Social Creation of Pro-Social Preferences for Collective Action

Avinash Dixit and Simon Levin*
Princeton University

July 10, 2016

Abstract

Pro-social preferences are thought to play a significant role in solving society's collective action problems of providing public goods and reducing public bads. Societies can benefit by deliberately instilling and sustaining such preferences in their members. We construct a theoretical model to examine an intergenerational education process for this. We consider both a one-time action of this kind and a constitution that establishes a steady state, and compare the two.

JEL classification codes: H41, D03

Corresponding author information:

Avinash Dixit
Department of Economics
Princeton University
Princeton, NJ 08544, USA

E-mail: dixitak@princeton.edu

*Funding was provided by the National Science Foundation grants GEO-1211972 and OCE-1426746, and by the Nordforsk-funded project Green Growth Based on Marine Resources: Ecological and Socio-Economic Constraints (GreenMAR) to Simon A. Levin.

1 Introduction and Motivation

Study of collective action to provide public goods was the focus of much of Richard Cornes' work. Attainment of aggregate efficiency in these situations has to overcome free riding by selfish participants. Exogenous and purely self-regarding preferences were the staple of traditional economic analysis. However, economics in recent years has increasingly recognized the reality of pro-social preferences, and is beginning to recognize that preferences are socially formed. In this paper we develop a model with these features.

Pro-social preferences and other-regarding behaviors more generally are a fact of life, though it is often puzzling how they are sustained (Henrich et al. 2001, Gintis 2003, Fehr and Gintis 2007, Akcay et al. 2009, Henrich et al. 2010). The most plausible explanation will combine genetic and evolutionary pathways with socio-cultural processes to incentivize and reinforce prosocialiity. In this paper we focus on one such societal process. Our basic framework builds on earlier work by the first author (Dixit 2009). The framework is a general one, where individuals allocate their efforts or resources between their own interests and the public good. The analysis applies equally to investments that limit the damage to common pool resources; examples are fisheries management, climate change and effectiveness of antibiotics (Hardin 1968, Ostrom 1990, Levin 1999, Smith et al. 2005).

In evolutionary biology, individuals are endowed with genes that determine the phenotypes that in turn 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.

Developmental psychologists have studied the process in more detail. Hoffman (2000, pp. 10-11) describes it thus: "When a child experiences, repeatedly, the sequence of transgression followed by a parent's induction[1] followed by child's empathetic distress and guilt feeling, the child forms Transgression → Induction → Guilt scripts . . . [That] may become strong enough with repetition, and when combined with cognitive development and peer pressure . . . may be effective. That is, peer pressure *compels* children to realize that others have claims; cognition *enables* them to understand others' perspectives; empathic distress and guilt *motivate* them to take others' claims and perspectives into account." (Emphasis in the original.) The importance of peer pressure, and an important role for schools (or similar settings where many unrelated children interact), is emphasized in sociological literature. Thus Boocock and Scott (2005, p. 84) find that "on a wide range of attitudes and behaviors, kids tend to become more like their friends and less like their parents."

Our purpose in this paper is to construct a simple model of such collective action to instill pro-social preferences in children.[2] We emphasize again that we do not claim this as the only or even the predominant way in which societies generate or sustain prosociality; it can coexist with other pathways of genetic or social evolution. But it is clearly a significant one in reality, and worth analysis and theoretical exploration on its own.

The model focuses on public good provision. Final output is produced using the public good and private effort. The public good increases the productivity of private effort. Therefore when each person contributes to the public good, this raises all individuals' personal benefit, which we call their private or *selfish utility.* When deciding whether and how much to contribute to the public good, purely selfish individuals would ignore the benefits that flow to others. Therefore they would contribute too little, to the detriment of all. This is the standard prisoners' dilemma of collective action.[3] One way to ameliorate it is for individuals

[1]Induction in this context is a mild form of discipline technique. Hoffman defines and explains it as follows: "When children harm or are about to harm someone – the parent, a sibling, a friend – . . . indicate[s] implicitly or explicitly that the act is wrong and that the child has committed an infraction. . . . This creates the condition for feeling empathy-based guilt. (Hoffman, 2000, pp. 150–151)

[2]The need for collective action for preference formation is the crucial respect in which our model differs from other models where individual parents shape their children's preferences, for example Bisin and Verdier (2001), Tabellini (2008).

