

1

RECIPROCAL ALTRUISM

It was the spring semester of 1969 and the population geneticist Richard Lewontin had come to Harvard to deliver a talk on a new methodology that promised to revolutionize the field of population genetics. Lewontin was then at nearly the height of his powers, and 250 people, myself included, were crammed into a lecture hall to hear his eagerly anticipated talk. It was my devout wish that he would fall flat on his face.

The reason for this is that I had just been introduced to him by E. O. Wilson, Harvard's fabulous social insect man, at the tea preceding his talk. Lewontin had at once proceeded to dump on me, a mere first-year graduate student. The previous fall I had written a very negative paper attacking the work of two theoretical ecologists, Robert MacArthur and Richard Levins, and Lewontin was a personal friend of Levins at the University of Chicago. A fellow graduate student had carried the manuscript to a conference and allowed someone to make a photocopy, and like a true pathogen it had spread rapidly around the globe. I even received a glowing letter from a well-known Australian ecologist. When I realized the manuscript had gone public I sent copies to MacArthur and Levins. MacArthur wrote a very nice letter in return; Levins did not deign to reply. Lewontin began by saying, "Oh yes, you're the fellow that wrote that wrong-headed paper on MacArthur and Levins," and dismissed it in a few sentences. He pointed to some equations on the board that he had apparently shown to Ed Wilson and told me that one of my criticisms was easily handled if you used these equations. As Wilson squinted myopically toward the board (and with only one good eye at that!) I thought I could see at a glance that Lewontin had merely complexified the problem—thereby hiding the error more deeply—but that the same problem remained in his formulation as in the work of Levins himself. I have learned in my life that my memory of what I regard as odious

behavior is always more negative than the actual behavior itself, but in any case Lewontin had a somewhat arrogant and condescending style, and his treatment of me, I later learned, was by no means out of character. I took an immediate dislike to him.

Ernst Mayr, Harvard's venerated evolutionary biologist, introduced Lewontin at the talk, and I remember Lewontin thanked "Ernst" for his introduction, which drew an almost audible gasp from the audience, who had never before heard Professor Mayr referred to in public by only his first name. Lewontin then went on to deliver a masterful talk. He began by describing a problem that he said lay at the heart of evolutionary biology, namely, the degree of genetic variation present in natural populations. There were two competing schools of thought: one, that there was a lot of variation; the other, that there was very little. Each view had important implications for other matters. He then described a hopeless methodological conundrum, one method biased to an unknown degree in one direction, and the alternative methodology biased to an unknown degree in the opposite. Enter from stage left Lewontin with his new methodology, gel electrophoresis. This method allowed a large number of loci to be sampled relatively easily for genetic variants. He described the method and then gave the first wave of results, which clearly supported one school of thought: there was a lot of genetic variation in natural populations. He then, as I remember it, presented a second wave of data that showed parallel geographical patterns of variation in closely related species, thus suggesting the ongoing action of natural selection (since random forces, acting independently, should not produce *parallel* distributions).

Some people claim that at the end of the lecture he shot his chalk twenty feet into the air and caught it in his suit pocket, but if this occurred I did not see it. The point is that he put on quite a show, both in content and in presentation. I remember halfway through the lecture feeling some intense internal pain. This organism—odious cretin though he was—was not about to fall on his face. This contradiction led me to some soul-searching about my paper on MacArthur and Levins. I realized that the paper was entirely negative; that is, I had nothing positive with which to replace their incorrect views—I had, at best, only a bag full of *their* errors. While this is not without its use in science, its long-term value is strictly limited, for the criticism will soon sink out of sight along with the views it criticizes. Did I have anything positive to say on the subject of theoretical ecology? I knew at once that I did not. To make progress in this area you needed mathematical skills and discipline that I was not about to develop, and you would also benefit greatly from twenty or thirty years of running around in the woods. That is, if you could combine intuition and real knowledge of nature with the kind of mathematics required to handle complex interactions, then perhaps you

could make some headway. I decided, in the space of a few minutes, not to try to publish this attack or to do any more work along those lines.

What positive thoughts did I have? Two immediately came to mind. One was reciprocal altruism—you scratch my back and I'll scratch yours—as an evolutionary problem. And the other was what became parental investment and sexual selection. As I recount in chapter 2, I had been watching pigeons for some time and knew that male pigeons practiced the same type of sexual double standard and had the same kind of attendant psychological problems (or at least behavioral ones) as did human males. This suggested some novel thoughts about the evolution of sex differences. I decided at once to concentrate on both of these problems and to tackle reciprocal altruism first, because the solution to that problem was more straightforward, but to continue my pigeon observations and associated thoughts.

I left Lewontin's talk, as I remember it, in high, good spirits, a burden lifted from my shoulders and my eyes firmly set on the prize ahead. For one thing, it was such a relief to kiss mathematical ecology good-bye! It was a small burden then but one that could easily have grown with time. I wonder how many academics happen to get into areas they neither really enjoy nor are particularly suited for, yet stay in those areas for many years. At the same time, by turning to what I really cared about and throwing the same kind of energy and effort into it as I had thrown into attacking MacArthur and Levins, I would really have something to show for my time.

It later often seemed ironic to me that Lewontin should have helped put me so firmly on a path that he came to detest so much himself. He was to return to Harvard as a full professor when I was an assistant professor, and he was to become chairman of Organismic and Evolutionary Biology when I came up for tenure. During that bitter dispute I heard by the grapevine that he had disparaged me to one group of students as an "intellectual opportunist." Of course I was an intellectual opportunist! What else made sense in this short life? The inability of biologists to think clearly on matter of social behavior and evolution for over a hundred years had left a series of important problems untackled. And it was a wonderful opportunity, especially if you had some social and psychological insight (my rough analog to twenty years in the woods). The area was so underdeveloped you hardly needed mathematics. Logic plus fractions would get you through most situations!

I had started to focus on reciprocal altruism largely because of the wonderful work of W. D. Hamilton on kinship and social behavior. Hamilton had shown that you could define a variable—degree of relatedness (or coefficient of relatedness) to others—that would have a strong effect on selection acting on social behavior between related individuals. In particular, an individual could sacrifice personal reproductive success (or fitness, as it

was usually called then) and still be favored by natural selection as long as this benefited relatives such that, when the benefit was devalued by the degree of relatedness, the resulting number was still larger than the cost suffered. I take up kinship theory in much greater depth in chapters 4 and 5.

Hamilton did something else in his famous 1964 paper that was deceptively simple—he defined the four major categories of social interaction in terms of their effects on the reproductive success of the two individuals involved. Thus, “altruistic” behavior was behavior that caused a loss to the actor and a benefit to the recipient where these were defined in terms of effects on their reproductive success. Selfish behavior was the reverse, while in cooperative behavior both parties benefited and in spiteful behavior neither party benefited: each suffered a cost. This fourfold classification of behavior, or social traits, more broadly, had the benefit of immediately stating how natural selection was acting on the interaction from the standpoint of each of the two individuals.

This was a use of language that social scientists and others would come to detest. When they were aroused in full antipathy toward “sociobiology,” you often read that this was a perversion of language. True altruism had to refer to pure internal motivations or other-directed internal motivations without thought or concern for self. To an evolutionist, this seemed absurd. You begin with the *effect* of behavior on actors and recipients; you deal with the problem of internal motivation, which is a secondary problem, afterward. If you made this point to some of these naysayers, they would often argue back that, if this were the case, then Hamilton should have chosen words that had no connotations in everyday language. This was also a very short-sighted view. In the extreme case, I suppose, Hamilton could have called the behaviors x, y, w, and z, so as to avoid any but alphabetical connotations. But this, of course, meant that you would always have to be translating those symbols into some verbal system that made sense to you before you could think clearly. Incidentally, when the great sociobiology controversy did roll forth, I soon came to see that the real function of these counterarguments was to slow down your work and, if possible, stop it cold. If you start with motivation, you have given up the evolutionary analysis at the outset. If you are forced to use arbitrary symbols, progress will be slowed for no good reason. Even the invitation to argue with them seemed to me to benefit them by wasting your time!

In any case, once you have conceptualized certain behaviors or traits as “altruistic” yet can show that they exist in nature, then you immediately have a problem, and a severe one, because in classic Darwinian theory, natural selection removes genes that tend to induce altruistic acts. The phrase “a gene for altruism” was also later subject to bitter attack. How could a single gene cause altruism? And again, the primary function of this objec-

tion was to make it very difficult for people to talk and think about the analysis in question. Lewontin, incidentally, who was one of those who raised this and related objections, I always felt fought sociobiology tooth and nail, because he had committed himself already to an alternative, marxist or pseudo-marxist system of social interpretation, and he would rather keep his evolutionary theory for population genetics and the kinds of problems he tackled in his office, while retaining marxist thinking for handling social interactions. In any case, by simply calling certain behaviors and acts altruistic and defining them the way he did, Hamilton brought into clear focus the problem involved. In his own kinship theory he solved many such cases in nature while at the same time producing much deeper implications as well. Again, I will leave this subject for later chapters.

When I came into biology at age twenty-two, never having had a course in biology and knowing next to nothing about animal behavior, my knowledge was almost entirely restricted to our own species. In adult humans it was obvious that, though kinship was a very important factor—blood being thicker than water—it could not explain all phenomena. We had strong positive feelings toward friends, and we were willing to act altruistically toward them and others. Kinship could not explain this. What could?

Well, reciprocity, in some form, could obviously do the trick—that is, you scratch my back and I'll scratch yours—but reciprocity required some thinking to get the argument right. When we are each scratching each other's backs we are simultaneously trading benefits and suffering costs. That does not create much of an evolutionary problem. But what about when we act nicely toward an individual and the return benefit, if any, must come later? This raised some interesting evolutionary problems. So, I saw that what in the human species was obviously a major area of life involving deep and complex behaviors was not explained by Hamilton's theory, and required some new explanation. Note that the use of the term "altruism" helped immediately in thinking about reciprocity or reciprocal altruism. Reciprocity, after all, can be negative—reciprocal spite—as Frans deWaal is fond of emphasizing. Anthropologists, I soon learned, were fond of talking about "reciprocity" but by using that term they usually bypassed the theoretical problem at once. They were even able to dream up formulations where parental care was an example of "reciprocity" between the generations, because the offspring would later invest parental care in its own offspring.

So one day I sat down and wrote out a short manuscript on the evolution of reciprocal altruism. I began with the evolutionary problem and tried to formulate the matter mathematically. I remember I had a locus with genes that affected altruism but I realized that I had better have a second locus in the other individual directing the return effort so that my analysis would

not become confused with kinship complications at that same locus. I wrote out a short and, I might add, feeble “mathematical-genetic” section. I had two examples in mind from nature, neither of which, in fact, was a good example of reciprocal altruism, though both were good examples of “return effect altruism.” That is, you act altruistic and a return benefit reaches you but not because another individual chooses to reciprocate your original altruism.

The first example was cleaning symbioses in fish. I used to read very widely in those days, and I remembered seeing an advertisement for a book entitled *The Biology of the Mouth*. I thought to myself, Who wouldn't want to understand the biology of the mouth? and I sent off for my copy. I was thoroughly disappointed in most of the articles, which gave no evolutionary understanding or overview of the mouth but dealt only with problems related to dentistry. One chapter, though, caught my eye, a chapter on cleaning symbioses in the ocean. These are symbioses in which a member of one species, the cleaner species, cleans ectoparasites from the body and sometimes the mouth of a member of the host species. Both parties benefit instantaneously. That is, with each bite, the cleaner gets a meal and the host loses an ectoparasite: evolution had clearly favored warm interactions between the two. But in the article, I read of behavior not so easily explained. In some cases, when a cleaner is cleaning the inside of a host fish's mouth, the host will spot a predator, and instead of (as I imagine I would have done) simply swallowing the cleaner and getting the hell out of there, the host closed its mouth and then opened it—as a warning to the cleaner to depart—and then took off running itself. Ahhh, here was a delay requiring explanation. Was there a return benefit? As you will see if you read the paper that follows, I did indeed gather evidence that the hosts often return to the same cleaners and that they probably benefit from doing so, and therefore, a concern for the welfare of the life of the cleaner, even at some cost to your own life, may pay its way. The individuals involved were members of different species, so kinship could be ruled out.

The second problem from nature was warning cries in birds. In many, many species of birds, individuals who spot a predator give an alarm call, which warns other individuals of a predator at some presumed cost, however small, to the individual giving the warning call. Indeed, Peter Marler (1955) had written a beautiful paper showing that the sound characteristics of warning calls were very different than those of territorial calls in a couple of species of birds: the latter had a wide frequency variation while the warning calls tended to be pure tone. Wide frequency variation permits listeners to locate, easily and quickly, the direction from which the sound is coming, while pure tones are almost *ventriloquial*. So, it was nice for the bird that its warning call did not reveal much about its location while its territorial assertion call did. But, from an evolutionary stand-

point, those pure tones had to evolve. They had to begin as calls with frequency variation. So that beautiful pure-tone picture that Marler published connoted to me so many dead birds to get to that pure pitch. That was indirect evidence of cost.

Now, there was in principle no reason not to explain all of bird calls as kin directed. On the other hand, nothing was known about the kin structure of any bird species, so there was no direct evidence in favor of this interpretation. How might one explain bird calls in my system? The obvious would be reciprocal altruism itself: I warn you when I spot the predator, you warn me when you spot the predator, and we both show a net gain over time. I saw a problem with this: it seemed a hopeless system to police. How are you going to handle cheaters? To put it in human terms, you can imagine a bird spotting a predator late, almost being eaten itself, and then going to its neighbor Fred and saying, "Why the hell didn't you warn me?" Fred throws up his wings and says, "I was as surprised as you, brother!" There was no way to identify cheaters, no way to punish them, except not calling yourself when the shoe was on the other foot. All I could see selection doing was silencing the birds throughout their range.

But there were plenty of opportunities for return-effect altruism to explain warning calls. I knew from my teacher William Drury that predators had been shown in some cases to form specific search images, in which experience killing one member of a species greatly increased their chance of spotting and killing other members of the same species. Imagine a highly cryptic (i.e., camouflaged) moth resting against the bark of a tree. You overlook many, but when you spot the first one as edible food and consume it, you become aware of their existence, rapidly learn characteristics to discriminate them from bark, and start cleaning up on moths resting on bark. In short, my neighbor not getting eaten might decrease the chance of a predator learning useful things when it turned to attack me. But some small familiarity with predator-prey interactions in animals immediately suggested more direct possibilities. When predators are sneaking up on prey and are spotted, and a warning cry is subsequently issued and everyone dashes to safety, often the predators then move to some other area where they have not yet been spotted. So, warning your neighbor that a predator is nearby may be the quickest way to get the predator to move on elsewhere. I tried to think through as many ways as I could of natural selection favoring alarm calls in birds through return effects. Since there was no direct evidence on any of these possibilities, my thinking at least had the virtue of generating alternative explanations to the kinship logic.

