

Socializing Education and Pro-Social Preferences*

by
Avinash Dixit
Princeton University

First draft August 2009
This version November 1, 2009

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, social and developmental psychology, and political economy. Human societies make collective choices to devote educational resources toward formation of pro-social preferences in children. I construct a model for this process, focusing on the context of provision of a public good through private effort. I find that a positive threshold of pro-social attitude must be exceeded to achieve a positive quantity of the public good. Steady state equilibria of the intergenerational transmission process can have poverty traps or limit cycles.

Address of author:

Department of Economics, Princeton University, Princeton, NJ 08544-1021, USA.

Phone: 609-258-4013. Fax: 609-258-6419.

E-mail: dixitak@princeton.edu

Web: <http://www.princeton.edu/~dixitak/home>

* I am grateful for comments received during presentations at the Institute for Advanced Study at Princeton, the fifth CSEG-IGIER Symposium in Anacapri, and the Stockholm School of Economics. Comments from Karla Hoff, Robert Keohane, Simon Levin, Alessandro Lizzeri, Sten Nyberg, Marco Pagano, Deborah Prentice, and Jörgen Weibull were especially useful. I thank the National Science Foundation for research support.

1 Introduction

Until recently, most of economic theory 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's behavior, in many situations of social and economic interactions, reflects pro-social concerns such as empathy and fairness (Camerer 2003, chapter 2). The second is peculiar to economics; sociologists and social psychologists have always allowed and even emphasized that individuals' preferences are endogenous, influenced by the society in which they live and function. Some recent work in economics has explored the possibility that failure of the first assumption is intrinsically linked to failure of the second, that is, pro-social preferences are formed or evolve endogenously. In this paper I explore a link that differs from, but is complementary to, the ones usually studied.

Some economists object to this whole line of inquiry by arguing that it is outside the scope of economics. For them, the economist's job is to take preferences, no matter how they arise, as given and work out their consequences for decisions and equilibria. However, such a boundary between disciplines is artificial, and should be crossed when the needs of a particular research question require a broader perspective. More importantly, such a separation suffers from a form of the Lucas Critique. If some economic policy or other exogenous shock affects the process of preference formation, then an analysis of the policy or the shock that takes preferences as given will yield erroneous conclusions. For example, suppose a public good is being provided through voluntary contributions based on pro-social preferences. The equilibrium is still short of the first best, and the government contemplates supplementing the quantity through tax-based public provision. This may reduce or destroy the pro-social preferences as the public comes to think: "Let the government do it." Ignoring this endogeneity would yield an overestimate of the benefit from the policy. Bar-Gill and Fershtman (2005) construct a model showing this.

The standard approach to endogenizing pro-social preferences comes from an obvious analogy with selection in evolutionary biology. The puzzle that needs to be explained is this: someone who expends resources to benefit others will fare worse than another who looks out for oneself, and therefore the pressure of selection should 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, 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 has been analyzed theoretically (Bowles and Gintis 2004, Weibull and Salomonsson 2005).

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 genes which 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.

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) describes and discusses these findings. More recent experiments with similar findings are reported by Bereby-Meyer and Fisk (2009).

Developmental psychologists study the process in more detail. Children are born with a sense of empathy, but it is somewhat limited in scope. Children at the age of a few weeks or months exhibit empathetic distress: a child will cry if it hears another child crying. At the age of a year or two, they realize that other children are different persons: a child that sees another child in distress will bring its own mother, and later still the other child’s mother, to comfort it. But from this it is a long way to the recognition of others’ perspectives and entitlements, and the resulting pro-social preferences (internalization of fairness and equity) and actions (sharing, helping). Hoffman (2000, pp. 10-11) describes the process thus: “When a child experiences, repeatedly, the sequence of transgression followed by a parent’s induction¹ followed by child’s empathetic distress and guilt feeling, the child forms Transgression → Induction → Guilt scripts ... When a script is formed for the first time in an actual situation involving conflict with others, its motive component may not be strong enough to overcome the prospect of egoistic gain. But it may become strong enough with repetition, and when combined with cognitive development and peer pressure it 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

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 – parents may take the victim’s perspective and show how the child’s behavior harms the victim. ... [I]nductions communicate the parent’s disapproval of the child’s act and indicate 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)

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.”

Existing economic theoretical literature on the process of preference formation in children considers transmission from parent to child. However, such individual decisions and their equilibria cannot fully explain the phenomenon. A parent would not want its child to become the only pro-social actor in contexts like contributing to a public good 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. As a matter of reality, schools 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. My purpose in this paper is to construct a simple model of such collective action to instill pro-social preferences in children.

The paper is organized as follows. In the next section I briefly review some relevant literature and place my contribution in context. Section 3 constructs a baseline model on the assumption that all individuals are identical. Sections 4 and 5 consider various aspects of heterogeneity. Section 6 offers concluding remarks and suggestions for future research.

