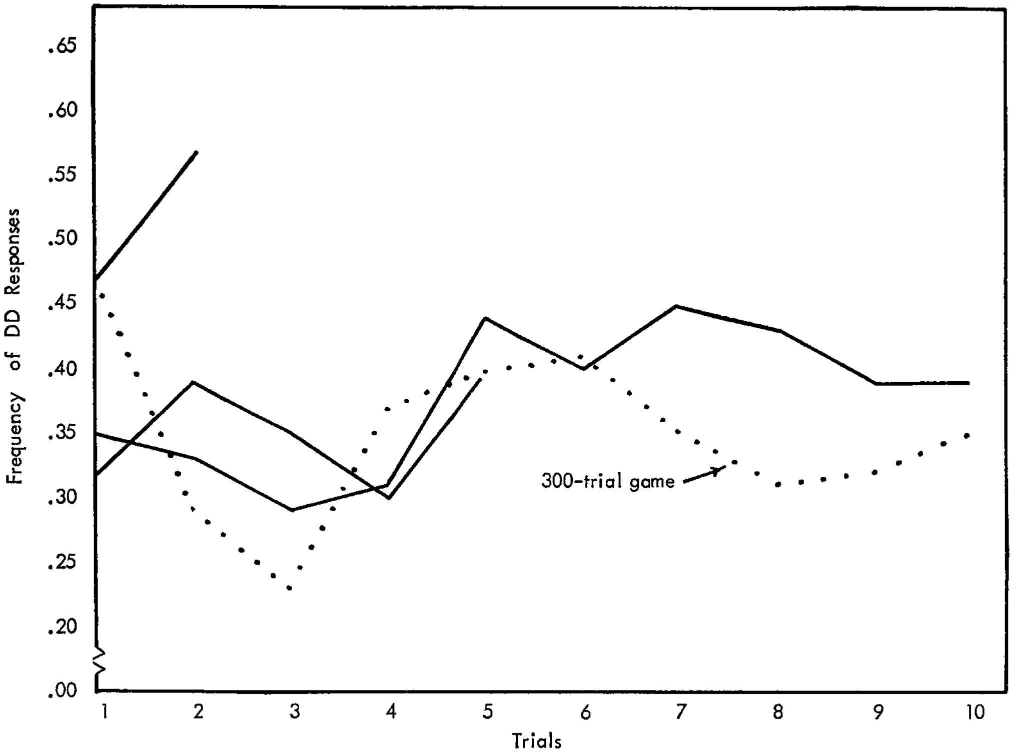

FIG. 3. Time courses for DD responses in two-, five-, ten-, and 300-trial games.

fect; if he defects and I cooperate, I will lose 10 points, but if I defect again I will lose only five points."

In slightly longer runs (five or ten trials) this reasoning is not as compelling, for there are more occasions for a player to show his intention to cooperate and to make up for a monetary loss on one trial. Whether or not one additional trial (a three-trial game) would eliminate this effect is a problem for further investigation.

To continue our analysis of two-trial games, consider for a moment the first two trials of a k-trial game. If a CC or DD outcome occurs on the first trial, then we would expect the players to follow suit on the second trial. This is supported by the high values of x and $1-w$. From the values of $1-y$ and $1-z$ we would expect that first-trial outcomes of CD or DC would lead to more DDs than CCs on the second trial, and so we would expect cooperation to decrease on the second trial. Two-trial and five-trial games support this expectation (Table 1), but the opposite occurs in ten-trial and 300-trial games. Although the foundation is admittedly weak, we suggest that the phenomenon of less cooperation on the second trial than on the first trial is more often true of very short-run games than it is of longer runs.

Still considering two successive trials, let us turn our attention to the last two trials in a game. For one-trial games the defection response is the "reasonable" solution to the Prisoner's Dilemma matrix if the

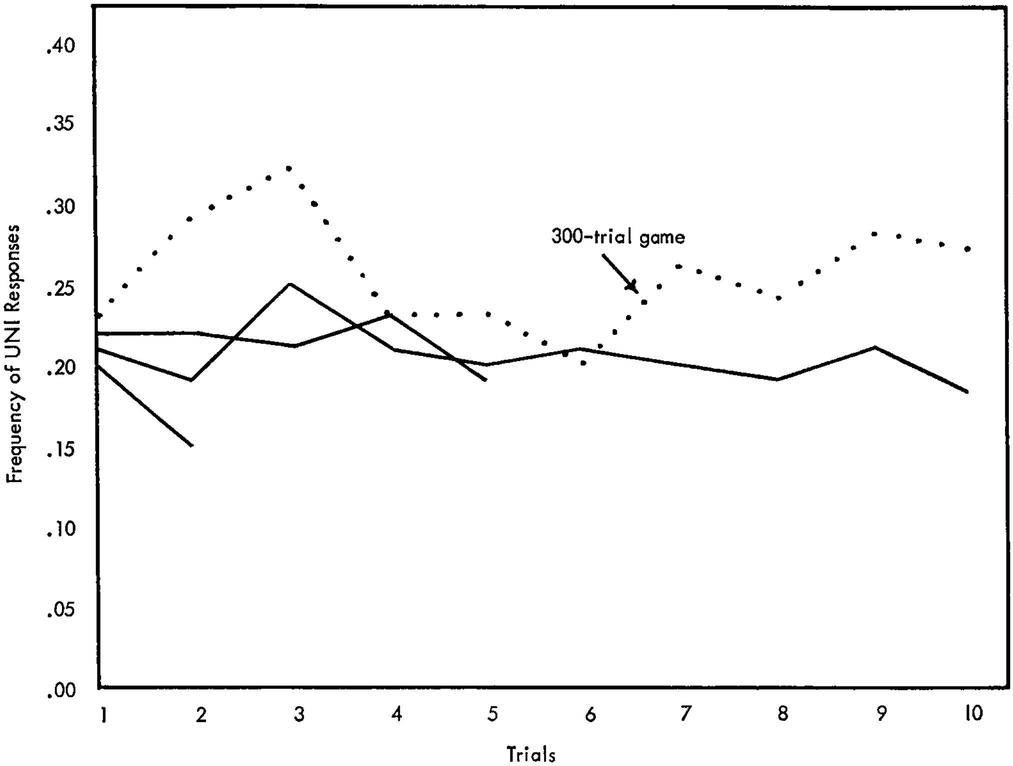

FIG. 4. Time courses for UNI responses in two-, five-, ten-, and 300-trial games.