[3]The problem of the commons, where the issue is how to dissuade individuals from creating a public bad, is the mirror image of this, and can be analyzed by similar methods. In other contexts, sufficiently prosocial preferences may lead to a coordination or "assurance" game. Such more complex interaction topologies are subject of our ongoing work.

to have pro-social preferences; we call such an objective the individual's *prosocial utility.*

Each parent cares about her children's welfare. Parents with such concerns would like the next generation to solve its prisoners' dilemma of public good provision; therefore they would like the next generation to have prosocial preferences. Of course any one parent would not want its child to become the only pro-social actor if all other children grow up to be selfish; that would make the child a sucker in a prisoners' dilemma game. However, each parent may be willing to vote for a tax that finances education for all children to instill pro-social preferences in all.[4] This is how we model the process. It fits with what we observe in reality: schools do devote a substantial amount of time and resources to socialization and to teaching concepts of civic duties, concern for and responsibility toward others, social norms of behavior, and so on.

The model is an extended numerical example to develop the ideas. We use special functional forms and parameter values chosen to facilitate solution. But the intuitions it creates are appealing and the qualitative results should remain valid in more general conditions.

2 Equilibrium with Given Pro-socialness

Begin with one generation, and examine how pro-social preferences improve the provision of the public good and the utilities of individuals. There are n individuals, labeled i = 1, 2, ... n. Each can exert two types of effort: private x_i, and public z_i. The public good may consist of the effort itself, for example volunteered time, or it may be a good or service produced one-for-one using aggregate public effort; either interpretation works equally well. Denote the average public effort by

$$\overline{z} = \frac{1}{n} \sum_{i=1}^{n} z_i \ . \tag{1}$$

Then the income of individual i is given by

$$y_i = (1 + \overline{z}) \ x_i \ . \tag{2}$$

[4] See Friedman (1962, p. 191) for an early argument along these lines in favor of government action to alleviate poverty.

Thus a higher $\overline{z}$ raises the (average and) marginal product of each individual's private effort, as mentioned above.[5]

The private or selfish utility of i:

$$u_i = y_i - \tfrac{1}{3}\left(x_i + z_i\right)^2 \ . \tag{3}$$

If individuals are selfish, the Nash equilibrium of their non-cooperative choices has no public effort:

$$x_i = 3/2, \qquad z_i = 0, \qquad y_i = 3/2, \qquad u_i = 3/4 \ . \tag{4}$$

The calculation, and similar calculations to follow, are simple but sometimes tedious, and do not contribute to the intuition; therefore they are relegated to the Appendix, section A.1.

Contrast this with the symmetric cooperative optimum where common effort levels x_i and z_i for all i are chosen to maximize total social surplus:

$$x_i = 2, \qquad z_i = 1, \qquad y_i = 4, \qquad u_i = 1 \ . \tag{5}$$

The contribution to public effort raises the incentive to make private effort, leading to much higher incomes, sufficiently higher to overcome the disutility of the greater effort.

Some societies may achieve this optimum by command and control. But this is often infeasible in open democratic societies; and even if feasible, many would prefer gentler methods. The one we consider is to change 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

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

where the parameter γ is assumed to be in the interval $(0, 1)$, and captures the idea that each individual includes some concern for the welfare of other individuals, but not as much as for her own. The pro-social utility of individuals 2, 3, $\ldots$ n is defined similarly. If

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

[5] By using the average $\overline{z}$, not the total $\sum_{i=1}^{n} z_i$, in (1), we are assuming that the public effort is a public good with congestion. The no-congestion case is worth separate analysis.

The Nash equilibrium is the same as in (4), with $z_i = 0$ for all i. 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 rises with n, but goes to 2/3, not 1, as $n \to \infty$: even in very large societies, the threshold is consistent with regarding others' utility worth less than one's own.

Use the abbreviation

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

With the assumption $\gamma < 1$, we have $\phi < 1$. For large n, $\phi = \gamma$ approximately; we will focus on this case because most public good situations involve large populations. Then (7) becomes simply $\phi \leq 2/3$.

If $\phi > 2/3$, the symmetric Nash equilibrium of individual contributions to public effort is:

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

The resulting pro-social utilities are

$$\begin{aligned} v_i &= [1 + (n-1)\,\gamma]\; u_i = n\,\phi\,u_i \\ &= n\,\frac{\phi^2\,(4-3\phi)}{(2-\phi)^2} \; . \end{aligned} \tag{10}$$

As ϕ increases from 2/3 to 1, (9) moves monotonically from the purely selfish (4) to the optimal (5). Therefore in this range, if everyone has more pro-social preferences, that raises everyone's *selfish* utility. Figure 1 shows this functional relationship.

