

Social Formation of Pro-Social Preferences

Avinash Dixit
Princeton University

Paper prepared for
“The Irrational Economist” Conference
and Book Writing in honor of Howard Kunreuther

December 4-5, 2008

Chapter to appear in
The Irrational Economist: Future Directions in Behavioral Economics and Risk Management
E. Michel-Kerjan and P. Slovic (eds.) *Public Affairs Press*. (forthcoming 2009)

Social Formation of Pro-Social Preferences *

Avinash Dixit

Princeton University

Abstract:

Research spanning psychology and economics has convincingly demonstrated that individuals' preferences deviate substantially from the traditional economic assumption of self-interest, and have pro-social components. The evolution and persistence of pro-social preferences can be justified by theories of group selection in evolutionary biology. I argue that a fuller analysis requires spanning sociology and political economy. Human societies make deliberate social effort to instill pro-social preferences in children. I suggest a model for this process.

Until recently economics rested on two entrenched assumptions about individual preferences, namely selfishness and exogeneity. The first has been largely overturned as numerous laboratory experiments and empirical studies have convincingly demonstrated that people have pro-social considerations such as empathy and fairness (Camerer 2003, chapter 2). The second assumption is also on its way out, and the purpose of this note is to suggest an exit route that is different from, but complementary to, the one usually followed.

The standard approach to endogenizing pro-social preferences comes from an obvious analogy with the question of selection in evolutionary biology. The puzzle that needs to be explained is: someone who expends resources to benefit others will fare worse than another who looks out for oneself, and therefore the pressure of selection will eliminate pro-social preferences. A part of the response is that societies whose members have pro-social preferences are better able to solve various collective action problems,

* For presentation at the Conference in Honor of Howard Kunreuther, to be held Dec 4-5, 2008 at the Wharton School, University of Pennsylvania. This version Nov 18. I thank the National Science Foundation for research support.

and therefore achieve higher payoffs for all their members, than can societies whose members are purely selfish.

However, this is subject to the same criticism as is the concept of group selection in evolutionary biology. A group with pro-social members may fare better than a group of egoists, but a mutant egoist invading a pro-social group (for example someone who free rides on the costly collective action undertaken by the others) will fare better still. The mutants will multiply faster; the pro-social strategy cannot be evolutionary stable.

Such invasion by selfish mutants can be defeated if the initial pro-social members of the group inflict punishment on the mutants. The act of punishment may even entail some personal cost; this is just another aspect of the pro-social behavior of the population. Thus the evolutionary stable strategy has two prongs: act in conformity with the social norm of good behavior, and punish, at some personal cost if necessary, anyone you see violating this norm. Such “strong reciprocator” behavior is indeed observed in experiments (Fehr and Gächter 2000) and analyzed theoretically (Bowles and Gintis 2004).

But there is more to socio-cultural evolution of preferences than can be captured in the biological parallel. In evolutionary biology, individuals are endowed with phenotypes that determine their strategies, and fitter strategies proliferate faster. In social settings, individuals are indeed born with some unchangeable behaviors, but they acquire many other behaviors during a long period of socialization that begins with families, extends for many years in schools, and continues at various levels of intensity into adulthood and indeed throughout life. The early years of life, when children are most impressionable and their preferences and behavior can be molded substantially, should be the most crucial phase in this long process.

This is no mere armchair reasoning. Experiments using ultimatum and dictator games that demonstrate the existence of pro-social preferences have been conducted on subjects of different ages. Up to the age of 5 or so, children are mostly quite self-interested. At the age of 6 or 7 they start to develop a norm of strict equality. More general ideas of equity and compromise emerge much later. (Camerer, 2003, pp. 65-67)

Many aspects of this process of preference-formation are results of conscious social decisions. How to raise children, and what to teach them in schools, are issues

constantly discussed in social and political forums. Different societies have different practices in these matters, with different outcomes. Understanding the reasons for these differences should be an important item in the research agenda.

The most obvious instance where preferences go beyond the purely selfish is that of caring about one's children and perhaps also successive generations; the literature on this, especially in macroeconomics and fiscal economics, is too vast to cite. Preferences for unrelated others and their consequences for equilibrium outcomes are a more recent addition to the literature; Rabin (1993) is among the most prominent. There are also models of parents' deliberate efforts to instill preferences in children. Lindbeck and Nyberg (2006) construct such a model of work norms and Tabellini (2008) models honesty in contract performance. Benabou and Tirole (2006) consider transmission of beliefs rather than preferences, and the choice variable is the level of taxation determined through a political process.