player has no idea what his partner will do (Luce, 1957). By the same reasoning we may expect cooperation to drop on the last trial of a game—a player may try to get the T payoff (T for "temptation," the reward for being the lone defector), since his partner will not be able to retaliate. The five-trial game supports this hypothesis, for total cooperation drops from .46 to .41 and bilateral defection rises from .30 to .40 between the fourth trial and the fifth (last). Between the last two trials in the 300-trial game, total cooperation drops from .16 to .13 and bilateral defection rises from .29 to .33—an effect which is not so strong but in the same direction. The ten-trial game, however, is an exception to this hypothesis, for cooperation between the last two trials rises slightly. In sum, the results tend to confirm the direction of our expectation but they are still far from conclusive.

In a two-trial game each player bases his first choice only on *a priori* reasoning. His second and last choice, however, is based on the nature of the outcome of the first trial. We have already considered the probability of a player's cooperating after particular outcomes on the first trial. But let us look at the second trial more closely by examining the frequency with which the unique successive outcomes are played. Given the four different outcomes that can occur on each of the two trials, there are 16 successive outcomes, but not all of them are unique. CC/CD is the same as CC/DC,

TABLE 4
PERCENTAGE OF UNIQUE OUTCOMES ON THE FIRST TWO TRIALS IN GAMES WITH RUNS OF DIFFERENT LENGTHS

Unique successive outcomes	2-trial (n = 96)	5-trial (n = 96)	10-trial (n = 96)	300-trial (n = 33)	Total (n = 321)
DD/DD	33%	16%	21%	15%	22%
CD/DD	23	19	11	9	17
CD/CD	10	11	17	24	14
DD/CD	10	14	11	18	12
CC/CC	8	13	18	9	12
CD/DC	4	13	7	9	8
CC/CD	4	7	3	6	5
DD/CC	3	3	3	3	3
CD/CC	2	1	7	6	4
CC/DD	1	4	2	0	2

and DD/CD is the same as DD/DC; likewise, all successive outcomes starting with CD are identical to those starting with DC. All together, there are ten unique successive outcomes.

Table 4 presents the unique successive outcomes in rank-order of their frequency of occurrence in two-trial games. By far the most frequent (56 percent of the subjects) is a bilateral defection after either unilateral or bilateral defection. This propensity to defect in the wake of defection is much stronger than the propensity to cooperate in the wake of cooperation.

The rank-order of preference remains approximately the same for five-, ten-, and 300-trial games. The mean rank intercorrelation among the four games is .78. Most of the variance among the games lies in the three least frequent unique successive outcomes. Subjects tend to play a two-trial game in much the same way they play the first two trials of a five-, ten-, or 300-trial game.

It is also interesting to compare the one-trial game with the first trial of other games. As mentioned earlier, the only basis for a subject's first choice is *a priori* reasoning. However, the number of trials

TABLE 5
MEAN COOPERATION ON THE FIRST TRIAL IN GAMES WITH RUNS OF DIFFERENT LENGTHS

Game	Mean C on first trial
1-trial	.33
2-trial	.33
5-trial	.46
10-trial	.44
300-trial	.33*

* See discussion of this result in the article by Rapoport and Dale in this issue.

he has in prospect may affect his choice on the first trial. Very tentatively we might conclude that the level of cooperation on the first trial does vary with the length of the run (see Table 5). This suggests again that people perceive different lengths of run in different ways.

REFERENCES

LUCE, R. DUNCAN, and HOWARD RAIFFA. *Games and Decisions.* New York: Wiley, 1957.

MOREHOUS, L. G. "The Effect of Authoritarianism on Cooperation." Unpublished undergraduate thesis, The University of Michigan, 1964.

RADLOW, ROBERT. "An Experimental Study of 'Cooperation' in the Prisoner's Dilemma Game," *Journal of Conflict Resolution,* 9, 2 (June 1965), 221–27.

RAPOPORT, ANATOL, and ALBERT M. CHAMMAH. "Sex Difference in Factors Contributing to the Level of Cooperation in the Prisoner's Dilemma Game," *Journal of Personality and Social Psychology,* 2, 6 (Dec. 1965).

———, ———, and CAROL J. ORWANT. *Prisoner's Dilemma: A Study in Conflict and Cooperation.* Ann Arbor: University of Michigan Press, 1965.

SAMPSON, EDWARD E., and MARCELLE KARDUSH. "Age, Sex, Class, and Race Differences in Response to a Two-Person Non-Zero-Sum Game," *Journal of Conflict Resolution,* 9, 2 (June 1965), 212–20.

SCODEL, A., J. S. MINAS, P. RATOOSH, and M. LIPETZ. "Some Descriptive Aspects of Two-Person Non-Zero-Sum Games," *Journal of Conflict Resolution*, 3, 2 (June 1959), 114–19.

TERHUNE, KENNETH W. "Psychological Studies of Social Interaction and Motives (SIAM): Phase 1. Two-Person Gaming." Buffalo, N.Y.: Cornell Aeronautical Laboratory, September 1965.