Nicholas Humphrey, 1997, “Varieties of altruism – and the common ground between them”, *Social Research*, 64, 199-209.

VARIETIES OF ALTRUISM – AND THE COMMON GROUND BETWEEN THEM[1]

Altruistic behaviour, where it occurs in nature, is commonly assumed to belong to one or other of two generically different types. Either it is an example of "kin selected altruism" such as occurs between blood relatives – a worker bee risking her life to help her sister, for example, or a human father giving protection to his child. Or it is an example of "reciprocal altruism" such as occurs between non-relatives who have entered into a pact to exchange favours – one male monkey supporting another unrelated male in a fight over a female, for example, or one bat who has food to spare offering it to another unrelated individual who is hungry.

The first kind of altruism was given a theoretical explanation by William Hamilton, who showed how a gene that predisposes its carrier to help a close relative can prosper in the population, provided the genetic relationship between the two individuals is such that the cost to the giver is more than made up for by the benefit to the recipient multiplied by the degree of relatedness.[2] The second kind was given an explanation by Robert Trivers, who showed how a gene that predisposes its carrier to help a like-minded friend can also prosper, provided the social relationship between the individuals is such that the costs to the giver are more than made up for by the benefits that can be expected to be received later in exchange.[3]

These two kinds of text-book altruism – the free gift of help to a relative, or the exchange of help between friends – have for twenty years now been considered biologically, conceptually and even morally distinct. In fact, so different are they supposed to be, that it has been claimed – on both sides – that only one of them should be counted genuinely "altruistic". Hamilton, for instance, in a recent review of the development of his own ideas has written: "My quest for biological altruism had carefully excluded anything I saw as possibly reciprocatory because it seemed that although behaviours of this category could mount a semblance of altruism, a donor always expected a benefit itself, at least in the long term: it was a semblance only. I still believe the reciprocal altruism that Trivers explained to me was misnamed".[4] Trivers, by contrast, in the opening lines of his original paper on reciprocation defined altruistic behaviour as behaviour occurring specifically between individuals who are "not closely related", and he played down altruistic acts within the family on the grounds that the altruist might "merely be contributing to the survival of his own genes." Models like Hamilton's, according to Trivers, were "designed to take the altruism out of altruism".[5]

Evidently the fathers of the two theories never doubted that the difference between them was a deep and important one. And, even with the moral question put aside, most later commentators have tended to agree. Although there is still some disagreement about the terms to use, almost everybody now accepts that there really are two very different things being talked about here: so that whenever we come across an example of helping behaviour in nature we can and ought to assign it firmly to one category or the other.

There have however been one or two dissenting voices. Stephen Rothstein, for example, in a little-known paper titled "Reciprocal altruism and kin selection are not clearly separable phenomena" argued that – for reasons I shall develop further in a moment – many examples of reciprocal altruism must actually involve some degree of kin selection.[6] He was building on an earlier idea of Richard Dawkins that two individuals who both carry copies of the same gene for altruism might be able to recognise one another by the very fact that they both tend to behave altruistically towards someone else – which in principle would allow them to promote their own genetic interests as altruists by selectively aiding each other (this being a special variant of what Dawkins called the "green beard effect").[7] Dawkins wrote of this merely as a hypothetical possibility. But Rothstein realised that in many cases of reciprocal altruism it may in effect be close to the reality, since reciprocal altruists do in fact make a point of choosing other reciprocal altruists as trading partners.

Clearly, if Rothstein is right, reciprocal altruism may actually shade into kin selection. In which case the distinction between the two cannot be anything like so absolute as we have been led to believe. But suppose, now, there were to be further arguments in the same vein. Suppose it could be argued that kin selection also shades into reciprocal altruism. In that case we might want to challenge the reality of the distinction altogether, and might even wonder whether it would not be best to start the theoretical discussion over again. It is the purpose of the present paper to propose just such a radical re-think. For I believe Rothstein's paper did indeed tell only half the story, and that, when the other side is included too, it becomes clear that we can no longer continue looking at the landscape of altruism in the way we have grown used to. Rothstein asked us to recognise that altruistic behaviour towards friends must in many cases end up benefiting shared genes: but, I shall now argue, it is just as important we should recognise that altruistic behaviour towards kin must in many cases end up bringing a return of benefits to the altruist himself.