3 Choosing Children's Pro-socialness

Now introduce a succession of generations. Each individual has one child. Each parent cares about the utility of her child. All parents in a cohort recognize that if their children's generation had enough prosociality, they would be able to achieve higher utility as in the previous section. The parents would be willing to sacrifice some of their own utility to achieve

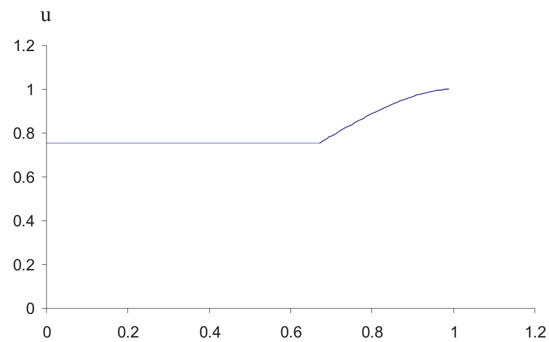

Figure 1: Selfish utility u in Nash equilibrium as function of pro-socialness parameter ϕ

such an outcome. But that requires coordinated action – just one parent bearing a cost to educate her own child to be prosocial would merely convey the benefit of her child's public effort to the children of other parents. We model the coordinated action as choosing a tax levied on all members of the parent generation to fund education that instills prosociality in the children's generation. We assume all individuals in a generation to be identical, therefore their choice is unanimous and made to maximize the utility (which includes the concern for the child's well-being) for each of them.

We consider two cases. In the first, this process happens just once. Generation 1, which had no prosociality instilled into it by generation 0, wakes up and realizes that it can improve the well-being of generation 2 through the socializing education. Generation 1 does not know whether generation 2 will continue to pursue any such actions, and ignores that possibility. This one-off model may be a purely hypothetical thought-experiment, but it serves a useful purpose of introducing the basic idea.

The other case is a steady state. At some time before the action starts, a cohort or generation of "founding mothers" meets to write a constitution. This specifies the tax to be levied on each member of every generation from that point on to educate the next generation. We leave aside the problem of what happens to the very first generation, or assume that this generation of the founding mothers somehow chooses to build the right level of prosocialness into itself in course of their constitutional deliberations. Thus we assume that the steady state is attained at once and continues; we do not examine dynamics starting from an arbitrary initial condition and the possibility that it may not converge to a steady state.

3.1 One-off Action

Write u^a for the selfish utility of any one adult in the generation that decides to educate its children for prosociality, and u^c for that of her child; each is defined as in (3). The adult's overall utility including that of her child's well-being is defined by

$$U^a = u^a + \delta\, u^c\ , \tag{11}$$

which includes concern for one's own child but not for anyone else. We assume $0 < \delta < 1$.

Education can give the child a social utility with parameter ϕ, related to the γ as in (8).We assume the following form for the cost of this per capita:

$$t = \frac{k}{1-\phi}, \quad \text{or} \quad \phi = 1 - \frac{k}{t} \qquad \text{for } t > k. \tag{12}$$

Observe that for t slightly above k, ϕ is close to 0; then t increases as a concave function of k and $t \to 1$ as $k \to \infty$. This makes intuitive sense: the threshold level k of expenditure is needed to instill any pro-socialness at all; thereafter the marginal cost of preference-formation is increasing; and it is infinitely costly to make each individual fully internalize social welfare.

Endowed with this ϕ, the children's generation will achieve a Nash equilibrium of their game of effort allocation between private and public uses, and that will yield its u^c as a function $u^c(\phi)$, as defined in equations (4) and (5).

The parent generation had no prosocialness installed in it; therefore its u^a is determined by the Nash equilibrium of the game with $\gamma = 0$, namely $u^a = 0.75$ as we found in Section 2. The tax is subtracted from this. Also, in our thought-experiment of this subsection, the parent generation acts as if the children's generation will not organize any prosocializing effort for the generation to follow, i.e. for this generation's grandchildren. Accordingly, no tax is anticipated or subtracted from u^c. Therefore this generation chooses t, or equivalently ϕ, to maximize U^a which is now a function of ϕ:

$$U^a \equiv f(\phi) = \begin{cases} \dfrac{3}{4} - \dfrac{k}{1-\phi} + \delta\, \dfrac{3}{4} & \text{for } \phi < 2/3 \\[2ex] \dfrac{3}{4} - \dfrac{k}{1-\phi} + \delta\, \dfrac{\phi\,(4-3\phi)}{(2-\phi)^2} & \text{for } 2/3 < \phi < 1 \end{cases} \tag{13}$$

