Chapter 13
Comparing Populations

OUR FIRST OBJECTIVE in undertaking the experimental program described here has been to gain some understanding of what goes on in long sequences of plays of Prisoner's Dilemma. We have attempted to gain this understanding by postulating a system going through a sequence of states and by attempting to formulate some mathematical models from which the dynamics of the system could be deduced. Once such a model is found, its parameters, properly interpreted, become the key terms in the emerging psychological theory. This strategy can be deemed successful, if the parameters so discovered are independent of the process itself, if they suggest further investigations, and if the further investigations, in turn, lead to a more inclusive theory.

For example, suppose we had found that the parameters x, y, z, and w were independent of the process and that, when the estimated values of these parameters were substituted into the Markov equations, the time courses of the four states were accurately predicted, as well as other important statistics of the process theoretically deduced from the generated stochastic model. Then the values of x, y, z, and w would be *the* parameters of the process. We could then ask questions about how these parameters are affected by, say, the payoffs. A model relating the payoffs to x, y, z, and w would then constitute an extension of the theory. We would also have a solid basis for comparing populations, namely, in terms of the values of these parameters. In the light of the psychological suggestiveness of the propensities x, y, z, and w, we could then say more specifically why there is more cooperation in one population than in another, for example because the one population is more "trustworthy" than another or more "trustful," or both, or perhaps less trustworthy but more trusting to the extent that the latter characteristic more than offsets the former, etc. We might find that two populations exhibit the same gross degree of cooperative behavior, but that their "profiles" in terms of propensities may be quite different. We could then predict a divergence between their behaviors under a different set of conditions.

Further, if extraneous circumstances brought about changes in performance, we could see where (in which parameters) these changes were brought to bear. Or, if extraneous circumstances brought about only transient changes at the start of the performance, we could explain this by pointing out that only the initial conditions were affected by the changes, not the system parameters themselves which govern the ultimate steady state characteristics of the system.

If, on the other hand, one of the adjustable parameter's models were most successful in accounting for the data, a different set of constants would be singled out for attention. These might be the constants of proportionality connecting the rate of adjustment to the gradient in the corresponding expected payoff, or the like. And in these models the positions of the unstable equilibria, rather than steady state equilibria, would be the most important features of the dynamic, as we have shown in Chapter 10.

At this time we cannot single out from the models proposed any one which is best in every respect. As we have seen, the task of comparing the various models, if taken seriously, is one of formidable difficulty, which we did not undertake to accomplish by a tour de force.

In our opinion, data much more voluminous than those we have gathered are required in order to establish confidence in a dynamic model, especially a stochastic one, possessing the degree of complexity which the situation requires.

We therefore will content ourselves with parameters which do not fulfill the requirement of being independent of the process. We shall use, as a basis for comparing populations, all the important variables and parameters which have entered our discussion. That is, we sacrifice parsimony (which would have been served had we been able to isolate the "basic" parameters) in order to get a descriptively "rich" comparison. In doing so, we shall be wary of the confusion which often results when seemingly unrelated statistics are piled up in describing or comparing phenomena. We shall try to avoid such confusion by fitting the indices of comparison into a more or less coherent picture.

1. *The frequency of cooperative responses.* The most natural index is, of course, the total relative frequency of cooperative responses in each of the games. We have seen that this index is very strongly affected by interaction. However, to the extent that we compare populations playing the game under identical conditions, we shall suppose that differences in the total frequency of cooperation found among different populations reflect a difference in some characteristic of the populations, whether the characteristic resides inherently in the individuals comprising each population or in the way these individuals interact. We shall therefore follow the established tradition in evaluating performance in Prisoner's Dilemma in terms of observed values of C, including its time course.

2. *The correlation indices* ρ_i. These indices measure the extent to which one player's choice is an imitation

of the other player's simultaneous choice, his choice on the last play, on the play before that, etc. Thus the ρ_i's form a "profile." We shall compare only values of ρ_0, ρ_1, and ρ_6.

