Chapter 15
Concluding Remarks

WE HAVE SUMMARIZED our experimental work on Prisoner's Dilemma through the end of 1963 and have given the rather sketchy outlines of a mathematical theory purporting to deal with the process consisting of repeated plays. As is frequently the case, the work did not follow a strict predesigned plan. To a great extent, lines of investigation were stimulated by some (to us) interesting results of an experiment just completed; certain lines were discontinued as too costly of time and of subjects. The depletion of the subject pool was a constantly threatening limitation. Once we decided to use naïve subjects, we had to restrict ourselves to using each subject only once. In the studies reported here 740 subjects were used. The switch to female subjects was made partly in order to tap a "virgin" population. But when the first "returns" began to come in, we quickly decided to utilize this opportunity of a cross-population comparative study.

As a rule, ideas for experiments forced themselves on our attention much more rapidly than experiments could be performed. Going at a normal pace, we could run two pairs per day during the five-day week. Thus, to complete one of our conditions typically took seven weeks.

In this final chapter, we shall outline some experiments, which, we feel, ought to be performed and which we intend to perform, but not soon enough to warrant further delays in publishing the results already obtained.

The Strategy as an Independent Variable

In our experiments we were not able to use the strategy of one player as an independent variable with which to ascertain the behavior of the other player. Obviously this could not be done as long as both players were freely-choosing subjects. A controlled experiment using one player's strategy as an independent variable can be performed only if that player is a confederate of the experimenter. The use of stooges in experimental psychology has certain disadvantages. Questions of ethics have been raised with regard to this practice. Quite aside from the ethical issue (if any), however, the use of stooges has certain practical drawbacks, an obvious one being the need for secrecy and the ever-present danger that the subject population will "get wise" to the ruse and so block an entire avenue of exploration.

In experiments such as ours the use of stooges does not present quite so serious a problem. For one thing, it may not even be necessary to conceal the fact that one of the players is "programmed" to play a certain way. For example, the experimenter himself may take the role of the other player. There are several degrees of knowledge that the subject can have about the other player's role. For example, he may not only be told that the other player is to play a prescribed strategy but even informed of the strategy that the other player will play. In this case, then, there is clearly a normative theory (that of simply maximizing one's own payoffs) with which the behavior of the subject can be compared. (Of course, if it is consistently found that the players do in fact maximize their payoffs when they know the other's strategy, the experiments cease to be psychologically interesting.) Next, the subject can be told that the other player has a preassigned strategy

but not what it is. This situation is no longer a simple maximization problem for the subject, and therefore the theory cannot be confined to a simple normative one: it must include psychological components. The disadvantage of this situation is that some subjects may design their choices so as to "experiment" in order to find out the strategy of the other player. Such trial-and-error explorations would be expected to mask the motivations we are interested in. It seems that the most suitable condition is the one in which the subject believes that the other player is also a bona fide subject, and so a certain amount of deception is unavoidable. However, the deception in this case is not "severe": the experiment is not something entirely different from what it purports to be. If word gets around that the "other player" is a confederate of the experimenter, this still does not make the experiment pointless, as it does in other situations with deception. For in our case the subject's problem is still to play as well as he can.

Assume, then, that the subject believes the other player to be a bona fide subject. We can now use the programmed strategy of our stooge as the independent variable. In particular consider the following class of strategies.

1. *The stooge always plays C.* One would expect that many subjects, perhaps the majority, will take advantage of such a player and exploit him. In other words, we may expect the response to a fully cooperative strategy to be mostly an uncooperative one.

2. *The stooge always plays D.* Here we can expect even with greater confidence that the subject will play mostly *D*. For to play *C* against *D* consistently takes an altogether ardent dedication to cooperation. At any rate, while "martyr runs" are seen to occur in

repeated games of Prisoner's Dilemma, very long such runs are quite rare.

3. *The stooge randomizes his choices*, playing a variable proportion of C. This proportion is then our independent variable. The two pure strategies mentioned above are special cases with $C = 1$ and $C = 0$.

