INDIVIDUAL HEURISTICS AND THE DYNAMICS OF COOPERATION IN LARGE GROUPS

MESSICK, DM; LIEBRAND, WBG

Published in:
Psychological Review

Publication date:
1995

Citation for published version (APA):
MESSICK, D. M., & LIEBRAND, W. B. G. (1995). INDIVIDUAL HEURISTICS AND THE DYNAMICS OF COOPERATION IN LARGE GROUPS. Psychological Review, 102(1), 131-145.

Individual Heuristics and the Dynamics of Cooperation in Large Groups

David M. Messick
Northwestern University

Wim B. G. Liebrand
University of Groningen

This article describes computer simulations in which pairs of "individuals" in large groups played a prisoners' dilemma game. The individual's choice to cooperate or not was determined by 1 of 3 simple heuristics: tit-for-tat; win-stay, lose-change; or win-cooperate, lose-defect. Wins and losses were determined through the comparison of a play's outcome with the average outcome of the individual's neighbors. The results revealed qualitative differences between small and large groups. Furthermore, the prevalence of cooperation in the population depended in predictable ways on the heuristic used, the values of the payoff matrix, and the details of the social comparison process that framed the outcomes as wins or losses.

This article deals with two enduring questions about social behavior. The first question concerns the processes that support self-sacrificial cooperative behavior, behavior that benefits others at a cost to the cooperator. This is the *problem of cooperation*, or altruism (Elster, 1989). The second question concerns the connections between individuals' behavior and the qualities of the groups to which the individuals belong. This is the *level of analysis problem*. The level of analysis problem includes the determination of the influence of individuals' actions on the group and the impact of the group on the individual. We present evidence derived from computer simulations that these two questions are interrelated in unexpected and surprising ways.

The question of cooperation has a long history (Rushton, 1980). Psychologists have investigated the developmental influences that promote self-sacrificial altruism; the types of reinforcement, both direct and vicarious, that maintain such behavior; and the types of personalities that are most likely to behave with a more or less selfless regard for other people (Krebs & Miller, 1985).

The problem also has roots in biology (Wilson, 1975). An altruistic trait is one that aids other conspecifics while handicapping the altruist. How can such traits evolve without violating the basic assumptions of the theory of evolution? The answers that have been offered to this question shed light on the

social nature of many species, including our own (Campbell, 1975).

The question of levels of analysis is not a new one either. In their important text, Kretch and Crutchfield (1948) posed the questions, "Can social phenomena be fruitfully investigated on the level of individual behavior? Are there other equally promising levels of social analysis? What are the appropriate units of analysis at each of these levels? What are the relations among these several levels of analysis?" (p. 14). These issues no longer lie in the mainstream of social psychological thinking, but it is, perhaps, time that they were revived. Important disparities have been shown between group and individual decision making (Allison & Messick, 1987) that merit further study. There is a growing interest in entomology in understanding how "intelligence" manifested by the behavior of army ant swarms is assembled from the behavior of the individual ants themselves (Franks, 1989). However, understanding the relationship between individual-level processes and collective outcomes promises to be as difficult as it is important. Schelling (1971) has pointed out that dramatic aggregate outcomes can be produced by apparently innocuous preferences and that the outcomes can be either desirable (the segregation of swimmers and surfers at a crowded beach) or undesirable (racially segregated housing). Related issues in economic theory concern the relationship between microeconomic models and market behavior. It may be, for example, that individuals suffer from biases in likelihood judgments, but that markets correct these biases. Thus it is a conceptual possibility that although individuals are irrational, markets are not. Camerer (1990) summarized evidence that this is not the case. Other perspectives suggest that although individuals may be rational, the resulting aggregate behavior may not be (Hardin, 1968).

Social dilemmas (Messick & Brewer, 1983) bind the question of cooperation to the problem of level of analysis. We use the well-known prisoner's dilemma, displayed in Figure 1, to illustrate. In this situation, the cooperate (C) choice is dominated by the defect (D) choice. No matter what choice is made by the other person, the chooser is better off making the D choice over C. The C choice costs one unit in that the chooser gets one unit more by making the D choice regardless of the other's action. However, the C choice provides two units to the other person.

Basic Prisoner's Dilemma

Neighbor's Choices

		Cooperate	Defect
Subject's Choices	Cooperate	2,2	0,3
	Defect	3,0	1,1

Figure 1. The basic prisoner's dilemma used in the simulations.

No matter what the other's choice, the other will get two units more if C is chosen. It is in each person's best interest to choose D, but it is in the best interest of the pair to choose C. Thus the group maximum is achieved when the individuals forego individual maximization. On a larger scale, groups with many individual maximizers (D choosers) will do more poorly than groups with few.

The specific problem that we address in this article is to describe the global consequences of heuristic choice strategies in the social situation embodied by the prisoners' dilemma. In particular, we examine what happens in large groups of individuals, all of whom use a well-defined decision heuristic for evaluating and making choices in the prisoner's dilemma game (PDG). Our approach resembles that of Axelrod (1984), who also used the PDG as a model situation for self-sacrificial cooperation, but it differs in a number of important ways. We need to make the differences clear at the outset because our goal is very different from his. First, Axelrod's tournament investigated the efficiency of various strategies for playing the PDG in an environment consisting of all the other strategies that had been submitted for the tournament. This was an environment that was maximally heterogeneous. Our interest is in the fate of cooperation in environments in which all persons use the same strategy. The environment for any strategy is simply a population of individuals using exactly the same strategy. Thus our goal is not to determine which of several strategies will perform better than others but to ask whether cooperation is sustainable in a population when all of its members make decisions in the same way.

A second important difference is that in Axelrod's work, each strategy played a PDG for 200 trials with each other strategy providing each interaction with a relatively long history. In our work, each individual played the PDG with another only once on any occasion. Moreover, each individual played only with neighbors and did not interact at all with more remote others. Thus our work assumes a "geography" that allows us to talk about neighborhoods and interactions among neighbors. Last, even among neighbors, those with whom one played on any given trial was determined randomly. Axelrod's procedures created a factorial environment in which the relative success of different strategies could be scientifically assessed. Our procedures bear a closer resemblance to a real social environment in which people interact most often with those close to them and not at all with remote persons. Moreover, each interaction was relatively brief.

Our goal is to explore the aggregate consequences of simple choice heuristics. The rules that we examine include two that have been described in previous research and one that seems psychologically realistic. The three are the tit-for-tat (TFT) strategy that has been the object of attention for years in research on cooperation (see e.g., Komorita, 1965; Pruitt, 1968; and Rapoport & Chammah, 1965) and that was the winner of Axelrod's tournament; the win-stay, lose-change (WSLC) strategy that was studied by Kelley, Thibaut, Radloff, and Mundy (1962) and Messick (1967); and a win-cooperate, lose-defect strategy (WCLD) that reflects the principle that people reciprocate affective states. Isen (1987), for instance, summarized research showing that people's willingness to be helpful increases after a positive outcome. Before discussing these strategies further, we describe the general nature of the environment in which we studied them.

The Simulation Environment

The simulation environment that was used for this research was the Warsaw Simulation System (WSS) developed in the Department of Psychology and the Computing Center of The University of Warsaw.[1] This system has been described by Gasik (1990). The system allows the user to define a rectangular group of up to 400 individuals and to specify the form of the interaction that will occur among them. In this simulation, the specified processes are executed, and prespecified dependent variables to describe the consequences of the simulation are recorded. Using a similar system, Nowak, Szamrej, and Latané (1990) studied the spread of an attitude through a group when the laws of attitude change that were implemented were those postulated by Latané's (1981) social impact theory. The simulation package thus permits the investigation of aggregate-level consequences of hypothesized individual-level psychological processes.

Our use of the WSS involves the following general processes. From a population of a specified size, one individual, the subject, is randomly selected. One of the subject's neighbors is then randomly selected. In our studies, a neighbor is one of the eight individuals (in the 3 × 3 grid) surrounding the selected individual. (Other definitions of *neighbor* are also possible and some of them will be explored in future papers.) The subject and neighbor "play" the two-person PDG described earlier. Each individual in the population is randomly assigned an initial response, either C or D, and this assignment determines the choice that the individuals make on the first encounter. The payoff to the subject from the interaction is recorded. The subject uses a choice heuristic that determines what the subject's next choice will be. The heuristic will be one of the three mentioned above. This terminates one play. Another individual is then randomly selected, and the process repeats itself. One "generation" is completed when a sample the size of the population has been

---

selected to play. Sampling is done with replacement, however, so some individuals in the population will have played more than once and some will not have played at all. The simulation can be run for a fixed number of generations or until some specified condition is met. One such condition is that the population becomes either homogeneously cooperative or homogeneously defecting.

The choice that characterizes each individual is the choice that the individual will use when the individual next plays, regardless of whether the play is as subject or neighbor. However, only subjects recalculate the choice after an interaction; neighbors do not. (We have run some simulations in which both subject and neighbor changed after an interaction. Unless otherwise noted, we found no major qualitative difference in the dynamics in this condition.) In our simulations, we assumed that each subject had eight neighbors, meaning that there were no edges or corners on the rectangle that defined the population. Thus a subject on the right-hand edge had a corresponding point on the left-hand edge as a neighbor, and subjects on the lower edge had corresponding points on the upper edge as neighbors and vice versa.