The details of the solution are in the Appendix, section A.2. To state it more compactly, define $\theta = (k/\delta)^{1/3}$. Then there is a critical level $\theta^* \approx 0.305$ such that when $\theta > \theta^*$ (corresponding to $k > 0.028\,\delta$ approximately), it is optimal to choose $\phi = 0$. If $\theta < \theta^*$, that is, $k < 0.028\,\delta$, the optimum choice is

$$\phi = \frac{2\,(1-\theta)}{2-\theta}\ . \tag{14}$$

Thus, if the cost of education is low relative to the regard for children's welfare, the parent generation will instill pro-social preferences in the next generation through education. The level instilled when k is just below its threshold of $\approx 0.028\,\delta$ (that is, when θ is just below $\theta^* \approx 0.305$) is $\phi \approx 0.82$, which significantly exceeds the threshold $\phi = \frac{2}{3}$ needed to induce a small positive public effort. This jump is due to the fixed cost feature of the education technology. If the fixed cost k becomes negligibly small, θ goes to zero and ϕ goes to 1; in this limit the full social optimum can be approached.

Since $\delta < 1$, $k \geq 0.028$ is sufficient to rule out prosocialness in the one-off situation. This will be contrasted with the steady-state case that follows.

3.2 Steady State

Now we consider the steady state of an ongoing succession of generations, where a stationary policy of taxation and education is chosen in advance in a binding constitution. The procedure is similar to that in Arrow and Levin (2009). Let $v(n)$ denote the utility of any one member of generation n, including the effects of the Nash equilibrium it achieves given its own prosociality and the tax it pays to educate generation $(n+1)$, but not the utility this person gets from the well-being of its own child. Let $V(n)$ denote the comprehensive utility of this person including this parental concern also. Then we have

$$V(n) = v(n) + \delta\, V(n+1)$$

We assume a steady state in which the constitution fixed in advance also fixes the tax to be the same for all generations, and then the equilibrium and utilities are the same for all generations n. Denoting these common values by v and V, we have

$$V = \frac{v}{1-\delta}$$

Maximizing V is therefore equivalent to maximizing v, independently of δ. The extent of each generation's concern for its children does not matter in the steady state; the founding mothers have taken it into account and made the best common choice for all.

If u is any one individual's utility and t the tax, we have

$$v = [1 + (n-1)\gamma]\,(u - t) = n\,\phi\,(u - t)$$

Observe that the prosocialness instilled into a parent leads her to include the effects both of the utility and the tax on her contemporaries.

Using equations (4), (10), and (12) gives the objective:

$$v = \begin{cases} n\,\phi\,\left[\dfrac{3}{4} - \dfrac{k}{1-\phi}\right] & \text{for } \phi < 2/3 \\[2ex] n\,\phi\,\left[\dfrac{\phi(4-3\phi)}{(2-\phi)^2} - \dfrac{k}{1-\phi}\right] & \text{for } 2/3 < \phi < 1 \end{cases} \tag{15}$$

A combination of analytical and numerical methods detailed in the Appendix section A.3 yields the following solution for the choice of ϕ to maximize v:

If $k < 1/12$ (= 0.8333), the function is increasing throughout the range $0 < \phi < 2/3$, and then has a peak in the range $2/3 < \phi < 1$. Thus it has a unique optimum in the range where sufficient prosociality is instilled to elicit public effort. Figure 2 graphs this for the upper extreme value of the range, $k = 1/12$, where the function just becomes flat approaching $\phi = 2/3$ from the left.

If (0.8333 =) $1/12 < k < 1/6$ (= 0.1667), the function has local peaks in each of the ranges $0 < \phi < 2/3$ and $2/3 < \phi < 1$. The former yields the optimum if $k < 0.1235$ and the latter if $k > 0.1235$. Figure 3 graphs this in the borderline case of $k = 0.1235$.

If (0.1667 =) $1/6 < k < 3/4$, the function has a peak in the range $0 < \phi < 2/3$ and is decreasing throughout the range $2/3 < \phi < 1$. Therefore it has a unique optimum in the former range, where some prosociality is instilled but not enough to elicit any public effort. We comment on this below. Figure 4 graphs the function for the lower extreme value of the range, $k = 1/6$, when the function is flat just to the right of $\phi = 2/3$.

For $k > 3/4$, the function is decreasing throughout, and $\phi = 0$ is the optimum.