If we are right in our conjecture that cooperation in response to both $C = 1$ and against $C = 0$ will be minimal,[40] it follows that maximal cooperation will be elicited by some intermediate value of C. It would be interesting to see what this optimum mixture (optimum in the sense of eliciting maximum cooperation) is. On the other hand, a mixture may be optimal in the sense of maximizing the payoff of the "mixer" (i.e., of the stooge) and this optimizing strategy may not be identical with the one which maximizes the subject's frequency of cooperative responses.

4. *The stooge plays a tit-for-tat strategy*. As the name implies, a tit-for-tat strategy is one which apes the other player: one plays whatever the other played the last time.[41] By convention we can agree that the first choice of this strategy is C. Departures can now be made from the tit-for-tat strategy in either the direction of greater cooperation or in the direction of greater defection. In the former case one always responds cooperatively to the other's C and, in addition, to a certain fraction of the other's D's. This fraction is now the independent variable. Deviating in the direction of defection, one always responds by defecting to the other's defection, and also to a certain fraction of the other's cooperative responses. This fraction is then the corresponding negative value of our variable. The tit-for-tat strategy is the special case where the value of this variable is zero.

5. An important class of strategies is one in which

the results of the first trials become the principal independent variable. For example, suppose one cooperates for the first N trials and thereafter plays a tit-for-tat strategy. Here N is the independent variable. Or suppose one defects for the first N trials and thereafter plays the tit-for-tat strategy. How large must N be for the tit-for-tat strategy to become ineffective in eliciting cooperation (we assume the one hundred percent tit-for-tat strategy will elicit cooperation rather effectively)? The initial N cooperative responses can also be combined with a subsequent completely defecting strategy in order to see how long the initial impact of cooperation takes to wear off. It is assumed that *initially* the one hundred percent cooperative strategy tends to induce cooperation, although not in the long run. Or one may ask, how large does the initial totally defecting run have to be in order for subsequent "therapy" (either cooperative or tit-for-tat) to become futile?

False Information about the Payoff Matrix

Somewhat more serious deception is involved when wrong information is given to the subjects about the payoffs. Obviously such wrong information can be given only about the payoffs of the other player. We have performed one experiment of this sort. Consider the game shown in Matrix 17.

	C	*D*
C	1,1	−2,50
D	2,−50	−1,−1

Matrix 17.

Here the game is no longer symmetric. For $S_1 \neq S_2$ and $T_1 \neq T_2$. In fact it appears to the first player that the second player's temptation parameter and the mag-

nitude of his sucker's payoff are very much greater than his own. In our experiment, both players had the same impression, i.e., they were shown a game matrix like Matrix 17, in which they were both supposed to be the row-chooser. In reality, however, the subjects were playing our Game IV (Matrix 10, p. 37). Thus it only appeared to each of them that the other's S and T were both much larger numerically than his own. To carry out this deception, the procedure had to be somewhat changed. The payoffs could not now be announced orally following each play of the game, since if this were done, the players would see the discrepancy between the announced payoffs of the other and the payoffs entered in the matrix. Consequently, the outcomes were announced by specifying which choice was made by each player, L (which corresponds to our C choice) or R (which corresponds to our D choice). After this announcement, the players presumably looked up the payoffs for that outcome in the game matrix.

The apparent matrix of this game, which we shall call IV-F, is apparently a mixture of our Games IV and V (cf. Matrices 10, 11). In fact, each player believes that he himself is playing Game IV while the other is playing Game V. Actually both are playing Game IV.

The idea behind this variant was to see how the *impression* of asymmetry would affect the performance. Using an actual asymmetric game would not do. We wanted the *objective* situation to be the same for both players, while each *imagined* the situation to be different for the other.

Psychologically the situation invites some interesting questions. For example, is the large temptation *attributed* to the other (and the concomitant large sucker's punishment) sufficient to bring the frequency of cooperation in this game substantially below its value

in Game IV, in spite of the fact that the actual payoffs which each player receives are exactly as those of Game IV? On the other hand, we would not expect the amount of cooperation in this game to be as low as in Game V, where the payoffs with large magnitudes, T and S, are actually realized by the players.