The questions we want to answer with these simulations are the following. What happens to the frequency of cooperation in the population over time? Does cooperation reach a stable level and stay there? How does the level depend on details of the psychological assumptions and on the payoff structure? Finally, can theory be offered to clarify the answers to the questions above? The feature that we are most interested in is the decision heuristic that is used to determine whether the subject will cooperate or defect on the subsequent trial.

In all of our simulations, we assume that the population is homogeneous and stable with regard to the heuristic used, but heterogeneous and changeable with regard to the behavior, cooperation or defection, that can be manifested. We want to assume a common psychological process or heuristic and investigate the behavioral consequences of that process. We could have assumed that the heuristics could change and that there was a "metaprocess" that governed heuristic selection. In this latter case, the common psychology would be the metaprocess, and the heuristic used would be part of the cognitive variability underlying the behavioral variability. We have limited ourselves here to the most basic level but recognize that the approach can be extended to more complex levels in which both heuristics and behaviors are heterogeneous and changing.

Decision Heuristics

TFT

Of the three decision heuristics that we explore, TFT is the best known and the simplest. Using TFT, the subject simply mimics the neighbor's choice on subject's subsequent play. It is important to note the difference between TFT as it is applied in this study and the way it has been used previously. In most previous studies, TFT has been shown to be effective in maintaining high levels of cooperation when the interaction was between the same two individuals. In our simulation, this is not the case. When the subject becomes cooperative after having interacted with a cooperative neighbor, the individual will be cooperative.

Whether selected as a neighbor (regardless of which neighbor is the subject) or again selected as subject, the cooperator will cooperate with any neighbor selected, not just the one that was previously cooperative.

Reciprocity in this case is not a direct one-to-one reciprocity but a generalized reciprocity in which the cooperator cooperates with any individual with whom it interacts so long as it is in the cooperative state. The same is true with noncooperation. The reciprocation is broadcast to the entire neighborhood.

Also, unlike the TFT strategy submitted by Anatol Rapoport to Axelrod's (1984) tournament, our version of TFT did not always begin with a cooperative choice. Half the population began with C and half began with D as their initial choice. Ours was not a "nice" TFT.

Clearly, if a population becomes homogeneously cooperative or homogeneously defecting with TFT, it will stay that way indefinitely. Most simply, TFT is simple mimicry. It cannot change what it is imitating.

TFT is the one strategy that is sensitive to whether one or both participants change strategy after an interaction. If both individuals change and mimic the other, the overall prevalence of cooperation can never change from the initial level. For each sampled dyad, the proportion of cooperative choices will always be the same because each merely copies the other.

WSLC

A WSLC process is one that has two components. The first is an evaluative component that determines what is a win and what is a loss; the second is the action component that dictates what should be done contingent on the output of the first component. In this abstract sense, elementary reinforcement mechanisms are WSLC processes. Such mechanisms differentiate positive from negative reinforcers and postulate that perseveration of behavior (stay) will tend to follow positive reinforcers while extinction (change) will follow negative reinforcers or none at all. Thus the WSLC process might be taken to be a primitive adaptation mechanism that steers the organism toward positive outcomes (approach) and away from aversive ones (avoidance).

In previous applications of this idea to social interaction, wins and losses were defined simply by the magnitudes of the outcomes themselves. The situation gets complicated, however, by the fact that in many cases, the coding of outcomes into wins and losses depends on how the outcomes are framed (Tversky & Kahneman, 1986). A central element of this framing has to do with the location of the reference point or the comparison level (Thibaut & Kelley, 1959) that is used to evaluate the outcome. The same outcome can be coded as either a win or a loss depending on whether it is above or below the relevant reference point. Moreover, it is well-known that in social contexts the reference point will be strongly influenced by one's knowledge of the outcomes of similar others. Whether an outcome is "good" or "bad," whether it is a "win" or "loss," will depend on the social comparison one makes with the relevant reference group (Pettigrew, 1967).

In our simulations, we want to embrace the realism that derives from assuming that outcomes are socially evaluated. As a result, we defined *win* and *lose* socially in terms of the payoffs

of the neighbors. Specifically, "win" results when the payoff to the subject is equal to or greater than the average of the payoffs of the subject's neighbors. After the subject and its neighbor play the game, the payoff to the subject is compared with a reference point that is calculated as follows: The payoff to the neighbor from that play is averaged together with the other seven neighbors' payoffs from the last time they were subjects. If the subject's payoff is at least as large as this average, the subject "wins." When the subject's payoff is smaller than this average of the neighbors, the subject "loses." After a win, the subject repeats its choice; after a loss, the subject changes to the other choice. (In our simulations, we initialize all subjects' initial payoffs at 1.5, the mean of the four numbers in the payoff matrix.)

Several implications of this rule emerge immediately. After a play in which the subject defected and the neighbor cooperated (DC), the subject's payoff is 3 and the neighbor's is 0. This outcome will always be coded as a win since 3 is the maximum payoff available (and the neighbors' average, including the 0, will have to be less than 3). Likewise, the combination of a C choice by the subject and a D choice by the neighbor (CD) will always be coded as a loss because the subject gets 0 and the average of the neighbors will include at least the one 3. Consequently, after both of these outcomes, the subject's next choice will be D, a "stay" after the DC interaction and a "change" after the CD play. Hence the only way the subject can make a C choice with this heuristic is after a CC (when the neighbors' average is no more than 2) or after a DD (when neighbors' average is equal to or above 1).

WCLD

The previous rule, WSLC, is content free in the sense that it is blind to the consequences of its choices. C and D are interchangeable, and the WSLC does not discriminate between them. WCLD is different in that it follows a win with a cooperative choice and a loss with defection. WCLD is a reciprocity rule whose justification comes from the matching of positive outcomes with positive outcomes. The C choice always provides a better payoff to the other but at a cost to the chooser. It is precisely this quality that has allowed the PDG to be used as a model of altruism. Because there is attention to the content of the choices, WCLD becomes an affective version of TFT rather than a mere mimic. In WCLD, the response to the positive outcome of winning is cooperating, giving a positive outcome (at some cost), and the response to the negative outcome of losing is defecting, making the dominating choice. Win and loss are defined, of course, as in WSLC. WCLD is a kind of reciprocation rule. Later we discuss its relationship to TFT.

There is evidence in the social psychological literature to support the principles on which this rule is based. It is known, for example, that positive moods enhance prosocial tendencies (Krebs & Miller, 1985) and that these tendencies are not restricted to the person or persons responsible for the positive mood. Being the beneficiary of a generous act by one person enhances the likelihood of behaving generously to another unrelated person. On the negative side, the tendency to punish one's spouse for the frustrations experienced in the workplace is commonplace. Reciprocity is affective, not imitative, and it may be directed at parties other than the instigator.

Table 1
Payoff and Next-Choice Consequences for Each Pair of Subject and Neighbor Choices for the Three Rules

Consequence	Choice pair (S, N)			
	C, C	C, D	D, C	D, D
Payoff to subject	2	0	3	1
Next choice by rule				
TFT	C	D	C	D
WSLC	C (win), D (lose)	D	D	D (win), C (lose)
WCLD	C (win), D (lose)	D	C	C (win), D (lose)

Note. S = subject; N = neighbor; C = cooperate; D = defect; TFT = tit-for-tat; WSLC = win-stay, lose-change; WCLD = win-cooperate, lose-defect.

As with WSLC, there are some immediate consequences of this rule. The DC and CD outcomes will always be coded as wins and losses, respectively, as we described above. However, the DC combination, which always gives a win, will lead to C as the next choice with WCLD.

The implications of the three rules are outlined in Table 1. In this table, we specify, if possible, what the payoff and choice consequences are for each combination of subject and neighbor choices. It is impossible to say what the next choice of the WSLC and WCLD will be after CC and DD choices because it depends on the neighbors' average payoff. All rules follow the CD choice with a D. The CC combination is more likely than the DD to be coded as a win because the payoff is 2 rather than 1. Both WSLC and WCLD will result in a choice of C if it is a win and D if not. Thus these two rules are identical in their response to the consequences of their own C choice. However, they are exactly opposite in their responses to the consequences of their D choices. WSLC and WCLD respond to the coding of the DD outcome in opposite ways—WSLC repeats D to a win and changes to C after a loss, whereas WCLD does the reverse. The two reciprocal rules, TFT and WCLD, follow the win of DC with a cooperative choice, while the WSLC repeats D. In fact, the two reciprocal rules have identical responses to the DC and CD plays, and if the neighbors' means were always between 1 and 2, these two rules would be identical. We return to this point later.