I then wrote, if I remember correctly, a very short few pages on human reciprocal altruism with no evidence cited or research results discussed.

I Meet W.D.

A little bit after this, perhaps in the fall of 1969, William Hamilton himself came to Harvard to lecture. He was coming from a "Man and Beast" symposium at the Smithsonian in Washington where he had presented some of his latest thinking, which I believe was the same talk he gave us. There were perhaps eighty or ninety people, almost filling a lecture hall, most of us with eager anticipation. Hamilton got up and gave one of the worst lectures, as a lecture, that I had ever heard. There was an emeritus Harvard professor who occasionally used to give a lecture, widely appreciated, on how to give a poor lecture. W. D. did not need any teaching in this regard and has generated some wonderful tricks of his own. I say this as a man who loved W. D., but his early troubles in this regard were sometimes very funny. For one thing, he lectured for a full fifty-five minutes without yet getting to the point. It was abstruse and technical; he often had his back to us while he was writing things on the board; you had difficulty hearing his voice; you did not get any overview of where he was going or why he was going there. When he realized that he was five minutes over time and still had not gotten to the point, or indeed very near it, he looked down at Ed Wilson, his host, and asked him

After his Harvard talk I got the chance to meet him in person. Mary Jane West-Eberhard, the celebrated student of wasps, was I believe a postdoc with Wilson at the time, and she held a small party for Hamilton. I brought along the little paper that I had written on reciprocal altruism to give him to read, and I think he expressed interest in meeting me (he had heard I had been thinking along these lines, or something like that). He was a shy man, very soft spoken. You often had to lean into him to catch what he was saying. He had a kind of horsey-looking face, as he would describe it to others (e.g., when he was meeting someone new at the train station). I remember thinking, at some point in my relationship with Bill, that if the argument ever became physical, the contest I would least like to be engaged in against Hamilton was a shoving contest. I felt that he would dig in his heels, that you would be unable to move him, and that he would lean forward and shove you slowly and stubbornly to wherever he wanted to get you.

Some time thereafter I received a letter from W. D. and I remembered liking the gentle tone with which he dealt with my efforts. He encouraged me to continue. He could see why I had chosen separate loci for actor and recipient, but he suggested (without telling me that my mathematical attempts were littered with errors, as they were) that it would perhaps be better to delete the “maths,” as he called it, from the paper. I only half followed his advice. I deleted the separate loci but tried to expand the “maths” and of course introduced new errors in the process. For example, you will see that I make assertions about the evolution of reciprocal altruism genes that in fact apply only when the gene for altruism is exactly 50% in frequency. This is not a very helpful formulation from an evolutionary standpoint. The evolutionary problem is to take a gene that begins at low frequency and get it to a high frequency—what is happening at the 50/50 mark is usually an irrelevancy: But I was very much trying to mimic, however feebly, Hamilton’s kind of thinking itself: try, if at all possible, to get the formulation down to the level of genes, to be more sure that you have got the argument right, and to give a quantitative form to it if possible. Hamilton also pointed out that my two nonhuman examples were not actually examples of reciprocal altruism and suggested that I rename the paper “the evolution of return-effect altruism.” This I was not about to do, though later I wished I had taken the opportunity to stress more clearly the differences between the two categories.

Human Reciprocal Altruism

The key to the paper, besides its evolutionary approach, was the section on human reciprocal altruism. Could I reorganize facts about human psychology around this new argument in a coherent and interesting way? If so, there would be immediate pay-off for the argumentation itself. This required that I become familiar with a literature from social psychology that I knew noth-

ing about. Naturally I sought shortcuts. I saw that in the Harvard catalog a man was teaching a course on moral behavior, and from the description in the catalog, this sounded just about right. So I said to myself that I would humble myself and take the entire course (as an auditor, of course) in order to learn this material, and at the end of the semester I should have the human section in hand, or at least be ready to write it. I attended, I believe, only two lectures. The problem with the course was that the teacher seemed to think it was immoral to teach a course on morality without first canvassing all the students to find out what they wanted to know about the subject. This was not what I wanted. I wanted someone to lay out the subject matter for me so that I could reinterpret it as I wished. However, I noted that his graduate teaching assistant, Dennis Krebs (now a Professor of Psychology at Simon Fraser University), was cited on the reading list as having written a review, soon to be published, of the literature on altruistic behavior. Ahh, this was perfect! As a graduate student of the professor, you could take it for granted that his review was going to be very thorough and very likely better than the professor himself would have done. I could skip the lectures entirely and simply read Krebs's paper, assuming he was willing to provide a copy. I went to Mr. Krebs and he very kindly provided me with a copy. I never saw the class or its professor again.

On the other hand, I did not immediately digest Krebs's paper, either. It was written in a different language, and I was required to master that language before I could understand exactly what he was saying. I read around in the paper to try to familiarize myself with the kinds of evidence that were available. One thing that amused me, right off the bat, was that social psychologists called altruistic behavior "pro-social behavior." Now, that seemed immediately somewhat ill-defined. Social is easily opposed to solitary, so social interaction is an interaction involving more than one organism. I can be fighting you and it is a social interaction. Am I pro-social if I am in favor of fights? An antisocial individual might be a hermit, but antisocial was being used to refer to someone whose behavior was bad for others, as the author saw it, and ought best to be curbed. Again, defining categories of behavior in this way seemed like a hopeless way to proceed if you were interested in building up a solid scientific approach to human behavior. There were, of course, other linguistic hurdles I had to overcome. Krebs had organized his paper along conventional distinctions within that discipline. For example, there were immediate precipitating variables where you were more or less to give a dollar to a beggar if snot was coming out of his nose. There were personality variables in the potential pro-socialists. There were situational variables that might affect the tendency. And, of course, all this research was guided by the goal of increasing pro-social behavior—not really understanding it, and certainly not at any deep level, but increasing its occurrence in society. What was missing was exactly what was missing from the disci-

pline itself: any functional understanding of the behaviors that they were discussing. Why did it make sense for the organism to do it? This was, of course, what evolutionary biology, and myself in particular, was set to provide. So, all I had to do was master the literature cited by Krebs and then reorganize it appropriately.

This sounded easier than it turned out to be, and the more I comprehended Krebs's paper, the more I dreaded having to study the works that he was citing. They would be written in a bizarre language and I was afraid to learn too much about the actual methodologies used. I read a few of the papers and then decided to take another shortcut and write the section based on Krebs's review. I piously told myself that I would, of course, read all the papers whose citations I would be lifting from Krebs's review, as by good scientific and academic procedure I should, but in fact I never did.

I was visiting my parents' home in the Midwest for two or three weeks and decided to take this opportunity to write the human section. I was depressed the first week and lay around doing little more than sleeping. Then I roused myself and said, You have two boring weeks ahead of you, however you slice it, so you might as well do this work, which has to be done anyway. I spread out the few papers I had actually copied, reread Krebs's account, and simply reorganized the information around obvious psychological categories, such as sympathy, gratitude, and moralistic aggression.

How to Write a Classic Paper

In retrospect, I think my paper on reciprocal altruism can be used to illustrate how one might go about writing a classic paper. Here is my recipe:

1. Pick an important topic.
2. Try to do a little sustained thinking on the topic, always keeping close to the task at hand.
3. Generalize outward from your chosen topic.
4. Write in the language of your discipline but, of course, try to do so simply and clearly.
5. If at all possible, reorganize existing evidence around your theory.

Pick an important topic. This is perhaps easier said than done, because you must pick an important topic on which progress can be made. But it still seems remarkable to me how often people bypass what are more important subjects to work on less important ones. Constructing a scientific understanding of human psychology and social behavior is an important task, and within that subject reciprocal altruism was, to me, an important topic when I began work on it. The very fact that so many strong emotions could be associated, in humans, with friendship, with helping others, with guarding

against being cheated, said to me that it was an important topic in human psychology. There was no reason, in advance, to believe that it would not be an important topic in at least some other species, certainly closely related monkeys and apes, but very likely other species as well.

Do some sustained thinking. To me this is easier than it sounds. There are no great intellectual gymnastics in my paper. I am not proving Fermat's Last Theorem or generating Goedel's Proof. I am only trying to think simply and clearly on an interesting and important subject. I was amazed when I went into academic work—and it still baffles me today—why so many people take the first available path off their main argument into trivia land. The sustained thinking must always be directed back to the key subject itself. I suppose it is easier, at first, to write a section on the semantics of discussing altruism, or sometimes to review previous failed efforts in considerable detail if these failed efforts are in front of you and easy to interpret. But this is all a waste of energy and effort. It diverts you from your main task. Get to the point and stick to the point. When I sent the paper to M. L. Roonwal, the great Indian student of termites and locusts (whom I had met on a very memorable ten-day monkey-viewing visit with Irv DeVore), he praised the paper for its "intellectual architecture." And I thought the phrase was very apt. I had constructed the paper much as you might build a house, with roughly the same kinds of mental operations required. Nothing brilliant or flashy, just the steady construction of a series of arguments and facts regarding reciprocal altruism.

Generalize outward. In most of my papers I took the road opposite the one biologists usually take, or in any case are usually accused of, which is arguing from a knowledge of animals to suppositions about humans. I usually begin with humans and then try to generalize outward to include as many other species and phenomena as I can. In this case, I took pains to formulate the argument with as few assumptions as possible. These only appear later as limitations. For example, both Charles Darwin and George C. Williams had a few words to say on human reciprocal altruism, but each presupposed that its appearance would require the kind of intellectual talents that we know exist in our species: to recognize individuals, to remember past interactions, to alter behavior appropriately, and so on. This limits the argument to human beings in advance, and in a way that is completely unnecessary. It seemed to me obviously preferable to avoid any such limitation unless absolutely forced to accept it. The mental processes they cite could easily evolve *after* the fact. If you could just get a little bit of reciprocal altruism going, selection pressures to spot cheaters, to reward especially good fellow altruists, and so on ought to evolve easily.

Write in the language of your discipline, but simply and clearly. I was trying to make an evolutionary argument within the field of evolutionary biology. You need to satisfy the criteria of your discipline to be accepted or noticed.

Of course, as I have admitted, I dressed up a pseudo-mathematical genetics section to look as if I were using the concepts and language of my discipline more than I was, but I am convinced that had the later material, on humans for example, been written without any citations or references, it could easily have been neglected. So, even though the literature that was cited was a good bit weaker than I would have liked, I think it was important for the seriousness with which people took the paper. Alas, it had very little effect, quite opposite to my supposition, on social psychology itself. I naively imagined that social psychologists would be delighted to see that their work could be given much more interest and meaning when put in an evolutionary framework, and that they would immediately see new ways of doing their work that would prove more fruitful than the paths that they were taking. For example, let us assume you show, as they did, that you are more likely to act nicely toward a person who resembles you more. As an evolutionist you immediately wonder whether this is because of kin selection, in which case the organism is unconsciously measuring degree of relatedness, or whether it might be due to a reciprocal mechanism in which the likelihood of exchanging benefits was more likely with closer resemblance. Yet this immediately raises another question since reciprocal altruism between individuals who are complementary in characteristics may give greater value in some contexts than when they are similar. Pursuing this line of thought would suggest experiments or observations that discriminated these possibilities as well. Nothing like this has, in fact, happened. I know of no social psychologists who have altered their work because of my paper (I would be very happy to hear of any) and have been very disappointed to discover that almost the only new work done is by people who first start with an evolutionary interest and then turn to trying to do something social psychological.

The virtues of writing simply and clearly should be obvious, certainly from the standpoint of the recipient, but attempting to do so has virtues for the writer as well, since it repeatedly forces you to think through your subject clearly. I was very fortunate at Harvard to have several professors who were excellent critical readers. The most useful from my standpoint was Irv DeVore, the famous Harvard "baboon man" and anthropologist. Irv is a superb stylist and through five or six rewrites of the paper he would continually improve the presentation. Similarly, my advisor, Ernest Williams, would continually force me to tighten the argument and clarify the presentation. Incidentally, when I brought Professor Williams the sixth draft of the paper I remember saying to him, "Is it finished Dr. Williams? Is it finished?" There was a small pause and then Ernest looked at me and said, "A paper is *never* finished, Bob, it is only abandoned." He then told me that he thought that this paper was ready to be abandoned.

All those drafts that the paper was put through earned a rich reward when it was actually published. There was an immediate, large, and very

welcome response. I began to receive reprint requests in the mail, mostly from the United States but in fact from all over the world. I had ordered six hundred reprints (this was in the days before wholesale photocopying of papers) and soon enough the pile was exhausted. It gave me immense pleasure, especially receiving the foreign reprint requests, to know that my thinking would be studied in countries around the world. I was especially gratified at the number of reprint requests that came from Soviet bloc or communist countries because, of course, there was otherwise so little exchange across the so-called Iron Curtain. After a while, it struck me that their socialist ideology and emphasis on the possibility of naturally cooperative behavior among humans would make this an interesting topic for them and, perhaps, even one that was socially acceptable to pursue. I also learned that in their literature, Peter Kropotkin was an early pioneer whom they would have expected me to cite.

Try to reorganize or reinterpret existing information. Many, many theoretical papers in biology and elsewhere fail to show, at the end of the paper, that there is anything out there in the real world to which the argumentation just given actually applies. It applies in principle, of course, but they are unable to reinterpret any existing information. A typical effort might discuss an important topic, provide some sustained mathematical thinking on some part of the problem, usually misrepresented as being central to the problem, and then derive results that not only cannot immediately be tested against reality, but also would require an awful lot of work to do so. The value of the work, of course, depends upon how restrictive the initial assumptions are. If they are highly restrictive, then even if your mathematics are correct, you have solved a highly specialized problem and it may be of very marginal interest to test whether your statements are true, especially as this may require very detailed and difficult measurements. Of course, if you have modeled something central to a problem and your results are not easy to test, that is another matter.

In my case, I think the last section gave life to the paper in a very important way because it showed, in principle, that the simple argument given could reinterpret information gathered in social psychology and give us a deeper understanding of similar kinds of facts known from our everyday experience.

A Sense of Justice

A very agreeable feature of my reciprocal altruism argument, which I had not anticipated in advance, was that a sense of justice or fairness seemed a natural consequence of selection for reciprocal altruism. That is, you could easily imagine that a sense of fairness would evolve as a way of regulating

reciprocal tendencies, a way of judging the degree to which other people were cheating you (and you them!). There seemed no plausible way in

were to prove prophetic. In ten short years, Hamilton himself, along with Robert Axelrod, a political scientist at the University of Michigan, were to do precisely what Ed had urged me to accomplish. Far from complexity, they isolated a simple rule of action, tit-for-tat, which was evolutionarily stable.



The Evolution of Reciprocal Altruism

ROBERT L. TRIVERS

Abstract. A model is presented to account for the natural selection of what is termed reciprocally altruistic behavior. The model shows how selection can operate against the cheater (non-reciprocator) in the system. Three instances of altruistic behavior are discussed, the evolution of which the model can explain: (1) behavior involved in cleaning symbioses; (2) warning cries in birds; and (3) human reciprocal altruism.

