Chapter 14
Summary of Results

IN OUR RESULTS, we have been principally interested in the coalescence of the performances of paired players. Whatever individual differences exist among the players (and it is difficult to believe other than that they exist) tend to be ironed out in the course of the interactions between them. The clearest manifestation of this effect is in the high values of $\rho_{C_1C_2}$ (cf. p. 59). Other correlation measures, e.g., ρ_i, $\rho_{x_1x_2}$, etc., while not as large as $\rho_{C_1C_2}$ are nevertheless overwhelmingly positive and point to the same effect.

The "cause" of this emerging similarity between the paired players (if one can speak of causes in this context) is the universally observed elimination of the unilateral responses. If no unilateral responses occurred, the players of a pair would be exactly like each other, because their protocols would be identical. It is only through the unilateral responses that the two members of a pair are distinguished. When these responses decline in frequency, the similarity of the players emerges.

Therefore the next question to ask is why the unilateral responses are eliminated. The answer which immediately suggests itself is that the payoff S is unbearable to the lone cooperator. He tends to switch from C to D upon receiving this payoff, and so regardless of what the other does (except when he happens to switch simultaneously from D to C) the unilateral response is eliminated. Of course, the unilateral response vanishes also when the defector is "converted," which happens much less frequently, but not too rarely, as indicated by the values of z.

This is not the whole story, however, since the unilateral responses could well appear as frequently as they disappear. A unilateral response could result from a player's defection from *CC* (in response to temptation) or from a player's attempt to escape from *DD*. These effects, however, are weak, as we see by examining the typical values of $1 - x$ and of w, which are the propensities associated with the events just mentioned. These values are both low. Consequently, not only are the unilaterals "killed" soon after they appear, but their "birthrate" is also low because of the lock-in on both *CC* and *DD*.

Typically, toward the end of the sessions over ninety percent of the responses are matched. It is this which accounts for the consistent similarity between paired players.

Therefore it makes sense to take the pair, rather than the individual, as the unit of a population. Variance among pairs is rather large; even within games (cf. Chapter 12). We have strong reason to believe, however, that much of this variance is accounted for not by the inherent propensities of the players to cooperate or not to cooperate, but rather by the characteristic instabilities of the dynamic process which governs the interactions in Prisoner's Dilemma. We have seen in Part II how these instabilities emerge as consequences of several mathematical models which we have proposed. To be sure, we cannot claim that any of these models has been corroborated, but our results are consistent with several of their aspects and we venture to suggest that the dynamics of interaction is of a somewhat similar sort as that inherent in our models.

Besides, our data provide evidence that the quality of interaction rather than inherent propensities is the important factor in the process. A case in point is the comparison between men and women. Whether playing against partners of their own or of the opposite sex, the paired men and women exhibited exactly the same average value of *C* on the first play, that is, before any interaction took place (cf. Table 25, column 34). The value of *C* on the second play (column 35) is still practically the same in men and women. Namely, it drops slightly less in *WW* than in *MM* and slightly more in *WM* than in *MW*. However, the protocols of *MM* and *WW* are widely different, whereas the protocols of *MW* and *WM* are quite similar. All this points to the interaction effects as being the chief determinants of the protocols.

The interaction effects are reflected in the time courses. The most typical feature of the time course of a Prisoner's Dilemma protocol is the initial decline in cooperation, followed eventually by a recovery. The time of onset of the recovery and its extent is the chief determinant of the overall amount of cooperation which will be observed. Thus the great difference between the Matrix Conditions and the No Matrix Conditions (cf. Chapter 2) is a result of the fact that in the latter the recovery comes much later and proceeds much more slowly than in the former. The same can be said for the difference between the *MM* and the *WW* populations.

We are thus led to the general conclusion about the psychological factors operating in the dynamics of Prisoner's Dilemma. There is initially either a reservoir of good intention or a lack of appreciation of the strategic structure of the game. At any rate the expectation that *DD* will predominate is not borne out on the first play. The fraction of cooperative responses is over fifty percent in all cases where the subjects see the matrix and is as high as sixty-three percent in pairs of mixed sex.

Thereafter this initial goodwill (or naïveté) decreases. The *DD* responses increase at the expense of the *CC* responses and especially of the unilateral responses whose initial rate of decline is typically the largest. This initial period, therefore, is a "sobering" period in which the "hard realities" of the Prisoner's Dilemma game become impressed on the players. However there is no percentage in *DD*, and sooner or later the players recognize this. Accordingly a recovery sets in. Typically the frequency of *CC* increases at the expense of the unilaterals. Hence, to put it another way, while in the beginning of the process the unilaterals tend to turn into *DD*'s (the cooperator defects), toward the end they tend to turn into *CC*'s (the defector cooperates). Thus learning goes both ways in Prisoner's Dilemma. First the subjects learn not to trust each other; then they learn to trust each other.

Payoffs have the expected effect on performance. As one would surmise by common sense, by and large the cooperation seen is related to the payoffs. However, the finer points of these effects have not yet been established. We do not know on the whole whether the reward for *CC* or the punishment for *DD* plays the more important role in eliciting cooperation. The consistently higher cooperation observed in Game XI compared with Game XII suggests that *R* is more important than *P*. However this is not borne out (nor contradicted) in the comparison between Games I and II.

It should be possible, in our opinion, to construct a mathematical model which would predict a greater effect of *P* when the magnitudes of *R* and *P* are large (as in Games I and II respectively) and a greater effect of *R* when these magnitudes are smaller (as in Games XI and XII respectively). This would account for the results in the Pure Matrix Condition. However the value of such a model would become manifest only if, in addition, it predicted also other characteristic peculiarities observed in the data.

This brings us to what we have already said is the principal valuable result of the investigation. The structure of the process generates questions of psychological interest and suggests the aspects of the data to be further examined to answer the questions. Many of the conjectures we have made cannot be substantiated without at least establishing the statistical significance of the results. But if the questions are of sufficient interest, motivation is provided to put the results to more severe tests.

For example, the differences between men and women in mixed pairs are all quite small (and probably statistically insignificant in relation to the sizes of the populations) but highly suggestive. We have conjectured, for example, on the basis of comparing ρ_1, ξ, and ω that men are more prone to give tit-for-tat than women. Two players who *always* give tit-for-tat would produce only three types of protocol, namely (1) all *CC*; (2) all *DD*; (3) all alternating *CD* and *DC*. However an *insight* into the fact that tit-for-tat is always given would invariably result in all-*CC* protocols, since these give the largest payoff to both. The other two types can be easily changed into an all-*CC* protocol by just one shift of phase from an alternating *CD-DC* protocol and just two shifts of phase from an all-*DD* protocol. Hence we would expect greater cooperation in a population of players more prone to give tit-for-tat, and this may be the principal reason why men cooperate more. At any rate, questions of this sort can be answered in principle thanks to the theoretical framework provided by a detailed structural analysis of the statistics generated by the data.

The opportunity of developing and refining such methods with a view of applying them broadly in experimental social psychology are, accordingly, offered as the principal result of our efforts.