For the WSLC and the WCLD rules, the decision to code outcome equality as a win was made for psychological reasons. The zero point on an evaluative outcome scale partitions the scale into positive and negative regions. We know of no research that poses the question of whether 0 itself acts more like a win or a loss. Evidence that the slope of a value function is steeper for losses than for gains (Kahneman & Tversky, 1979), and evidence that in interpersonal comparisons it is more aversive to get a unit less than another gets than it is attractive to get a unit more (Lowenstein, Thompson, & Bazerman, 1989; Messick & Sentis, 1985; Messick & Thorngate, 1967), suggests that equality is more reasonably treated as a gain or a win than as a loss. Nevertheless, we present evidence bearing on the importance of the assumption. It is clear *a priori*, for example, that it makes a difference in terms of the stability of homogeneous populations for WCLD. All D would not be a stable outcome with the

WCLD, because as soon as all outcomes were the same in a neighborhood, the subject would switch to C. All C will be stable, of course.

Theory

The descriptions of the heuristics in the last section constitute the individual social psychological processes under investigation. In this section, we focus on the generalization of these processes to the aggregate or population level. Because the TFT heuristic is qualitatively different from and simpler than WSLC and WCLD, we concentrate on these latter two heuristics.

WSLC

The elements that are needed for the generalization are presented in Table 2 for WSLC. In this table, we assume that the overall level of cooperation in the population is a. (We make the simplifying assumption of independence among the members of the population to carry out the analysis.) The second row of Table 2 presents the probability of each of the possible choice pairs, and the third row presents the conditional probabilities for WSLC of a cooperative choice following each choice pair. As we noted in the last section, cooperation will never follow CD or DC with WSLC, so those conditional probabilities are 0. Following elementary rules of probability theory, we can establish the equilibrium equation below. This equation expresses the relationship between the overall level of cooperation, a, and the conditional probabilities of cooperation following mutual cooperation and mutual defection.

$$P(c) = a = w_1 a^2 + w_2(1 - a)^2. \tag{1}$$

Equation 1 is easily solved for w_1. Equation 2 shows that the relationship between the two conditional probabilities is a family of negative linear functions whose slopes and intercepts depend on a.

$$w_1 = a^{-1} - ((1 - a)/a)^2 w_2. \tag{2}$$

Before examining the implications of Equation 2, it is useful to consider the conditional probabilities w_1 and w_2 in more detail. First, w_1 is the probability that joint cooperation is followed

Table 2

Unconditional Probabilities of Choice Pairs and Conditional Probabilities of Cooperation for WSLC and WCLD

Probability	Choice pair (S, N)			
	C, C	C, D	D, C	D, D
Probability of pair	a^2	$a(1 - a)$	$(1 - a)a$	$(1 - a)^2$
$P(C/S, N)$ for WSLC	w_1	0	0	w_2
$P(C/S, N)$ for WCLD	w_1	0	1	w_3

Note. WSLC = win-stay, lose-change; WCLD = win-cooperate, lose-defect; S = subject; N = neighbor; C = cooperate; D = defect; a = overall prevalence of cooperation; w_1 = probability that joint cooperation is followed by a win or rewarded; w_2 = probability that joint defection is followed by a loss or punished; w_3 = probability of cooperation following mutual defection.
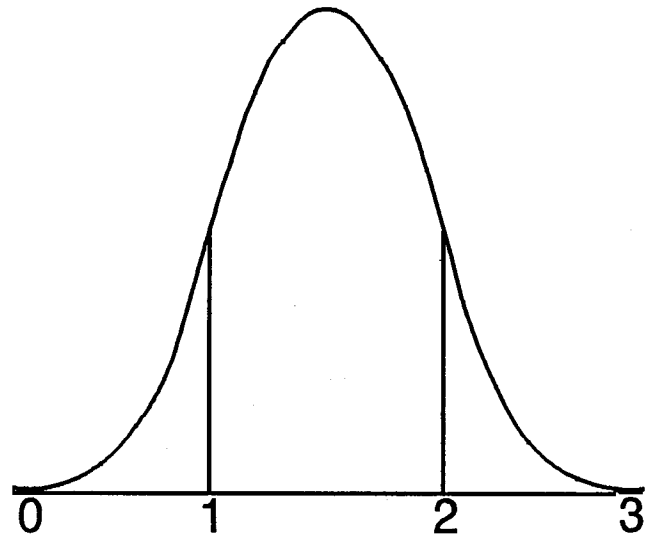

Figure 2. The hypothetical distribution of neighbors' mean payoffs and the cutoff values for the conditional probabilities, w_1 and w_2. The probability w_1 is the area above and to the left of the value 2, and w_2 is the area to the right of the value 1.

by a win or reward, and w_2 is the probability that joint defection is followed by a loss or punished. Imagine sampling all the points in the population and, for each point, calculating the mean of the neighbors' payoffs. The distribution of these means is the distribution from which the conditional probabilities are derived. Figure 2 displays a hypothetical distribution of these means. It is simple to show that the mean of this distribution is equal to $(1 + a)$ for the payoff values we have used.[2] The greater the level of cooperation, the higher the average payoff in the group. When $a = .50$, the distribution will be approximately symmetric. When a is either 0 or 1, the mean payoff will be 1 (uniform defection) or 2 (uniform cooperation) with zero variance.

The variable w_1 is the probability that the sampled mean is equal to or smaller than 2, the payoff for joint cooperation. That is, w_1 is the fraction of this distribution that falls on and to the left of 2 in Figure 2. The variable w_2 is the probability that the sampled mean is greater than 1, the payoff for joint defection. It is the fraction of the distribution in Figure 2 that falls to the right of 1. Making these probabilities explicit will aid in interpreting the theory and the effects of the two experiments to be presented shortly.

One of the immediate and surprising implications of Equation 2 is that it does not have solutions for $a > .50$ consistent with w_1 and w_2 being probabilities, that is for $0 \le w_1, w_2 \le 1$. Thus, the first theoretical result is that *WSLC cannot generate more than 50% cooperation in the population.*

A second feature of the equilibrium equation is that a depends more on w_2 than on w_1. This is perhaps best visualized in

---

[2] If the probability of cooperation is a, and if subject's and neighbor's choices are independent, then the mean outcome will be given by: mean outcome = $2a^2 + 3a(1 - a) + (1 - a)^2 = 1 + a$.
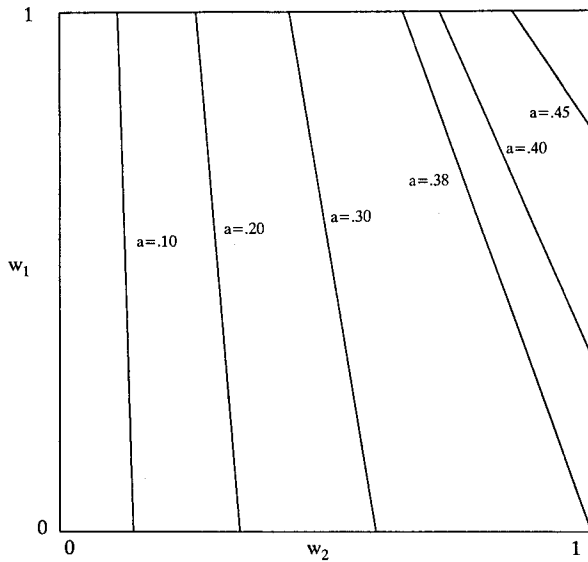
Figure 3. Equal cooperation contours for win-stay, lose-change. The variable w_1 is the probability of cooperation following mutual cooperation, w_2 is the probability of cooperation following mutual defection, a is the overall prevalence of cooperation.

Figure 3 in which we have plotted the equicooperation contours or the linear functions relating the two conditional probabilities for different values of a. The main aspect to notice is that the contours are mainly vertical, indicating that changes in the horizontal axis, w_2, are more important in determining a than are changes in the vertical axis. The importance of w_1 increases (the slope becomes more horizontal) as a increases. We note that the function for $a = .38$ passes through the lower right-hand corner of the figure, when $w_1 = 0$ and $w_2 = 1$. Thus, when $w_2 = 1$, .38 $< a < .50$. This is an important fact in interpreting our findings.

We can be more specific and say that factors that increase or decrease w_2 will increase or decrease a, the global level of cooperation (whose maximum, recall, is .5 and whose minimum is 0). The variable w_2 is the probability that the mean of the neighbors' outcomes is greater than the DD payoff, 1. In the simulations that we report, we examine factors that influence this probability. We examine one parameter of the evaluation process and one associated with the payoff structure.

WCLD

Using the same approach as above, cross multiplying and summing conditional and unconditional probabilities, we find the equilibrium equation for WCLD as below.

$$P(c) = a = a^2 w_1 + a(1 - a) + (1 - a)^2 w_3. \qquad (3)$$

In this expression, w_3 is the complement of w_2 in Equation 2. It is the area above and to the *left* of one in Figure 2. A win following mutual defection with WCLD results in cooperation, while with WSLC it results in defection. Some trivial tweaking of this expression yields an explicit relationship between the two conditional probabilities in terms of the parameter a.

$$w_1 = 1 - ((1 - a)/a)^2 w_3. \qquad (4)$$

Again, the two conditional probabilities are linearly and negatively related. Equation 4 may be displayed as a family of functions, as in Figure 4, passing through the point $w_1 = 1$, $w_3 = 0$ in the northwest corner of the figure. Unlike in Figure 3, it is clear that (nearly) all values of a are possible. (Homogeneous defection, $a = 0$, is not possible. If the mean of all neighbors' outcomes were 1 and the subject's payoff were 1, the comparison would be coded as a win, and subject would next cooperate.) All else equal, an increase (or decrease) in either conditional probability increases (or decreases) a. Finally, there is a different asymmetry in this case from WSLC. Complete cooperation, $a = 1$, can only result if $w_1 = 1$, and complete defection can be obtained only if $w_3 = 0$. These are necessary but not sufficient conditions for homogeneous cooperation or defection, respectively.