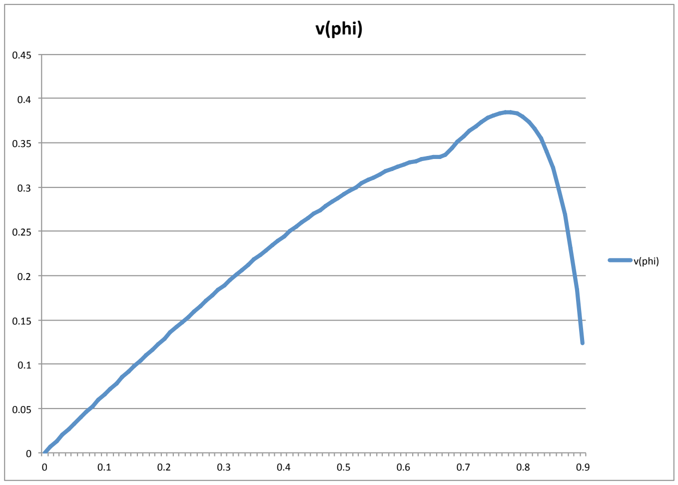

Figure 2: Steady state: case of low k

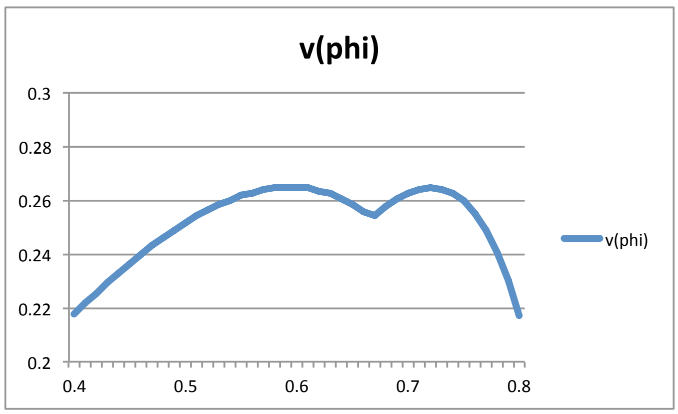

Figure 3: Steady state: case of mid-range k

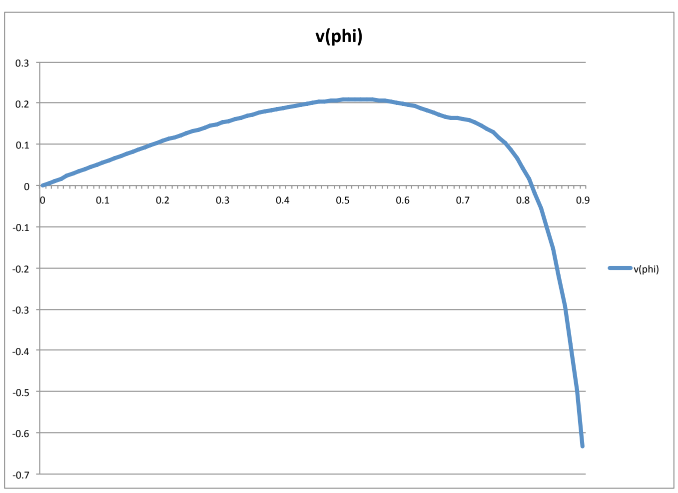

Figure 4: Steady state: case of high k

Some features of the solution require comment: (1) Prosociality is instilled for a much larger range of k in the steady state than in the one-off case. Recall that in the latter, $k = 0.028\,\delta < 0.028$ was the upper limit of the cost parameter for which enough prosociality to induce public effort would be instilled. Now that limit is $k = 0.123$. (2) As remarked earlier, the extent of prosociality is independent of δ. The founding mothers once and for all figure out what is best for every generation and specify it as a rule in the constitution. (3) If $0.123 < k < 0.75$, some prosociality is instilled but not enough to lead to any public effort. This is because the inclusion of other people's utility in each person's own social utility has value in itself. For example, if $k = 0.125$, it turns out that the optimum is $\phi = 0.6$. Then $t = 0.125/(1 - 0.6) = 0.31$ and each person's selfish utility is $u = 0.75 - 0.31 = 0.44$. But the social utility is $v = 0.44 + (n-1)\,\gamma \times 0.44 = n\,\phi\, \times 0.44$. (4) If k increases across 0.1235, the optimum undergoes a discrete downward jump, in Figure 3 from about $\phi = 0.72$ to 0.59, and in the Nash equilibrium the extent of public effort drops discretely to zero. As in the one-off case, the reason is the fixed cost nature of the education technology.