The arguments – Rothstein's and this new one of my own – are mirror images of each other, and, at the risk of being pedantic, I shall lay them out one after the other so as to show the structural similarities. But I shall begin with the new argument about how kin selected altruism must often involve a degree of reciprocation, before giving my version of Rothstein's original argument about how reciprocal altruism involves a degree of kin selection.

Let us begin then by taking a new look at the case of kin selected altruism: the case where we are dealing with an individual who has a gene that predisposes him to give help to a relative. Hamilton's famous point here was that every time this individual helps his relative he is benefiting any copy of the altruistic gene that the relative himself may happen to be carrying. So that, provided the cost, **C**, to the altruist is sufficiently small, and the benefit to the recipient, **B**, is sufficiently large, and there is in fact a sufficient degree of relationship between the two of them, **r**, – provided, to be precise, that $\mathbf{C} < \mathbf{Br}$ – the altruistic act will have provided a net gain in fitness to the gene. Which is why, in many circumstances, the gene is likely to evolve.

Hamilton's point is, of course, both valid and important. Nonetheless I would now suggest that there has always been something nearly as important being ignored by any such simple analysis. Namely, that every time the kin selected altruist helps his relative he is also helping keep alive another individual, who, assuming he does have a copy of the altruistic gene, can be counted on to behave the same way towards *his* relatives - among whom is the original altruist himself. Hence the kin altruist, by helping his relative, is in effect increasing the pool of extant individuals from whom *he himself* may one day receive help.

Suppose, for example, that I have a gene that predisposes me to save my brother from drowning. Hamilton pointed out that in performing this altruistic act I have benefited any copy of my gene that is carried by my brother. But the new point is that I have in addition ensured the continued survival of someone who when the occasion arises is quite likely to save me from drowning. Therefore the pay-off to my fitness as a kin altruist will very likely come not only indirectly through the benefit to any copy of the altruism gene carried by my relative but also directly through the benefit to my own gene.

It is true, as Hamilton would no doubt have wished to stress, that if I actually go so far as to commit suicide in order to save my brother, I will then miss out on any future return of benefits directly to myself. The same must in effect be true if I am already so old that I cannot expect to live for long or to bear any more children in the future. In such cases the return of benefits to me will only be able to come as a proxy benefit to one or other of my surviving descendants. Thus if I save a brother much younger than myself from drowning, it may be not me myself but, for example, my son (his nephew) who stands to benefit from his survival. But even here there is still going to be the chance of a significant return to my own direct line of genes.

The chances of either the altruist himself or any of his descendants getting this return of benefits is, of course, bound to depend on whether the relative who has been helped stays around long enough in the vicinity to be in a position to do his own bit of helping if and when the need arises. But this is likely to be much less of a problem than it might seem to be, since the very fact that the relative has received the earlier help is bound to increase his loyalty to the place and context in which it happened, and thus increase the chances that he will stay close to the altruist and/or his descendants. The fact that my brother, for example, has been saved by me from drowning is bound to encourage him to maintain close contact with me and my family in future years.

Let's be clear that it is not necessary to suppose that any kind of reciprocal-altruism-like "bargain" is being struck in such cases. With these cases of kin altruism the original helpful act can unquestionably be justified on Hamiltonian grounds alone, even if it never does bring any return to the altruist himself. The altruist certainly need not have any "expectation" of getting anything in return, and the recipient need not feel under any "moral compunction" to return the favour. Nonetheless my point is that it will often so happen that the altruist *will* get the return.

Indeed maybe it will so often so happen that a major part of the cost of the original altruistic act will as a matter of fact get repaid directly to the altruist. In which case it means that Hamilton's equation setting out the conditions under which this kind of altruism can be expected to evolve – his $\mathbf{C < Br}$ – has always been unduly pessimistic. For, in reality, the true net cost, $\mathbf{C}$, of the original altruistic act will often work out in the long run to be much lower than at first it seems.