2 Some Related Literature

The literature closest to the tradition of economic analysis of individual preference and choice analyzes games between one’s present and future self. The current self takes foresighted strategic actions to alter choices of the future self. Kohn and Shavell (1974) model search that reveals possibilities and preferences currently unknown to oneself; the recommendation “try it, you’ll like it” fits this framework. Schelling’s (1984) analysis of commitment is followed by Becker and Murphy (1988) on rational choices to acquire or avoid addiction, Gul and Pesendorfer (2001) on self-control, and Bénabou and Tirole (2002) on strategic use of selective memory to acquire self-confidence or will-power.

At the next level, the actions to alter preferences are still taken by individuals, but these are influenced by, and in turn influence, the environment in which other agents are taking other actions, and the equilibria of such interactions are studied. Examples of such models are Gerber and Jackson (1993) where individuals’ political preferences respond to the competing parties’ ideological positions, Palacios-Huerta and Santos (2004) where individuals’ risk preferences are affected by market incompleteness, and Bénabou and Tirole (2003) where intrinsic motivation is affected by the availability of extrinsic incentives.

Next consider the intergenerational transmission of preferences. Bisin and Verdier (2001) construct a model where a child inherits a parent’s preferences with a probability that the parent can affect through costly effort, and with the complementary probability acquires preferences at random from the population. They show that the dynamics of such a system yields a polymorphic equilibrium with cultural heterogeneity. Under other assumptions, only one dominant culture may survive. This model has been modified and applied to many other contexts; for example Hauk and Saez-Marti (2002) use it to study propagation of a

culture of corruption. Other models where parents can directly transmit norms to their children include Lindbeck and Nyberg (2006) on work norms, and Tabellini (2008) on norms of honesty in contractual performance.

Bénabou and Tirole (2006) model the transmission of beliefs. The specific belief they consider is that of a “just world,” namely that personal effort is well rewarded by the market. A society where such a belief dominates will vote for low taxes. This will lead to an economic equilibrium in which the after-tax marginal product of personal effort is high, thereby confirming the belief in the next generation. Conversely, there can be a self-fulfilling belief that one’s personal reward from private effort is low, leading to a vote for more redistributive taxation. Thus the model combines actions at the level of the individual parent (belief formation) and the level of social choice (voting on taxation).

In all of these contributions, preferences are transmitted from parent to child on an individual basis. This mechanism is obviously important, but collective action has an additional role to play in the formation of preferences. Many social interactions are prisoners’ dilemmas or assurance games. In dilemma games, an individual parent would not want his or her own child to be a cooperator if all others cheat, while if all others are cooperators, an individual parent still wants his or her child to be a free rider. In assurance games there can be multiple equilibria, and the one where all cooperate is better for everyone. But for that outcome to be selected, expectations must converge (common knowledge) on the cooperative action, and collective action through schooling may be the best way to ensure this. Therefore a collective action mechanism for preference transmission should be examined, to supplement or complement individual parent-to-child mechanisms in the literature. In this paper I construct a simple exploratory model with that aim.

3 The Baseline Model

The model focuses on public good provision. Final output is produced using the public good and private effort; the two inputs are complements. The public good is provided through contributions of individuals. If individuals have a pro-social component in their preferences that internalizes the welfare of others to some extent, then the contribution equilibrium has a larger quantity of the public good. This raises the marginal product of private effort (the complements assumption), and therefore raises the level even of the part of each individual’s utility that depends on his/her own income and effort, which I call selfish utility.

The process of instilling pro-social preferences in children works through an interaction between two types of other-regarding preferences. One is each parent’s concern for his/her own children; this already exists as in the usual Ricardian or dynastic preference model. The second is inclusion of strangers’ welfare, which leads to greater provision of the public good and therefore to higher selfish utilities in the manner explained in the previous paragraph. Therefore instilling the second kind of other-regarding preference in the next generation will increase the (selfish) utility of each member of that generation. That will increase the dynastic utility of each member of the parent generation. But this requires collective action:

being the only pro-social person in its generation won't be good for one's child.²

The model is really an extended numerical example to develop this idea, with special functional forms and parameter values chosen to facilitate solution. But the intuitions it builds on and creates are appealing and the qualitative results should remain valid in more general conditions.

3.1 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 $i = 1, 2, \dots, n$. 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

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

Then the income of individual i is given by

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

Thus a higher \bar{z} raises the (average and) marginal product of each individual's private effort; this is the complementarity mentioned above.³

The private or selfish utility of i :

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

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

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

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.