Regarding human reciprocal altruism, it is shown that the details of the psychological system that regulates this altruism can be explained by the model. Specifically, friendship, dislike, moralistic aggression, gratitude, sympathy, trust, suspicion, trustworthiness, aspects of guilt, and some forms of dishonesty and hypocrisy can be explained as important adaptations to regulate the altruistic system. Each individual human is seen as possessing altruistic and cheating tendencies, the expression of which is sensitive to developmental variables that were selected to set the tendencies at a balance appropriate to the local social and ecological environment.

Altruistic behavior can be defined as behavior that benefits another organism, not closely related, while being apparently detrimental to the organism performing the behavior, benefit and detriment being defined in terms of contribution to inclusive fitness. One human being leaping into water, at some danger to himself, to save another distantly related human from drowning may be said to display altruistic behavior. If he were to leap in to save his own child, the behavior would not necessarily be an instance of "altruism"; he may merely be contributing to the survival of his own genes invested in the child.

Models that attempt to explain altruistic behavior in terms of natural selection are models designed to take the altruism out of altruism. For example, Hamilton (1964) has demonstrated that degree of relationship is an important parameter in predicting how selection will operate, and behavior which appears altruistic may, on knowledge of the genetic relationships of

the organisms involved, be explicable in terms of natural selection: those genes being selected for that contribute to their own perpetuation, regardless of which individual the genes appear in. The term “kin selection” will be used in this paper to cover instances of this type—that is, of organisms being selected to help their relatively close kin.

The model presented here is designed to show how certain classes of behavior conveniently denoted as “altruistic” (or “reciprocally altruistic”) can be selected for even when the recipient is so distantly related to the organism performing the altruistic act that kin selection can be ruled out. The model will apply, for example, to altruistic behavior between members of different species. It will be argued that under certain conditions natural selection favors these altruistic behaviors because in the long run they benefit the organism performing them.

The Model

One human being saving another, who is not closely related and is about to drown, is an instance of altruism. Assume that the chance of the drowning man dying is one-half if no one leaps in to save him, but that the chance that his potential rescuer will drown if he leaps in to save him is much smaller, say, one in twenty. Assume that the drowning man always drowns when his rescuer does and that he is always saved when the rescuer survives the rescue attempt. Also assume that the energy costs involved in rescuing are trivial compared to the survival probabilities. Were this an isolated event, it is clear that the rescuer should not bother to save the drowning man. But if the drowning man reciprocates at some future time, and if the survival chances are then exactly reversed, it will have been to the benefit of each participant to have risked his life for the other. Each participant will have traded a one-half chance of dying for about a one-tenth chance. If we assume that the entire population is sooner or later exposed to the same risk of drowning, the two individuals who risk their lives to save each other will be selected over those who face drowning on their own. Note that the benefits of reciprocity depend on the unequal cost/benefit ratio of the altruistic act, that is, the benefit of the altruistic act to the recipient is greater than the cost of the act to the performer, cost and benefit being defined here as the increase or decrease in chances of the relevant alleles propagating themselves in the population. Note also that, as defined, the benefits and costs depend on the age of the altruist and recipient (see *Age-dependent changes* below). (The odds assigned above may not be unrealistic if the drowning man is drowning because of a cramp or if the rescue can be executed by extending a branch from shore.)

Why should the rescued individual bother to reciprocate? Selection

would seem to favor being saved from drowning without endangering oneself by reciprocating. Why not cheat? ("Cheating" is used throughout this paper solely for convenience to denote failure to reciprocate; no conscious intent or moral connotation is implied.) Selection will discriminate against the cheater if cheating has later adverse effects on his life which outweigh the benefit of not reciprocating. This may happen if the altruist responds to the cheating by curtailing all future possible altruistic gestures to this individual. Assuming that the benefits of these lost altruistic acts outweigh the costs involved in reciprocating, the cheater will be selected against relative to individuals who, because neither cheats, exchange many altruistic acts.

This argument can be made precise. Assume there are both altruists and nonaltruists in a population of size N and that the altruists are characterized by the fact that each performs altruistic acts when the cost to the altruist is well below the benefit to the recipient, where cost is defined as the degree to which the behavior retards the reproduction of the genes of the altruist and benefit is the degree to which the behavior increases the rate of reproduction of the genes of the recipient. Assume that the altruistic behavior of an altruist is controlled by an allele (dominant or recessive), a_2 , at a given locus and that (for simplicity) there is only one alternative allele, a_1 , at that locus and that it does not lead to altruistic behavior. Consider three possibilities: (1) the altruists dispense their altruism randomly throughout the population; (2) they dispense it nonrandomly by regarding their degree of genetic relationship with possible recipients; or (3) they dispense it nonrandomly by regarding the altruistic tendencies of possible recipients.

(1) *Random dispensation of altruism*

There are three possible genotypes: a_1a_1 , a_2a_1 , and a_2a_2 . Each allele of the heterozygote will be affected equally by whatever costs and benefits are associated with the altruism of such individuals (if a_2 is dominant) and by whatever benefits accrue to such individuals from the altruism of others, so they can be disregarded. If altruistic acts are being dispensed randomly throughout a large population, then the typical a_1a_1 individual benefits by $(1/N)\sum b_i$, where b_i is the benefit of the i th altruistic act performed by the altruist. The typical a_2a_2 individual has a net benefit of $(1/N)\sum b_i - (1/N)\sum c_j$, where c_j is the cost to the a_2a_2 altruist of his j th altruistic act. Since $-(1/N)\sum c_j$ is always less than zero, allele a_1 will everywhere replace allele a_2 .

(2) *Nonrandom dispensation by reference to kin*

This case has been treated in detail by Hamilton (1964), who concluded that if the tendency to dispense altruism to close kin is great enough, as a

function of the disparity between the average cost and benefit of an altruistic act, then a_2 will replace a_1 . Technically, all that is needed for Hamilton's form of selection to operate is that an individual with an "altruistic allele" be able to distinguish between individuals with and without this allele and discriminate accordingly. No formal analysis has been attempted of the possibilities for selection favoring individuals who increase their chances of receiving altruistic acts by appearing as if they were close kin of altruists, although selection has clearly sometimes favored such parasitism (e.g., Drury and Smith, 1968).

(3) *Nonrandom dispensation by reference to the altruistic tendencies of the recipient*

What is required is that the net benefit accruing to a typical a_2a_2 altruist exceed that accruing to an a_1a_1 nonaltruist, or that

$$(1/p^2) (\sum b_k - \sum c_j) > (1/q^2) \sum b_m,$$

where b_k is the benefit to the a_2a_2 altruist of the k th altruistic act performed toward him, where c_j is the cost of the j th altruistic act by the a_2a_2 altruist, where b_m is the benefit of the m th altruistic act to the a_1a_1 nonaltruist, and where p is the frequency in the population of the a_2 allele and q that of the a_1 allele. This will tend to occur if $\sum b_m$ is kept small (which will simultaneously reduce $\sum c_j$). And this in turn will tend to occur if an altruist responds to a "nonaltruistic act" (that is, a failure to act altruistically toward the altruist in a situation in which so doing would cost the actor less than it would benefit the recipient) by curtailing future altruistic acts to the nonaltruist.

Note that the above form of altruism does not depend on all altruistic acts being controlled by the same allele at the same locus. Each altruist could be motivated by a different allele at a different locus. All altruistic alleles would tend to be favored as long as, for each allele, the net average benefit to the homozygous altruist exceeded the average benefit to the homozygous nonaltruist; this would tend to be true if altruists restrict their altruism to fellow altruists, regardless of what allele motivates the other individual's altruism. The argument will therefore apply, unlike Hamilton's (1964), to altruistic acts exchanged between members of different species. It is the *exchange* that favors such altruism, not the fact that the allele in question sometimes or often directly benefits its duplicate in another organism.

If an "altruistic situation" is defined as any in which one individual can dispense a benefit to a second greater than the cost of the act to himself, then the chances of selecting for altruistic behavior, that is, of keeping $\sum c_j + \sum b_m$ small, are greatest (1) when there are many such altruistic situations in the lifetime of the altruists, (2) when a given altruist repeatedly interacts with the same small set of individuals, and (3) when pairs of altruists are

exposed “symmetrically” to altruistic situations, that is, in such a way that the two are able to render roughly equivalent benefits to each other at roughly equivalent costs. These three conditions can be elaborated into a set of relevant biological parameters affecting the possibility that reciprocally altruistic behavior will be selected for.

(1) *Length of lifetime.* Long lifetime of individuals of a species maximizes the chance that any two individuals will encounter many altruistic situations, and all other things being equal one should search for instances of reciprocal altruism in long-lived species.

(2) *Dispersal rate.* Low dispersal rate during all or a significant portion of the lifetime of individuals of a species increases the chance that an individual will interact repeatedly with the same set of neighbors, and other things being equal one should search for instances of reciprocal altruism in such species. Mayr (1963) has discussed some of the factors that may affect dispersal rates.

(3) *Degree of mutual dependence.* Interdependence of members of a species (to avoid predators, for example) will tend to keep individuals near each other and thus increase the chance they will encounter altruistic situations together. If the benefit of the mutual dependence is greatest when only a small number of individuals are together, this will greatly increase the chance that an individual will repeatedly interact with the same small set of individuals. Individuals in primate troops, for example, are mutually dependent for protection from predation, yet the optimal troop size for foraging is often small (Crook, 1969). Because they also meet the other conditions outlined here, primates are almost ideal species in which to search for reciprocal altruism. Cleaning symbioses provide an instance of mutual dependence between members of different species, and this mutual dependence appears to have set the stage for the evolution of several altruistic behaviors discussed below.

(4) *Parental care.* A special instance of mutual dependence is that found between parents and offspring in species that show parental care. The relationship is usually so asymmetrical that few or no situations arise in which an offspring is capable of performing an altruistic act for the parents or even for another offspring, but this is not entirely true for some species (such as primates) in which the period of parental care is unusually long. Parental care, of course, is to be explained by Hamilton’s (1964) model, but there is no reason why selection for reciprocal altruism cannot operate between close kin, and evidence is presented below that such selection has operated in humans.

(5) *Dominance hierarchy.* Linear dominance hierarchies consist by definition of asymmetrical relationships; a given individual is dominant over another but not vice versa. Strong dominance hierarchies reduce the extent to which altruistic situations occur in which the less dominant individual is capable of performing a benefit for the more dominant which the

more dominant individual could not simply take at will. Baboons (*Papio cynocephalus*) provide an illustration of this. Hall and DeVore (1965) have described the tendency for meat caught by an individual in the troop to end up by preemption in the hands of the most dominant males. This ability to preempt removes any selective advantage that food-sharing might otherwise have as a reciprocal gesture for the most dominant males, and there is no evidence in this species of any food-sharing tendencies. By contrast, Van Lawick-Goodall (1968) has shown that in the less dominance-oriented chimpanzees more dominant individuals often do not preempt food caught by the less dominant. Instead, they besiege the less dominant individual with "begging gestures," which result in the handing over of small portions of the catch. No strong evidence is available that this is part of a reciprocally altruistic system, but the absence of a strong linear dominance hierarchy has clearly facilitated such a possibility. It is very likely that early hominid groups had a dominance system more similar to that of the modern chimpanzee than to that of the modern baboon (see, for example, Reynolds, 1966).

(6) *Aid in combat.* No matter how dominance-oriented a species is, a dominant individual can usually be aided in aggressive encounters with other individuals by help from a less dominant individual. Hall and DeVore (1965) have described the tendency for baboon alliances to form which fight as a unit in aggressive encounters (and in encounters with predators). Similarly, vervet monkeys in aggressive encounters solicit the aid of other, often less dominant, individuals (Struhsaker, 1967). Aid in combat is then a special case in which relatively symmetrical relations are possible between individuals who differ in dominance.

The above discussion is meant only to suggest the broad conditions that favor the evolution of reciprocal altruism. The most important parameters to specify for individuals of a species are how many altruistic situations occur and how symmetrical they are, and these are the most difficult to specify in advance. Of the three instances of reciprocal altruism discussed in this paper only one, human altruism, would have been predicted from the above broad conditions.

The relationship between two individuals repeatedly exposed to symmetrical reciprocal situations is exactly analogous to what game theorists call the Prisoner's Dilemma (Luce and Raiffa, 1957; Rapoport and Cham-mah, 1965), a game that can be characterized by the payoff matrix

	A ₂	C ₂
A ₁	R, R	S, T
C ₁	T, S	P, P

where $S < P < R < T$ and where A_1 and A_2 represent the altruistic choices possible for the two individuals, and C_1 and C_2 , the cheating choices (the first letter in each box gives the payoff for the first individual, the second letter the payoff for the second individual). The other symbols can be given the following meanings: R stands for the reward each individual gets from an altruistic exchange if neither cheats; T stands for the temptation to cheat; S stands for the sucker's payoff that an altruist gets when cheated; and P is the punishment that both individuals get when neither is altruistic (adapted from Rapoport and Chammah, 1965). Iterated games played between the same two individuals mimic real life in that they permit each player to respond to the behavior of the other. Rapoport and Chammah (1965) and others have conducted such experiments using human players, and some of their results are reviewed below in the discussion of human altruism.

W. D. Hamilton (pers. commun.) has shown that the above treatment of reciprocal altruism can be reformulated concisely in terms of game theory as follows. Assuming two altruists are symmetrically exposed to a series of reciprocal situations with identical costs and identical benefits, then after $2n$ reciprocal situations, each has been "paid" nR . Were one of the two a non-altruist and the second changed to a nonaltruistic policy after first being cheated, then the initial altruist would be paid $S + (n - 1)P$ (assuming he had the first opportunity to be altruistic) and the nonaltruist would receive $T + (n - 1)P$. The important point here is that unless $T \gg R$, then even with small n , nR should exceed $T + (n - 1)P$. If this holds, the nonaltruistic type, when rare, cannot start to spread. But there is also a barrier to the spread of altruism when altruists are rare, for $P > S$ implies $nP > S + (n - 1)P$. As n increases, these two total payoffs tend to equality, so the barrier to the spread of altruism is weak if n is large. The barrier will be overcome if the advantages gained by exchanges between altruists outweigh the initial losses to nonaltruistic types.

Reciprocal altruism can also be viewed as a symbiosis, each partner helping the other while he helps himself. The symbiosis has a time lag, however; one partner helps the other and must then wait a period of time before he is helped in turn. The return benefit may come directly, as in human food-sharing, the partner directly returning the benefit after a time lag. Or the return may come indirectly, as in warning calls in birds (discussed below), where the initial help to other birds (the warning call) sets up a casual chain through the ecological system (the predator fails to learn useful information) which redounds after a time lag to the benefit of the caller. The time lag is the crucial factor, for it means that only under highly specialized circumstances can the altruist be reasonably guaranteed that the casual chain he initiates with his altruistic act will eventually return to him and confer, directly or indirectly, its benefit. Only under these conditions will the cheater be selected against and this type of altruistic behavior evolve.

