

## THE ROOTS OF COOPERATION

*Hanna Newcombe*

There are six roots of cooperation in human societies: kinship, reciprocity, contract, utility, equity, and universality. These roots range from minimal to maximal in motivation, but it is interesting that even at the reciprocity level, cooperation can be firmly and robustly established in a group.

Theoretically, the kinship principle is rooted in sociobiology, the wish to help the survival of one's genes ("extended fitness"). I share approximately half the genes with my parents and my children, a high proportion with my siblings, less with my cousins, uncles and aunts, nieces and nephews, much less with other members of my tribe, hardly any with foreigners. I will therefore practice most altruism and cooperation with the closest members of my family, somewhat less with members of the extended family, only limited cooperation with the rest of the tribe, and none (or even hostility) to strangers. This principle produces cooperation only in a limited group, not universal; concentric circles, like a Mandala, bright at the centre and quickly shading through greyness into the outer darkness. On this is based family solidarity, and in the larger realm tribalism, patriotism and nationalism. "I against my brothers; I and my brothers against our cousins; my brothers and my cousins against outsiders."

There are some difficulties in this. The

family application may be clear; the biological lines of similarities are obvious and easily traced. But a tribe may have received some immigrants (it certainly has by exogenous marriage), and no nation on Earth is racially pure. The application to nationalism is then definitely suspect, and based more on artificially created fiction than on natural reality. The "purpose" (excuse the teleological abbreviation) of sexual reproduction is, after all, to churn up the gene pool, not to leave the lines of descent clear and distinct. And the human gene pool has certainly been churned up, through immigration, refugees, travel and commerce (increasingly in modern times), and also through war (distant armies) (as someone quipped "soldiers drop their genes").

Secondly, the proportion of genes shared with relatives (one half, one quarter, or whatever) are not right. This applies only to the highly variable regions which defines the "self" in immunology; but after all, all humans share by far most of their genome universally, because those parts define the species rather than the self. The truth is that I share 99.9% of the genes (or maybe I am thinking of DNA base sequences) even with a stranger, and maybe 99.99% with my sister; the whole proposition then becomes far less meaningful. If I truly want to preserve my genes, I must practice altruism

Hanna Newcombe

to the whole human race. And I also share maybe 80% of my genes with other mammals, and so on down the line of all creation. This other brand of sociobiology would lead straight into universality - but I am getting ahead of myself.

The second root of cooperation is reciprocity, which is social rather than biological. The idea is when I do someone a favour, I expect the favour to be returned at some future time, and this is why I do it; and if someone does me a favour, I feel obligated to return it at the next opportunity. The other face of reciprocity is that when someone is nasty to me, I will be nasty right back, and vice versa, I expect such punishment from others, which may deter me from being nasty. Reciprocity has therefore been called "tit for tat", and Robert Axelrod has written a wonderful book about it, "The Evolution of Cooperation".

Axelrod's book deals with a "game" situation called Prisoner's Dilemma (PD), which simulates the very common social situations in which "defection" (non-cooperation) pays off in the short run (like telling a lie or cheating), but if both participants defect, they receive less benefit than if both cooperate (tell the truth and practice honesty); but the biggest loser is the player who cooperates while the other defects. Thus there is a temptation to defect in the short run, but if players expect to meet repeatedly over the long run, it pays to cooperate.

Axelrod arranged the encounters between

different strategies of playing PD (supplied by experts in the field) in a computer round-robin tournament, in which every strategy would meet every other strategy many times. It turned out, in repeated experiments, that the simplest strategy submitted, tit for tat, won (got the most points) every time. This strategy never defects first (is "nice"), always retaliates immediately after the other player's defection (is "provocable"), but retaliates only once and then returns to cooperation if the other player does (is "forgiving"). (It is taking "an eye for an eye" to signal that it will not be exploited, but never takes two eyes for one eye, i.e. does not escalate.) It is also easy to figure out by the other player (is "clear"), so that mutual learning can take place.