The Simulations

Basic Elements

Certain default initializations characterize all our simulations unless we state otherwise. We begin all runs by randomly assigning cooperative and defecting strategies with equal probability. We begin with an initial payoff of 1.5 to each individual. We selected this value because 1.5 is the average of the four payoffs in the payoff matrix.

To provide the reader with a visual image of the behavior of the groups that we simulated, we first present trial by trial results for one simulation for a population of 100 for each of the choice heuristics we investigated. In Figure 5 we display the
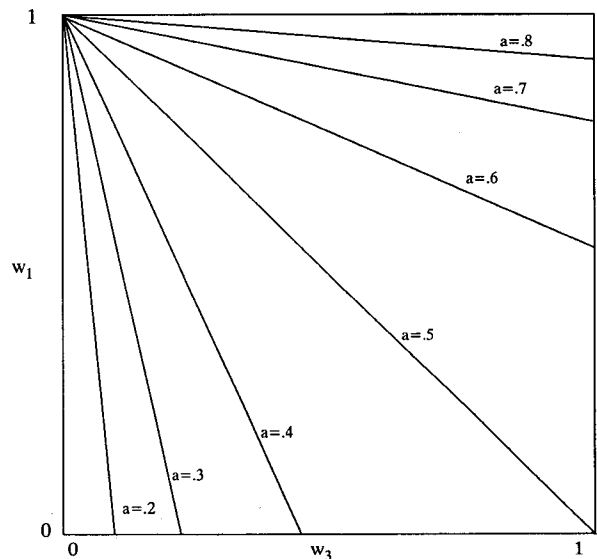

Figure 4. Equal cooperation contours for win-cooperate, lose-defect. The variable w_1 is the probability of cooperation following mutual cooperation and w_3 is the probability of cooperation following mutual defection; w_3 is the area to the left of the value 1 in Figure 2. The variable a is the overall prevalence of cooperation.
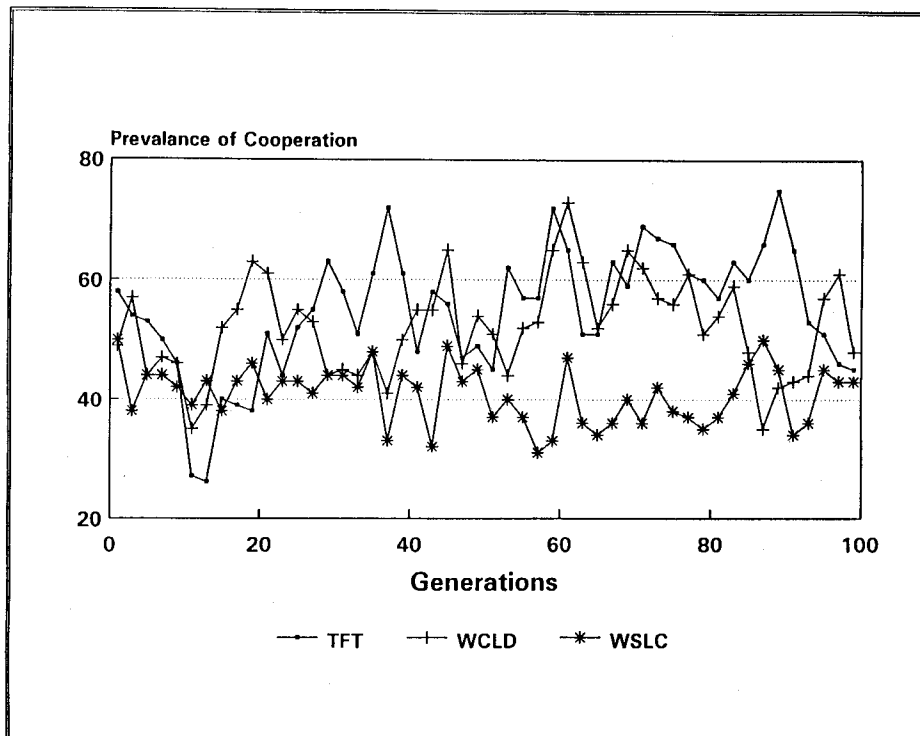
Figure 5. Prevalence of cooperation over generations for the three choice rules. TFT = tit-for-tat; WCLD = win-cooperate, lose-defect; WSLC = win-stay, lose-change.

number of cooperators in the population for even-numbered generations between 0 and 100 generations. The first and most important thing to notice about this figure is that cooperation does not disappear for any of the choice rules. The number of cooperators in each generation seems to vary considerably for the TFT rule, somewhat less so for WSLC, and least for WCLD. Furthermore, the number of cooperators appears to be consistently greater for the WCLD rule than for WSLC. Subsequent runs confirm this impression of both means and variability.

A description solely in terms of the number of cooperators in the population would be an incomplete summary of the structure of cooperation in these groups. We can describe the population in terms of how well the individuals did, whether there is a difference between cooperators and defectors in terms of their winnings, and whether there is a tendency for the population to become segregated so that cooperators tend to be close to other cooperators, and likewise for defectors. We have examined all of these factors, but for the current article, we restrict ourselves to the prevalence of cooperation as the primary measure of interest. None of the analyses that we omit qualify the generalizations we reach in this article.

Before turning to our major goal, the description of the prevalence and dynamics of cooperation in large groups, we must comment on some crucial differences we discovered between large groups and small ones. We do not claim that the theory developed previously is applicable to small groups because small groups are more dependent in that a given person is more likely to be a neighbor of any other person than in larger groups.

In groups of nine, for instance, each individual is a neighbor of every other individual.

Small groups versus large groups. We take a group of size nine to be the smallest group with which we will be concerned. This choice is based on the symmetrical structure that, in a group of nine, each individual has eight neighbors and each individual is neighbor to every other individual. We have found important differences between groups of size nine and larger groups.

The major finding with groups of nine is that such groups always become permanently homogeneous, either with cooperation or defection. For a group to become permanently homogeneous, it must be that once the group is in the state it will never leave, and there must be a positive probability that the group can enter the state. With the TFT strategy, the probability is one half that the group will become all cooperative and one half that it will become all defecting. Because this strategy is purely imitative, there are no forces that favor either cooperation or defection. Our experiments, which have now involved thousands of runs of TFT, confirm that the all-C and all-D outcomes occur equally often. It is important to understand that once a population using TFT has become homogeneous it will remain so forever. In fact, TFT with a nine-person group can be summarized as a 10-state Markov chain (states are defined by the number of cooperators) with stationary transition probabilities and two absorbing states, all-C or all-D.

What is less obvious is that with nine-person groups, both WSLC and WCLD also become permanently homogeneous.

WSLC becomes all-D and WCLD becomes all-C. With the latter, one can see that if the group entered the state in which all members were C so that they exchanged 2 points per play, they would receive outcomes that are coded as wins because they are equal to the mean outcomes of their neighbors, and they would continue to cooperate.

The properties of WSLC are somewhat more subtle. First, it is not possible to go into an all-C state. The all-C state must be preceded by an eight-C and one-D state in which the D individual is subject. The D subject will always receive 3 points interacting with any neighbor, and the 3 points will always be a win, causing the D to stay. It is possible for the group to become all D, on the other hand. In an eight-D, one-C group, when the C is subject, the C will receive 0 points, which will always be coded as a loss, causing the cooperator to change to defection. However, the new all-D group will only be stable if the mean of the payoffs is one or less, causing the payoff for mutual defection (1 unit) to be coded as a win.

The discussion above suggests that the processes by means of which small groups become homogeneous are different for the three different heuristics. One manifestation of these differences is in the speed with which the groups become homogeneous. In a series of simulations, we estimated that TFT is the fastest, requiring an average of 5.42 generations to enter an absorbing state. WCLD is next, requiring an average of 15.04 generations, and by far the slowest is WSLC, which took an average of 65.3 generations to converge. These means are clearly significantly different, and in this case, as in other cases where differences are obvious, we do not report statistical tests or significance levels.

The discovery that all of these rules converge in 9-person groups suggests two generalizations. First, there may be absorbing states for groups of all sizes, and with enough generations, groups of any size will converge to one of the two homogeneous states. We cannot offer a formal proof of this suggestion, but it seems plausible that the principles that allow small groups to converge would also allow large groups to converge, even though it might take very long to do so. We have some evidence to support this suggestion. In experiments in which we have measured trials to convergence, we have found that the TFT, as one would expect, rather quickly converges for all group sizes we have studied up to 400. In Figure 6 we display the estimated generations to convergence for this rule for (square) population sizes from 9 to 100. (Each mean is based on 50 runs in which the simulation was stopped when the population converged.) The generations to convergence increase as a linear function of the size of the population. When the population is 400, the mean number of generations to convergence is 397.