Although the preconditions for the evolution of reciprocal altruism are specialized, many species probably meet them and display this type of altruism. This paper will limit itself, however, to three instances. The first, behavior involved in cleaning symbioses, is chosen because it permits a clear discrimination between this model and that based on kin selection (Hamilton, 1964). The second, warning calls in birds, has already been elaborately analyzed in terms of kin selection; it is discussed here to show how the model presented above leads to a very different interpretation of these familiar behaviors. Finally, human reciprocal altruism is discussed in detail because it represents the best documented case of reciprocal altruism known, because there has apparently been strong selection for a very complex system regulating altruistic behavior, and because the above model permits the functional interpretation of details of the system that otherwise remain obscure.

Altruistic Behavior in Cleaning Symbioses

The preconditions for the evolution of reciprocal altruism are similar to those for the operation of kin selection: long lifetime, low dispersal rate, and mutual dependence, for example, tend to increase the chance that one is interacting with one's close kin. This makes it difficult to discriminate the two alternative hypotheses. The case of cleaning symbiosis is important to analyze in detail because altruistic behavior is displayed that cannot be explained by kin selection, since it is performed by members of one species for the benefit of members of another. It will be shown instead that the behavior can be explained by the model presented above. No elaborate explanation is needed to understand the evolution of the mutually advantageous cleaning symbiosis itself; it is several additional behaviors displayed by the host fish to its cleaner that require a special explanation because they meet the criteria for altruistic behavior outlined above—that is, they benefit the cleaner while apparently being detrimental to the host.

Feder (1966) and Maynard (1968) have recently reviewed the literature on cleaning symbiosis in the ocean. Briefly, one organism (e.g., the wrasse, *Labroides dimidiatus*) cleans another organism (e.g., the grouper, *Epinephelus striatus*) of ectoparasites (e.g., caligoid cope-pods), sometimes entering into the gill chambers and mouth of the "host" in order to do so. Over forty-five species of fish are known to be cleaners, as well as six species of shrimp. Innumerable species of fish serve as hosts. Stomach analyses of cleaner fish demonstrate that they vary greatly in the extent to which they depend on their cleaning habits for food, some apparently subsisting nearly entirely on a diet of ectoparasites. Likewise, stomach analyses of host fish reveal that cleaners differ in the rate at which they end up in the stomachs of their

hosts, some being apparently almost entirely immune to such a fate. It is a striking fact that there seems to be a strong correlation between degree of dependence on the cleaning way of life and immunity to predation by hosts.

Cleaning habits have apparently evolved independently many times (at least three times in shrimps alone), yet some remarkable convergence has taken place. Cleaners, whether shrimp or fish, are distinctively colored and behave in distinctive ways (for example, the wrasse, *L. dimidiatus*, swims up to its host with a curious dipping and rising motion that reminds one of the way a finch flies). These distinctive features seem to serve the function of attracting fish to be cleaned and of inhibiting any tendency in them to feed on their cleaners. There has apparently been strong selection to avoid eating one's cleaner. This can be illustrated by several observations. Hediger (1968) raised a grouper (*Epinephelus*) from infancy alone in a small tank for six years, by which time the fish was almost four feet in length and accustomed to snapping up anything dropped into its tank. Hediger then dropped a small live cleaner (*L. dimidiatus*) into the grouper's tank. The grouper not only failed to snap up the cleaner but opened its mouth and permitted the cleaner free entry and exit.

Soon we watched our second surprise: the grouper made a movement which in the preceding six years we had never seen him make: he spread the right gill-covering so wide that the individual gill-plates were separated from each other at great distances, wide enough to let the cleaner through. (translated from Hediger, 1968, p. 93)

When Hediger added two additional *L. dimidiatus* to the tank, all three cleaned the grouper with the result that within several days the grouper appeared restless and nervous, searched out places in the tank he had formerly avoided, and shook himself often (as a signal that he did not wish to be cleaned any longer). Apparently three cleaners working over him constantly was too much for him, yet he still failed to eat any of them. When Hediger removed two of the cleaners, the grouper returned to normal. There is no indication the grouper ever possessed any edible ectoparasites, and almost two years later (in December, 1968) the same cleaner continued to "clean" the grouper (pers. observ.) although the cleaner was, in fact, fed separately by its zoo-keepers.

Eibl-Eibesfeldt (1959) has described the morphology and behavior of two species (e.g., *Aspidontus taeniatus*) that mimic cleaners (e.g., *L. dimidiatus*) and that rely on the passive behavior of fish which suppose they are about to be cleaned to dart in and bite off a chunk of their fins. I cite the evolution of these mimics, which resemble their models in appearance and initial swimming behavior, as evidence of strong selection for hosts with no intention of harming their cleaners.

Of special interest is evidence that there has been strong selection not to

eat one's cleaner even after the cleaning is over. Eibl-Eibesfeldt (1955) has made some striking observations on the goby, *Elacatinus oceanops*:

I never saw a grouper snap up a fish after it had cleaned it. On the contrary, it announced its impending departure by two definite signal movements. First it closed its mouth vigorously, although not completely, and immediately opened it wide again. Upon this intention movement, all the gobies left the mouth cavity. Then the grouper shook its body laterally a few times, and all the cleaners returned to their coral. If one frightened a grouper it never neglected these forewarning movements. (translated from Eibl-Eibesfeldt, 1955, p. 208)

Randall has made similar observations on a moray eel (*Gymnothorax japonicus*) that signalled with a "sharp lateral jerk of the eel's head," after which "the wrasse fairly flew out of the mouth, and the awesome jaws snapped shut" (Randall, 1958, 1962). Likewise, Hediger's Kasper Hauser grouper shook its body when it had enough of being cleaned.

Why does a large fish not signal the end to a cleaning episode by swallowing the cleaner? Natural selection would seem to favor the double benefit of a good cleaning followed by a meal of the cleaner. Selection also operates, of course, on the cleaner and presumably favors mechanisms to avoid being eaten. The distinctive behavior and appearance of cleaners have been cited as evidence of such selection. One can also cite the distinctive behavior of the fish being cleaned. Feder (1966) has pointed out that hosts approaching a cleaner react by "stopping or slowing down, allowing themselves to assume awkward positions, seemingly in a hypnotic state." Fishes sometimes alter their color dramatically before and while being cleaned, and Feder (1966) has summarized instances of this. These forms of behavior suggest that natural selection has operated on cleaners to avoid attempting to clean fish without these behaviors, presumably to avoid wasting energy and to minimize the dangers of being eaten. (Alternatively, the behaviors, including color change, may aid the cleaners in finding ectoparasites. This is certainly possible but not, I believe, adequate to explain the phenomenon completely. See, for example, Randall, 1962.)

Once the fish to be cleaned takes the proper stance, however, the cleaner goes to work with no apparent concern for its safety: it makes no effort to avoid the dangerous mouth and may even swim inside, which as we have seen, seems particularly foolhardy, since fish being cleaned may suddenly need to depart. The apparent unconcern of the cleaner suggests that natural selection acting on the fish being cleaned does not, in fact, favor eating one's cleaner. No speculation has been advanced as to why this may be so, although some speculation has appeared about the mechanisms involved. Feder advances two possibilities, that of Eibl-Eibesfeldt (1955) that fish come to be cleaned only after their appetite has been satisfied, and one of his own, that the irritation of ectoparasites may be sufficient to inhibit hun-

ger. Both possibilities are contradicted by Hediger's observation, cited above, and seem unlikely on functional grounds as well.

A fish to be cleaned seems to perform several "altruistic" acts. It desists from eating the cleaner even when it easily could do so and when it must go to special pains (sometimes at danger to itself) to avoid doing so. Furthermore, it may perform two additional behaviors which seem of no direct benefit to itself (and which consume energy and take time); namely, it signals its cleaner that it is about to depart even when the fish is not in its mouth, and it may chase off possible dangers to the cleaner:

While diving with me in the Virgin Islands, Robert Schroeder watched a Spanish hogfish grooming a bar jack in its bronze color state. When a second jack arrived in the pale color phase, the first jack immediately drove it away. But later when another jack intruded on the scene and changed its pale color to dark bronze it was not chased. The bronze color would seem to mean "no harm intended; I need service." (Randall, 1962, p. 44)

The behavior of the host fish is interpreted here to have resulted from natural selection and to be, in fact, beneficial to the host because the cleaner is worth more to it alive than dead. This is because the fish that is cleaned "plans" to return at later dates for more cleanings, and it will be benefited by being able to deal with the same individual. If it eats the cleaner, it may have difficulty finding a second when it needs to be cleaned again. It may lose valuable energy and be exposed to unnecessary predation in the search for a new cleaner. And it may in the end be "turned down" by a new cleaner or serviced very poorly. In short, the host is abundantly repaid for the cost of its altruism.

To support the hypothesis that the host is repaid its initial altruism, several pieces of evidence must be presented: that hosts suffer from ectoparasites; that finding a new cleaner may be difficult or dangerous; that if one does not eat one's cleaner, the same cleaner can be found and used a second time (e.g., that cleaners are site-specific); that cleaners live long enough to be used repeatedly by the same host; and if possible, that individual hosts do, in fact, reuse the same cleaner.

(1) *The cost of ectoparasites.* It seems almost axiomatic that the evolution of cleaners entirely dependent on ectoparasites for food implies the selective disadvantage for the cleaned of being ectoparasite-ridden. What is perhaps surprising is the effect that removing all cleaners from a coral reef has on the local "hosts" (Limbaugh, 1961). As Feder (1966) said in his review:

Within a few days the number of fishes was drastically reduced. Within two weeks almost all except territorial fishes had disappeared, and many of these had developed white fuzzy blotches, swellings, ulcerated sores, and frayed fins. (p. 366)

Clearly, once a fish's primary way of dealing with ectoparasites is by being cleaned, it is quickly vulnerable to the absence of cleaners.

(2) *The difficulty and danger of finding a cleaner.* There are naturally very few data on the difficulty or danger of finding a new cleaner. This is partially because, as shown below, fish tend repeatedly to return to familiar cleaners. The only observation of fish being disappointed in their search for cleaners comes from Eibl-Eibesfeldt (1955): "If the cleaners fail to appear over one coral in about half a minute, the large fishes swim to another coral and wait there a while" (translated from p. 210). It may be that fish have several alternative cleaning stations to go to, since any particular cleaning station may be occupied or unattended at a given moment. So many fish tend to be cleaned at coral reefs (Limbaugh, 1961, observed a cleaner service 300 fish in a 6-hour period), that predators probably frequent coral reefs in search of fish being cleaned. Limbaugh (1961) suggested that good human fishing sites are found near cleaning stations. One final reason why coming to be cleaned may be dangerous is that some fish must leave their element to do so (Randall, 1962):

Most impressive were the visits of moray eels, which do not ordinarily leave their holes in the reef during daylight hours, and of the big jacks which swam up from deeper water to the reef's edge to be "serviced" before going on their way. (p. 43)

(3) *Site specificity of cleaners.* Feder (1966) has reviewed the striking evidence for the site specificity of cleaners and concludes:

Cleaning fishes and cleaning shrimps have regular stations to which fishes wanting to be cleaned can come. (p. 367)

Limbaugh, Pederson, and Chase (1961) have reviewed available data on the six species of cleaner shrimps, and say:

The known cleaner shrimps may conveniently be divided into two groups on the basis of behavior, habitat and color. The five species comprising one group are usually solitary or paired. . . . All five species are territorial and remain for weeks and, in some cases, months or possibly years within a meter or less of the same spot. They are omnivorous to a slight extent but seem to be highly dependent upon their hosts for food. This group is tropical, and the individuals are brightly marked. They display themselves to their hosts in a conspicuous manner. They probably rarely serve as prey for fishes. A single species, *Hippolytina californica*, comprises the second group. . . . This species is a gregarious, wandering, omnivorous animal . . . and is not highly dependent upon its host for survival. So far as is known, it does not display itself to attract fishes. (p. 238)

It is *H. californica* that is occasionally found in the stomachs of at least one of its hosts. The striking correlation of territoriality and solitariness with

cleaning habits is what theory would predict. The same correlation can be found in cleaner fish. *Labroides*, with four species, is the genus most completely dependent on cleaning habits. No *Labroides* has ever been found in the stomach of a host fish. All species are highly site-specific and tend to be solitary. Randall (1958) reports that an individual *L. dimidiatus* may sometimes swim as much as 60 feet from its cleaning station, servicing fish on the way. But he notes,

This was especially true in an area where the highly territorial damsel fish *Pomacentris nigricans* (Lepede) was common. As one damsel fish was being tended, another nearby would assume a stationary pose with fins erect and the *Labroides* would move on to the latter with little hesitation. (p. 333)

Clearly, what matters for the evolution of reciprocal altruism is that the same two individuals interact repeatedly. This will be facilitated by the site specificity of either individual. Of temperate water cleaners, the species most specialized to cleaning is also apparently the most solitary (Hobson, 1969).

(4) *Lifespan of cleaners.* No good data exist on how long cleaners live, but several observations on both fish and shrimp suggest that they easily live long enough for effective selection against cheaters. Randall (1958) repeatedly checked several ledges and found that different feeding stations were occupied for "long periods of time," apparently by the same individuals. One such feeding station supported two individuals for over three years. Of one species of cleaner shrimp, *Stenopus hispidus*, Limbaugh, Pederson, and Chase (1961) said that pairs of individuals probably remain months, possibly years, within an area of a square meter.

(5) *Hosts using the same cleaner repeatedly.* There is surprisingly good evidence that hosts reuse the same cleaner repeatedly. Feder (1966) summarizes the evidence:

Many fishes spend as much time getting cleaned as they do foraging for food. Some fishes return again and again to the same station, and show a definite time pattern in their daily arrival. Others pass from station to station and return many times during the day; this is particularly true of an injured or infected fish. (p. 368)

Limbaugh, Pederson, and Chase (1961) have presented evidence that in at least one species of cleaner shrimp (*Stenopus scutellus*), the shrimp may re-serve the same individuals:

One pair was observed in the same football-sized coral boulder from May through August 1956. During that period, we changed the position and orientation of the boulder several times within a radius of approximately seven meters without disturbing the shrimp. Visiting fishes were momentarily disturbed by the changes, but they soon relocated the shrimps. (p. 254)

Randall (1958) has repeatedly observed fish swimming from out of sight directly to cleaning stations, behavior suggesting to him that they had prior acquaintance with the stations. During two months of observations at several feeding stations, Eibl-Eibesfeldt (1955) became personally familiar with several individual groupers (*Epinephelus striatus*) and repeatedly observed them seeking out and being cleaned at the same feeding stations, presumably by the same cleaners.

In summary, it seems fair to say that the hosts of cleaning organisms perform several kinds of altruistic behavior, including not eating their cleaner after a cleaning, which can be explained on the basis of the above model. A review of the relevant evidence suggests that the cleaner organisms and their hosts meet the preconditions for the evolution of reciprocally altruistic behavior. The host's altruism is to be explained as benefiting him because of the advantage of being able quickly and repeatedly to return to the same cleaner.