4 Concluding Comments

Most societies put considerable effort into socializing youngsters to achieve better behavior and outcomes. We have constructed a simple mathematical model to capture this process. We hope this will prove a useful step in the endeavor of incorporating insights not only from psychology but also from sociology or social psychology into economics, and thereby improve our understanding of how collective action is organized.

References

Akcay, E., J. Van Cleve, M. W. Feldman, and J. Roughgarden. 2009. A theory for the evolution of other-regard integrating proximate and ultimate perspectives. Proceedings of the National Academy of Sciences of the United States of America 106:19061–19066.

Arrow, Kenneth J. and Simon Levin. 2009. Intergenerational resource transfers with random offspring numbers. Proceedings of the National Academy of Sciences of the United

States of America 106:13702–13706.

Bisin, Alberto and Thierry Verdier. 2001. The Economics of Cultural Transmission and the Dynamics of Preferences. Journal of Economic Theory 97(2): 298–319.

Boocock, Sarane Spence, and Kimberley Ann Scott. 2005. Kids in Context: The Sociological Study of Children and Childhoods. Lanham, MD: Rowan and Littlefield.

Dixit, Avinash. 2009. Governance Institutions and Economic Activity. American Economic Review 99:5-24.

Fehr, Ernst, and Herbert Gintis. 2007. Human motivation and social cooperation: Experimental and analytical foundations. Annual Review of Sociology 33:43-64.

Friedman, Milton. 1962. Capitalism and Freedom. University of Chicago Press, Chicago, IL.

Gintis, Herbert. 2003. Solving the puzzle of prosociality. Rationality and Society 15:155-187.

Hardin, Garrett. 1968. The tragedy of the commons. Sciences 162:1243-1248.

Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath. 2001. In search of Homo economicus: Behavioral experiments in 15 small-scale societies. American Economic Review 91:73-78.

Henrich, J., S. J. Heine, and A. Norenzayan. 2010. The weirdest people in the world? Behavioral and Brain Sciences 33:61-135.

Hoffman, Martin L. 2000. *Empathy and Moral Development.* Cambridge, UK: Cambridge University Press.

Levin, Simon A. 1999. Fragile Dominion: Complexity and the Commons. Perseus Books, Reading, MA.

Ostrom, Elinor. 1990. Governing the Commons: The Evolution of Institutions for Collective Action. Cambridge University Press, Cambridge.

Rabin, Matthew. 1993. Incorporating Fairness into Game-Theory and Economics. American Economic Review 83:1281-1302.

Smith, D. L., S. A. Levin, and R. Laxminarayan. 2005. Strategic interactions in multi-institutional epidemics of antibiotic resistance. Proceedings of the National Academy of Science, USA 102:3153-3158.

Tabellini, Guido. 2008. The Scope of Cooperation: Values and Incentives. Quarterly Journal of Economics 123(3): 905–950.

Appendix – Mathematical Derivations

A.1 Nash equilibrium with prosociality

Begin with the purely selfish case, set out in Section 2, equations (1)–(3). The Kuhn-Tucker conditions for individual i's choice of (x_i, z_i) to maximize u_i are

$$\frac{\partial u_i}{\partial x_i} = 1 + \overline{z} - \tfrac{2}{3}\,(x_i + z_i) \leq 0, \quad x_i \geq 0\,, \tag{A.1}$$

$$\frac{\partial u_i}{\partial z_i} = \tfrac{1}{n}\,x_i - \tfrac{2}{3}\,(x_i + z_i) \leq 0, \quad z_i \geq 0\,, \tag{A.2}$$

with complementary slackness in each equation. Note that in (A.2) we have used $\partial\overline{z}/\partial z_i = 1/n$. The matrix of second-order partials is

$$\begin{pmatrix} -\frac{2}{3} & -\frac{2}{3} + \frac{1}{n} \\ -\frac{2}{3} + \frac{1}{n} & -\frac{2}{3} \end{pmatrix}.$$

The diagonal elements of this are negative, and the determinant is

$$\left(\tfrac{2}{3}\right)^2 - \left(\tfrac{2}{3} - \tfrac{1}{n}\right)^2 > 0\,.$$

Therefore the matrix is negative definite, so the second-order sufficient conditions are met and the Kuhn-Tucker conditions yield the global maximum of u_i. (This will continue to be so in all the variants of the model considered here, and will not be mentioned further.)