3. *The correlation coefficient* $\rho_{C_1C_2}$ *over a population of pairs.* This is a grosser measure than the ρ_i; it measures the overall similarity of pair members with respect to their cooperative frequencies.

4. *The response-conditioned propensities* ξ, η, ζ, *and* ω. Of these ξ and ω are measures of responsiveness to the other's choices, while η and ζ are measures of "response" to one's own choices. (N.B.: the two pairs are not independent since one's own choices always occur in conjunction with the other's choices [cf. p. 68].)

5. *The state-conditioned propensities* x, y, z, *and* w. These are propensities similar to the preceding ones but now separated by reference to the four mutually exclusive response categories. It will be useful also to compare x, y, $1 - z$, and $1 - w$ (cf. p. 84), which are measures of persistence in the *same* response in each of the four states.

6. *The ratios* $r^{(i)}$ *discussed in Chapter 11.* The behavior of the sequences of these ratios ($i = 1,2, \ldots$) will tell us something about the lock-in effect, whether it is operating (if the $r^{(i)}$ increase) or not, or whether perhaps an "antilock-in effect" is operating (if the $r^{(i')}$ decrease).

7. Next, there are a few special indices of interest. Consider, for example, the index

$$M = \frac{(1 - y_1)(1 - z_2)}{y_1 z_2} \quad \text{or} \quad \frac{(1 - y_2)(1 - z_1)}{y_2 z_1}.$$

This index is related to the "martyr" runs (runs of unilateral states). The numerator of M represents the probability that a unilateral state passes to DD, i.e., the "martyr" gives up while the defector continues to defect. The denominator represents the probability that the defector starts to cooperate while the "martyr" continues to cooperate. Thus *M* represents the ratio of "failures" to "successes" of such runs, that is, the ratio of the number of times such runs turn into double defecting or double cooperative responses.

8. Next we wish to examine the fractions of compared populations choosing *C* on the very first play [*C*(1)] and on the second play [*C*(2)].

9. Finally, we shall examine the fractions of compared populations which have locked in on the *CC* response (L_{CC}) and on the *DD* response (L_{DD}) in the last twenty-five plays. Our criterion of lock-in will be arbitrarily taken as twenty-three of twenty-five responses in the category in question.

We shall compare three populations, namely, *MM*: 70 male pairs; *WW*: 70 female pairs; *MW*: 70 mixed pairs.

All the populations played Prisoner's Dilemma in the Pure Matrix Condition. Therefore for our *MM* population we shall take the one already examined.

In our comparison, games will not be differentiated. Since each population played under identical conditions we shall be concerned only with the gross indices averaged over each of the entire populations of seventy pairs.

In the *MM* and *WW* populations the individual players in a pair (i.e., the players labeled 1, 2) will not be differentiated. In the *MW* population they will be. Thus in Table 25 the row *MW* contains indices pertaining to men playing opposite women, while row *WM* contains indices pertaining to women playing against men. For example, the *CD* in *MW* represents cooperation by the man and defection by the woman, while in *WM* it represents cooperation by the woman and defection by the man. Obviously, sym-

TABLE 25

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
MM	.51	.08	.09	.32	.59	.97	.46	.51	.36	.76	.78	.25	.25	.85	.40	.38	.20	.79	−.10	−.25	.48
MW	.40	.10	.09	.41	.50	.91	.35	.39	.28	.68	.71	.24	.25	.80	.40	.33	.21	.58	.12	.02	.72
WM	.40	.09	.10	.41	.48	.91	.35	.35	.24	.65	.69	.22	.27	.78	.44	.29	.22	.58	.12	.02	.72
WW	.23	.11	.11	.55	.34	.87	.31	.34	.22	.55	.60	.17	.19	.75	.37	.26	.15	.34	−.29	−.20	.46
	CC	*CD*	*DC*	*DD*	*C*	ρC_1C_2	ρ_0	ρ_1	ρ_6	ξ	η	ζ	ω	x	y	z	w	$\rho_{x_1x_2}$	$\rho_{y_1y_2}$	$\rho_{z_1z_2}$	$\rho_{w_1w_2}$