The situation is somewhat different for the other two rules. For WCLD, which converged in an average of 15.08 generations when the population was 9, the mean was 102.3 for a population of 16 and 613.95 for a population of 25. Even over this small range, time to convergence is not linear in population size, and the increase in time to convergence is much more dramatic than that for TFT. With the WSLC rule, the situation is even more precipitous. This rule took an average of 65.3 generations to converge with a 9-person population and 1,645 generations to converge, on average, with a population of 16. We were not able to estimate the mean time to convergence for populations of 25. For technical and practical reasons we did not run simulations

longer than 30,000 generations. We ran a series in which the simulation stopped when the number of generations reached 30,000 or when the population converged, whichever happened first. In 40 runs, only 8 resulted in convergence within the 30,000 generations.

The first generalization is that there may, in principle, be absorbing states for all populations. The second is that with large populations, populations of 50 or 100 or more, the likelihood of convergence is so small that it can effectively be treated as 0. If time to convergence is essentially infinite, then the interesting questions have to do with the dynamics of cooperation in heterogeneous populations. Are there stable levels of cooperation supported by WSLC and WCLD? If so, what are those levels and how do they depend on factors like the size of the population, the initial distribution of cooperators and defectors, the values of the payoff matrix, and the nature of the criterion distinguishing a gain from a loss? It is to these issues that we now turn.

Cooperation in large groups. Of the scores of simulations that we have conducted, we summarize the results of two major experiments. We have selected these two experiments to describe because they are psychologically interesting and because they provide tests of the theory developed in the previous section. We first outline some basic findings regarding the global consequences of WSLC and WCLD, rules that do not become homogeneous in populations of 100, which will be our basic size.

Basic findings. We can summarize our basic findings by describing an experiment in which we manipulated the size of the population and the heuristic rule, either WSLC or WCLD, used to generate choices. In this experiment, we varied the population from 36 (6^2), 64 (8^2), 100 (10^2), 144 (12^2), to 196 (14^2). For each population size, each rule, following a random 50-50 initial assignment of cooperation and defection, was run for 50 generations. The status of the population at the end of the 50 generations was measured. Each rule was replicated 50 times for each population size, yielding a total of 500 simulations, each of which was 50 generations long. This experiment allows us to investigate the joint effects of population size and to rule on the prevalence and distribution of cooperation.

In this as in the other experiments, we do not list values of F ratios or present significance levels for most of the results we describe. We discuss only effects that are large and significant. Our statistical tests are all extremely powerful because the noise levels are moderate at worst and because the experiments usually have a minimum of 500 degrees of freedom for the error terms. The standard error for the means in the graphs that will be plotted (based on 50 replications) are .013 or smaller, so the narrow confidence intervals of $|.052|$ or less are not shown.

The only factor influencing the prevalence of cooperation was the rule. WSLC generated an average of 40.2% cooperation, whereas WCLD led to 55.4% cooperation. These levels have been replicated in hundreds of further simulations. Mean levels of cooperation for the first 30 generations are presented in Figure 7. There was no qualitative change thereafter. These data from our basic experiment provide a baseline against which a variety of changes can be compared. It is to these other questions that we now turn our attention.

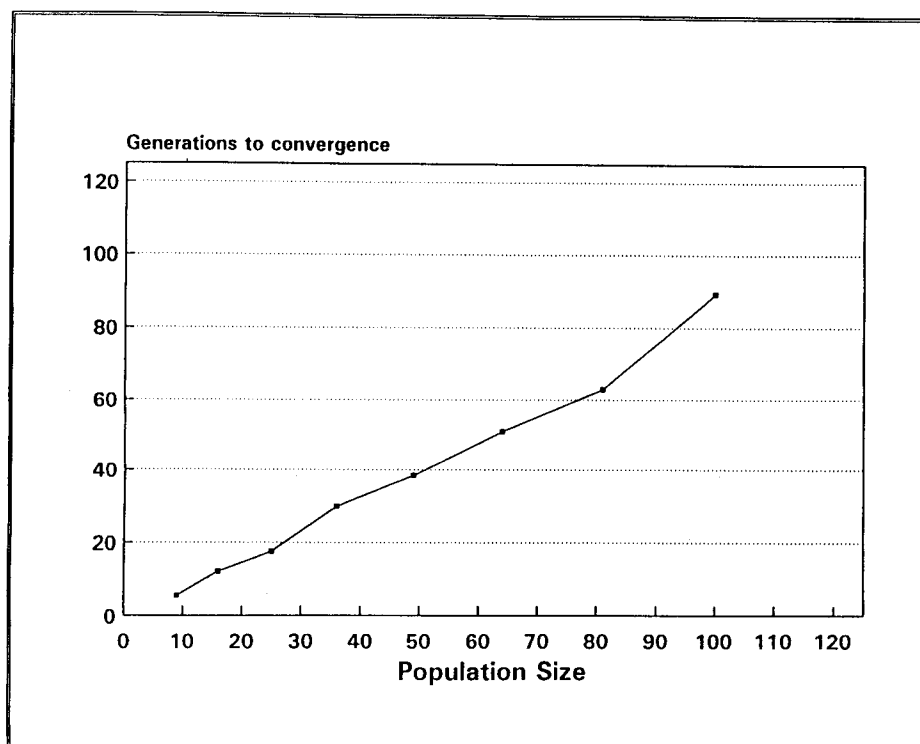*Temptation payoff.* One of the most commonly studied fea-

Figure 6. Mean generations to convergence for tit-for-tat rule as a function of population size. Each mean is based on 50 simulations.

tures of the PDG is the payoff structure itself (Pruitt & Kimmel, 1977). How do changes in the payoffs influence the levels of cooperation? We are interested in the same question, but we remind the reader that our simulations are very different from the traditional two-person experiments that have been reported in the experimental social psychological literature. These latter studies typically involved two persons interacting exclusively with each other over a long series of trials. Outcome evaluations in the two-person setting will be restricted to comparisons with the other's outcomes (see Messick & Thorngate, 1967, for evidence of the importance of this type of comparison), and this type of comparison will be different from comparison with all neighbors. Our simulations are not intended to represent the dynamics of cooperation in one-on-one encounters. The whole point of our simulations is to investigate global consequences for cooperation under very different conditions. There are many ways in which the payoff structure of a PDG can be manipulated. Cooperative indices have been defined in terms of the payoffs in the matrix (Rapoport & Chammah, 1965), concepts of fear and greed have been defined as functions of the matrix (Coombs, 1973), and Messick and McClintock (1968) showed how additive matrices, like the one used here, can be decomposed into additive components.

We describe the results of an experiment involving the so-called "temptation" parameter (T), the payoff received by a defector when the other person cooperates. This payoff will always be the largest in a PDG matrix, and the goal of the study we conducted was to assess the impact of increasing the temptation parameter from its canonical value of 3. In this experiment, we

let T vary from 3 to 19 in steps of 4. We simultaneously varied the population size through three levels: 49, 100, and 225. (Even though our basic experiments had shown no effects for population size, we occasionally included it as a factor to check for the possibility that size might interact with some of the other factors we examined.) Of course, we also varied the heuristic choice rule. Each replication ran for 50 generations, and there were 50 replications of each combination of rule, population size, and T level.

How will increases in the magnitude of T influence cooperation? The impact of this change will come about by changing the distribution of the mean neighbors' score. By increasing T, the maximum value of the outcomes will be increased and the mean of the distribution will also be increased. In other words, the distribution will be shifted to the right, and the larger T, the greater the shift to the right. As the distribution is shifted to the right, w_1 will decrease (from its baseline level) and w_2 will increase. The larger T becomes, the greater the change.

The effects for WSLC are clear. The baseline condition is approximately along the line labeled $a = .40$ in Figure 3. So the movement is in the direction of the lower right-hand corner. By the time $T = 13$, if only one neighbor has an outcome of 13, the mean of the neighbors will exceed 1, so as T increases, w_2 will approach 1. The prediction is, then, that for WSLC, the level of cooperation will decline as T increases, approaching a value of about 38%. Thus we do not expect large changes in a for WSLC.