Axelrod also performed repeated rounds of the same tournament, in which low-scoring strategies (receiving less than a predetermined number of points) were eliminated at the end of each round, while the high-scoring strategies (receiving more than a certain upper limit of points) were allowed to "reproduce", i.e. be represented by multiple copies in the next round of the tournament. This experiment simulated the evolution of the various strategies, with the unfit going extinct and the fittest showing reproductive success. Some strategies went to the wall almost immediately; mainly the ones that were not "nice" (i.e. defected first), because they lost too many points in being punished by the retaliators. So greed does not pay. Some other strategies increased for a while but then decreased, because the strategies against which they did

## The Roots of Cooperation

well had gotten eliminated. Again, tit for tat was the most robust survivor. (Anatol Rapoport, who submitted this strategy to the contest, explains that tit for tat does not really "win" over any other strategy; in fact, it cannot, because it only follows the other's lead. But the other strategies that can win, kill each other off, and meek little tit for tat inherits the earth. Or as Axelrod puts it, tit for tat does not win any single rounds in the sense of getting more points than the other player, but excels in eliciting cooperation from the other player, enabling both to win points. After all, PD is a non-zero-sum game, not pure conflict. You don't have to win OVER the other as in chess. But life is more like PD than like chess, most of the time.)

Axelrod derives certain interesting propositions:

1. If the discount parameter  $w$  (which measures the frequency or the probability of the two players meeting again in the future) is large enough, there is no best strategy independent of the strategy used by the other player. (In other words, it depends on the social environment in which you find yourself.)
2. Tit for tat is collectively stable (i.e. it cannot be invaded by any other strategy once it is firmly established so that most players use it), ONLY IF  $w$  (the probability of players meeting again) is large enough. In the short term, when you don't expect to meet again very much, it is best to play all-D (defect on every move), i.e. "war of all against all".
3. The above may apply to other strategies that are "nice", i.e. cooperate on the first

move and are never the first to defect. Again, the expectation of meeting again has to be large enough.

4. For a "nice" strategy to be collectively stable, it must be provoked to retaliation by the very first defection by the other player.
5. The "All-D" strategy is always collectively stable, for any value of  $w$  (i.e. whether we meet again or not). This means that a single copy of any other strategy, including tit for tat, could not successfully spread in such an environment and would be eliminated. An isolated cooperator cannot get a foothold in a society of "meanies". Nor can several isolated cooperators.
6. However, some strategies can invade All-D if they are in a cluster, i.e. if they interact more frequently with each other than with the more numerous "meanies", so that they can give each other some winning points. The strategies which can do this cluster invasion of All-D most easily and quickly are ones that are maximally discriminating, i.e. cooperate with their own kind and retaliate against the "meanies". Tit for tat is a maximally discriminating strategy, and can therefore successfully invade All-D in clusters.
7. If a "nice" strategy, such as tit for tat, cannot be invaded by a single individual copy of another strategy, then it cannot be invaded by a cluster of such strategies either. In particular, that means that a world of tit for tat players (acting as cooperators) cannot be invaded by All-D, either single or in clusters. Tit for tat is "robust", in addition to all its other properties (nice, provokable forgiving, clear).

I have given a rather lengthy summary of

Hanna Newcombe

this because Axelrod's work is not too well known yet, but it deserves to be. Summarizing the summary: In a persistent long-range interaction situation (as in a society), "nice" but "provocable" (though "forgiving") cooperation is collectively stable. A Hobbesian situation of "war of all against all" is also collectively stable, in both short- and long-lasting interactions; while it cannot be invaded by single cooperators, it can be successfully invaded by clusters of cooperators, which in time will transform the Hobbesian situation into a cooperating society. Once established, such a cooperating society cannot be successfully invaded by either single "meanies" or clusters of "meanies".

It should be remembered that this result was obtained with Prisoner's Dilemma, a game whose "rational" outcome for egoists (individual utility maximizers) is mutual defection (All-D). If anything, the experiment was biased against cooperation, and yet cooperation emerged, was able to establish itself and grow, and then resist being displaced. This is because mutual cooperation is in the long run more advantageous, even to egoists. The overall conclusion then is that cooperation can grow and be robust even in a world of egoists without central authority. Thus Hobbes' solution of the Leviathan is not necessary to cure the Hobbesian dilemma of the war of all against all. There is a "natural" (as opposed to coerced) tendency to cooperation in human societies (at least in computer simulations), as the Anarchists have always maintained.