Now let us turn to the other side of the picture and take a closer look – as Rothstein did – at the case of the reciprocal altruist: the case where we are dealing with an individual who has a gene that predisposes him to help not a relative but rather a friend whom he trusts to pay him back. Trivers's famous point here was that every time this individual helps his friend he is adding to the stock of favours that are owed him. Hence, provided the cost of giving help to the friend in need is in general less than the benefit of receiving it when the altruist is in need himself, the exchange will have provided a net gain in fitness to the gene. Which is why, in many circumstances, this gene is likely to evolve.

Trivers's point, again, is true and important. But – and this was precisely Rothstein's argument – there has again been an important factor ignored by this analysis. Namely, that every time the reciprocal altruist helps his friend he is also, so long as he has chosen wisely, increasing the chances of survival of another individual who is himself carrying the gene for reciprocal altruism. That is to say, he is increasing the chances of survival of another individual who, by carrying this gene, is in *this* respect a relative.

Suppose, for example, that I have a gene that predisposes me to save someone whom I think of as my friend from drowning. Trivers would say that in performing this altruistic act I have behaved in a way likely to provide the friend with an incentive to return the favour to me in the future. But the new point is that I have also behaved in a way likely to bring immediate benefit to the gene that both my friend and I are carrying that has predisposed us to be friends to start with. Therefore the pay-off to my fitness as a reciprocal altruist is coming not only directly through the future benefit to my own gene but also indirectly through the immediate benefit to his copy of it.

It is true, as Trivers himself would wish to stress, that if my friend were to happen to be a member of another species, a dog, say, rather than a human, this aspect of the pay-off would be wasted, since the dog's gene for reciprocal altruism is not part of my own species' gene pool and – even if it is functionally equivalent – it is presumably not transferable. Such cross-species friendships, in so far as they do occur in nature, must clearly be considered an exception to the point that is now being made. But this does nothing to weaken the argument as it relates to the much more common case of within-species friendships.

It needs to be said that, even with the within-species friendships, the copy of the gene for reciprocal altruism that can be assumed to be carried by each of the friends need not necessarily be the same gene by virtue of descent from a common ancestor – as it would be with true blood relatives. But there is no reason whatever why this should matter to natural selection. Indeed, for all that natural selection cares, one or other of the friends might actually be a first generation reciprocal altruist who has acquired the gene by random mutation. All that matters is that the genes of the two friends have equivalent effects at the level of behaviour – and to suppose that only genes shared by common descent can count as being "related" would, I think, be to fall into the conceit that philosophers have sometimes called "origins chauvinism".

Let's be clear again that it is not necessary to suppose that the probability of a kin-selection-like genetic pay-off in the case of friends has to be any part of their explicit motivation. The individual's act of reciprocal altruism can unquestionably be justified in the way that Trivers did originally, in terms of its expected return, without reference to any other possible effects on the fitness of the gene. Nonetheless my point – and Rothstein's – is that in reality the indirect effect *will* often be there.

So much so that, again, as in the case of kin selection, it means that the standard model for how reciprocal altruism might evolve by natural selection may have seriously underestimated what there is going for it. In particular, the existence of an indirect benefit to the reciprocal altruism gene means that even if – because of bad luck or bad management – a particular altruistic act yields no return to the altruist, the effort put into it need still not have been entirely wasted. Trivers and his followers have tended to regard any such unrequited act of altruism as a disaster, and have therefore emphasised "cheater-detection" as one of the primary concerns of social life. Yet the present analysis suggests that the system may in reality prove considerably more tolerant and more forgiving.

So, where does this leave us? We have clearly arrived at a rather different picture of the possibilities for altruism from the one that Hamilton and Trivers handed down. Instead of there being two fundamentally different types of altruistic behaviour, sustained by different forms of selection, we have discovered that each type typically has features of the other one. Kin altruism, even if primarily motivated by disinterested concern for the welfare of a relative, is often being selected partly because of the way it redounds to the altruist's own personal advantage. Reciprocal altruism, even if primarily motivated by the expectation of future personal reward, is often being selected partly because of the way it promotes the welfare of a gene-sharing friend.

In which case there may be a further lesson to be learned. For if there is so much overlap between the two types of altruism, why should we continue to think in terms of *two* types at all? Might it not make more sense to suppose that at bottom all examples of altruism have a common formal structure – and a common basis at the level of the gene?