For WCLD the situation is very different. The conditional probabilities will change in the same manner, and w_3 will decrease. (Recall it is the complement of w_2.) Thus movement in
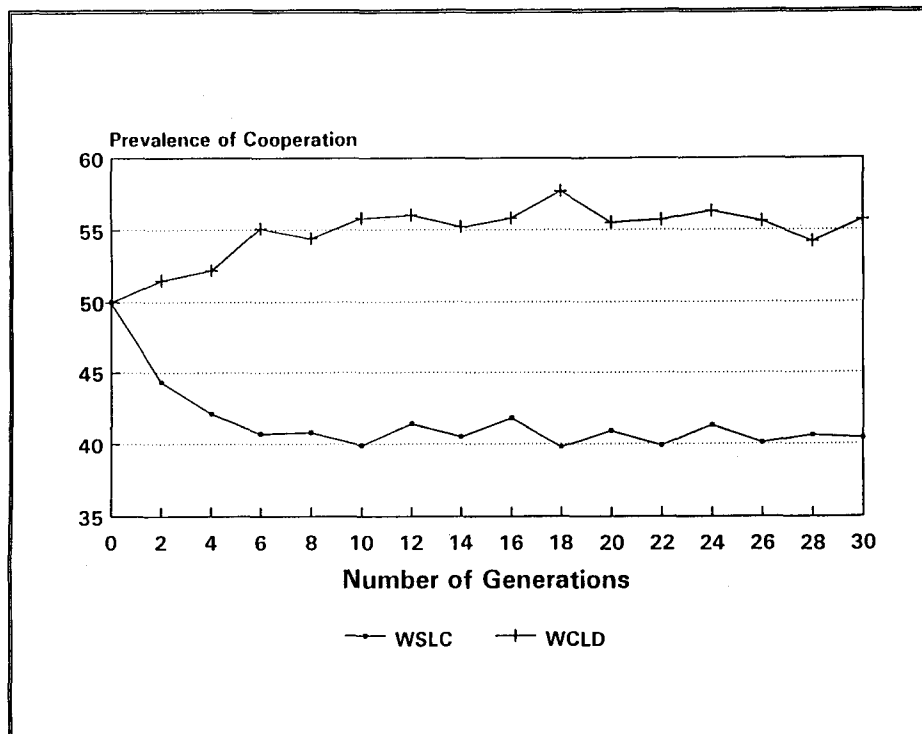
Figure 7.   Mean prevalence of cooperation for WSLC and WCLD as a function of generations. Each mean is based on 50 simulations. WSLC = win-stay, lose-change; WCLD = win-cooperate, lose-defect.

the parameter space in Figure 4 will be from the hypothetical line of a = .55 in the direction of the lower left-hand corner. Cooperation should decrease. As we noted earlier, WCLD cannot become homogeneously defecting because equality in the comparison process is coded as a win. Hence we can predict that cooperation will decrease with WCLD, but we cannot say by how much.

The main findings of this experiment with regard to the prevalence of cooperation are displayed in Figure 8, where we have plotted the prevalence of cooperation for both rules against the temptation parameter. Cooperation decreases as T increases, as we predicted, and the impact of changing T is much greater for WCLD than it is for WSLC. For WSLC, the value of a for the larger levels of T was 36%, quite close to the value of 38% that we identified theoretically.

WCLD is most sensitive to changes in T when T is small. Cooperation drops from the basic rate of about 56% at T = 3 to about 33% when T is increased to 7, and the rate of cooperation falls only marginally thereafter to about 30%. We see neither an *a priori* nor an *a posteriori* reason why WCLD stabilizes at this level.

Evaluation processes.   The evaluation process is the psychological heart of our simulations. It is here where we make assumptions about the intraindividual psychological processes that lead to cooperative or noncooperative choices. Therefore, it is important to investigate the extent to which changes in our assumptions lead to changes in the prevalence of cooperation in groups.

The variable that we examine concerns the criterion that

differentiates a positive outcome (win) from a negative one (loss). In other words, we vary the way wins and losses are framed. In the simulations that we have described so far, a win is coded when the outcome is equal to or above the reference point defined by the average of the neighbors' outcomes. We could, on the other hand, make the nonsocial assumption that the reference point is a constant that does not depend on the neighbors' outcomes. Alternatively, we might consider it a win only if we were strictly better than our neighbors. Or we might need to be better than our neighbors by some positive (or negative) amount. In this section we examine all of these issues.

The first question is what difference does it make if we use a fixed, nonsocial reference point to differentiate wins from losses. Assume that the relevant reference point is the average of the four scores that are available in the payoff matrix (namely, 1.5) so that a win is an outcome that is above 1.5 and a loss is one that falls below this boundary. This process makes the outcomes of 2 and 3 wins and of 0 and 1 losses. With this classification, the WCLD rule simply becomes TFT because one wins when the neighbor chooses cooperatively and loses when the neighbor defects. Thus the solitary feature that differentiates TFT from WCLD is the use of a local social reference point for WCLD and the use of a fixed one for TFT. The large-group implications of this difference are, as we have seen, immense. The changes in the dynamics are qualitative, not quantitative.

With WSLC, using a fixed reference point that falls between 1 and 2 produces a different type of process as well. In this case, it is simple to show that the rule resulting from WSLC with a fixed reference is to cooperate when the subject and the neigh-
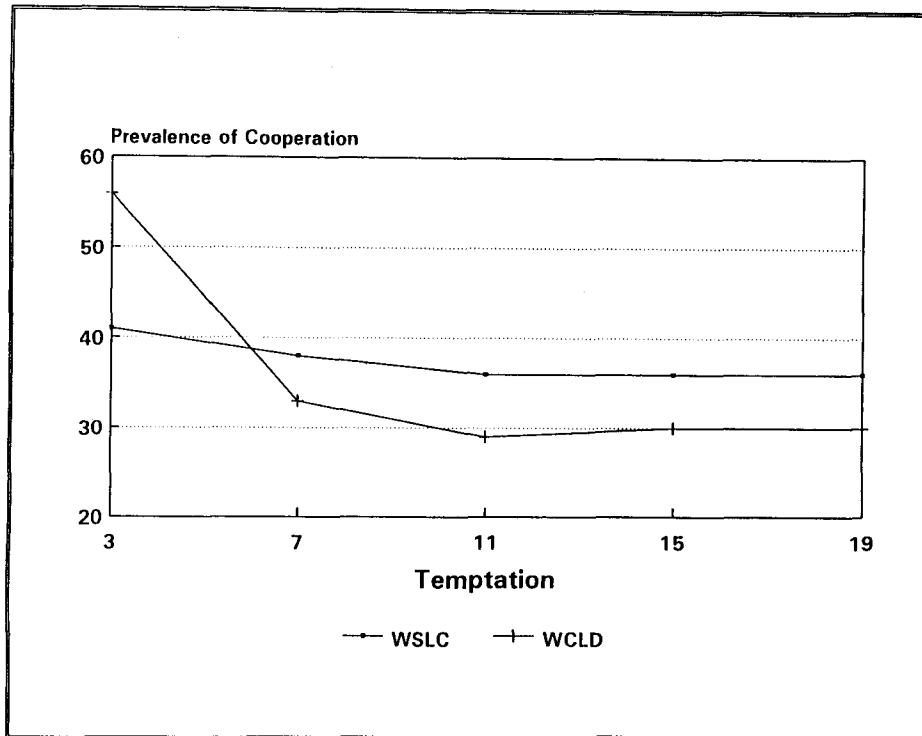
Figure 8. Mean prevalence of cooperation for WSLC and WCLD as a function of the temptation parameter. Each mean is based on 50 simulations. WSLC = win-stay, lose-change; WCLD = win-cooperate, lose-defect.

bor make the same choice and to defect when they make different choices. This rule could never lead to the homogeneous defection we observed in small groups with WSLC because mutual defection leads the subject to cooperate. Because mutual cooperation leads to cooperation, if the group ever reached a homogeneously cooperative state, it would remain there. However, under the rules of our simulations, the population can never become homogeneously cooperative. To see why this is so, imagine a population that is all cooperators except one. When the single defector is selected as the subject, it will interact with a cooperative neighbor, leading to defection the next time around. Under slightly different simulation procedures, all-cooperate would be a feasible outcome. For example, if both the sampled subject *and* the sampled neighbor changed their choices according to this rule, then homogeneous cooperation would result. Again, changing from a social to a fixed reference point changes the dynamics of cooperation qualitatively.

Earlier in the article, we defended our decision to associate equality of outcomes with winning. Now we can ask whether this decision makes a major difference in the results of the simulation. If we code it as a loss rather than a win when the subject's outcome is identical to the neighbors' mean, w_1 will decrease and w_2 will increase. We cannot be more specific about the magnitudes of the changes, but in Figure 3 these changes map into movement down and to the right, relative to the $a = .40$ line. We can only predict that with WSLC, changing the coding of equality will not make a large difference because the two changes tend to be offsetting. As w_2 is the more important

parameter, we might expect cooperation to increase somewhat with this coding. (In many contexts, the coding of equality may not be important because exact equality of outcome and standard would be very unlikely. However, when the standard is calculated as the mean of eight integer outcomes, the likelihood of the standard taking the value of either 1 or 2 is considerable.)

WCLD should display a very different effect. Because both w_1 and w_3 are decreased, movement in Figure 4 will be down and leftward, implying that cooperation should decrease. By encoding fewer outcomes as wins, WCLD should display less cooperation.

We conducted several experiments in which we varied the population size, the rule used, and whether equality of outcomes was defined as a win or a loss. The results indicate that the placement of equality makes little difference for the WSLC rule. When equality is associated with a loss, the frequency of cooperation increases slightly from 41% to about 43%. For the WCLD rule, however, there is a larger impact. When equality is associated with a loss, cooperation drops from the 55% level to about 43%. As predicted, cooperation dropped noticeably. Population size had no effect.