Thus reciprocity is a powerful mechanism for cooperation, more subtle and convincing than the biologically-based kinship model. However, there is an interesting example which illustrates the simultaneous operation of both kinship and reciprocity. This is the mutual feeding of female vampire bats, as described by Gerald Wilkinson in *Scientific American*, February 1990. These bats have to obtain a blood meal from horses or cows each night - if a bat fails to obtain a meal even two nights in succession, she starves to death. A bat who has missed a meal begs another bat to share hers, by regurgitation. The benefactor bat may comply or not; it is an act of altruism on her part, because she increases her risk of starvation if she fails to obtain a meal the next night; but she is more likely to do it if either the begging bat is a relative (especially her own child, where this is done routinely), or if the two bats live in the same tree colony and therefore interact continuously over time, so that reciprocation may be expected in the future. Reciprocation in meal sharing then increases the survivability of both cooperating bats.

The third root of cooperation is social contract, usually applied and administered by what Axelrod calls "central authority", i.e. government. Although cooperation can develop spontaneously by reciprocity, as just shown, it is useful to confirm and legitimize it by an explicit contract. Contractarian theories (e.g. Hobbes, Locke, Rousseau, Rawls) deal in terms of a sociological fiction: that there was a time when humans lived together without a social contract; that they realized that all would benefit if there

### The Roots of Cooperation.

was a contract; and that they came together to negotiate the terms and created and ratified a contract. I think the authors all realize that this is not a historical account of events, but a convenient "creation myth". In reality, an implicit contract may have been dimly in the consciousness of a few leaders, who then started gradually enforcing it by their own absolute power based on force, either sheer military force or the authority of religion. Enlightened tyrants of this type would enforce a contract that would benefit their whole society, not just themselves. They would realize that in the long-range game of politics (if it is a Prisoner's Dilemma) it pays to cooperate with one's subjects as long as they cooperate, rather than exploit them mercilessly, because they might revolt. But never mind what REALLY happened; we are interested in the contractarian fiction, because it clarifies conceptually what a social contract involves.

In this fiction, humans "in the wild" ("in a state of nature") are either intent on continual mutual robbery, murder and mayhem (as Hobbes would have it), or are "noble savages" who trust and cooperate (according to Rousseau). In either case, they are assumed to be rational, i.e. capable of making decisions to improve their condition. They decide (one and all) that they wish to form a society, and are seeking some rational rules by which that society will operate. They act "behind a veil of ignorance", in the sense that they don't know what position in the society each would get to occupy - powerful or lowly, rich or poor. They therefore strive to make it "fair" for everybody, because it is in their

interest to do so, in case they end up on the bottom rung of the pecking order. The strive for "justice" in this sense of "fairness", according to John Rawls; yet not necessarily equality, because they want to balance equality with freedom, and also inequality might provide more incentives to work hard and raise the standard of living. (I find it amusing to visualize these noble or ignoble savages as already so knowledgeable about capitalist theory - budding Reaganites and Thatcherites before their time.)

In any case, the original society as created by the contract does not yet have a government; that is the next step. After that primordial contract fixes the underlying ideology (egalitarian or not; libertarian or not), the founders must then write a constitution to describe the detailed structures of government. The constitution is fundamental with respect to the particular legislation that will follow, but the social contract itself is even more fundamental than the constitution. According to John Locke, "civil society" pre-dates government. This has recently become the inspiration of the non-violent revolutionaries of Central and Eastern Europe in the fall of 1989. After having been crushed by a very top-heavy government for 40-45 years, they now want more power to go to spontaneous citizen organizations like Civic Forum in Czechoslovakia. The Helsinki Citizens Assembly wants this to prevail throughout the 35 nations of the Conference on Security and Cooperation in Europe (Helsinki Agreement). I can only wish them well.