		22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37
MM	1)	.68	.79	.86	.87	.69	.69	.71	.74								
	2)	.66	.78	.80		.68	.67	.68		.34	.39	.32	2.4	.53	.48	.57	.13
	3)	.65	.75			.67	.62			.27	.21						
MW	1)	.74	.75	.85	.85	.64	.70	.71	.79								
	2)	.68	.69	.74		.62	.69	.67		.31	.40	.35	3.7	.63	.56	.39	.21
	3)	.64	.62			.60	.65			.32	.28						
WM	1)	.74	.75	.85	.85	.64	.70	.71	.79								
	2)	.68	.69	.74		.62	.69	.67		.41	.39	.23	2.6	.63	.54	.39	.21
	3)	.64	.62			.60	.65			.36	.25						
WW	1)	.69	.66	.73	.68	.69	.71	.72	.73								
	2)	.68	.61	.64		.68	.71	.71		.33	.37	.27	4.8	.53	.49	.17	.30
	3)	.63	.53			.68	.70			.30	.25						
		$r^{(1)}_{CC}$	$r^{(2)}_{CC}$	$r^{(3)}_{CC}$	$r^{(4)}_{CC}$	$r^{(1)}_{DD}$	$r^{(2)}_{DD}$	$r^{(3)}_{DD}$	$r^{(4)}_{DD}$	$r^{(1)}_{CD}$	$r^{(2)}_{CD}$	$r^{(3)}_{CD}$	M	$C(1)$	$C(2)$	L_{CC}	L_{DD}

1) Average over pairs having runs at least five long.
2) Average over pairs having runs at least four long.
3) Average over pairs having runs at least three long.

metric indices like *CC* or $\rho_{x_1x_2}$ will be identical in both these rows.

We turn to the results.

First we examine the distributions of the four states. The most striking difference is between the male and the female populations. There is a clear indication that males cooperate more than females, as can be seen directly by comparing the respective *C*'s.[38]

However, when men play against women there is no perceptible difference between them with respect to the total frequency of cooperative choices. The one percent difference in favor of the men observed in the unilateral responses cannot be significant in view of the fact that a difference of this magnitude is observed also between *CD* and *DC* responses of the all-male population, where it must reflect only a statistical fluctuation, since here the difference is only between players labeled 1 and 2.

Next we compare the performance of men playing against men with that of men playing against women and also the performance of women playing against women with that of women playing against men. We find that women are "pulled up" when playing against men, that is, they play more cooperatively against men than against women. Men, on the contrary, are "pulled down" when playing against women as compared with their performance against players of their own sex. However, the men are not pulled down as much as the women are pulled up. In general, the performance of mixed groups is squarely between the performances of the men and of the women, rather nearer that of men.

Next we note that this difference between the sexes is not observed at all at the very beginning of the process. (Columns 34 and 35 show the fraction of subjects choosing *C* on the first and second plays respectively.) The behavior of men and of women is practically identical both in the homogeneous pairs and in mixed pairs. Therefore, we cannot say that the pronounced difference in the performances is due to some *initial* difference in the propensities to cooperate. We must look for the roots of the difference in the interaction effects.

Accordingly, we examine next the conditional propensities, ξ, η, ζ, ω, x, y, z, and w. Throughout we observe the same effect: the women's propensities are brought up when they play against men; the men's propensities are brought down when they play against women. When men play against women, the propensities are practically equal when averaged over the entire session.

Comparing $1 - \omega$ with ξ we see that men playing men are somewhat more likely to respond cooperatively to the other's cooperative choice than to retaliate against the other's defecting choice ($\xi > 1 - \omega$). However, women playing against women are much more likely to retaliate against the other's defecting response than to respond cooperatively against the other's cooperative response ($\xi < 1 - \omega$). When men play against women, the retaliating tendency is slightly greater than cooperative responsiveness in both.