Comparison of this game (IV-F) with Games IV and V in the Pure Matrix Condition is shown in Table 26.

TABLE 26

Game	IV	IV–F	V
CC	.56	.50	.21
CD	.10	.10	.06
DC	.10	.12	.06
DD	.24	.28	.67
C	.66	.61	.27

Although IV-F does fall between IV and V, it is much closer to IV than to V.

Comparison of propensities is shown in Table 27.

TABLE 27

	IV	IV–F	V
x	.962	.952	.962
y	.453	.500	.352
z	.377	.391	.229
w	.189	.136	.047

One might interpret these results as an indication that one's own payoffs have a stronger influence on

the conduct of the game than the payoffs one attributes to the other. On the other hand, we must keep in mind that because of the deception, the payoffs of the other were not *announced* by the experimenter (as they were in the Pure Matrix Condition). The other's payoffs had to be "looked up" in the matrix. To be sure, the matrix was in front of the subjects all the time, but we have no assurance that subjects looked up both their own payoffs and those of the other player every time. Some might have paid attention only to their own payoffs and so were effectively playing Game IV. It might be interesting to perform the same experiment with an apparatus which displays the payoffs separately to each subject to see whether the absence of announcements might have been responsible for our results.

It should also be kept in mind that the nature of Prisoner's Dilemma is such that an argument can be made for either of two contrary results. In this case, the same features of the game which make it appear "severe" (inducing a strong motivation to defect) can also be interpreted to make the game seem mild. The first line of reasoning which may occur to a player goes something like this:

"He stands to gain 50 if he defects. Also, he stands to lose 50 if he cooperates alone. Therefore he will probably not cooperate. Consequently I will not cooperate either."

However, the reasoning might also go as follows:

"If I defect and he cooperates, I gain 2, but he loses 50. This does not seem fair. Besides, this will make him angry, and he is sure to retaliate. Better to cooperate to show him that I am not taking advantage of his precarious position."

Especially if the other does make a cooperative choice, the interpretation might go this way:

"He risks to lose 50 if he cooperates alone. Nevertheless he did cooperate. He must be a decent fellow. I will go along with him."

As we have said, it is typical of arguments in support of a particular style of play in Prisoner's Dilemma that the features of the game which support the argument can be turned around to support the opposing argument. We had intended to build a strong self-fulfilling assumption into Game IV-F. But in doing so, we may have brought an opposite self-fulfilling assumption into play which all but offsets the first one.

If the arguments we have just developed are valid, they ought to be reflected in the propensities x, y, z, and w. We would expect x in Game IV-F to be greater than in Game IV and w to be smaller. This is because the lock-ins (results of self-fulfilling assumptions) ought to be stronger in IV-F. On the other hand, both y and z ought to be greater in IV-F, especially the latter ("repentance") if the conscience-motivated abstention from continued defection is a fact. Comparing the propensities in IV-F and IV (Table 26), we see that our conjecture is corroborated with respect to y, z, and w, but not with respect to x. The question, therefore, remains open.

The Case of $S \neq -T$ and Asymmetrical Games

One question which has remained entirely unanswered in our investigations is whether the temptation to defect or the fear of being left holding the bag is the stronger motive in inducing defection (or inhibiting cooperation). In all our experiments we had $S = -T$. Consequently any changes in one were always accompanied by the same changes in the other. To assess the effects of each separately, we should keep one of the parameters constant while we vary the other. The follow-

ing set of games could suggest the answer to our question about the relative importance of T and S.

	C	D
C	1,1	−3,2
D	2,−3	−1,−1

Matrix 18.

	C	D
C	1,1	−2,3
D	3,−2	−1,−1

Matrix 19.

Each of these two games is to be compared with Game IV (Matrix 10). In one of them the game is made more "severe" by increasing the magnitude of S while keeping T constant; in the other by increasing the magnitude of T while keeping S constant.

Games with Third Choice

An interesting modification of Prisoner's Dilemma can be made by adding "third choices." An example is shown in Matrix 20.