Warning Calls in Birds

Marler (1955, 1957) has presented evidence that warning calls in birds tend to have characteristics that limit the information a predator gets from the call. In particular, the call characteristics do not allow the predator easily to determine the location of the call-giver. Thus, it seems that giving a warning call must result, at least occasionally, in the otherwise unnecessary death of the call-giver, either at the hands of the predator that inspired the call or at the hands of a second predator formerly unaware of the caller's presence or exact location.

Given the presumed selection against call-giving, Williams (1966) has reviewed various models to explain selection for warning cries:

(1) Warning calls are functional during the breeding season in birds in that they protect one's mate and offspring. They have no function outside the breeding season, but they are not deleted then because "in practice it is not worth burdening the germ plasm with the information necessary to realize such an adjustment" (Williams, 1966, p. 206).

(2) Warning calls are selected for by the mechanism of group selection (Wynne-Edwards, 1962).

(3) Warning calls are functional outside the breeding season because there is usually a good chance that a reasonably close kin is near enough to be helped sufficiently (Hamilton, 1964; Maynard Smith, 1964). Maynard Smith (1965) has analyzed in great detail how closely related the benefited kin must be, at what benefit to him the call must be, and at what cost to the caller, in order for selection to favor call-giving.

The first is an explanation of last resort. While it must sometimes apply

in evolutionary arguments, it should probably only be invoked when no other explanation seems plausible. The second is not consistent with the known workings of natural selection. The third is feasible and may explain the warning calls in some species and perhaps even in many. But it does depend on the somewhat regular nearby presence of closely related organisms, a matter that may often be the case but that has been demonstrated only as a possibility in a few species and that seems very unlikely in some. A fourth explanation is suggested by the above model:

(4) Warning calls are selected for because they aid the bird giving the call. It is disadvantageous for a bird to have a predator eat a nearby conspecific because the predator may then be more likely to eat him. This may happen because the predator will

- (i) be sustained by the meal,
- (ii) be more likely to form a specific search image of the prey species,
- (iii) be more likely to learn the habits of the prey species and perfect his predatory techniques on it,
- (iv) be more likely to frequent the area in which the birds live, or
- (v) be more likely to learn useful information about the area in which the birds live.

In short, in one way or another, giving a warning call tends to prevent predators from specializing on the caller's species and locality.

There is abundant evidence for the importance of learning in the lives of predatory vertebrates (see, for example, Tinbergen, 1960; Leyhausen, 1965; Brower and Brower, 1965). Rudebeck (1950, 1951) has presented important observations on the tendency of avian predators to specialize individually on prey types and hunting techniques. Owen (1963) and others have presented evidence that species of snails and insects may evolve polymorphisms as a protection against the tendency of their avian predators to learn their appearance. Similarly, Kuyton (1962; cited in Wickler, 1968) has described the adaptation of a moth that minimizes the chance of its predators forming a specific search image. Southern (1954), Murie (1944), and numerous others have documented the tendency of predators to specialize on certain localities within their range. Finally, Blest (1963) has presented evidence that kin selection in some cryptic saturnid moths has favored rapid, post-reproductive death to minimize predation on the young. Blest's evidence thus provides an instance of a predator gaining useful information through the act of predation.

It does not matter that in giving a warning call the caller is helping its non-calling neighbors more than it is helping itself. What counts is that it outcompetes conspecifics from areas in which no one is giving warning calls. The non-calling neighbors of the caller (or their offspring) will soon find themselves in an area without any caller and will be selected against relative to birds in an area with callers. The caller, by definition, is always in an area

with at least one caller. If we assume that two callers are preferable to one, and so on, then selection will favor the spread of the warning-call genes. Note that this model depends on the concept of *open* groups, whereas “group selection” (Wynne-Edwards, 1962) depends partly on the concept of closed groups.

It might be supposed that one could explain bird calls more directly as altruistic behavior that will be repaid when the other birds reciprocate, but there are numerous objections to this. It is difficult to visualize how one would discover and discriminate against the cheater, and there is certainly no evidence that birds refrain from giving calls because neighbors are not reciprocating. Furthermore, if the relevant bird groupings are very fluid, with much emigration and immigration, as they often are, then cheating would seem to be favored and no selection against it possible. Instead, according to the model above, it is the mere fact that the neighbor survives that repays the call-giver his altruism.

It is almost impossible to gather the sort of evidence that would discriminate between this explanation and that of Hamilton (1964). It is difficult to imagine how one would estimate the immediate cost of giving a warning call or its benefit to those within earshot, and precise data on the genetic relationships of bird groupings throughout the year are not only lacking but would be most difficult to gather. Several lines of evidence suggest, however, that Hamilton’s (1964) explanation should be assumed with caution:

- (1) There exist no data showing a decrease in warning tendencies with decrease in the genetic relationship of those within earshot. Indeed, a striking feature of warning calls is that they are given in and out of the breeding season, both before and after migration or dispersal.
- (2) There do exist data suggesting that close kin in a number of species migrate or disperse great distances from each other (Ashmole, 1962; Perdeck, 1958; Berndt and Sternberg, 1968; Dhont and Hublé, 1968).
- (3) One can advance the theoretical argument that kin selection under some circumstances should favor kin dispersal in order to avoid competition (Hamilton, 1964, 1969). This would lead one to expect fewer closely related kin near any given bird, outside the breeding season.

The arguments advanced in this section may also apply, of course, to species other than birds.

Human Reciprocal Altruism

Reciprocal altruism in the human species takes place in a number of contexts and in all known cultures (see, for example, Gouldner, 1960). Any complete list of human altruism would contain the following types of altruistic behavior:

- (1) helping in times of danger (e.g., accidents, predation, intraspecific aggression);
- (2) sharing food;
- (3) helping the sick, the wounded, or the very young and old;
- (4) sharing implements; and
- (5) sharing knowledge.

All these forms of behavior often meet the criterion of small cost to the giver and great benefit to the taker.

During the Pleistocene, and probably before, a hominid species would have met the preconditions for the evolution of reciprocal altruism: long lifespan; low dispersal rate; life in small, mutually dependent, stable, social groups (Lee and DeVore, 1968; Campbell, 1966); and a long period of parental care. It is very likely that dominance relations were of the relaxed, less linear form characteristic of the living chimpanzee (Van Lawick-Goodall, 1968) and not of the more rigidly linear form characteristic of the baboon (Hall and DeVore, 1965). Aid in intraspecific combat, particularly by kin, almost certainly reduced the stability and linearity of the dominance order in early humans. Lee (1969) has shown that in almost all Bushman fights which are initially between two individuals, others have joined in. Mortality, for example, often strikes the secondaries rather than the principals. Tool use has also probably had an equalizing effect on human dominance relations, and the Bushmen have a saying that illustrates this nicely. As a dispute reaches the stage where deadly weapons may be employed, an individual will often declare: "We are none of us big, and others small; we are all men and we can fight; I'm going to get my arrows," (Lee, 1969). It is interesting that Van Lawick-Goodall (1968) has recorded an instance of strong dominance reversal in chimpanzees as a function of tool use. An individual moved from low in dominance to the top of the dominance hierarchy when he discovered the intimidating effects of throwing a metal tin around. It is likely that a diversity of talents is usually present in a band of hunter-gatherers such that the best maker of a certain type of tool is not often the best maker of a different sort or the best user of the tool. This contributes to the symmetry of relationships, since altruistic acts can be traded with reference to the special talents of the individuals involved.

To analyze the details of the human reciprocal-altruistic system, several distinctions are important and are discussed here.

(1) *Kin selection.* The human species also met the preconditions for the operation of kin selection. Early hominid hunter-gatherer bands almost certainly (like today's hunter-gatherers) consisted of many close kin, and kin selection must often have operated to favor the evolution of some types of altruistic behavior (Haldane, 1955; Hamilton, 1964, 1969). In general, in attempting to discriminate between the effects of kin selection and what

might be called reciprocal-altruistic selection, one can analyze the form of the altruistic behaviors themselves. For example, the existence of discrimination against non-reciprocal individuals cannot be explained on the basis of kin selection, in which the advantage accruing to close kin is what makes the altruistic behavior selectively advantageous, not its chance of being reciprocated. The strongest argument for the operation of reciprocal-altruistic selection in man is the psychological system controlling some forms of human altruism. Details of this system are reviewed below.

(2) *Reciprocal altruism among close kin.* If both forms of selection have operated, one would expect some interesting interactions. One might expect, for example, a lowered demand for reciprocity from kin than from nonkin, and there is evidence to support this (e.g., Marshall, 1961; Balicki, 1964). The demand that kin show some reciprocity (e.g., Marshall, 1961; Balicki, 1964) suggests, however, that reciprocal-altruistic selection has acted even on relations between close kin. Although interactions between the two forms of selection have probably been important in human evolution, this paper will limit itself to a preliminary description of the human reciprocally altruistic system, a system whose attributes are seen to result only from reciprocal-altruistic selection.

(3) *Age-dependent changes.* Cost and benefit were defined above without reference to the ages, and hence reproductive values (Fisher, 1958), of the individuals involved in an altruistic exchange. Since the reproductive value of a sexually mature organism declines with age, the benefit to him of a typical altruistic act also decreases, as does the cost to him of a typical act he performs. If the interval separating the two acts in an altruistic exchange is short relative to the lifespans of the individuals, then the error is slight. For longer intervals, in order to be repaid precisely, the initial altruist must receive more in return than he himself gave. It would be interesting to see whether humans in fact routinely expect "interest" to be added to a long overdue altruistic debt, interest commensurate with the intervening decline in reproductive value. In humans reproductive value declines most steeply shortly after sexual maturity is reached (Hamilton, 1966), and one would predict the interest rate on altruistic debts to be highest then. Selection might also favor keeping the interval between act and reciprocation short, but this should also be favored to protect against complete non-reciprocation. W. D. Hamilton (pers. commun.) has suggested that a detailed analysis of age-dependent changes in kin altruism and reciprocal altruism should show interesting differences, but the analysis is complicated by the possibility of reciprocity to the kin of a deceased altruist (see *Multiparty interactions* below).

(4) *Gross and subtle cheating.* Two forms of cheating can be distinguished, here denoted as gross and subtle. In *gross cheating* the cheater fails to reciprocate at all, and the altruist suffers the costs of whatever altruism he has

dispensed without any compensating benefits. More broadly, gross cheating may be defined as reciprocating so little, if at all, that the altruist receives less benefit from the gross cheater than the cost of the altruist's acts of altruism to the cheater. That is, $\sum_i C_{ai} > \sum_j b_{aj}$, where c_{ai} is the cost of the i th altruistic act performed by the altruist and where b_{aj} is the benefit to the altruist of the j th altruistic act performed by the gross cheater; altruistic situations are assumed to have occurred symmetrically. Clearly, selection will strongly favor prompt discrimination against the gross cheater. *Subtle cheating*, by contrast, involves reciprocating, but always attempting to give less than one was given, or more precisely, to give less than the partner would give if the situation were reversed. In this situation, the altruist still benefits from the relationship but not as much as he would if the relationship were completely equitable. The subtle cheater benefits more than he would if the relationship were equitable. In other words,

$$\sum_{i,j} (b_{ai} - c_{qj}) > \sum_i (b_{qi} - c_{ai}) > \sum_{i,j} (b_{aj} - c_{ai})$$

where the i th altruistic act performed by the altruist has a cost to him of c_{ai} and a benefit to the subtle cheater of b_{qi} and where the j th altruistic act performed by the subtle cheater has a cost to him of c_{qj} and a benefit to the altruist of b_{aj} . Because human altruism may span huge periods of time, a lifetime even, and because thousands of exchanges may take place, involving many different "goods" and with many different cost/benefit ratios, the problem of computing the relevant totals, detecting imbalances, and deciding whether they are due to chance or to small-scale cheating is an extremely difficult one. Even then, the altruist is in an awkward position, symbolized by the folk saying, "half a loaf is better than none," for if attempts to make the relationship equitable lead to the rupture of the relationship, the altruist, assuming other things to be equal, will suffer the loss of the substandard altruism of the subtle cheater. It is the subtlety of the discrimination necessary to detect this form of cheating and the awkward situation that ensues that permit some subtle cheating to be adaptive. This sets up a dynamic tension in the system that has important repercussions, as discussed below.

(5) *Number of reciprocal relationships.* It has so far been assumed that it is to the advantage of each individual to form the maximum number of reciprocal relationships and that the individual suffers a decrease in fitness upon the rupture of any relationship in which the cost to him of acts dispensed to the partner is less than the benefit of acts dispensed toward him by the partner. But it is possible that relationships are partly exclusive, in the sense that expanding the number of reciprocal exchanges with one of the partners may necessarily decrease the number of exchanges with another. For example, if a group of organisms were to split into subgroups for much of the day (such as breaking up into hunting pairs), then altruistic exchanges will be more likely between members of each subgroup than between members

of different subgroups. In that sense, relationships may be partly exclusive, membership in a given subgroup necessarily decreasing exchanges with others in the group. The importance of this factor is that it adds further complexity to the problem of dealing with the cheater and it increases competition within a group to be members of a favorable subgroup. An individual in a subgroup who feels that another member is subtly cheating on their relationship has the option of attempting to restore the relationship to a completely reciprocal one or of attempting to join another subgroup, thereby decreasing to a minimum the possible exchanges between himself and the subtle cheater and replacing these with exchanges between a new partner or partners. In short, he can switch friends. There is evidence in hunter-gatherers that much movement of individuals from one band to another occurs in response to such social factors as have just been outlined (Lee and DeVore, 1968).

(6) *Indirect benefits or reciprocal altruism?* Given mutual dependence in a group it is possible to argue that the benefits (nonaltruistic) of this mutual dependence are a positive function of group size and that altruistic behaviors may be selected for because they permit additional individuals to survive and thereby confer additional indirect (nonaltruistic) benefits. Such an argument can only be advanced seriously for slowly reproducing species with little dispersal. Saving an individual's life in a hunter-gatherer group, for example, may permit nonaltruistic actions such as cooperative hunting to continue with more individuals. But if there is an optimum group size, one would expect adaptations to stay near that size, with individuals joining groups when the groups are below this size, and groups splitting up when they are above this size. One would only be selected to keep an individual alive when the group is below optimum and not when the group is above optimum. Although an abundant literature on hunter-gatherers (and also nonhuman primates) suggests that adaptations exist to regulate group size near an optimum, there is no evidence that altruistic gestures are curtailed when groups are above the optimum in size. Instead, the benefits of human altruism are to be seen as coming directly from reciprocity—not indirectly through nonaltruistic group benefits. This distinction is important because social scientists and philosophers have tended to deal with human altruism in terms of the benefits of living in a group, without differentiating between nonaltruistic benefits and reciprocal benefits (e.g., Rousseau, 1954; Baier, 1958).