Next we examine the correlation measures. The ρ's of the men (playing against men) are consistently higher than those of the women (playing against women), which is to say that the men tend to imitate each other more than the women. The values of the ρ's in the mixed groups indicate that there men tend to imitate women more than women tend to imitate men (columns 8 and 9). In short, men are inclined to play tit-for-tat more than women.

From column 18 we conjecture that men tend to become more like each other with regard to the propensity x (to cooperate following *CC*) than women. Here the value of $\rho_{x_1x_2}$ in the mixed groups is again intermediate between that in the male and in the female pairs. With respect to $\rho_{w_1w_2}$, however, the tendency to become like each other is strongest in the mixed groups. We shall not venture to interpret this result.

Next we look at the dynamics of the state-conditioned propensities. We have already seen (cf. Chapter 11) that the mean probability of continuing a given state is not necessarily a constant (as implied by the four-state Markov model) but appears to be a function of the number of times the state in question has just occurred. In particular, we conjecture that the lock-in effect is due primarily to the fact that the more times in succession the *CC* or the *DD* state occurs, the more likely it is to be repeated at least in the pairs in whose protocols sufficiently long runs occur. This effect, if it occurs, is shown in columns 22 through 25. Comparing the probabilities of the continuation of *CC* runs, we see that in the male pairs these steadily increase and that the increase is still seen even when the pairs without runs longer than three plays are included. In the female pairs this effect is not observed. Even in the pairs where *CC* runs at least five plays long occur, r_{CC} actually declines. This is to say, when women play women the average probability of a *CC* response following two consecutive *CC*'s is actually less than the average probability of a *CC* response following a single *CC*. This indicates that the lock-in effect on *CC* does not operate in the average female pair. In mixed pairs, the lock-in effect is observed in pairs containing *CC* runs of at least five. It is still observed when pairs with runs not longer than four are included and is lost when pairs are included which contain no runs longer than three.

With regard to *DD* runs, the picture is reversed.

In the male pairs, the lock-in effect is observed only in the selected pairs with *DD* runs at least five long, is lost as soon as pairs with no runs longer than four are included, and is actually reversed when pairs with runs no longer than three are included. In the female pairs, on the other hand, the effect is observed even when pairs with runs no longer than three are included. The mixed pairs behave like the male pairs with respect to the *DD* runs. We conjecture that women become more prone than men to lock-in on *DD* as a *DD* run continues.

With respect to r_{CD} the picture is about the same throughout. In general $r_{CD}^{(2)} > r_{CD}^{(1)} > r_{CD}^{(3)}$ (with the curious exception of women against men), when pairs with "martyr runs" of at least four are included. When pairs are included with runs no longer than three, $r_{CD}^{(2)} < r_{CD}^{(1)}$. On the whole, this means that following two unilateral cooperative plays the probability of the next such play *decreases* (an antilock-in effect), although it increases in pairs which have unilateral runs longer than three. Even in these pairs the probability of the fourth unilateral cooperative choice decreases markedly.

Turning to our index $M = (1 - y)(1 - z)/yz$, which, we recall, is a measure of the extent to which "martyr runs" fail, we see another striking difference between men and women. Martyr runs of men playing against men end in failure about two and one-half times more frequently than in success (column 33). Martyr runs of women playing against women end in failure almost five times as frequently as in success. Looking at the *MW* population we see that with respect to this index the difference between men and women is still pronounced (although it is erased in most other respects). The martyr runs of *women* playing against men end in failure about two and a half times more frequently than in success.[39] That is to say, a man gets "converted" by a woman's martyr run with about probability .28 (a value close to .29, the probability of his being converted by a man's martyr run). But a woman gets converted by a man's martyr run with probability of only .21. This is a larger probability than .17 with which a woman gets converted by a woman's martyr run but still significantly smaller than .28.