	C	S	D
C	5,5	−1,−1	−10,10
S	−1,−1	−1,−1	−1,−1
D	10,−10	−1,−1	−5,−5

Matrix 20.

In this game either player can escape from the Dilemma situation by playing strategy S ("Sanctuary"). For in that case, there is nothing the other can do to change the outcome. Thus each player can on any given play "refuse" to play. The payoff of this "sanctuary" strategy can now be taken as an independent variable. It can, in particular, be decreased to almost P, which each player can guarantee himself in the two-choice Prisoner's Dilemma, and we can see how unattractive,

albeit still sought after, the sanctuary can become. On the other hand, the sanctuary payoff can be increased to almost R. Interesting questions arise if the sanctuary payoff becomes less than P or greater than R. Failure to take advantage of the sanctuary in the latter case clearly indicates an "irrational" choice or else points to considerations other than payoffs, e.g., the attraction of the game itself. On the other hand, if the sanctuary payoff is less than P, or even less than S, its use may indicate acts of revenge, where a player punishes the other (and incidentally himself) more severely than he could if he simply chose D, possibly as a demonstration of disapproval of the other's failure to cooperate.

Sanctuary payoffs need not be equal for the row and the column player. A difference offers the opportunity to observe the differential effect in the case of linked subjects, while symmetric sanctuaries but with different payoffs in different games can be compared in the case of independent subjects. Thus the effect of linkage can be assessed. Is there an imitative effect in choosing the sanctuary?

In a comparative study made by E. Travis (unpublished) on fifteen pairs of mental hospital patients diagnosed as schizophrenic and fifteen pairs diagnosed as nonschizophrenic, the results indicate that the frequency of choosing the "sanctuary" or "escape" strategy S was about twice as great among the schizophrenics as among the nonschizophrenics.

As another example of a game with a third choice, consider the game represented by Matrix 21.

We note that this is a modification of the well-known divide-the-dollar game. In the latter game, each of two subjects names the fraction of the dollar which he claims for himself. The payoffs are determined by the amounts named. If these amounts add up to a dollar or less, each gets what he has claimed. If, however, the two amounts

add up to more than a dollar, neither gets anything. Matrix 21 would be a representation of this game if all the three entries below the secondary diagonal (the three lower right entries) were zero. (The players are

	25	50	75
25	25,25	25,50	25,75
50	50,25	50,50	−25,75
75	75,25	75,−25	0,0

Matrix 21.

assumed to be confined in their choices to 25¢, 50¢, and 75¢.) Experimental evidence indicates that in the divide-the-dollar game, subjects predominantly choose 50¢, which is the so-called "prominent" solution proposed by T. C. Schelling (1960).

In the present modification, the player who claims 50¢ is penalized if the other claims 75¢. Moreover, the player who claims 75¢ gets his 75¢ if the other claims the equitable share of 50¢. We see from Matrix 21 that if we delete the first row and the first column, a Prisoner's Dilemma game results. Thus the game is a sort of cross between divide-the-dollar and Prisoner's Dilemma.

Consider now the three responses of the column chooser to the row chooser's claim of 75¢. The column player can "give in" to the claim by settling for 25¢. Let us call this the accommodating response. Or the column player can counter with his own claim of 75¢ and so deprive both of any gain. Let us call this the tit-for-tat response. Finally the column chooser can make the equitable claim of 50¢. This results in the greatest loss for him and the greatest gain for the other. One would conjecture that such outcomes would be rare compared with the outcomes (25,75), (50,50), (75,25), and (0,0). Nevertheless they may occur, and it is inter-

esting to look for a psychological explanation if they do. Responses of 25¢ to the other's persistent claims of 75¢ are easily explained. Here the more ruthless player has bullied the more accommodating one into submission, that is, has made him accept the lesser evil. The outcomes (0,0) are also easily explainable: they are the results of "confrontation." The outcome (75, −25) appears as simply the result of misplaced trust. Its *persistence*, however, could indicate something else, namely the insistence of the player who suffers the loss that he is neither giving in nor punishing the other for his greed, that he stands pat on his equitable choice, because that is where the mutually beneficial outcome lies if both should make this choice. This attitude can, of course, also be attributed to the "martyr" in the Prisoner's Dilemma, but it is more pronounced in the three-choice game, because the nonaggressive individual has the additional choice of accommodating to the aggressive player's demand and saving a portion of his share of the reward, which he did not have in the Prisoner's Dilemma game. This game could be subjected to the same variations as the previous games.