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:

$$x_i = 2, \quad z_i = 1, \quad y_i = 4, \quad u_i = 1 . \quad (5)$$

² In fact each parent may be tempted to follow a Machiavellian strategy: vote for taxes to finance pro-social education in schools, and negate its effects on one's own child by telling it at home that the teaching in school is all nonsense and free-riding is optimal! I do not consider this strategy, mainly because it is unlikely to work. The influence of teachers and peer pressure seems to be stronger than that of parents. For example, children of smokers usually bring home the anti-smoking lessons they learn in school and urge their parents to quit; they are much less likely to tell their teachers or fellow students that their parents' behavior proves that smoking is just fine. For more systematic evidence on this, see Boocock and Scott (2005, Chapter 6).

³By using the average \bar{z} , not the total $\sum_{i=1}^n z_i$, in (1), I am assuming that the public effort is a public good with congestion. The no-congestion case is worth separate analysis.

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 I 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

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

and similarly for other individuals. If

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

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 $n \rightarrow \infty$: even in very large societies, the threshold is consistent regarding others' utility worth less than one's own.

Use the abbreviation

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

For large n , $\phi = \gamma$ approximately; I will usually focus on this case. Then (7) becomes simply $\phi \leq 2/3$.

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

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

The resulting pro-social utilities are

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

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.

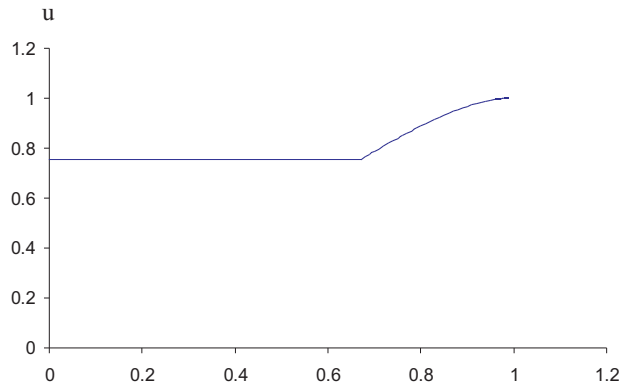


Figure 1: Selfish utility u as function of pro-socialness parameter ϕ

3.2 Choosing Children's Pro-socialness

Now introduce a succession of generations. Each individual has one child. Write u_i^a for the selfish utility of adult i , and u_i^c for that of his child; each is defined as in (3). The adult's selfish dynastic utility is defined by

$$U_i^a = u_i^a + \delta u_i^c . \quad (11)$$

By constructing pro-social utilities v_i^a and v_i^c from the u_i^a and u_i^c as in (6), we can define the adult's pro-social dynastic utility

$$V_i^a = v_i^a + \delta v_i^c . \quad (12)$$

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

$$t = \frac{k}{1 - \phi}, \quad \text{or} \quad \phi = 1 - \frac{k}{t} . \quad (13)$$

This has intuitively appealing properties: A threshold level k of expenditure is needed to instill any pro-social preference; the marginal cost of preference-formation is increasing; and it is infinitely costly to make each individual fully internalize social welfare.

I consider two models of socializing education. One regards it as purely instrumental. Even though each adult may have pro-social preferences instilled by the previous generation's education policy, when it comes to considering such policy for the next generation, the adult takes into account only his selfish dynastic utility. So the benefit of providing socializing education for the child's generation is the increase in the child's selfish utility as in (9), and the cost is the adult's personal tax payment.

In the second model, pro-socialness is comprehensive. Each adult takes into account his pro-social dynastic utility. Therefore the benefit of providing socializing education for the child's generation is the increase in the child's pro-social utility (including the child's internalization of the increases in the utilities of his peers through the child's pro-socialness

parameter γ^c , which is the object of the current choice), and the cost is the adult's pro-social valuation of the taxes paid by him and his peers (based on the adult's pro-socialness parameter γ^a , which was fixed by the education policy adopted by the previous generation).

Purely instrumental pro-socialness

Now consider two generations, and the parent generation's problem of choosing t , the level of educational expenditure per capita, for the children's generation. The resulting pro-socialness ϕ is given by (13). If t is so low that $\phi \leq 2/3$, then the equilibrium in the children's generation is the selfish (4), with utilities $u_i^c = 3/4$, contributing δ times that to the parent's dynastic utility (11). This is the same for all t , so t is best kept at 0. If t is high enough to achieve $\phi > 2/3$, the equilibrium is as in (9). The contribution to the parent's dynastic utility, net of the tax, is

$$\delta \frac{\phi(4-3\phi)}{(2-\phi)^2} - t = \delta \frac{\phi(4-3\phi)}{(2-\phi)^2} - \frac{k}{1-\phi}.$$

Therefore the parent's choice of t is equivalent to choosing ϕ to maximize the function $f(\phi)$ defined by

$$f(\phi) = \begin{cases} \frac{3}{4} \delta & \text{with } \phi = 0 < \frac{2}{3} \\ \delta \frac{\phi(4-3\phi)}{(2-\phi)^2} - \frac{k}{1-\phi} & \text{if } \phi > \frac{2}{3}. \end{cases} \quad (14)$$

The appendix verifies that this function is single-peaked. Since all parents are identical, the ϕ that maximizes each parent's selfishly dynastic utility net of the tax will also emerge as the Condorcet winner of their social choice.