In this paper I offer a very simple model of a political process by which pro-social preferences to promote collective action can be formed. It combines the various features of the models cited in the previous paragraph. The key new idea is an interaction between two aspects of unselfish preferences: naturally existing ones toward one's own progeny and created one toward strangers. Collective action is made feasible by the latter, and it benefits everyone, so one can benefit one's own progeny by giving them pro-social preferences that include strangers in their own generation. However, this requires that everyone in the next generation has such preferences; therefore collective rather than individual action in one's own generation – schooling rather than family upbringing – is needed.

I use the simplest functional forms that enable analytic solutions and convey the ideas, and later suggest ways in which the model can be generalized and built upon.

The society consists of n individuals. Each individual denoted by i can expend two types of effort, private, denoted by x_i and public, denoted by z_i . The average public effort is

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i. \quad (1)$$

Then the income of individual i is given by

$$y_i = (1 + \bar{z}) x_i. \quad (2)$$

Thus public effort creates a public input that raises the average and marginal product of private input. The private or selfish utility of individual i is

$$u_i = y_i - \frac{1}{3}(x_i + z_i)^2. \quad (3)$$

First consider the Nash equilibrium where individuals choose their effort levels noncooperatively. A little care is needed about corner solutions, but it is easy to verify that

$$x_i = 3/2, \quad z_i = 0, \quad y_i = 3/2, \quad u_i = 3/4. \quad (4)$$

Next consider the symmetric cooperative optimum where common effort levels x and z are chosen to maximize each individual's utility. This yields

$$x = 2, \quad z = 1, \quad y = 4, \quad u = 1. \quad (5)$$

If the society cannot compel individuals to exert the public effort directly, it may induce them to do so by changing their preferences to include a pro-social element. Begin by examining the effect of pro-social preferences and then consider how they can be instilled. Suppose individual 1's pro-social utility is

$$v_1 = u_1 + \gamma \sum_{i=2}^n u_i, \quad (6)$$

and similarly for other individuals. If

$$\gamma \leq \frac{2n-3}{3(n-1)}, \quad (7)$$

the Nash equilibrium is still a corner solution with $z_i = 0$ for all i , the same as (4). The right hand side of (7) is the minimum threshold of pro-social preference needed to induce positive public effort. Thus just a little pro-socialness does not work; this is similar to the result of Rabin (1993) in the context of fairness. The threshold is an increasing function of n . But as $n \rightarrow \infty$, the threshold goes to $2/3$, not 1: even in the largest society, the threshold is consistent regarding others' utility worth less than one's own.

When γ exceeds the threshold, the solution can be written compactly using the abbreviation

$$\phi = \frac{1 + \gamma(n-1)}{n}. \quad (8)$$

Then (7) translates to $\phi \leq 2/3$. For $\phi > 2/3$, we get the symmetric Nash equilibrium

$$x_i = \frac{2}{2-\phi}, \quad z_i = \frac{3\phi-2}{2-\phi}, \quad y_i = \frac{4\phi}{(2-\phi)^2}, \quad u_i = \frac{\phi(4-3\phi)}{(2-\phi)^2}. \quad (9)$$

As ϕ increases from $2/3$ to 1 , the solution in (9) moves monotonically from the purely selfish Nash equilibrium (4) to the purely cooperative optimum (5). A higher ϕ in this range raises everyone's *private* utility.

From here on, to keep the algebra simple, I will consider a large population, so n is large and $\phi = \gamma$ to a good approximation.

Suppose there is a succession of generations. Each individual has one child, whose selfish utility is denoted by u_i^c defined similarly to (3) above. The parent has private family utility

$$f_i = u_i + \delta u_i^c. \quad (10)$$

Education can instill a social utility with parameter ϕ into each child. The cost of this per capita is t given by

$$t = \frac{k}{1-\phi}, \quad \text{or} \quad \phi = 1 - \frac{k}{t}. \quad (11)$$

This has several sensible properties: a threshold level of education k is needed to instill any pro-social preference at all; the marginal cost of preference-formation is increasing; and it is infinitely costly to make each individual fully internalize the social welfare.