Games with Communication of Intent

Some investigators have directed their attention to questions related to the role of communication in nonzero-sum games (Deutsch, 1958). For example, one might ask how the structure of Prisoner's Dilemma is affected if one of the players announces his choice of strategy to the other. If the announcement is binding, it amounts to a move in a game, in which the first player has a choice of two moves, and the second player a choice of two replies to each of the first player's choices. Note that this game is not at all of the sort we have called Prisoner's Dilemma. In order to be compared, the two

games both must be in so-called "normal form," i.e., both must be reduced to a form in which each of the players has only one move and their moves must be made *independently of each other.* One of the fundamental results of game theory is that this reduction can always be carried out provided the situation depicted meets the criteria of a game, as games are formally defined. That is to say, regardless of how many moves there may be in a game described as a sequence of moves, where the choices open to one player are contingent on the choices made up to that point (which is the usual situation in games of strategy), the game may be represented in normal form as a matrix, whose rows and columns represent the possible *single* choices of the respective players. Each player needs to choose only once in a play of the game: he chooses a *strategy* (not a move) and this choice (it is shown in game theory) contains all the possible choices, including the contingencies of these choices, which the player might make when the game is played sequentially.

Now Prisoner's Dilemma, as we have been studying it, is already in normal form, because each player must choose only once per play of the game, and the choices must be simultaneous (that is, independent). If one player is asked to choose first, the game is no longer in normal form. However, as we shall now show, it is now a new game which can also be reduced to a normal form.

The first player now has two strategies, which for him are equivalent to the two moves open to him, namely

S_1: Choose C.
S_2: Choose D.

The second player, however, now has *four* strategies at his disposal, namely

S_1': Choose C regardless of what player 1 chooses.
S_2': Choose C if he chooses C, otherwise D.
S_3': Choose D if he chooses C, otherwise C.
S_4': Choose D regardless of what he chooses.

In normal form, this game is now represented by a matrix with two rows and four columns, as shown in Matrix 22. The payoffs are those corresponding to our Game IV.

	S_1'	S_2'	S_3'	S_4'
S_1	1,1	1,1	−2,2	−2,2
S_2	2,−2	−1,−1	2,−2	−1,−1

Matrix 22.

We see that this is not at all a Prisoner's Dilemma game. For one thing, the number of strategies available to one player is not equal to that available to the other. And, what is perhaps more important, the essential feature of Prisoner's Dilemma is missing: Player 1 has no dominating strategy. Note that if the column chooser can announce his choice of strategy (he now has four), he can force the row chooser to take S_1 by announcing S_2' (his tit-for-tat strategy). To the row chooser, on the other hand, no advantage accrues from the privilege of announcing his strategy. If we now assume that the column player has the privilege of announcing his strategy, *this* game, reduced to normal form makes four strategies available to the column chooser and *sixteen* to the row chooser, as shown in Matrix 23.

This is again a different game. Here the best choice for the row chooser is strategy S_{0100} and for the column chooser S_2'. Translated into English, this solution states the following.