The reference point separating winning from losing does not have to be at the average of the neighbors. For instance, one might believe that one deserved more than the average neighbor, or that one was happy, one won, with somewhat less than this benchmark. We explored the effects of shifting this boundary by systematically varying a bias parameter so that win is coded when the outcome plus a bias parameter (bias) is equal to or
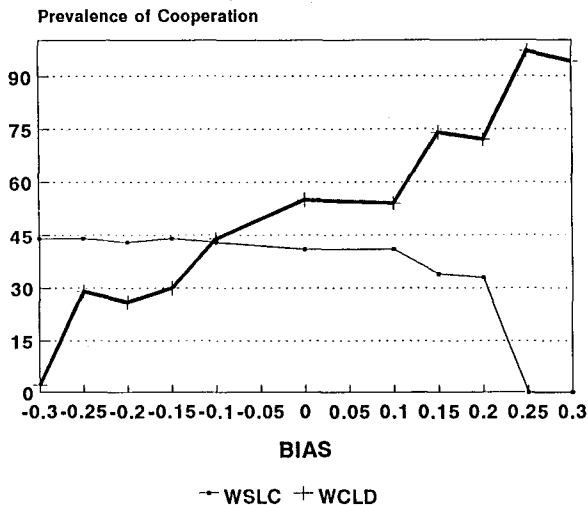
Prevalence of Cooperation

Figure 9. Mean prevalence of cooperation for win-stay, lose-change (WSLC) and win-cooperate, lose-defect (WCLD) as a function of additive bias. (Win is when subject's payoff + the bias parameter [bias] is at least as large as mean last payoff of neighbors.) Each mean is based on 50 simulations.

greater than the mean of the neighbors (and loss is coded otherwise). When bias is positive, winning is more inclusive than when bias is negative. The larger bias, therefore, the more frequent stay should be with WSLC and the more frequent should be with WCLD.

In theoretical terms, increasing bias increases w_1 and decreases w_2. Because w_2 is more important than w_1 for WSLC, we can expect that increasing bias will lead to a decrease in cooperation for WSLC. Decreasing bias should increase cooperation closer to .50 than the base rate of .41. With WCLD, increasing bias increases both w_1 and w_3, implying that cooperation will increase. Likewise, negative values of bias should lead to decreases in cooperation from the base rate of .55.

The data that we discuss result from varying bias from −.30 to .30 for both rules. We selected this range initially because it seemed small relative to the average value of the outcomes in our basic experiment. That value was about 1.4 for WSLC and 1.55 for WCLD.

The effect of manipulating bias is displayed in Figure 9. Each of the means plotted in the figure is the average of 50 simulations, each of which was 50 generations long. As anticipated, as bias increases, the prevalence of cooperation increases for WCLD. It decreases for WSLC. With bias of .30 we find homogeneous defection for WSLC and nearly homogeneous cooperation for WCLD.

Perhaps the most striking aspect of these results is the sensitivity of WCLD to bias. Over the range of this variable that we explored, cooperation rises from nearly 0 when bias = −.30 to nearly 100% when it is .30. The WSLC rule, on the other hand, is relatively insensitive to bias until it gets large, that is to say greater than .20. For bias = −.10 or less, we observe the 43% to 44% cooperation that we saw when equality was associated with losing. From 0 to .1 it is at the basic level of about 40%. It then

drops to 33% to 34% for bias = .15 and .2 and thereafter goes to 0, strongly suggesting a discontinuity.

The function for WCLD also hints that the relationship between bias and cooperation in large groups may be discontinuous. Cooperation jumps from nearly 0 to about 28% between −.30 and −.25, but between −.25 and −.15 inclusive, it is essentially flat. It then jumps to 44% at −.1, which is the level we observed when equality was associated with losing. At bias = 0 and .1, we observe the level of cooperation that we saw in our basic experiment and that we have witnessed scores of times since, namely about 55% cooperation, but at .15 and .20 the level jumps to about 73%. At .25 and .30 the prevalence of cooperation is about 95%, and most of the simulations have become homogeneously cooperative.[3]

The bias parameter could represent subjective biasing processes as well as shifts in criterion placement. For instance, for positive bias, one could think of the bias as resulting from tendencies to belittle the outcomes of others, to exaggerate the value of one's own outcomes, or both. Negative biases may represent the opposite tendencies: to demean one's own outcomes and to inflate the worth of others'. There is ample evidence for positive subjective bias in the experimental literature, and negative biases have often been associated with problems resulting from low self-esteem (Taylor, 1990). We have more to say about this in the last section of the article.

Conclusions

The most important conclusion to be drawn from the work we have described is this: In the highly competitive PDG, using a competitive type of social comparison for performance evaluation, cooperative behavior is maintained in large groups. In none of the rules that we studied, TFT, WSLC, nor WCLD, did the prevalence of cooperation regularly approach 0. Specifically, we believe that the cooperative consequences of WSLC and WCLD, rules that have had other applications in psychological theory, represent the discovery of a novel mechanism or process through which self-sacrificial cooperative behavior can be maintained in large groups of interacting individuals. We have shown that it is sufficient to have (a) a simple rule specifying the conditions for either changing one's response or for making a self-sacrificial choice for another person as a function of the evaluation of one's outcome and (b) an evaluation process that compares the outcome to a reference point that is based on the average outcome of neighboring individuals. Furthermore, we have outlined a theoretical approach that provides some guidance in making predictions about the global consequences of individual behavioral processes.

We have also shown that there are important differences between large and small groups. All of the rules lead to homogeneous small groups, but only TFT leads to homogeneity in large groups. The differences that we have found add a new dimension to discussions of group size effects. Some authors (e.g., Thibaut & Kelley, 1959) have suggested that there are qualitative

---

[3] The suggestion of a discontinuous function is not an illusion, but the discontinuity is not of psychological importance. The mean of the neighbors' outcomes can change only by discrete units of ⅛. Therefore, discontinuities in bias will be found in units of ⅛, or .125.

changes in group sizes between two and three (where coalitions become possible), but that there is nothing conceptually new above three. Although we cannot pinpoint the size at which small groups become large groups, we can confidently say that it is larger than three. In terms of our simulations, a group of size nine is the largest in which no group member can "hide," that is to say in which one might not be involved in the play, either through being the subject or being the neighbor of the subject.

For WSLC and WCLD there seems to be a critical size above which the qualitative aspects of the structure of cooperation change. Small groups become homogeneous, whereas cooperation remains everlastingly mixed with defection in larger groups, as far as we have been able to determine. However, once the group size has passed the critical value, it ceases to be important. We find scant evidence that size interacts with any of the other variables that we have studied in influencing the prevalence of cooperation.

We have found some factors that make a qualitative difference in the dynamics of cooperation, and these we take to be of extreme importance. We already mentioned group size as one such factor. Another is the nature of the evaluative criterion. When the dividing line between winning and losing is fixed, the dynamics of cooperation are very different than when the reference point is the mean of the neighbors. A fixed reference point between 1 and 2 makes WCLD behave exactly like TFT, which means (a) that it will become homogeneous rather quickly and (b) that it could become homogeneously defecting as well as cooperative. A similar fixed reference point for WSLC prevents that rule from becoming all defecting, even in small groups. It is thus clear that the fluid, local, social reference point is essential for preserving the quality of the dynamics of these rules.

A third factor that makes a qualitative change is the bias parameter. For extreme values of bias (.30 to −.30), WCLD becomes essentially either all cooperative or all defecting, and for large values of bias (.25 and up), WSLC becomes all defecting. Bias represents a continuous psychological variable that makes qualitative changes in the aggregate level of cooperation.

We have also explored many variations that do not make a qualitative difference in the dynamics of cooperation. We have varied the size of the neighborhood, the initial prevalence of cooperation, and the topology of the space—whether there are edges in the population. These factors change little. We have run simulations in which both the subject and the neighbor update their choices following interaction, and we have allowed the bias parameter to be an individual difference rather than a homogeneous constant. Both of these alterations produce small changes in the final levels of cooperation, but not qualitative differences from the results reported here. None of these factors had dramatic effects on the speed of convergence.

One of the two basic issues engaged in this work concerns the connections between individual-level behavior and aggregate-level outcomes. Many variations of this theme are well-known: Schelling's (1971) examination of the collective consequences of individual choices, Camerer's (1990) investigation of market reflections of individual cognitive biases, and Allison and Messick's (1985, 1987) comparison of individual and group decision-making processes. What the present work has shown is that simple heuristic rules, coupled with a realistic social evaluation

process and some rules for sampling participants, generate stable levels of cooperation in interacting groups. When we began this research, this fact was not at all apparent. After the fact, it seems reasonable. The rules that we selected were not selected because we thought that they would "work" in producing cooperation, but because they were a priori reasonable and plausible and because at least two of them, TFT and WSLC, had been examined in previous research. We view the cooperation that results from our simulations as a type of "emergent" social phenomenon that cannot be adequately described by reference to individual psychological processes. Likewise, a description of aggregate behavior will provide an inadequate insight into individual choice processes. An observer looking at two different asymptotic states of TFT, one of which had become all defecting and the other of which was all cooperative, would have great difficulty concluding that precisely the same process had produced both phenomena.