Although each parent cares only about the family utility, he/she recognizes that instilling pro-social preference in all children will increase his/her own child's private utility as seen in . Each generation votes on a tax t per capita to finance the education. Since there is no heterogeneity, the t that maximizes the parent's net private family utility $f_i - t$ will emerge as the Condorcet winner. This is equivalent to choosing ϕ to maximize the net increase in the parent's family utility:

$$g(\phi) = \delta \left[\frac{\phi(4-3\phi)}{(2-\phi)^2} - \frac{3}{4} \right] - \frac{k}{1-\phi}. \quad (12)$$

This maximization presents two special issues. [1] There may be a corner solution at $\phi = 0$ with $g'(0) \leq 0$. This happens if $\delta \leq 8k$. [2] Even if $\delta > 8k$, a positive solution to the first-order condition for ϕ may be only a local optimum. Calculation shows that this requires $\delta > 36k$ approximately; and then the optimum choice of ϕ is given by

$$\phi = \frac{2(1-\theta)}{2-\theta} \quad \text{where } \theta = \left(\frac{k}{\delta}\right)^{1/3}. \quad (13)$$

When δ is at its lower limit $\delta \cong 36k$, we have $\theta \cong 0.305$, and then $\phi \cong 0.82$. At the other extreme, when k is very small compared to δ , θ is close to 0 and ϕ close to 1.

Thus a sufficiently patient parent generation can instill pro-social preferences in the next generation through education. The level instilled at the minimum of the patience threshold $\delta \cong 36k$, namely $\phi \cong 0.82$, is in excess of the threshold $\phi = 2/3$ needed to induce positive public effort; this is due to the fixed cost feature of the education technology. If the fixed cost is negligibly small, full social welfare can be approached.

This simple model can be generalized and enriched in numerous ways. At a minimum, the functional forms can be generalized to produce “a theory” rather than “an example”. However, I don’t think that will produce any major new insights. Here are two suggestions for more substantive generalization:

[1] I have considered just one generation of parent and child. This can be extended to a proper overlapping-generations structure and a stationary state equilibrium can be analyzed.

[2] Change the production technology (2) to $y_i = (1 + \alpha \bar{z}) x_i$, where α is an unknown parameter. Each generation gets a signal, either $\alpha = 1$ or $\alpha = 0$, with specified probabilities. Then the generation will or will not spend on its children’s education depending on the signal. Depending on the signal process, there can be multiple equilibria, one with investment, and one without, where a generation that gets a wrong signal $\alpha = 0$, does not spend on education, and therefore has no opportunity to observe the true effect of public effort on output. Such multiple equilibria have been found in other contexts, for example work ethic and stigma (Lindbeck, Nyberg and Weibull 1999), and belief about the justness of market outcomes (Benabou and Tirole 2006).

[3] Individuals can differ in their parameters of pro-socialness and patience. There is a genetically determined distribution of γ in the population, and education shifts this distribution to the right in the sense of first-order stochastic dominance. Individuals with high patience δ are willing to spend a higher amount t per capita on education, and the median voter's most preferred t is implemented.

REFERENCES

- Battaglini, Marco, Roland Bénabou, and Jean Tirole. 2005. "Self-control in peer groups." *Journal of Economic Theory* 123, 105–134
- Benabou, Roland and Jean Tirole. 2006. "Belief in a just world and redistributive politics." *Quarterly Journal of Economics* 121(2), May, 699-746.
- Bowles, Samuel and Herbert Gintis. 2004. "The evolution of strong reciprocity: Cooperation in heterogeneous populations." *Theoretical Population Biology* 65(1), February, 17-28.
- Camerer, Colin F. 2003. *Behavioral Game Theory: Experiments in Strategic Interaction*. New York: Russell Sage Foundation & Princeton, NJ: Princeton University Press.
- Fehr, Ernst and Simon Gächter. 2000. "Fairness and retaliation: The economics of reciprocity." *Journal of Economic Perspectives* 14(2), Spring, 159-181.
- Lindbeck, Assar, Sten Nyberg and Jörgen Weibull. 1999. "Social norms and economic incentives in the welfare state." *Quarterly Journal of Economics* 114(1), February, 1-35.
- Lindbeck, Assar and Sten Nyberg. 2006. "Raising children to work hard: Altruism, work norms, and social insurance." *Quarterly Journal of Economics* 121 (4), November, 1473-1503.
- Rabin, Matthew. 1993. "Incorporating fairness into game theory and economics." *American Economic Review* 83(5), December, 1281-1302.
- Tabellini, Guido. 2008. "The scope of cooperation: Values and incentives." *Quarterly Journal of Economics* 123(3), August, 905-950.