If the second player has a choice of announcing one of the four strategies S_1', S_2', S_3', S_4', he cannot do better

	S'_1	S'_2	S'_3	S'_4
S_{0000}	2,−2	−1,−1	2,−2	−1,−1
S_{0001}	2,−2	−1,−1	2,−2	−2,2
S_{0010}	2,−2	−1,−1	−2,2	−1,−1
S_{0011}	2,−2	−1,−1	−2,2	−2,2
S_{0100}	2,−2	1,1	2,−2	−1,−1
S_{0101}	2,−2	1,1	2,−2	−2,2
S_{0110}	2,−2	1,1	−2,2	−1,−1
S_{0111}	2,−2	1,1	−2,−2	−2,2
S_{1000}	1,1	−1,−1	2,−2	−1,−1
S_{1001}	1,1	−1,−1	2,−2	−2,2
S_{1010}	1,1	−1,−1	−2,2	−1,−1
S_{1011}	1,1	−1,−1	−2,2	−2,2
S_{1100}	1,1	1,1	2,−2	−1,−1
S_{1101}	1,1	1,1	2,−2	−2,2
S_{1110}	1,1	1,1	−2,2	−1,−1
S_{1111}	1,1	1,1	−2,2	−2,2

Matrix 23.

The row player has decided in advance what he will do in case the column player announces each of his four strategies. For example, S_{1001} means that the row player has decided to play C if the column player announces either S'_1 or S'_4, but he will play D if the column player announces either S'_2 or S'_3.

than announce S_2'. The first player cannot do better than to decide, "I will play cooperatively if and only if you announce your strategy S_2'" (which is to say, "I will play cooperatively if and only if you will").

If games are to be analyzed strictly formally (as is done in game theory), the reduction to normal form is mandatory. Analysis without such reduction has been undertaken by some investigators. This sort of analysis of necessity remains incomplete. There is no reason, however, why it should not be pursued in the context of *psychological* rather than game-theoretical investigations of games. After all, reduction of a game to normal form is only a formalistic device and needs to have no counterpart in the players' views of the game. Communication, on the other hand, is very definitely a human phenomenon with rich psychological implications. We shall therefore depart from formal game-theoretical analysis and introduce communication, not formally as moves in the game (as must be done in game-theoretical analysis), but as something sui generis imposed upon the game.

Let us suppose Prisoner's Dilemma played in the following way. Before the choice is made, each player reveals to the other (simultaneously) what choice he is going to make, with the understanding that this revelation is not binding. No payoffs are associated with either announcement.

We now have essentially two games played consecutively. The first game is represented by Matrix 24, and the second game is Prisoner's Dilemma. From the purely strategic point of view the choice between C and D in

	C_2	D_2
C_1	0,0	0,0
D_1	0,0	0,0

Matrix 24.

the first game is obviously arbitrary and has no bearing on the choices in Prisoner's Dilemma which follows. Psychologically, however, the situation is very different. The first game gives the players an opportunity to declare their intentions without cost. In repeated plays of Prisoner's Dilemma, the way one plays can also be used to communicate intentions, but in that case there are costs attached, for example, to unreciprocated cooperative choices.

Suppose now both players announce *C*. Is the choice of *C* by either player on the next play more or less likely than if either or both had announced *D*? Or, to put it more generally, in a sequence of plays in which announcements alternate with payoff games, is there likely to be more or less cooperation than in a sequence of payoff games, such as the ones we have been studying?

As with all questions related to Prisoner's Dilemma, one can argue in favor of either answer. On the one hand, it stands to reason that there should be more, rather than less, cooperation when announcements alternate with games, since the declarations of intent serve the purpose of facilitating collusion. On the other hand, the temptation to renege on the declaration of intent is always there. (The rules of the game must specify that the declaration of intent is not binding.) If players renege on their declarations, the resulting mistrust may be even more severe than in the case of a sequence of payoff games, since such a switch is more likely to be interpreted as a doublecross than a straight *D* response, which can be attributed to "self defense." All the other psychological questions involving the conditional response probabilities, etc., remain in force.

Interpreted Games

So far we have assumed that Prisoner's Dilemma is presented to the subjects as a parlor game played for money. Several variants are possible in which monetary gains

and losses are replaced by other kinds of rewards and punishments. Along with these changes, one can introduce different interpretations of Prisoner's Dilemma. The original anecdote involving the two prisoners is one such interpretation (cf. p. 24). Several other interpretations are possible. For example, the two players can be asked to imagine that they are two firms in competition. Each has a choice of selling its product at one of two price levels. If one firm sells at a high level while the other sells at a low level, the second firm reaps the profits (by winning the market). If both sell at a high level, both profit (though not as much as when competition is eliminated). If both sell at a low level, both lose money. Clearly, this situation is isomorphic to Prisoner's Dilemma. Or, the players can imagine that they are rival power blocs who have made a disarmament agreement. The cooperative choice now means keeping the agreement; the defecting choice, breaking it. There is supposedly an advantage accruing to the bloc which breaks the agreement unilaterally, etc.