First try the purely selfish solution where $x_i > 0$ and $z_i = 0$ for all i. Then $\overline{z} = 0$ also, and the conditions (A.1) and (A.2) are

$$1 - \tfrac{2}{3}\,x_i = 0, \qquad \left(\tfrac{1}{n} - \tfrac{2}{3}\right)x_i \leq 0\,,$$

or

$$x_i = \tfrac{3}{2}, \qquad \tfrac{1}{n} \leq \tfrac{2}{3}\,.$$

The inequality is true for $n \geq 2$. Therefore the solution $x_i = 3/2$, $z_i = 0$ in (4) is verified. The y_i and u_i are easily computed.

Next consider the symmetric social optimum. Let $x_i = x$ and $z_i = z$ for all i. Then $\overline{z} = z$, and

$$u_i = x\,(1+z) - \tfrac{1}{3}\,(x+z)^2$$

for all i. The first-order conditions for maximization of u_i are

$$\begin{aligned} \frac{\partial u_i}{\partial x} &= 1 + z - \tfrac{2}{3}\,(x+z) = 0\,, \\ \frac{\partial u_i}{\partial z} &= x - \tfrac{2}{3}\,(x+z) = 0\,. \end{aligned}$$

These yield the solution $x = 2$, $z = 1$ in (5). The resulting y_i and u_i are easily found.

Next consider equilibria where people have the pro-social utility (6), with the same γ for all. The Kuhn-Tucker conditions for person 1 are:

$$\frac{\partial v_1}{\partial x_1} = 1 + \overline{z} - \tfrac{2}{3}\,(x_1 + z_1) \leq 0, \quad x_1 \geq 0\,, \tag{A.3}$$

$$\frac{\partial v_1}{\partial z_1} = \tfrac{1}{n}\,x_1 - \tfrac{2}{3}\,(x_1 + z_1) + \gamma\,\sum_{j=2}^{n} \tfrac{1}{n}\,x_j \leq 0, \quad z_i \geq 0\,, \tag{A.4}$$

with complementary slackness in each. Similar conditions obtain for the other individuals.

See if the selfish solution with $x_i > 0$, $z_i = 0$ still works. The conditions (A.3) and (A.4) become

$$1 - \tfrac{2}{3}\,x_i = 0, \qquad \tfrac{1}{n}\,x_1 - \tfrac{2}{3}\,x_1 + \gamma\,\sum_{j=2}^{n} \tfrac{1}{n}\,x_j \leq 0\,,$$

or

$$x_i = \tfrac{3}{2}, \qquad \tfrac{3}{2}\left(\tfrac{1}{n} - \tfrac{2}{3} + \gamma\tfrac{n-1}{n}\right) \leq 0\,.$$

The inequality becomes

$$\phi = \frac{1 + (n-1)\,\gamma}{n} \leq \frac{2}{3}\,,$$

which is equivalent to (7) in the text.

When this condition is not met, look for a symmetric Nash equilibrium with $x_i = x > 0$ and $z_i = z > 0$ for all i. The conditions (A.3) and (A.4) become

$$\begin{aligned} 1 + z &= \tfrac{2}{3}\,(x + z)\,, \\ \phi\, x &= \tfrac{2}{3}\,(x + z)\,. \end{aligned}$$

These yield the solution (9) in the text.

Writing u for the common level of selfish utility, it is then mechanical to verify

$$\frac{du}{d\phi} = \frac{8\,(1-\phi)}{(2-\phi)^2}\,.$$

Therefore u is an increasing function of ϕ over the range $(\frac{2}{3}, 1)$. Thus more pro-socialness achieves higher selfish utilities all round.

A.2 One-time education for pro-socialness

In the range $0 \le \phi \le 2/3$, it is obviously best to set $\phi = 0$ and get $U^a = 0.75\,(1+\delta)$. In the range $2/3 \le \phi \le 1$, use $\theta = (k/\delta)^{1/3}$, to substitute $k = \delta\,\theta^3$, and write the formula defining the function for $\phi \ge \frac{2}{3}$ as

$$f(\phi) = 0.75 + \delta\,\left[\frac{\phi\,(4-3\phi)}{(2-\phi)^2} - \frac{\theta^3}{1-\phi}\right]. \tag{A.5}$$

It is then mechanical to verify

$$f'(\phi) = \frac{\delta}{(1-\phi)^2}\,\left[\,8\,\left(\frac{1-\phi}{2-\phi}\right)^3 - \theta^3\,\right].$$