There are different versions of the social

Hanna Newcombe

contract in contractarian theories. Hobbes considers even a tyrannical "Leviathan" to be preferable to "the war of all against all", which it well might be - except that I doubt that that particular "state of nature" ever existed in history or nature. As somebody remarked, if humans were so intent on killing each other off, they would have done it long ago and we would not be writing about it.

Rousseau considered his initial society as being "close-coupled", as I would now call it. That is, the individual wills of citizens are assumed to have merged into "the general will"; and in voting, each citizen will be trying to second-guess what that general will is and vote accordingly, rather than defending his own selfish interests. Rousseau's society is thus diametrically opposed to an alternative model, in which special interests (e.g. farmers, workers, industrialists, professionals, etc.) organize, and the collective decisions that emerge reflect their relative strength or power or clout - through some process of hard bargaining and mutual trade-offs. Rousseau's model seems more idealistic, but as so often happens with utopian ideologies, it goes sour in practice.

The General Will is too centralizing, too oppressive; it becomes the object of idolatry. In the French Revolution it was most prominently embraced by the Jacobins, and it led to the Terror. Any dissenter from the revolutionary state was obviously not attuned to the General Will, and was therefore put to death as an enemy of society. This is a familiar pattern to us in

the 20th century for various brands of totalitarianism, both left and right. Rousseau thought that he was writing a blueprint for a perfect democracy and total social harmony, but the human tendency to fanaticism when fired up with an idea led straight to absolute tyranny.

There was an alternative, even then, at the end of the 18th century. The Girondins represented federalism and the diffusion of power. The French Revolution could have taken a different turn. Closer to our own time, Stalin did not have to follow Lenin in the Russian Revolution; it could have been Trotsky. Yet I have some doubts that the Terror in both cases was the result of the depravity or fanaticism of one man, Robespierre or Stalin; it is more likely that the fault lies in the ideology itself, and the fanaticism it arouses. Trotsky would not have been any less oppressive, I would think.

So one has to be very careful about the social contract. It might turn out to be Leviathan after all. But there are more benign versions. I shall look at only one more, as presented in John Rawls' book on "Justice". There are two parts to the social contract in this version. In the first part, people agree to maximize individual freedom, i.e. to permit every member of society to do as he or she pleases as long as it does not infringe on the freedom of others. (This is the same principle as that expounded by J.S. Mill in "On Liberty".) In the second part, only as much inequality is permitted as would leave even the most disadvantaged member of society no worse

## The Roots of Cooperation

off than he/she would be under conditions of equality. The assumption here is that some inequality is needed to spur production to increase the total wealth of society. So with a sufficient amount of inequality to provide incentives, the overall size of the "pie" will be greater. The poorest will receive a smaller slice of the pie because of the inequality; but the principle stipulates that this slice that goes to the poorest must be greater than their larger proportion of the smaller pie would be under equality. This is why, under the veil of ignorance, every member of the original contract would choose this limited degree of inequality over complete equality.

This model is thus constructed along the lines of liberal capitalism and is obviously under the influence of that ideology. Justice is seen as a compound of freedom and equality, as is quite common in political ideologies; but freedom and equality can form compounds in different proportions, like carbon and oxygen forming either CO or CO<sub>2</sub>, and this gives us the basis of different ideologies, like capitalism or socialism. Rawls opts for one possibility which has a certain appeal; it is not unbridled capitalism because the degree of inequality permitted is limited, and there is concern for the welfare of the poor. But the thesis of Adam Smith that only selfish incentives will be a spur to production is accepted without question; and the benefit of more production (i.e. growth) is also implicit. But it is now questionable whether growth beyond a certain point produces more happiness, or whether it rather depletes resources and pollutes the

environment. All these considerations are not included by Rawls, and therefore his conception of justice remains just one of a series of alternatives, not necessarily only one or the best one.

The fourth root of cooperation, utilitarianism, gave its name to Utilitarianism, the theory of the greatest happiness of the greatest number, elaborated by Jeremy Bentham and John Stuart Mill. But first I must give an introduction to the last three roots, because in going from Root 3 to Root 4, we are crossing the boundary from theories based on selfishness to theories placed on an ethical basis. From now on we do not necessarily ask "What's in it for me?" though we do not abandon self-interest altogether. There was already a beginning of this in contractarianism, in the concept of the General Will, but now it will become more pervasive.