The central point of interest in these interpretations is the question of whether the pattern of play will change markedly in each variant, and, if so, in what direction. In other words, one can design experiments to tap the relative strength of the subjects' ethical convictions or their cognitive sets on various matters. In the case of the original Prisoner's Dilemma, the question is either whether one should snitch on one's partner in crime or what the subjects think is the likelihood that two prisoners in such a situation will trust each other to hold out. In the case of business competition, the question pertains to the likelihood of tacit price fixing (in the subjects' estimation). In the case of international relations, the question pertains to the role of trust or mistrust (again, of course, in the subjects' estimation, not necessarily in reality).

Note that the experiments provide answers to these questions entirely in behavioral terms, not in terms of verbal responses as in questionnaires. However, there is also an opportunity to compare behavioral data with verbal responses from the same population and to note whether behavioral patterns correlate with verbally expressed convictions.

We have already indicated that attempts to correlate frequency of cooperative responses with independently assessed personality characteristics of *individuals* in repeated plays cannot be expected to yield much information, because of the strong interaction effects. However, we have also suggested that this difficulty may be overcome by "distilling" out other variables, which are more immune to interaction and so may be expected to reflect more faithfully individual personality characteristics. Also it is entirely feasible to correlate behavior patterns in the Prisoner's Dilemma game to personality characteristics of *populations* (as distinct from individuals). The experiments can thus become a source of information about the hierarchy of values in a given population. Our comparison of male and female populations was an example of such an approach.

Prisoner's Dilemma with Group Decisions

As a final example of an intriguing variant, consider Prisoner's Dilemma played by two teams each consisting of three individuals. We shall suppose that there is no communication either between the opposing teams or among the members of each team. The decisions to play *C* or *D* are made by a silent vote. The players can see how their own teammates are voting but not how the individual members of the other team are voting. (Of course they know the result of that vote because the outcome is announced after each play, as in the preceding versions.)

In addition to all the other information about this game, we now obtain a great deal more. Previously, we were concerned with the way one's own previous choice and the other's previous choice affected the probability of cooperative (or noncooperative) choice on the next play. Now we have also the effect of one's team's choice, which is not necessarily identical with one's own choice. Here, then, we have an opportunity to observe not only "martyrs" (players who choose unilateral cooperation persistently) but also nonconformists, i.e., players who consistently are in the minority of one in their own group.

Another interesting question is whether introducing communication among teammates makes a difference. Preliminary investigations (Martin, unpublished) suggest that the effect of intra-team communication is quite large and in the direction of greater cooperation between the two teams (in spite of the fact that no inter-team communication is allowed).

One might explain this effect as follows. Suppose the two teams are locked in on *DD*. Suppose one member of one of the teams gets the idea that it might be possible to break out of it. (The probability that this idea will occur to two or more members *simultaneously* is small, since in the absence of communication the events are independent and the corresponding probabilities must be multiplied.) The lone would-be cooperator's vote does not change the team vote, and so his persistence (if he continues to vote for *C*) fails to produce any response from the other team. Thus the lone vote for *C* remains doubly futile and is given up more quickly than in the case of two individuals playing Prisoner's Dilemma. Besides, once given up, the attempt to get one's own team to cooperate may be quite unlikely to be made again, because it has failed on two counts—getting the other team to cooperate and getting one's own team to

start cooperating. Thus it may be more difficult to break out of the *DD* trap in the team-vs.-team game.