The Psychological System Underlying Human Reciprocal Altruism

Anthropologists have recognized the importance of reciprocity in human behavior, but when they have ascribed functions to such behavior they have

done so in terms of group benefits, reciprocity cementing group relations and encouraging group survival. The individual sacrifices so that the group may benefit. Recently psychologists have studied altruistic behavior in order to show what factors induce or inhibit such behavior. No attempt has been made to show what function such behavior may serve, nor to describe and interrelate the components of the psychological system affecting altruistic behavior. The purpose of this section is to show that the above model for the natural selection of reciprocally altruistic behavior can readily explain the function of human altruistic behavior and the details of the psychological system underlying such behavior. The psychological data can be organized into functional categories, and it can be shown that the components of the system complement each other in regulating the expression of altruistic and cheating impulses to the selective advantage of individuals. No concept of group advantage is necessary to explain the function of human altruistic behavior.

There is no direct evidence regarding the degree of reciprocal altruism practiced during human evolution nor its genetic basis today, but given the universal and nearly daily practice of reciprocal altruism among humans today, it is reasonable to assume that it has been an important factor in recent human evolution and that the underlying emotional dispositions affecting altruistic behavior have important genetic components. To assume as much allows a number of predictions.

(1) *A complex, regulating system.* The human altruistic system is a sensitive, unstable one. Often it will pay to cheat: namely, when the partner will not find out, when he will not discontinue his altruism even if he does find out, or when he is unlikely to survive long enough to reciprocate adequately. And the perception of subtle cheating may be very difficult. Given this unstable character of the system, where a degree of cheating is adaptive, natural selection will rapidly favor a complex psychological system in each individual regulating both his own altruistic and cheating tendencies and his responses to these tendencies in others. As selection favors subtler forms of cheating, it will favor more acute abilities to detect cheating. The system that results should simultaneously allow the individual to reap the benefits of altruistic exchanges, to protect himself from gross and subtle forms of cheating, and to practice those forms of cheating that local conditions make adaptive. Individuals will differ not in being altruists or cheaters but in the degree of altruism they show and in the conditions under which they will cheat.

The best evidence supporting these assertions can be found in Krebs' (1970) review of the relevant psychological literature. Although he organizes it differently, much of the material supporting the assertions below is taken from his paper. All references to Krebs below are to this review. Also, Hartshorne and May (1928–1930) have shown that children in experimental situations do not divide bimodally into altruists and “cheaters” but are

distributed normally; almost all the children cheated, but they differed in how much and under what circumstances. ("Cheating" was defined in their work in a slightly different but analogous way.)

(2) *Friendship and the emotions of liking and disliking.* The tendency to like others, not necessarily closely related, to form friendships and to act altruistically toward friends and toward those one likes will be selected for as the immediate emotional rewards motivating altruistic behavior and the formation of altruistic partnerships. (Selection may also favor helping strangers or disliked individuals when they are in particularly dire circumstances.) Selection will favor a system whereby these tendencies are sensitive to such parameters as the altruistic tendencies of the liked individual. In other words, selection will favor liking those who are themselves altruistic.

Sawyer (1966) has shown that all groups in all experimental situations tested showed more altruistic behavior toward friends than toward neutral individuals. Likewise, Friedrichs (1960) has shown that attractiveness as a friend was most highly correlated among undergraduates with altruistic behavior. Krebs has reviewed other studies that suggest that the relationship between altruism and liking is a two-way street: one is more altruistic toward those one likes and one tends to like those who are most altruistic (e.g., Berkowitz and Friedman, 1967; Lerner and Lichtman, 1968).

Others (Darwin, 1871); Williams, 1966; and Hamilton, 1969) have recognized the role friendship might play in engendering altruistic behavior, but all have viewed friendship (and intelligence) as prerequisites for the appearance of such altruism. Williams (1966), who cites Darwin (1871) on the matter, speaks of this behavior as evolving,

in animals that live in stable social groups and have the intelligence and other mental qualities necessary to form a system of personal friendships and animosities that transcend the limits of family relationships. (p. 93)

This emphasis on friendship and intelligence as prerequisites leads Williams to limit his search for altruism to the Mammalia and to a "minority of this group." But according to the model presented above, emotions of friendship (and hatred) are not prerequisites for reciprocal altruism but may evolve after a system of mutual altruism has appeared, as important ways of regulating the system.

(3) *Moralistic aggression.* Once strong positive emotions have evolved to motivate altruistic behavior, the altruist is in a vulnerable position because cheaters will be selected to take advantage of the altruist's positive emotions. This in turn sets up a selection pressure for a protective mechanism. Moralistic aggression and indignation in humans was selected for in order

- (a) to counteract the tendency of the altruist, in the absence of any reciprocity, to continue to perform altruistic acts for his own emotional rewards;

- (b) to educate the unreciprocating individual by frightening him with immediate harm or with the future harm of no more aid; and
- (c) in extreme cases, perhaps, to select directly against the unreciprocating individual by injuring, killing, or exiling him.

Much of human aggression has moral overtones. Injustice, unfairness, and lack of reciprocity often motivate human aggression and indignation. Lee (1969) has shown that verbal disputes in Bushmen usually revolve around problems of gift-giving, stinginess, and laziness. DeVore (pers. commun.) reports that a great deal of aggression in hunter-gatherers revolves around real or imagined injustices—inequities, for example, in food-sharing (see, for example, Thomas, 1958; Balikci, 1964; Marshall, 1961). A common feature of this aggression is that it often seems out of all proportion to the offenses committed. Friends are even killed over apparently trivial disputes. But since small inequities repeated many times over a lifetime may exact a heavy toll in relative fitness, selection may favor a strong show of aggression when the cheating tendency is discovered. Recent discussions of human and animal aggression have failed to distinguish between moralistic and other forms of aggression (e.g., Scott, 1958; Lorenz, 1966; Montague, 1968; Tinbergen, 1968; Gilula and Daniels, 1969). The grounds for expecting, on functional grounds, a highly plastic developmental system affecting moralistic aggression are discussed below.

(4) *Gratitude, sympathy, and the cost/benefit ratio of an altruistic act.* If the cost/benefit ratio is an important parameter in determining the adaptiveness of reciprocal altruism, then humans should be selected to be sensitive to the cost and benefit of an altruistic act, both in deciding whether to perform one and in deciding whether, or how much, to reciprocate. I suggest that the emotion of gratitude has been selected to regulate human response to altruistic acts and that the emotion is sensitive to the cost/benefit ratio of such acts. I suggest further that the emotion of sympathy has been selected to motivate altruistic behavior as a function of the plight of the recipient of such behavior; crudely put, the greater the potential benefit to the recipient, the greater the sympathy and the more likely the altruistic gesture, even to strange or disliked individuals. If the recipient's gratitude is indeed a function of the cost/benefit ratio, then a sympathetic response to the plight of a disliked individual may result in considerable reciprocity.

There is good evidence supporting the psychological importance of the cost/benefit ratio of altruistic acts. Gouldner (1960) has reviewed the sociological literature suggesting that the greater the need state of the recipient of an altruistic act, the greater his tendency to reciprocate; and the scarcer the resources of the donor of the act, the greater the tendency of the recipient to reciprocate. Heider (1958) has analyzed lay attitudes on altruism and finds that gratitude is greatest when the altruistic act does good. Tesser, Gatewood, and Driver (1968) have shown that American undergraduates

thought they would feel more gratitude when the altruistic act was valuable and cost the benefactor a great deal. Pruitt (1968) has provided evidence that humans reciprocate more when the original act was expensive for the benefactor. He shows that under experimental conditions more altruism is induced by a gift of 80 per cent of \$1.00 than 20 per cent of \$4.00. Aronfreed (1968) has reviewed the considerable evidence that sympathy motivates altruistic behavior as a function of the plight of the individual arousing the sympathy.

(5) *Guilt and reparative altruism.* If an organism has cheated on a reciprocal relationship and this fact has been found out, or has a good chance of being found out, by the partner and if the partner responds by cutting off all future acts of aid, then the cheater will have paid dearly for his misdeed. It will be to the cheater's advantage to avoid this, and, providing that the cheater makes up for his misdeed and does not cheat in the future, it will be to his partner's benefit to avoid this, since in cutting off future acts of aid he sacrifices the benefits of future reciprocal help. The cheater should be selected to make up for his misdeed and to show convincing evidence that he does not plan to continue his cheating sometime in the future. In short, he should be selected to make a reparative gesture. It seems plausible, furthermore, that the emotion of guilt has been selected for in humans partly in order to motivate the cheater to compensate his misdeed and to behave reciprocally in the future, and thus to prevent the rupture of reciprocal relationships.

Krebs has reviewed the evidence that harming another individual publicly leads to altruistic behavior and concludes:

Many studies have supported the notion that public transgression whether intentional or unintentional, whether immoral or only situationally unfortunate, leads to reparative altruism. (p. 267)

Wallace and Sadalla (1966), for example, showed experimentally that individuals who broke an expensive machine were more likely to volunteer for a painful experiment than those who did not, but only if their transgression had been discovered. Investigators disagree on the extent to which guilt feelings are the motivation behind reparative altruism. Epstein and Hornstein (1969) supply some evidence that guilt is involved, but on the assumption that one feels guilt even when one behaves badly in private, Wallace and Sadalla's (1966) result contradicts the view that guilt is the only motivating factor. That private transgressions are not as likely as public ones to lead to reparative altruism is precisely what the model would predict, and it is possible that the common psychological assumption that one feels guilt even when one behaves badly in private is based on the fact that many transgressions performed in private are *likely* to become public knowledge. It should often be advantageous to confess sins that are likely to be

discovered before they actually are, as evidence of sincerity (see below on detection of mimics).

(6) *Subtle cheating: the evolution of mimics.* Once friendship, moralistic aggression, guilt, sympathy, and gratitude have evolved to regulate the altruistic system, selection will favor mimicking these traits in order to influence the behavior of others to one's own advantage. Apparent acts of generosity and friendship may induce genuine friendship and altruism in return. Sham moralistic aggression when no real cheating has occurred may nevertheless induce reparative altruism. Sham guilt may convince a wronged friend that one has reformed one's ways even when the cheating is about to be resumed. Likewise, selection will favor the hypocrisy of pretending one is in dire circumstances in order to induce sympathy-motivated altruistic behavior. Finally, mimicking sympathy may give the appearance of helping in order to induce reciprocity, and mimicking gratitude may mislead an individual into expecting he will be reciprocated. It is worth emphasizing that a mimic need not necessarily be conscious of the deception; selection may favor feeling genuine moralistic aggression even when one has not been wronged if so doing leads another to reparative altruism.

Instances of the above forms of subtle cheating are not difficult to find. For typical instances from the literature on hunter-gatherers see Rasmussen (1931), Balıkcı (1964), and Lee and DeVore (1968). The importance of these forms of cheating can partly be inferred from the adaptations to detect such cheating discussed below and from the importance and prevalence of moralistic aggression once such cheating is detected.

(7) *Detection of the subtle cheater: Trust-worthiness, trust, and suspicion.* Selection should favor the ability to detect and discriminate against subtle cheaters. Selection will clearly favor detecting and countering sham moralistic aggression. The argument for the others is more complex. Selection may favor distrusting those who perform altruistic acts without the emotional basis of generosity or guilt because the altruistic tendencies of such individuals may be less reliable in the future. One can imagine, for example, compensating for a misdeed without any emotional basis but with a calculating, self-serving motive. Such an individual should be distrusted because the calculating spirit that leads this subtle cheater now to compensate may in the future lead him to cheat when circumstances seem more advantageous (because of unlikelihood of detection, for example, or because the cheated individual is unlikely to survive). Guilty motivation, insofar as it evidences a more enduring commitment to altruism, either because guilt teaches or because the cheater is unlikely not to feel the same guilt in the future, seems more reliable. A similar argument can be made about the trustworthiness of individuals who initiate altruistic acts out of a calculating rather than a generous-hearted disposition or who show either false sympathy or false gratitude. Detection on the basis of the underlying psycho-

logical dynamics is only one form of detection. In many cases, unreliability may more easily be detected through experiencing the cheater's inconsistent behavior. And in some cases, third party interactions (as discussed below) may make an individual's behavior predictable despite underlying cheating motivations.

The anthropological literature also abounds with instances of the detection of subtle cheaters (see above references for hunter-gatherers). Although I know of no psychological studies on the detection of sham moralistic aggression and sham guilt, there is ample evidence to support the notion that humans respond to altruistic acts according to their perception of the motives of the altruist. They tend to respond more altruistically when they perceive the other as acting "genuinely" altruistic, that is, voluntarily dispatching an altruistic act as an end in itself, without being directed toward gain (Leeds, 1963; Heider, 1958). Krebs (1970) has reviewed the literature on this point and notes that help is more likely to be reciprocated when it is perceived as voluntary and intentional (e.g., Goranson and Berkowitz, 1966; Lerner and Lichtman, 1968) and when the help is appropriate, that is, when the intentions of the altruist are not in doubt (e.g., Brehm and Cole, 1966; Schopler and Thompson, 1968). Krebs concludes that, "When the legitimacy of apparent altruism is questioned, reciprocity is less likely to prevail." Lerner and Lichtman (1968) have shown experimentally that those who act altruistically for ulterior benefit are rated as unattractive and are treated selfishly, whereas those who apparently are genuinely altruistic are rated as attractive and are treated altruistically. Berscheid and Walster (1967) have shown that church women tend to make reparations for harm they have committed by choosing the reparation that approximates the harm (that is, is neither too slight nor too great), presumably to avoid the appearance of inappropriateness.

Rapoport and Dale (1967) have shown that when two strangers play iterated games of Prisoner's Dilemma in which the matrix determines profits from the games played there is a significant tendency for the level of cooperation to drop at the end of the series, reflecting the fact that the partner will not be able to punish for "cheating" responses when the series is over. If a long series is broken up into subseries with a pause between subseries for totaling up gains and losses, then the tendency to cheat on each other increases at the end of each subseries. These results, as well as some others reported by Rapoport and Chammah (1965), are suggestive of the instability that exists when two strangers are consciously trying to maximize gain by trading altruistic gestures, an instability that is presumably less marked when the underlying motivation involves the emotions of friendship, of liking others, and of feeling guilt over harming a friend. Deutsch (1958), for example, has shown that two individuals playing iterated games of Prisoner's Dilemma will be more cooperative if a third individual, disliked by both, is

present. The perceived mutual dislike is presumed to create a bond between the two players.

It is worth mentioning that a classic problem in social science and philosophy has been whether to define altruism in terms of motives (e.g., real vs. "calculated" altruism) or in terms of behavior, regardless of motive (Krebs, 1970). This problem reflects the fact that, wherever studied, humans seem to make distinctions about altruism partly on the basis of motive, and this tendency is consistent with the hypothesis that such discrimination is relevant to protecting oneself from cheaters.

(8) *Setting up altruistic partnerships.* Selection will favor a mechanism for establishing reciprocal relationships. Since humans respond to acts of altruism with feelings of friendship that lead to reciprocity, one such mechanism might be the performing of altruistic acts toward strangers, or even enemies, in order to induce friendship. In short, do unto others as you would have them do unto you.