Why should any of us consider the interests of others, for their own sake, not just for reasons which benefit ourselves (such as kinship, reciprocity, or contract)? The answer is not easy to express in words, though I feel it strongly, almost like an axiom in geometry. (Of course it is not an axiom, because some people do not find it to be self-evident.) It is this: other people, all people, have the same essence that I have, consist mentally and spiritually of the same "stuff" as I, and therefore deserve the same care and consideration that I lavish on myself. Just to have a name for it, I will call it "the Principle of the Shared Essence" (PSE). Quakers might say "There is that of God in every person", but I don't

Hanna Newcombe

want to make this dependent on religion.

Returning to utilitarianism: in saying that the aim of our actions should be the securing of the greatest happiness of the greatest number of people, we have widened the meaning of "utility" from the selfish utility maximization of the "rational player" of Prisoner's Dilemma (or of Adam Smith's "economic man") to the maximization of collective utility. If people played PD in this frame of mind, they would cooperate automatically, without any dilemma, and Axelrod's elaborate mechanisms would be unnecessary, as would any contract or central authority. This would be part of Level 6 on Kohlberg's scale of moral judgment, namely principled conscience. (The lower levels, starting with 1, are fear of punishment, seeking approval, conformity, respect for law and order, and social contract. Some of these may correlate with the "roots" in our previous discussion, e.g. conformity with reciprocity, social contract with social contract.)

Some precautions are needed in fostering other people's happiness. The main one is that we should not give them what we think is good for them, but what they themselves tell us they want for their own happiness. It would be no use giving classical records to someone who likes rock and roll, or serve a nice juicy steak to a vegetarian. My friend Don Bryant, who has done a lot of thinking and writing about the philosophy of altruism, recommends leaving a portion of our "sphere of interests" empty (as a "crypt of disinterest") for other people to fill - only this would be true altruism.

However, there is a more fundamental problem. In striving for the greatest happiness of the greatest number, we are trying to maximize the sum of everybody's utilities. I will leave aside the vexing problem of whether individual utilities can be intersubjectively compared or added together; this problem is rather abstract. I am more concerned with a practical problem: I may be able to maximize the sum by taking away some happiness from one person in order to give a greater amount of happiness to another person; but is this "fair"? Or in the extreme case, I may be able to increase the security of the state by sending a few recalcitrants to the guillotine - and here we go again down the slippery slope to terror and tyranny. For the sake of greatest happiness, mind you! Something is terribly wrong.

The trouble is that maximizing the sum of everyone's utilities is not enough, we must also have some rules about the distribution of happiness among the people. We might add some absolute prohibitions, in the form of a few core human rights: freedom from extra-judicial or arbitrary killing, freedom from torture and cruel treatment, freedom from arbitrary arrest and detention, freedom from hunger, freedom from discrimination; or my even shorter list: no genocide, no torture, no disappearances, no death squads, no starvation, no apartheid. The trouble with lists is that you are apt to forget some important item. And it is no use lengthening the list too much either; the Universal Declaration of Human Rights has 30 articles, but not all can be fulfilled in practice, especially by very poor countries.



## The Roots of Cooperation

Alternatively, we might proclaim that human life is sacred and must never be touched - an absolute tabu. But we would run into problems in medical ethics: which patient will get the organ transplant when these are scarce, which patient will get scarce resources like a kidney machine, and will a patient be allowed to die without heroic measures if he/she wants to? In practice, there must be some trade-offs sometimes.

Another approach might be to say that we should never take an action that makes even one person even a little bit less happy than before; we are allowed only to take actions that increase someone's happiness while leaving others at least the same as before. But this may be too stringent; if we are on a sight-seeing tour and my two-year-old starts to cry that he wants to go home, I should not spoil everyone else's enjoyment by insisting that the tour bus go back at once; I bundle up the child and take him along anyway, knowing that he will soon forget the incident and be happy again.