The WCLD rule has the interesting property that it gives a count in each generation of the number of individuals in the group who positively compared their outcome with the mean of their neighbors. That is to say it gives a count of the proportion who are "at least as good as the neighbors." The base rate for this percentage is about 55%. This figure conflicts with the intuition that the frequency of "winners" in a population should be about the same as the frequency of "losers." Clearly the reason for the conflict is that winning and losing in our simulations are not one-to-one contests that match a loser to every winner.

In social psychology there has been a great deal of research on the generic phenomenon that "most people think that they are above average" (see for example Myers & Ridl, 1979), and there have been extensive investigations into the "biases" that produce these cognitive distortions (Taylor, 1990). Our results prompt two observations that are relevant to this line of research. First, there is no inconsistency in having more than half the population reporting that they are "at least as good" or "at least as well off" as their neighbors. In our simulations, we know precisely when the comparers are and are not biased, and we consistently get more than half of the group making positive reports where there is no bias.

Second, it may not take a very large bias to lead nearly everyone in the group to the conclusion that they are winners. In our bias experiments with WCLD, we get close to 100% cooperation, which means that nearly everyone is comparing with neighbors favorably, when the bias parameter is .25. These findings suggest that a relatively small devaluation of the outcomes of others and/or a relatively small enhancement of own outcomes, plus the dynamics embodied in the simulation, may be all that is needed to lead to the perception that "everyone is above average" (Myers & Ridl, 1979). With the WCLD rule, taking a rosy view of one's outcomes leads to very high levels of cooperation in large groups, suggesting a possible social advantage of egocentric evaluative biases.

There is one issue left to be discussed, and this concerns the theoretical description of our simulations. We have offered an approximation to the statistical properties of our simulations. The resulting equations provided some insight into the behavior of the system created by the combination of the heuristics and the "sociological" rules governing the interaction. The theory yielded some qualitative and some quantitative predictions,

which were confirmed. Yet, the theory is still elementary and needs further development.

In simulations of the aggregate consequences of individual models of social impact, Nowak, Szamrej, and Latané (1990) and Lewenstein, Nowak, and Latané (in press) have argued that the field of statistical mechanics might provide theoretical insights. Statistical mechanics attempts to describe the behavior of complex physical systems that are created from simple underlying processes. Wolfram (1986), for example, has shown how complicated two-dimensional patterns such as self-replicating fractals (patterns that have themselves as their components and subcomponents) can be generated by the repeated application of very simple rules. Theoretically, our work seems closely related to these cellular automata with random components. Some of the processes that are created by the repeated application of simple rules, however, are so complicated that the only way that has been found to study them is through computer simulation. So it may be overly optimistic to expect the resulting theories to be simple.

There are two simple illustrations of this type of approach in previous social psychological experiments. Kelley et al. (1962) analyzed the ability of a WSLC strategy to lead to the emergence of mutual cooperation in a mutual-fate control game referred to as the "minimal social situation." The analysis consisted of applying this rule to the choices of two (or more) interacting partners and observing whether the end state of the process was mutual cooperation. These authors discovered, among other things, that WSLC would lead to mutual cooperation, but only when choices were made simultaneously (not alternatively) and when the group size was a power of two, that is, groups of size two, four, eight, and so forth.

A second example was offered by Messick (1967), who explored the heuristic strategies human subjects might use to play a competitive, zero-sum game against a computer using supposedly "normative" decision principles. To understand how the human did so much better than the computer strategies, Messick (1967) showed that the use of a simple variant of WSLC would lead to recurrent patterns of choices by the computer that the subject could exploit to be able to win on most of the trials. The WSLC variant, coupled with the computer's strategy, generated a repetitive periodic sequence of choices that the subjects mostly won. However, the dynamics of this two-party exchange and the WSLC analysis of Kelley et al. (1962) are orders of magnitudes simpler than the large-group problem that we have described here.

The work that we have described shows how simple behavioral rules, rules instantiating the most rudimentary forms of adaptive social interaction, can lead to unexpected global patterns in large groups. We have sketched a theoretical path that can be taken to generalize from individual rules to collective phenomena, but we have also identified areas where more theoretical insight is needed. Most important, perhaps, we have shown that through the use of computer simulation, it is possible to explore the relationship between individual and group behavior and to address, thereby, one of social psychology's oldest puzzles.

References

Allison, S. T., & Messick, D. M. (1985). The group attribution error. *Journal of Experimental Social Psychology, 21*, 563–579.

Allison, S. T., & Messick, D. M. (1987). From individual inputs to groups outputs and back again: Group processes and inferences about members. In C. Hendricks (Ed.), *Review of Personality and Social Psychology* (Vol. 8, pp. 111–143). Beverly Hills, CA: Sage.

Axelrod, R. (1984). *The evolution of cooperation.* New York: Basic Books.

Camerer, C. (1990). Do markets correct biases in probability judgments: Evidence from market experiments. In L. Green & J. H. Kagel (Eds.), *Advances in behavioral economics* (Vol. 2, pp. 126–172). Norwood, NJ: Ablex.

Campbell, D. T. (1975). On the conflict between biological and social evolution and between psychology and moral tradition. *American Psychologist, 30*, 1103–1126.

Coombs, C. H. (1973). A re-parameterization of the prisoners' dilemma game. *Behavioral Science, 18*, 424–428.

Elster, J. (1989). *The cement of society.* Cambridge, England: Cambridge University Press.

Franks, N. R. (1989). Army ants: A collective intelligence. *American Scientist, 77*, 139–145.

Gasik, S. (1990). *Definition of the Warsaw Simulation Language.* Unpublished manuscript.

Hardin, G. (1968). The tragedy of the commons. *Science, 162*, 1243–1248.

Isen, A. M. (1987). Positive affect, cognitive processes, and social behavior. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 203–254). New York: Academic Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–291.

Kelley, H. H., Thibaut, J. W., Radloff, R., & Mundy, D. (1962). The development of cooperation in the "minimal social situation." *Psychological Monographs, 76* (Whole No. 19).

Komorita, S. S. (1965). Cooperative choice in a prisoner's dilemma game. *Journal of Personality and Social Psychology, 2*, 741–745.

Krebs, D. L., & Miller, D. T. (1985). Altruism and aggression. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 1). Hillsdale, NJ: Erlbaum.

Kretch, D., & Crutchfield, R. (1948). *Theory and problems of social psychology.* New York: McGraw-Hill.

Latané, B. (1981). The psychology of social impact. *American Psychologist, 36*, 343–356.

Lewenstein, M., Nowak, A., & Latané, B. (in press). Statistical mechanics of social impact. *Physics Review "A".*

Lowenstein, G., Thompson, L., & Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology, 57*, 426–441.

Messick, D. M. (1967). Interdependent decision strategies in zero-sum games. *Behavioral Science, 12*, 33–48.

Messick, D. M., & Brewer, M. B. (1983). Solving social dilemmas: A review. In L. Wheeler & P. Shaver (Eds.), *Review of personality and social psychology* (Vol. 4, pp. 11–44). Beverly Hills, CA: Sage.

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology, 4*, 1–25.

Messick, D. M., & Sentis, K. P. (1985). Estimating social and nonsocial utility functions from ordinal data. *European Journal of Social Psychology, 15*, 389–399.

Messick, D. M., & Thorngate, W. B. (1967). Relative gain maximization in experimental games. *Journal of Experimental Social Psychology, 3*, 85–101.

Myers, D. G., & Ridl, J. (1979). Can we all be better than average? *Psychology Today, 13*, 89–92.

Nowak, A., Szamrej, J., & Latané, B. (1990). From private attitude to public opinion: A dynamic theory of social impact. *Psychological Review, 97*, 362–376.

Pettigrew, T. (1967). Social evaluation theory: Convergences and applications. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15, pp. 241–311). Lincoln, NE: University of Nebraska Press.

Pruitt, D. G. (1968). Reciprocity and credit building in a laboratory dyad. *Journal of Personality and Social Psychology, 8,* 143–147.

Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology, 28,* 363–392.

Rapoport, A., & Chammah, A. M. (1965). *Prisoner's Dilemma: A study in conflict and cooperation.* Ann Arbor: University of Michigan Press.

Rushton, J. P. (1980). *Altruism, socialization, and society.* Englewood Cliffs, NJ: Prentice Hall.

Schelling, T. C. (1971). Dynamic models of segregation. *Journal of Mathematical Sociology, 1,* 143–186.

Taylor, S. E. (1990). *Positive illusions.* Basic Books: New York.

Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups.* New York: Wiley.

Tversky, A., & Kahneman, D. (1986). Rational choice and the framing of decisions. *Journal of Business, 59,* S251–S278.

Wilson, E. O. (1975). *Sociobiology: The new synthesis.* Cambridge, MA: Harvard University Press.

Wolfram, S. (1986). *Theory and applications of cellular automata.* Singapore: World Scientific.