To be brief, I would suggest that the most revealing way of looking at the landscape of altruism is indeed to see it not as islands of kin altruism and reciprocal altruism, but as a continuum of possibilities that all have their roots in just one genetic trait: namely, a trait that is nothing else than *a trait for behaving altruistically to others who share this trait*. Or, to put this more expansively, *a trait for being helpful to those who can be expected to be helpful to those who can be expected to be helpful to those* . . . The recursiveness here is real and significant. It reflects precisely what happens when kin selection and reciprocation get combined. And it must add considerably to the chances of the trait becoming an evolutionary success.

The fact that all cases of altruism might be based on this one trait, however, should not lead us to expect that all cases should look alike in practice. For it is important to appreciate that the trait, as defined above, is only a semi-abstract formal disposition that still has to be realised at the level of behaviour. In particular, it still has to be decided how the possessor of the trait is going to be able to recognise who else counts as "another individual who shares this trait" – who else counts if you like as "one of us" – and hence who precisely should be the target of his or her own altruism.

The possibilities are various – and encompass both types of classical altruism we met with earlier. If, say, the target were to be identified solely on the basis of evidence of blood relationship ("there's a good chance she's one of us because she's my half sister"), it would amount to an example of classical kin altruism. If, on the other hand, the target were to be identified solely on the basis of evidence of willingness to participate in friendly exchanges with oneself ("she's proved herself one of us by returning all the favours I've offered her"), it would amount to an example of classical reciprocal altruism.

But these would only be the two extremes, and in between would lie a range of other variations on the basic theme. There might be, for instance, a particular strain of altruists who identify their targets on the basis of evidence of altruistic behaviour directed to a third party ("she must be one of us because she's being so generous to them"). Another strain might identify them on the basis of the fact that they are already the targets of other altruists' behaviour ("she must be one of us because others of us are treating her as one of theirs"). And in a population where the altruistic trait has already evolved nearly to the point of fixation, it could even be that most altruists would identify their targets simply on the basis that they have not yet shown evidence of *not* being altruistically inclined ("let's assume she's one of us until it turns out otherwise").

Not all these varieties of altruism would be evolutionarily stable under all conditions, and the two classical varieties probably do represent the two strategies that are evolutionarily safest. Nonetheless others would prove adaptive at least in the short term. And the best policy of all for any individual altruist would presumably be to mix and match different criteria for choosing targets, according to conditions.

We should therefore expect, in theory, to find altruism occurring in nature at many levels and in many different forms. This is a satisfying conclusion because – as must be obvious to anyone who can think in terms of more than the two original categories – in practice it is just what we do find. Frans de Waal in his compelling new book, *Good Natured*, has detailed how wide-ranging and rich are the cooperative and succourant behaviours that are to be observed among non-human animals[8] – going far beyond what would seem to have been "justified" by Hamilton's and Trivers's models. The same is more true still for human beings.

"The loveliest fairy in the world," Charles Kingsley wrote in *The Water Babies*, "is Mrs Doasyouwouldbedoneby".[9] And she is also, as it happens, one of the most versatile and most successful.

1. First published as Nicholas Humphrey, 1997, “Varieties of altruism – and the common ground between them”, *Social Research*, 64, 199-209.

2. W. D. Hamilton,1963, "The Evolution of Altruistic Behaviour," *The American Naturalist*, 97, 354-6.

3. Robert L. Trivers, 1971, “The evolution of reciprocal altruism”, *Quarterly Review of Biology*, 46, 35-57.

4. W. D. Hamilton, 1996, *Narrow Roads of Gene Land: The Collected Papers of W.D.Hamilton*, Vol. 1, p. 263, Oxford: W.H.Freeman.

5. Robert L. Trivers, 1971, op. cit.

6. Steven I. Rothstein, 1980, “Reciprocal altruism and kin selection are not clearly separable phenomena”, *Journal of Theoretical Biology*, 87, 255-261.

7. Richard Dawkins, 1976, *The Selfish Gene*, p. 96, Oxford: Oxford University Press; also 1982, *The Extended Phenotype*, p. 155, Oxford: Oxford University Press.

8. Frans de Waal, 1996, *Good Natured: the Origins of Right and Wrong in Humans and Other Animals*, Cambridge Ma.: Harvard University Press.

9. Charles Kingsley, 1863, *The Water Babies*, Ch. 5, London.