IV Egoism II: A Posteriori Issues

The upshot of the last session is that we can make some sense of the idea of psychological egoism in terms of I-desires. 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 all egoistic. In fact many 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 sceptical we should be that we can tell by introspection that they are not altruistic. In the moral domain there is some evidence that people like to see themselves as acting morally; this is an I-desire, but it is unclear if it is a selfish one (we'll come back to this when we think about self-deception).

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 tendencies. (Disregard any 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 group selection emerges (e.g. Wade's work on *tribolium castaneum*—flour beetles—where selecting *populations* for fecundity was shown to have different results to selecting individuals); but it's not clear whether this involves Simpson style effects, or how common it is.

There is another possibility, namely that altruistic behaviour is not directly the result of evolution, since it is *learned* behaviour. All that evolution gives us is the *capacity to learn*. That capacity would need to be compatible with altruistic behaviour; but it need not dictate it. So we need to look more carefully at the behaviour that we actually find.

Findings from Empirical Psychology

(i) Batson

Batson finds that what he calls 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.)

(ii) Economic Games

Background: The prisoners' dilemma

> Two members of a burglary gang are arrested Each is kept in solitary confinement with no means of communicating with the other. The police lack sufficient evidence to convict either of the pair for burglary, but they have enough to convict each on a lesser charge of possessing tools for housebreaking. They offer each prisoner a bargain. Each is given the opportunity to betray their accomplice by testifying that the other has committed burglaries; alternatively they might reject the police offer and cooperate with their accomplice by remaining silent. The possible outcomes are:
>
> If A and B each betray the other, each of them serves two years in prison
>
> If A betrays B but B remains silent, A will be set free as a reward, and B will serve three years in prison (and *vice versa*)
>
> If A and B both remain silent, both of them will serve only one year in prison (on the lesser charge).

The rational strategy on a single play looks to be to betray. Why? Because whatever the other person does, you will get a better reward by betraying the. Betrayal is thus the *dominant* strategy in that sense. What happens when you iterate the game? Axelrod did a computer simulation here, and found that tit-for-tat strategy—cooperate on the first play, and then do whatever the other person did on the previous play—bring the biggest reward. This plausibly gets its success from a reputation effect: the other player comes to realize that you will cooperate if they do, and that you will not cooperate if they don't. This gives rise to a form of instrumental altruism (what Fehr and Fischbacher call 'Reciprocal Altruism') involving short-term sacrifice for long-term benefit. So that might show how we could get some apparently altruistic behaviour going even with a prisoners' dilemma structure. However, there are limits to this: it doesn't work with large numbers of players. There is typically a decline of trust in iterated public goods games with many players. Most people cooperate on the first play; then they see that a number of people have benefitted by betraying, and so they start to betray; after four or five plays, virtually no one is cooperating. If we want to see effective altruism arising in games, something more than this is needed.

Fehr and Fischbacher define strong reciprocity as 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. Note that there are really two things going on here: we're looking for cases in which the apparently altruistic action can't be explained as merely instrumentally altruistic (or at least not obviously so); and we are examining games that involve not just the possibility of altruistic behaviour, but also of punishment. Prisoners' dilemmas, and their multi-person equivalents,

give no real possibility of direct reward or punishment of others. But other games have different structures, and so are more like real world interactions. Consider the ultimatum game: One player is given a sum of money (say $10) to share with someone else; they can make any division they like. The other player then has the choice of rejecting or accepting the deal. If they accept, the money is distributed in line with the first persons' choice; if they reject, no one gets anything. If agents were motivated purely by a desire for money, then in anonymous single play games they should accept any share greater than zero. But in fact they tend to reject almost all shares less than 25%, and a great number below 50%. Even more impressive, third parties watching such interactions are prepared to spend their own money to punish those who they judge to have behaved badly. These tendencies are stronger when there is a reputation effect. But they remain even when this is absent: for instance in single-shot anonymous games. These could be I-desires (subjects might want to be the one who does the punishment); but equally (more plausibly?) they are non-I-desires: 'they deserve to be punished, and no one else is going to do it, so will'. Recall that it is their own money that they are spending.

Moreover, following up on the earlier point there is some evidence that this is *learned* behaviour: it differs across different societies; and it is more prevalent in older individuals, or those who have played more games.