The mechanism hypothesized above leads to results inconsistent with the assumption that humans always act more altruistically toward friends than toward others. Particularly toward strangers, humans may initially act more altruistically than toward friends. Wright (1942) has shown, for example, that third grade children are more likely to give a more valuable toy to a stranger than to a friend. Later, some of these children verbally acknowledged that they were trying to make friends. Floyd (1964) has shown that, after receiving many trinkets from a friend, humans tend to *decrease* their gifts in return, but after receiving many trinkets from a neutral or disliked individual, they tend to *increase* their gifts in return. Likewise, after receiving few trinkets from a friend, humans tend to increase their gifts in return, whereas receiving few trinkets from a neutral or disliked individual results in a decrease in giving. This was interpreted to mean that generous friends are taken for granted (as are stingy non-friends). Generosity from a non-friend is taken to be an overture to friendship, and stinginess from a friend as evidence of a deteriorating relationship in need of repair. (Epstein and Hornstein, 1969, provide new data supporting this interpretation of Floyd, 1964.)

(9) *Multiparty interactions.* In the close-knit social groups that humans usually live in, selection should favor more complex interactions than the two-party interactions so far discussed. Specifically, selection may favor learning from the altruistic and cheating experiences of others, helping others coerce cheaters, forming multiparty exchange systems, and formulating rules for regulated exchanges in such multiparty systems.

(i) *Learning from others.* Selection should favor learning about the altruistic and cheating tendencies of others indirectly, both through observing interactions of others and, once linguistic abilities have evolved, by hearing about such interactions or hearing characterizations of individuals (e.g., "dirty, hy-

pocritical, dishonest, untrustworthy, cheating louse"). One important result of this learning is that an individual may be as concerned about the attitude of onlookers in an altruistic situation as about the attitude of the individual being dealt with.

(ii) *Help in dealing with cheaters.* In dealing with cheaters selection may favor individuals helping others, kin or non-kin, by direct coercion against the cheater or by everyone refusing him reciprocal altruism. One effect of this is that an individual, through his close kin, may be compensated for an altruistic act even after his death. An individual who dies saving a friend, for example, may have altruistic acts performed by the friend to the benefit of his offspring. Selection will discriminate against the cheater in this situation, if kin of the martyr, or others, are willing to punish lack of reciprocity.

(iii) *Generalized altruism.* Given learning from others and multiparty action against cheaters, selection may favor a multiparty altruistic system in which altruistic acts are dispensed freely among more than two individuals, an individual being perceived to cheat if in an altruistic situation he dispenses less benefit for the same cost than would the others, punishment coming not only from the other individual in that particular exchange but from the others in the system.

(iv) *Rules of exchange.* Multiparty altruistic systems increase by several-fold the cognitive difficulties in detecting imbalances and deciding whether they are due to cheating or to random factors. One simplifying possibility that language facilitates is the formulation of rules of conduct, cheating being detected as infraction of such a rule. In short, selection may favor the elaboration of norms of reciprocal conduct.

There is abundant evidence for all of the above multiparty interactions (see the above references on hunter-gatherers). Thomas (1958), for example, has shown that debts of reciprocity do not disappear with the death of the "creditor" but are extended to his kin. Krebs has reviewed the psychological literature on generalized altruism. Several studies (e.g., Darlington and Macker, 1966) have shown that humans may direct their altruism to individuals other than those who were hurt and may respond to an altruistic act that benefits themselves by acting altruistically toward a third individual uninvolved in the initial interaction. Berkowitz and Daniels (1964) have shown experimentally, for example, that help from a confederate leads the subject to direct more help to a third individual, a highly dependent supervisor. Freedman, Wallington, and Bless (1967) have demonstrated the surprising result that, in two different experimental situations, humans engaged in reparative altruism only if it could be directed to someone other than the individual harmed, or to the original individual only if they did not expect to meet again. In a system of strong multiparty interactions it is possible that in some situations individuals are selected to demonstrate generalized altruistic tendencies and that their main concern when they have harmed

another is to show that they are genuinely altruistic, which they best do by acting altruistic without any apparent ulterior motive, e.g., in the experiments by acting altruistic toward an uninvolved third party. Alternatively, A. Rapoport (pers. commun.) has suggested that the reluctance to direct reparative altruism toward the harmed individual may be due to unwillingness to show thereby a recognition of the harm done him. The re-direction serves to allay guilt feelings without triggering the greater reparation that recognition of the harm might lead to.

(10) *Developmental plasticity.* The conditions under which detection of cheating is possible, the range of available altruistic trades, the cost/benefit ratios of these trades, the relative stability of social groupings, and other relevant parameters should differ from one ecological and social situation to another and should differ through time in the same small human population. Under these conditions one would expect selection to favor developmental plasticity of those traits regulating altruistic and cheating tendencies and responses to these tendencies in others. For example, developmental plasticity may allow the growing organism's sense of guilt to be educated, perhaps partly by kin, so as to permit those forms of cheating that local conditions make adaptive and to discourage those with more dangerous consequences. One would not expect any simple system regulating the development of altruistic behavior. To be adaptive, altruistic behavior must be dispensed with regard to many characteristics of the recipient (including his degree of relationship, emotional makeup, past behavior, friendships, and kin relations), of other members of the group, of the situation in which the altruistic behavior takes place, and of many other parameters, and no simple developmental system is likely to meet these requirements.

Kohlberg (1963), Bandura and Walters (1963), and Krebs have reviewed the developmental literature on human altruism. All of them conclude that none of the proposed developmental theories (all of which rely on simple mechanisms) can account for the known diverse developmental data. Whiting and Whiting (in prep.) have studied altruistic behavior directed towards kin by children in six different cultures and find consistent differences among the cultures that correlate with differences in child-rearing and other facets of the cultures. They argue that the differences adapt the children to different adult roles available in the cultures. Although the behavior analyzed takes place between kin and hence Hamilton's model (1964) may apply rather than this model, the Whittings' data provide an instance of the adaptive value of developmental plasticity in altruistic behavior. No careful work has been done analyzing the influence of environmental factors on the development of altruistic behavior, but some data exist. Krebs has reviewed the evidence that altruistic tendencies can be increased by the effects of warm, nurturant models, but little is known on how long such effects endure. Rosenhan (1967) and Rettig (1956) have shown a correlation between

altruism in parents and altruism in their college-age children, but these studies do not separate genetic and environmental influences. Class differences in altruistic behavior (e.g., Berkowitz, 1968; Ugurel-Semin, 1952; Almond and Verba, 1963) may primarily reflect environmental influences. Finally, Lutzker (1960) and Deutsch (1958) have shown that one can predict the degree of altruistic behavior displayed in iterated games of Prisoner's Dilemma from personality typing based on a questionnaire. Such personality differences are probably partly environmental in origin.

It is worth emphasizing that some of the psychological traits analyzed above have applications outside the particular reciprocal altruistic system being discussed. One may be suspicious, for example, not only of individuals likely to cheat on the altruistic system, but of any individual likely to harm oneself; one may be suspicious of the known tendencies toward adultery of another male or even of these tendencies in one's own mate. Likewise, a guilt-motivated show of reparation may avert the revenge of someone one has harmed, whether that individual was harmed by cheating on the altruistic system or in some other way. And the system of reciprocal altruism may be employed to avert possible revenge. The Bushmen of the Kalahari, for example, have a saying (Marshall, 1959) to the effect that, if you wish to sleep with someone else's wife, you get him to sleep with yours, then neither of you goes after the other with poisoned arrows. Likewise, there is a large literature on the use of reciprocity to cement friendships between neighboring groups, now engaged in a common enterprise (e.g., Lee and DeVore, 1968).

The above review of the evidence has only begun to outline the complexities of the human altruistic system. The inherent instability of the Prisoner's Dilemma, combined with its importance in human evolution, has led to the evolution of a very complex system. For example, once moralistic aggression has been selected for to protect against cheating, selection favors sham moralistic aggression as a new form of cheating. This should lead to selection for the ability to discriminate the two and to guard against the latter. The guarding can, in turn, be used to counter real moralistic aggression: one can, in effect, *impute* cheating motives to another person in order to protect one's own cheating. And so on. Given the psychological and cognitive complexity the system rapidly acquires, one may wonder to what extent the importance of altruism in human evolution set up a selection pressure for psychological and cognitive powers which partly contributed to the large increase in hominid brain size during the Pleistocene.

ACKNOWLEDGMENTS

I thank W. H. Drury, E. Mayr, I. Nisbet, E. E. Williams and E. O. Wilson for useful comments on earlier drafts of this paper, and I thank especially I. DeVore and W. D.

Hamilton for detailed comment and discussion. I thank A. Rapoport and D. Krebs for access to unpublished material. This work was completed under a National Science Foundation pre-doctoral fellowship and partially supported by grant number NIMH 13156 to I. DeVore.

LIST OF LITERATURE

- Almond, G. A., and S. Verba. 1963. *The Civic Culture*. Princeton University Press, Princeton, N.J.
- Aronfreed, J. 1968. *Conduct and Conscience*. Academic Press, N.Y.
- Ashmole, M. 1962. Migration of European thrushes: a comparative study based on ringing recoveries. *Ibis*, 104: 522-559.
- Baier, K. 1958. *The Moral Point of View*. Cornell University Press, Ithaca, N.Y.
- Balikci, A. 1964. Development of basic socio-economic units in two Eskimo communities. National Museum of Canada Bulletin No. 202, Ottawa.
- Bandura, A., and R. H. Walters. 1963. *Social Learning and Personality Development*. Holt, Rhinehart and Winston, N.Y.
- Berkowitz, L. 1968. Responsibility, reciprocity and social distance in help-giving: an experimental investigation of English social class differences. *J. Exp. Soc. Psychol.*, 4: 664-669.
- Berkowitz, L., and L. Daniels. 1964. Affecting the salience of the social responsibility norm: effects of past help on the response to dependency relationships. *J. Abnorm. Soc. Psychol.*, 68: 275-281.
- Berkowitz, L., and P. Friedman. 1967. Some social class differences in helping behavior. *J. Personal. Soc. Psychol.*, 5: 217-225.
- Berndt, R., and H. Sternberg. 1968. Terms, studies and experiments on the problems of bird dispersion. *Ibis*, 110: 256-269.
- Berscheid, E. and E. Walster. 1967. When does a harm-doer compensate a victim? *J. Personal. Soc. Psychol.*, 6: 435-441.
- Blest, A. D. 1963. Longevity, palatability and natural selection in five species of New World Saturniid moth. *Nature*, 197: 1183-1186.
- Brehm, J. W., and A. H. Cole, 1966. Effect of a favor which reduces freedom. *J. Personal. Soc. Psychol.*, 3: 420-426.
- Brower, J.V.Z., and L. P. Brower. 1965. Experimental studies of mimicry. 8. *Am. Natur.*, 49: 173-188.
- Campbell, B. 1966. *Human Evolution*. Aldine, Chicago.
- Crook, J. H. 1969. The socio-ecology of primates. In J. H. Crooke (ed.), *Social Behavior in Birds and Mammals*, p. 103-166. Academic Press, London.
- Darlington, R. B., and C. E. Macker. 1966. Displacement of guilt-produced altruistic behavior. *J. Personal. Soc. Psychol.*, 4: 442-443.
- Darwin, C. 1871. *The Descent of Man and Selection in Relation to Sex*. Random House, N.Y.
- Deutsch, M. 1958. Trust and suspicion. *J. Conflict Resolution*, 2: 267-279.
- Dhont, A. A., and J. Hublé. 1968. Fledging-date and sex in relation to dispersal in young Great Tits. *Bird Study*, 15: 127-134.
- Eibl-Eibesfeldt, I. 1955. Über Symbiosen, Parasitismus und andere besondere zwischenartige Beziehungen tropischer Meeresfische. *Z. f. Tierpsychol.*, 12: 203-219.
- . 1959. Der Fisch *Aspidontus taeniatus* als Nachahmer des Putzers *Labroides dimidiatus*. *Z. f. Tierpsychol.*, 16: 19-25.

- Epstein, Y. M., and H. A. Horstein. 1969. Penalty and interpersonal attraction as factors influencing the decision to help another person. *J. Exp. Soc. Psychol.*, 5: 272-282.
- Feder, H. M. 1966. Cleaning symbioses in the marine environment. In S. M. Henry (ed.), *Symbiosis*, Vol. 1, p. 327-380. Academic Press, N.Y.
- Fisher, R. A. 1958. *The Genetical Theory of Natural Selection*. Dover, N.Y.
- Floyd, J. 1964. Effects of amount of award and friendship status of the other on the frequency of sharing in children. Unpublished doctoral dissertation. University of Minnesota. (Reviewed in Krebs, 1970).
- Freedman, J. L., S. A. Wallington, and E. Bless. 1967. Compliance without pressure: the effect of guilt. *J. Personal. Soc. Psychol.*, 7: 117-124.
- Friedrichs, R. W. 1960. Alter versus ego: an exploratory assessment of altruism. *Am. Sociol. Rev.*, 25: 496-508.
- Gilula, M. F., and D. N. Daniels. 1969. Violence and man's struggle to adapt. *Science*, 164: 395-405.
- Goranson, R., and Berkowitz. 1966. Reciprocity and responsibility reactions to prior help. *J. Personal. Soc., Psychol.* 3: 227-232.
- Gouldner, A. 1960. The norm of reciprocity: a preliminary statement. *Am. Sociol. Rev.*, 47: 73-80.
- Haldane, J.B.S. 1955. Population genetics. *New Biology*, 18: 34-51.
- Hall, K.R.L., and I. DeVore. 1965. Baboon social behavior. In I. DeVore (ed.), *Primate Behavior: Field Studies of Monkeys and Apes*, p. 53-110. Holt, Rhinehart and Winston, N.Y.
- Hamilton, W. D. 1964. The genetical evolution of social behavior. *J. Theoret. Biol.*, 7: 1-52.
- . 1966. The moulding of senescence by natural selection. *J. Theoret. Biol.*, 12: 12-45.
- . 1969. Selection of selfish and altruistic behavior in some extreme models. Paper presented at "Man and Beast Symposium" (in press, Smithsonian Institution).
- Hartshorne, H., and M. A. May. 1928-1930. *Studies in the Nature of Character*. Vol. 1, *Studies in Deceit*; Vol. 2, *Studies in Self-Control*; Vol. 3, *Studies in the Organization of Character*. Macmillan, N.Y.
- Hediger, H. 1968. Putzer-fische im aquarium. *Natur und Museum*, 98: 89-96.
- Heider, F. 1958. *The Psychology of Interpersonal Relations*. Wiley, N.Y.
- Hobson, E. S. 1969. Comments on certain recent generalizations regarding cleaning symbioses in fishes. *Pacific Science*, 23: 35-39.
- Kohlberg, L. 1963. Moral development and identification. In H. W. Stevenson (ed.), *Yearbook of the National Society for the Study of Education. Part 1. Child Psychology*, p. 277-332. University of Chicago Press, Chicago.
- Krebs, D. 1970. Altruism—an examination of the concept and a review of the literature. *Psychol. Bull.*, 73: 258-302.
- Kuyton, P. 1962. Verhalten-beobachtungen an der Raupe des Kaiseratlas. *Z. d. Entomol.*, 72: 203-207.
- Lee, R. 1969. !Kung Bushman violence. Paper presented at meeting of American Anthropological Association, Nov. 1969.
- Lee, R., and I. DeVore, 1968. *Man the Hunter*. Aldine, Chicago.
- Leeds, R. 1963. Altruism and the norm of giving. *Merrill-Palmer Quart.*, 9: 229-240.
- Lerner, M. J., and R. R. Lichtman. 1968. Effects of perceived norms on attitudes