Therefore

$$f'(\phi) > 0 \quad \text{iff} \quad 2\,\frac{1-\phi}{2-\phi} > \theta, \quad \text{i.e.} \quad \phi < \phi^* = \frac{2\,(1-\theta)}{2-\theta}\,. \tag{A.6}$$

Therefore $f(\phi)$ is single-peaked, and its maximum occurs where $f'(\phi) = 0$, that is, at $\phi = \phi^*$. Substituting and simplifying, the maximum value is

$$f(\phi^*) = 0.75 + \delta\,\left(\,\theta^3 - 3\,\theta^2 + 1\,\right).$$

If this exceeds $0.75\,(1+\delta)$, then ϕ^* maximizes $f(\phi)$ in (13); otherwise $\phi = 0$ yields the maximum of $f(\phi)$.

We also need to restrict $\phi^* > 2/3$ to have an equilibrium that results in the utilities that enter the construction of $f(\phi)$. From the definition in (6), we see that $1 \geq \phi^* > 2/3$ corresponds to $0 \leq \theta < 1/2$. Now define

$$h(\theta) = \theta^3 - 3\,\theta^2 + 1 - 0.75 = \theta^3 - 3\,\theta^2 + 0.25\,.$$

We have

$$h'(\theta) = 3\;\theta^2 - 6\,\theta = 3\,\theta\,(\theta - 2)\,,$$

which is negative over the interval $(0, \frac{1}{2})$. Therefore $h(\theta)$ is a decreasing function throughout this range. Also $h(0) = 1/4 > 0$ and $h(\frac{1}{2}) = -3/8 < 0$. Therefore there is a unique θ^* in the interval such that $h(\theta) > 0$ for $\theta < \theta^*$ and $h(\theta) < 0$ for $\theta > \theta^*$. Numerical calculation shows that $\theta^* \approx 0.305$. This completes the proof of the statements in the text leading to (14).

A.3 Education for pro-socialness in steady state

Write the objective function in (A.7) as $v = n\,g(\phi)$ where

$$g(\phi) = \begin{cases} 0.75\;\phi - \frac{k\,\phi}{1-\phi} & \text{for } \phi < 2/3 \\ \frac{\phi^2\,(4-3\phi)}{(2-\phi)^2} - \frac{k\,\phi}{1-\phi} & \text{for } 2/3 < \phi < 1 \end{cases} \tag{A.7}$$

For $0 \leq \phi < 2/3$ we have

$$g'(\phi) = 0.75 - \frac{k}{(1-\phi)^2}$$

and

$$g''(\phi) = \frac{-2\,k}{(1-\phi)^3} < 0$$

Therefore $g(\phi)$ is concave in this range. Also

$$g'(0) = 0.75 - k > 0 \quad \text{if } k < 0.75$$

and

$$g'(2/3) = 0.75 - 9\,k < 0 \quad \text{if } k > 1/12$$

For $2/3 < \phi < 1$ we have, after some tedious algebra,

$$g'(\phi) = \frac{\phi\,(16 - 18\;\phi + 3\;\phi^2}{(2-\phi)^3} - \frac{k}{(1-\phi)^2}$$

Then

$$g'(2/3) = 3/2 - 9\,k > 0 \quad \text{if } k < 1/6$$

and

$$g'(\phi) \to -\infty \quad \text{as} \quad \phi \to 1$$

In general, $g(\phi)$ is not concave throughout the range $2/3 < \phi < 1$. (Specifically, the first term on the right hand side of (A.7) is convex for $2/3 < \phi < 4/5$ and concave for $4/5 < \phi < 1$, so $g(\phi)$ can be convex for a sub-range to the right of 2/3 if k is small.) But numerical calculation shows that

$$(1-\phi)^2\; g'(\phi) = \frac{\phi\,(1-\phi)^2\,(16 - 18\;\phi + 3\;\phi^2}{(2-\phi)^3} - k$$

is a decreasing function of ϕ. It equals $1/6 - k$ for $\phi = 2/3$ and $-k$ for $\phi = 1$. Therefore, if $k < 1/6$, $g'(\phi)$ is positive for $\phi <$ a critical value ϕ^* (which is uniquely defined for each k and of course depends on k), and negative for $\phi > \phi^*$. Then $g(\phi)$ is increasing for $2/3 < \phi < \phi^*$ and decreasing for $\phi^* < \phi < 1$, i.e. $g(\phi)$ has a unique interior maximum at ϕ^* in the range $2/3 < \phi < 1$.

Putting all this information together gives the statements in the text.