

III Our Moral Natures I

The upshot of the last session is that we can make some sense of the idea of psychological egoism. But we haven't yet seen a good *a priori* argument that it's true. Are there *a posteriori* arguments? Two oft-mentioned possibilities:

- (1) arguments from unconscious motives; Freud etc.
- (2) an arguments from evolutionary biology.

Unconscious Motives

Even for Freud it is unclear that the unconscious motives are really egoistic. Contemporary psychology recognizes plenty of unconscious motives. But they are not obviously egoistic. In fact most are too domain specific for notions of egoism and altruism to have much application. Moreover, the more we insist on the importance of unconscious motives, the more skeptical we should be that we can tell by introspection that they are not altruistic.

Evolutionary Arguments

There is an argument for thinking that altruists *must* do worse compared to egoists, and so must in time be eliminated from the population. Assume that egoism and altruism are inherited. (Disregard mutation). We can accept that groups containing more altruists will do better than groups containing fewer, and so will benefit at the cost of those groups. Nonetheless there is an argument that the egoists will do better than the altruists within those groups, and so in time will squeeze them out:

Suppose that altruism benefits the whole population, at a cost to those who practice it. Accept for the sake of argument that the benefit benefits the whole population: indeed even for the altruists themselves it outweighs the cost to them. Still, the *net* benefit to the egoists will be greater, (since they are paying no costs), and that gives them comparative advantage. Comparative, not absolute, advantage is what matters in a competitive situation. So the egoists will win out in the long-run.

A possible response: Simpson's Paradox. An example based on a real case in the University of California:

Imagine that 90 women and 10 men apply to a department with a 30% acceptance rate. This department does not discriminate and therefore accepts 27 women and 3 men. Another department, with a 60% acceptance rate, receives applications from 10 women and 90 men. This department doesn't discriminate either and therefore accepts 6 women and 54 men. Considering both departments together, 100 men and 100 women applied, but only 33 women were accepted, compared with 57 men.

Sober and Wilson *Unto Others*

In each *subgroup* there is no discrimination against women; but in the *total* group women do less well. (Discussion question: does this constitute *de facto* discrimination against women?) Could something analogous happen in the competition between egoists and altruists: could it be that

in each subgroup egoists do better, but in the total group the proportion of altruists remains stable? For example:

Imagine that a population of 200, equally divided among altruists and egoists, is split into two equal sized groups, the tough group and the soft group. The tough group contains 90 egoists and 10 altruists. As a result of the tough conditions, at the end of the breeding cycle the group has declined so that it only has 90 members, 85 egoists and 5 altruists (the egoists have done comparatively better). The soft group starts with 10 egoists and 90 altruists. As a result of the soft conditions, at the end of the cycle it has grown to contain 110 members, 15 egoists and 95 altruists (the egoists have again done comparatively better). The population still contains 100 altruists and 100 egoists.

Conditions for getting Simpson's paradox effects along these lines:

- (i) there must be isolated groups for breeding;
- (ii) the groups must vary in their proportion of altruists;
- (iii) those with more altruists must have more offspring;
- (iv) the breeding groups must come together again to form a pool from which new breeding groups form; the initial proportions in the breeding groups must be roughly the same in each cycle.

How realistic are these conditions? The most obviously problematic is the last. But perhaps this is not so strange. Imagine people choosing who they want to associate with: everyone wants to associate with altruists, and avoid egoists, and they are fairly good at recognizing each other but not perfect. There is some experimental work that suggests that in some non-human populations Simpson's effects emerge (e.g. Wade's work on tribolium beetles); but it's not clear how common it is.

There is another possibility, namely that altruistic behaviour is not the result of evolution, since it is learned behaviour. All that evolution gives us is the *capacity to learn*.

Empirical Work

Batson: empathy (i.e. something like an aversive emotional state caused by the perception of another's suffering) is important (perhaps necessary) for altruistic action. This can be promoted by many simple methods: getting the subject to think through the sufferer's suffering, telling them that they have something in common with them, giving them similar experiences. Batson's claim then: even so, altruistic action doesn't seem to be the result of people feeling uncomfortable, so that they move to get rid of their own discomfort. Given a choice between helping, and removing oneself so that one doesn't feel the discomfort any more, (take an electric shock in place of someone one is watching suffer them, or simply stop watching) the empathetic prefer to help. Could this be because they think they will feel guilty? Probably not; Stock's finding that they still prefer to help even if they think that they will forget the whole thing. (See Stich et al for a discussion of all this: pp. 34–58.)

The prisoners' dilemma. What happens when you iterate? Axelrod's experiment, and the strength of tit-for-tat. Two kinds of reputation effect: that you will cooperate if the other does;

that you will not cooperate if the other doesn't. So this can be explained as a form of reciprocal altruism: short-term sacrifice for long-term benefit. The limits to this: doesn't work with large numbers. The typical decline of trust in iterated public goods games. Something more is needed. Strong reciprocity: a readiness to reward altruism and punish egoism in others; and to do this moreover even when there is no long-term benefit to be gained. Prisoners' dilemmas, and their multi-person equivalents, give no real possibility of reward or punishment; one can only withdraw cooperation in general. But other games have different structures, and so are more like real world interactions. Punishment: the ultimatum game. Reward: the trust game. Third party punishment (recall Butler). These tendencies are stronger when there is a reputation effect. But there is good evidence that people will reward and punish even when this is absent: for instance in single-shot anonymous games. Neurological evidence that we get pleasure from cooperating (Rilling); and that the desire to punish may be driven by emotional factors.

There is evidence that this is *learned* behaviour: it differs across different societies; it is more prevalent in older individuals, or those who have played more games. In fact though the evidence is more complicated than that. It appears that altruistic behaviour comes in two stages; and that much of the learned behaviour looks more like a reciprocal activity than pure altruism.