However, we cannot on this basis alone lay a greater blame for the failures of martyr runs on the woman. For the failure of a martyr run can be ascribed as much to the martyr's giving up his unilateral cooperation as to the defector's failure to respond. Indeed, comparing the y's of men and women in mixed pairs we see that women's are greater, indicating a greater persistence in the martyr runs. Also the women's ω is slightly higher than men's, indicating a higher overall propensity to respond cooperatively to man's defection. The slightly higher value of w in the woman (playing against a man) indicates a slightly higher propensity to break out of *DD*, i.e., to *initiate* a martyr run. In short, the greater frequency of failures of men's martyr runs is as much due to the fact that the man is somewhat more prone to give up as to the fact that the woman is somewhat less prone to switch from successful defection to cooperation, both tendencies being a reflection of man's greater propensity to give tit-for-tat.

Turning to the final lock-ins (columns 36 and 37), we see that seventy percent of the male pairs end the sessions locked in, and that of these over four times as many pairs have locked in on *CC* than on *DD*. Of the female pairs less than fifty percent end the sessions locked in and of these almost twice as many have locked in on *DD* than on *CC*. The mixed pairs are again in between: sixty percent have locked in and of these twice as many end the sessions cooperatively rather than uncooperatively.

Finally we look at the values of *C* on the first and second plays (columns 34 and 35). Here we see no sig-

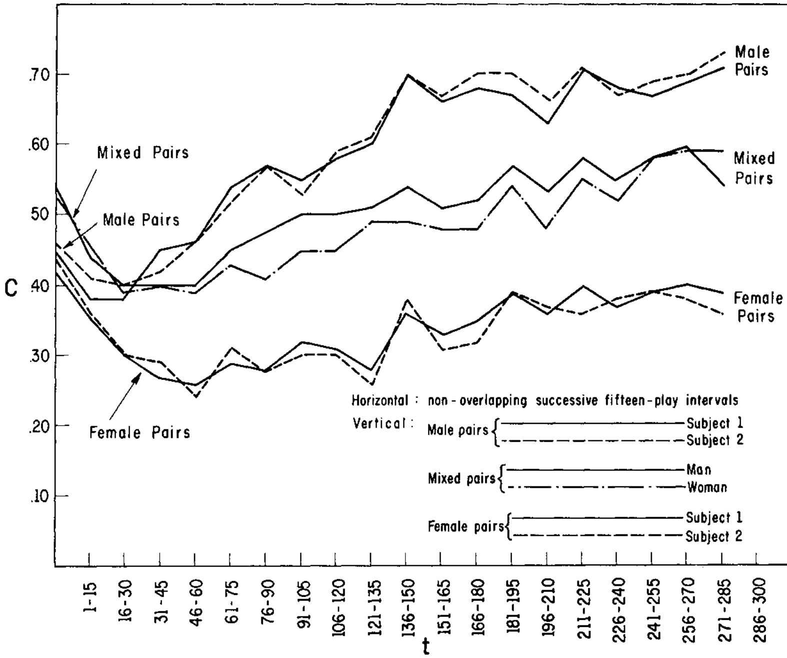

Figure 41. The time course of *C*. The differences between subjects 1 and 2 in the male and female pairs are evidently due to statistical fluctuations. In the case of mixed pairs, the higher values of *C* in men seem to be fairly consistent. These are, however, numerically small. For evaluation of significance, see Appendix III.

nificant differences between men and women either when they play against partners of the same sex or of the other.

We see, however, a remarkable result in the initial high value of *C* when men play against women. The difference between .53 and .63 in a population sample of 140 individuals is about five standard deviations (assuming the largest value of a standard deviation of C binomially distributed when $C = D = .5$). An interesting conjecture suggests itself that the *initial* propensity to cooperate is greatest in mixed pairs. The fact that the overall cooperative level attained by mixed pairs is below that of the male pairs must therefore be attributed to interaction effects.

All in all, when men play opposite women, in the long run the women are pulled up toward the men's levels of cooperation and in the amount of interaction (as reflected in the various interaction indices) while the men are slightly pulled down on both counts. These effects are easily seen in graphical representation as shown in Figure 41.