Suppose now intra-team communication is introduced. The players who see the advantage of cooperation may induce the team to adopt a cooperative strategy on a trial basis. It may be agreed to try a few cooperative plays to be continued or discontinued depending on whether they are or are not reciprocated by the other side. Thus at least attempts to break out of the *DD* trap will be more likely. Because of the inherent instability of the game, these attempts may be all it takes to throw the plays into a *CC* run.

It may be argued at this point that the same reasoning applies to the *CC* run. Suppose an idea occurs to one of the players to defect from *CC* (recall that the temptation is always there). The lone vote in favor of *D* cuts no more ice than the lone vote in favor of *C*, and so the stability at *CC* ought to be greater in teams (without intra-team communication) than in individuals. This may very well be the case and yet the balance may be in favor of more defecting responses in the case of non-communicating teams. The reason for this is that *CC* runs are already quite stable in individuals. Making them still more stable will not make a great difference. On the other hand, the *DD* runs are not as stable in the case of individuals as *CC* runs. Strengthening the lock-in effect on *DD* may be expected to induce greater changes in the overall play patterns than strengthening the already strong lock-in effect on *CC*.

It would be interesting to note whether the results are opposite in the case of populations in which the lock-in effect on *DD* is stronger. This seems to be the case in women subjects. The conjecture, therefore, is that teams of women playing Prisoner's Dilemma without intra-team communication ought to show more cooperation than individuals in contrast to men subjects.

If this turns out to be the case, the sort of explanation we have offered gains credibility.

Other Related Nonzero-Sum Games

Consider the game shown in Matrix 25.

	C_2	D_2
C_1	1,1	−2,2
D_1	2,−2	−5,−5

Matrix 25.

It shares an important feature with Prisoner's Dilemma. Each player is tempted to play D in order to get a bigger payoff, but if both do so, both get punished. However, the basic inequality in terms of which we have defined Prisoner's Dilemma, namely $T > R > P > S$ (cf. p. 34) is violated in Matrix 25, since in that game $T > R > S > P$. We see, also, that a fundamental feature of Prisoner's Dilemma is absent: D no longer dominates C. Strategy D is still best against the other's strategy C, but not against the other's strategy D. The motivations of this game are, therefore, different. According to strategic analysis, C responses on the part of one player should elicit D responses on the part of the other; and vice versa. In Prisoner's Dilemma, strategic analysis indicates that both C and D responses on the part of one player ought to elicit D responses on the part of the other (because of the fact that D is dominant). In practice, this is not at all the case: predominantly C responses elicit C responses, while D responses elicit D responses. Will the game-theoretical prediction be borne out in the case of games represented by Matrix 25? If so, the ρ_i and the $\rho_{C_1C_2}$ associated with such games (cf. Chapter 3) should be substantially smaller than in the case of Prisoner's Dilemma games, perhaps even negative.

Let us examine more closely the psychology of the game represented by Matrix 25. The interesting feature of this game is in the opportunity it offers for "preemption." Suppose one of the players plays D persistently. In the case of Prisoner's Dilemma, the other had no choice but respond with D to D in self-defense (if he were guided entirely by the payoff matrix). In the game now considered, the other player has a choice of either punishing the D response (and, of course, himself as well) by responding with D or of "giving in" in the face of a "determined stand" on the part of the first player. It is clear that the game is psychologically isomorphic to the now famous game of Chicken. Here the C response stands for "chicken" and the D response for "daring." The DD outcome can also be interpreted as "disaster." Matrix 26 presents Chicken in an especially severe form:

	C_2	D_2
C_1	1,1	−2,2
D_1	2,−2	−100,−100

Matrix 26.

Like Prisoner's Dilemma, Chicken in its symmetric form has four independent parameters, which we may now label R (reward for prudence), T (temptation to bank on the other's giving in), S (surrender), and P (perdition). The inequality to be satisfied must be $T > R > S > P$. The experimental program described here and all the methods of analysis can be directly applied to this interesting class of games.

It appears, then, that the nonzero-sum game offers excellent opportunities to develop programs of laboratory experiments and mathematical theories in which clearly psychological concepts and interpretations are coupled with the generation and analysis of hard, replicable data.