We need some measure of "fairness" or "equity" to supplement our utility principle. And this takes us to the fifth root of cooperation which we call "equity". Note that we are not replacing the maximum utility principle, but adding to it. The sense of equity as fairness is actually very commonly observed; people play some non-zero-sum games not as selfish maximizers exclusively, but they do sometimes what would be fair to the other player, especially if it does not cost them much. And children at play are constantly shrieking "that's not

fair", assuming that the other child **know** what that means.

Under the utility principle, the rule **was** to maximize the sum of all utilities. **What** we now need to do under the **equity** principle is to minimize the difference between individual utilities. We do not **say** that everyone's utilities should be equal which probably cannot be done - but that **we** should strive to minimize the differences. For example, someone with a disability, let us say in a wheelchair, will be less happy than a walking person (if other things are equal); so we try to even things out for the disabled person by providing wheelchair access to all buildings, even though this costs taxpayers some money.

Now, maximizing the sum of two numbers while simultaneously minimizing their difference means that we maximize the product. Try it:  $1+9$ ,  $2+8$ ,  $3+7$ ,  $4+6$ ,  $5+5$  all add up to 10; but the products are 9, 18, 21, 24, 25. The largest product is obtained when the difference is smallest (For the same reason, of all quadrilaterals with the same perimeter, the square has the largest area.) Though we will not prove it here, the same rule applies when we are adding or multiplying more than two numbers. Maximizing the product is associated with "the Nash solution" in economics and decision theory, and this is what we are adopting here.

The sixth and final root of cooperation is universality, which simply means including everybody in the calculations of utility and equity, not only those "playing the game a

Hanna Newcombe

the time", i.e. narrowly and locally concerned. Universality is already implied in the PCE (Principle of Common Essence) - ALL humans are made of the same stuff, and deserve the same respect, as the present players in a situation. But PCE is also strongly implied in equity (the innate sense of fairness) and in collective utility. PCE encompasses all three of the higher roots of cooperation.

We have discussed universality as extending to all humans, but it should also be extended to all living creatures, with appropriate discounting factors, because their "essence", while similar to ours, is not quite identical. The discounting factors would differ from one species to another. They might depend on the degree of their consciousness or intelligence (except that we only guess at this and don't know it), or the degree of relatedness to us (sociobiology again). We have to leave these concepts a bit vague, but insist that all living creatures be included in some way in our utility calculations, in order to satisfy the requirement of universality.

With universality capping collective utility and equity, we have a PCE-based complex which can be called Agape (universal love). It is "rational agape" in the sense that we do not emotionally love all those people out there - we cannot, because we have not even met them. But we do recognize their claim to life and to quality of life as being just as valid as our own. We do this rationally and out of a sense of fairness. This has been expressed in various moral rules, such as "Do onto others what you

want them to do onto you"; this is the Christian wording of the Golden Rule, but the same thought has been formulated by all the higher religions. Immanuel Kant's "categorical imperative" ("so act that you would be willing to have everyone act that way") is a way of explicitly universalizing the moral value of your actions. Kant also said "Never use any human being as a means; always treat him/her as an end". Martin Buber would add "Every human being is a Thou, not an It", which is an admirable way of stating the Principle of Common Essence. J.S. Woodsworth, the founder of the CCF party in Canada (now the NDP) said "What we want for ourselves, we want for all." This is the same principle as Gandhi's Sarvodaya, which means "the welfare of all". Needless to say, any of these rules would abolish the dilemma in PD without any "strategies" being necessary. As guides to action, they all simply say "cooperate".

However, rational agape can be further intensified if we add empathy, meaning that we rejoice when others are happy and mourn when others suffer. This is a feeling, not reason. Most people are genuinely saddened or outraged or moved to action (helping behaviour) when they see television pictures of starving Ethiopians. People's spirits are lifted in watching a joyous celebration with music and happy faces. We can summarize this by saying that rational agape plus empathy equals integral agape, the highest motivation for cooperation that we can have. It involves not only reason, but feeling and will as well - in fact, our whole being. It would be Kohlberg's stage

The Roots of Cooperat:

7, which Kohlberg did not even describe. As Paul said "The greatest of these is Love."