- and altruistic behavior toward a dependent other. *J. Personal. Soc. Psychol.*, 9: 226-232.
- Leyhausen, P. 1965. Über die Funktion der relativen Stimmungshierarchie (dasgestellt am Beispiel der phylogenetischen und ontogenetischen Entwicklung des Beutefangs von Raubtieren. *Z. f. Tierpsychol.*, 22: 412-494.
- Limbaugh, C. 1961. Cleaning symbioses. *Scient. Am.*, 205: 42-49.
- Limbaugh, C., H. Pederson, and F. Chase, 1961. Shrimps that clean fishes. *Bull. Mar. Sci. Gulf Caribb.*, 11: 237-257.
- Lorenz, K. 1966. *On Aggression*. Harcourt, Brace and World, N.Y.
- Luce, R. D., and H. Raiffa. 1957. *Games and Decisions*. Wiley, N.Y.
- Lutzker, D. 1960. Internationalism as a predictor of cooperative game behavior. *J. Conflict Resolution*, 4: 426-435.
- Marler, P. 1955. The characteristics of certain animal calls. *Nature*, 176: 6-7.
- . 1957. Specific distinctiveness in the communication signals of birds. *Behavior*, 11: 13-39.
- Marshall, L. K. 1959. Marriage among !Kung Bushmen. *Africa*, 29: 335-365.
- . 1961. Sharing, talking and giving: relief of social tension among !Kung Bushmen. *Africa*, 31: 231-249.
- Maynard, E.C.L. 1968. Cleaning symbiosis and oral grooming on the coral reef. In P. Person (ed.), *Biology of the Month*, p. 79-88. Philip Person, American Association for the Advancement of Science, Wash., D.C.
- Maynard Smith, J. 1964. Kin selection and group selection. *Nature*, 201: 1145-1147.
- . 1965. The evolution of alarm calls. *Amer. Natur.*, 99: 59-63.
- Mayr, E. 1963. *Animal Species and Evolution*. Belknap Press, Cambridge.
- Montagu, F.M.A. 1968. *Man and Aggression*. Oxford University Press, N.Y.
- Murie, A. 1944. *The Wolves of Mount McKinley*. Fauna of National Parks, Fauna Series #5; Wash., D.C.
- Owen, D. F. 1963. Similar polymorphisms in an insect and a land snail. *Nature*, 198: 201-203.
- Perdeck, A. 1958. Two types of orientation in migrating starlings, *Sturnus vulgaris* L., and chaffishes, *Fringilla coeleus* L., as revealed by displacement experiments. *Ardea*, 46: 1-35.
- Pruitt, D. G. 1968. Reciprocity and credit building in a laboratory dyad. *J. Personal. Soc. Psychol.*, 8:143-147.
- Randall, J. E. 1958. A review of the Labrid fish genus *Labroides* with descriptions of two new species and notes on ecology. *Pacific Science*, 12: 327-347.
- . 1962. Fish service stations. *Sea Frontiers*, 8: 40-47.
- Rapoport, A., and A. Chammah. 1965. *Prisoner's Dilemma*. University of Michigan Press, Ann Arbor.
- Rapoport, A., and P. Dale. 1967. The "end" and "start" effects in iterated Prisoner's Dilemma. *J. Conflict Resolution*, 10: 363-366.
- Rasmussen, K. 1931. The Netsilik Eskimos: social life and spiritual culture. Report of the Fifth Thule Expedition 1921-1924, Vol. 8(1,2). Gyldendalske Boghandel, Copenhagen.
- Rettig, S. 1956. An exploratory study of altruism. *Dissert. Abstr.*, 16: 2220-2230.
- Reynolds, L. 1966. Open groups in hominid evolution. *Man*, 1: 441-452.
- Rosenhan, D. 1967. *The Origins of Altruistic Social Autonomy*. Educational Testing Service, Princeton, N.J.
- Rousseau, J. J. 1954. *The Social Contract*. Henry Regnery Co., Chicago.

- Rudebeck, G. 1950. The choice of prey and modes of hunting of predatory birds with special reference to their selective effort. *Oikos*, 2: 65–88.
- . 1951. The choice of prey and modes of hunting of predatory birds, with special reference to their selective effort (*cont.*) *Oikos*, 3: 200–231.
- Sawyer, J. 1966. The altruism scale: a measure of cooperative, individualistic, and competitive interpersonal orientation. *Am. J. Social.* 71: 407–416.
- Schopler, J., and V. T. Thompson. 1968. The role of attribution process in mediating amount of reciprocity for a favor. *J. Personal. Soc. Psychol.*, 10: 243–250.
- Scott, J. P. 1958. *Aggression*. University of Chicago Press, Chicago.
- Southern, H. N. 1954. Tawny owls and their prey. *Ibis*, 96: 384–410.
- Struhsaker, T. 1967. Social structure among vervet monkeys (*Cercopithecus aethiops*). *Behavior*, 29: 83–121.
- Tesser, A., R. Gatewood, and M. Driver. 1968. Some determinants of gratitude. *J. Personal. Soc. Psychol.*, 9: 232–236.
- Thomas, E. M. 1958. *The Harmless People*. Random House, N.Y.
- Tinbergen, L. 1960. The dynamics of insect and bird populations in pine woods. 1. Factors influencing the intensity of predation by song birds, *Arch. Neerl. de Zool.*, 13: 265–343.
- Tinbergen, N. 1968. On war and peace in animals and man. *Science*, 160: 1411–1418.
- Ugurel-Semin, R. 1952. Moral behavior and moral judgment of children. *J. Abnorm. Soc. Psychol.*, 47: 463–474.
- Van Lawick-Goodall, J. 1968. A preliminary report on expressive movements and communication in the Gombe Stream chimpanzees. In P. Jay (ed.), *Primates*, p. 313–374 Holt, Rhinehart and Winston, N.Y.
- Wallace, J., and E. Sadalla. 1966. Behavioral consequences of transgression: the effects of social recognition. *J. Exp. Res. Personal.*, 1: 187–194.
- Wickler, W. 1968. *Mimicry in Plants and Animals*. McGraw-Hill, N.Y.
- Williams, G. C. 1966. *Adaptation and Natural Selection*. Princeton University Press, Princeton, N.J.
- Wright, B. 1942. Altruism in children and the perceived conduct of others. *J. Abnorm. Soc. Psychol.*, 37:218–233.
- Wynne-Edwards, V. C. 1962. *Animal Dispersion in Relation to Social Behavior*. Hafner, N.Y.



Postscript

About the time that my reciprocal altruism paper was published I ran into a friend at the Museum of Comparative Biology at Harvard, Allen Greer, an Australian herpetologist who was a couple of years ahead of me as a graduate student. He asked me what I was up to now that my reciprocal altruism paper was done. I told him parental investment and sexual selection. He said, “No, no—that’s not the way science is done. You should now perform an experiment or two on reciprocal altruism and then later write

a review paper, maybe hold a conference. You should make the area your own and stay on top of it, so to speak." My reply was that I was not really interested in reciprocal altruism per se but in social theory based on natural selection, of which reciprocal altruism was one part. In any case, I am certainly glad that I never took the road he suggested, because I believe that I have not had a fresh thought on the subject since! On the other hand, I have enjoyed a response to my paper that has greatly enriched the subject both empirically and theoretically.

One of the most agreeable features of writing a paper that is widely read by others is that, soon enough, you get an empirical and/or theoretical response, which enlarges your own understanding of the subject. The empirical response was almost immediate in coming. I had wanted, in my original paper, to include the possibility of baboon reciprocal altruism because I had learned from my mentor, Professor DeVore, about his famous "central hierarchy." In baboons it was known that adult males in a group arrived separately from elsewhere and thus are unlikely to be closely related, yet three or four adult males often seem to act as a unit against other adult males in certain kinds of interactions. Support is especially dramatic when one of the central males is in consort with a single female and this relationship is challenged by a male who may be individually more dominant but who is outside the so-called central hierarchy. In this case another member of the central group may rush in and help the male in consort retain the consortship. But there was, in fact, no evidence of reciprocity, only the supposition that such reciprocity was likely in nature, so the situation seemed too weak to merit even comment in my paper. But in 1977 Craig Packer published evidence that in baboon troops one male's frequency of soliciting another male's help in such situations was positively correlated with the other male's tendency to solicit help from him. This did not quite demonstrate reciprocal altruism, but rather reciprocal solicitation toward altruism, but that was a very welcome start. (For the most recent work on reciprocal altruism in primates, see de Waal 1997a, b, and de Waal and Berger 2000; for revenge in primates, see Aureli et al 1992; for reconciliation, see Aureli and van Schaik 1991.)

Similar advances came soon enough as well. Leigh (1984) showed that hermaphroditic sea bass were reciprocal in their couplings. When they mate one acts as a female, releasing eggs, and the other acts as a male, releasing sperm. Since the latter is much less expensive than the former, selection would favor hermaphrodites that spend most of their time being males, if they can get away with it. Instead, reciprocal egg trading has evolved. Fish that have as many as two hundred eggs to spawn in an afternoon release them not all at once but instead in small numbers, say, three to five at a time, and wait for the reciprocal act before continuing the relationship. A nice feature of this work was that two species were compared that differed in their degree of reciprocity. One had a high tendency to reverse roles, and

the other had a significant tendency to reverse roles but not as often as the first. Reciprocal egg trading permits the hermaphrodite over evolutionary time to save energy on investment in testes and, sure enough, both species have smaller testis/ovary size ratios than do mass spawners, and the more reciprocal species has the smallest testis/ovary size ratio.

Axelrod and Hamilton (1981) Arrives by Mail

In 1981 I was a professor at the University of California and somewhat out of it in more ways than one. I was not keeping up with the new journals, and so it came as a welcome surprise to receive one day in the mail a reprint of Axelrod and Hamilton (1981) from Hamilton. The paper was inscribed, "To Bob, Tit-for-Tat or Hamilton's Revenge, Bill." The paper described how a strategy for playing iterated games of Prisoner's Dilemma could win out against competing strategies under very robust conditions. This strategy was called tit-for-tat. In a computer tournament that Axelrod had run, with strategies submitted by invitation, the simplest strategy had turned out to be the winner. Cooperate on the first move and do whatever your partner did on the previous move. So, if your partner also cooperated you would continue to cooperate, but if your partner failed to cooperate you would then cease cooperating yourself, tit-for-tat. My first reaction to the inscription was that this paper was the tit for the tat of Trivers and Hare (1976) (see chap. 5). That is, I had stolen a portion of Hamilton's thunder on the haplodiploid Hymenoptera and he was now returning the favor on reciprocal altruism. But Bill has a very subtle mind and my first wife, Lorna, was the first to draw my attention to the frequency with which Bill's statements had double and even triple meanings. In this regard, his mental operations reminded me of the dreaded knight on the chess board, which moves in a forking motion, simultaneously attacking to the left and the right. Earlier I had on the one hand prevented Bill from publishing his first formulation of game theory applied to reciprocal altruism (when he asked permission to include it in his "Men and Beast" paper), and then had turned around and stolen it back (to be sure while citing Hamilton), and this was my tat for that tit, as well. In any case, I sat down at around eight o'clock at night, turned on some classical music, read the paper, and, as I later wrote Bill, "my heart soared." For one wild moment, I kidded him, I actually believed that there was progress in science! What Axelrod and Hamilton had done was to *prove* that in these repeated games of the Prisoner's Dilemma, there were only two evolutionary stable strategies, tit-for-tat and perpetual defection, or noncooperation. To me the paper had almost biblical proportions. That is, you could see how a kind of social heaven and social hell could evolve right here on earth. The social hell was perpetual isolation, perpetual inability to link

up with others in a positive way, never being cheated by others to be sure, but at the cost of eternal loneliness. The social heaven was not heavenly in some naïve way, dancing around the mulberry bush together without regard for selfish possibilities. Instead, cooperation required perpetual vigilance to enjoy its fruits, but tit-for-tat, a very simple mechanism that could apply even to bacteria, could bring about this cooperative world. Even the first sentence of the paper set the subsequent tone: "The benefits of life are disproportionately available to cooperating creatures."

It was the simplicity of the rule that was so beguiling. When Axelrod had first held his computer tournament, he solicited an entry from me. That is, I was offered the opportunity to send in my own strategy, which would then compete against strategies submitted by others, the computer doing the dirty work. I am embarrassed to say that my first reaction was that the matter would be complicated and that some kind of complex formula for responding to the play of your partner would succeed. And, since at the time I also lacked any ability to work on a computer, the invitation to submit my entry in computer form also stood as a barrier. So I sent in nothing.

Axelrod and Hamilton, of course, had the same theoretical bias that I had, which was to try to state the argument in as general a form as possible and to apply it as broadly in nature as one could. The simplicity of the tit-for-tat strategy bypassed, in one step, the cognitive complexity that was often assumed to be required to get reciprocity going in our own species. A single-celled organism, as they pointed out, that could respond to chemicals produced by neighbors was, in principle, in a position to perform a tit-for-tat strategy, producing a cooperative chemical on the first move and an uncooperative one in response to an uncooperative one from the neighbor. There was, however, one moment in the paper where they pushed the argument one step beyond what I could then envision. Since this involved genetics, it almost certainly came from Hamilton (so I will now write as if it did). He argued that the paired chromosomes in germ cells such as oocytes might be able to respond to a selfish maneuver by returning a selfish maneuver in kind. In particular, he was thinking about meiotic drive in which one chromosome gains an advantage in reproduction at the expense of its paired chromosome. In this case, one chromosome may cause itself, for example, to land in the egg cell instead of the polar body at the relevant division, improving over the 50/50 chance that the fair rules of meiosis were supposed to produce. If I understood them right, they were arguing that if this defection occurred in one oocyte, a paired chromosome in a neighboring oocyte might be "aware" of it and do the same itself. This could easily result in both chromosomes ending up in the egg cell with the unfortunate result that the offspring would be trisomic, that is, have three copies of the chromosome (such as is found in Down syndrome, trisomy 21). I could not imagine that the neighboring oocyte could get the information that this argument required.

I remember the sensation vividly as if it were yesterday. It was as if Ham-