The details of the solution are in the appendix. 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 $\delta < 35.2k$ approximately), it is optimal to choose $\phi = 0$. If $\theta < \theta^*$, that is, $\delta > 35.2k$, the optimum choice is

$$\phi = \frac{2(1-\theta)}{2-\theta}. \quad (15)$$

Thus a sufficiently patient parent generation can instill pro-social preferences in the next generation through education. The level instilled when δ just exceeds its threshold of $\approx 35.2k$ (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 becomes negligibly small, θ goes to zero and ϕ goes to 1; in this limit the full social optimum can be approached.

The pro-socialness parameter γ (or ϕ) instilled in the next generation for purely instrumental reasons depends on the current generation's dynastic preference parameter δ , but not on its own pro-socialness parameter. However, the more comprehensive pro-socialness considered in the second variant just below does so depend. Also, in Section 5 I consider the case where γ is transmitted from one generation to the next even in the absence of any schooling, but can be further increased by schooling, where such dependence will appear.

Comprehensive pro-socialness

The adult generation gets its pro-socialness parameter γ^a (or equivalently ϕ^a) from the education policy chosen by *its* parent generation; this is now exogenous. It chooses the education policy to instill the pro-socialness parameter γ^c (or equivalently ϕ^c) in its children. The cost of this is t^c per capita, given by (13). Then each child will have the pro-social utility v^c . The parent recognizes this benefit. Everyone in the adult generation pays the cost t^c , and each parent internalizes the others' cost using γ^a . Therefore the choice of education policy is made to maximize

$$\delta v^c - [1 + (n - 1)\gamma^a] t^c = \delta v^c - n \phi^a t^c = \delta v^c - n \frac{k \phi^a}{1 - \phi^c}.$$

For $\phi^c \geq 2/3$, v^c is calculated using (10). For $\phi^c < 2/3$, we have $u^c = 3/4$ from (4), and then

$$v^c = [1 + (n - 1)\gamma^c] \frac{3}{4} = n \phi^c \frac{3}{4}.$$

Therefore we can write the objective function as

$$g(\phi^c) = \begin{cases} n \left[\frac{3}{4} \delta \phi^c - \frac{k \phi^a}{1 - \phi^c} \right] & \text{if } \phi^c \leq \frac{2}{3} \\ n \left[\delta \frac{(\phi^c)^2 (4 - 3\phi^c)}{(2 - \phi^c)^2} - \frac{k \phi^a}{1 - \phi^c} \right] & \text{if } \phi^c > \frac{2}{3} \end{cases} \quad (16)$$

This is different in one respect from the objective (14) of the purely instrumental case: in the range $\phi^c \leq \frac{2}{3}$ it may be desirable to increase ϕ^c to give one's child greater pro-social utility. However, in the appendix we find that in this case it is desirable to push ϕ^c farther to the right, into the region where pro-socialness provides enough collective action to raise private utilities also.

A qualitative feature of the solution is easy to see: an increase in ϕ^a lowers the marginal product of ϕ^c , namely $g'(\phi^c)$. Therefore an adult generation with higher pro-socialness will instill lower pro-socialness in the child generation. This is because each adult attaches more importance to the taxes paid by his/her peers. However, numerical calculation shows this effect to be small. For example, with $\delta = 0.9$ and $k = 0.03$, an increase in ϕ^a from 0.80 to 0.90 reduces ϕ^c only from 0.87 to 0.86. If this process runs from one generation to the next, the pro-socialness parameter will quickly converge to a stationary level of 0.865, with an oscillatory dynamics like that shown for a different context in the left hand panel of Figure 6. Numerical solutions show that ϕ^c also responds quite slowly to changes in δ and k . For example, holding $k = 0.03$ and $\phi^a = 0.85$, as δ increases from 0.7 to 0.9, ϕ^c increases from 0.845 to 0.865. And holding $\delta = 0.9$ and $\phi^a = 0.85$, as k increases from 0.1 to 0.5 (a factor of five), ϕ^c decreases from 0.92 to 0.82.

Henceforth I will focus on the case where parents intend socializing education to create pro-socialness in their children for purely instrumental reasons, because that isolates in a minimal context the effect I wish to highlight here.

4 Heterogeneous δ

Now suppose different individuals have different dynastic preference parameters δ . They vote over the tax rate t , which will endow each member of the next generation with the pro-socialness parameter ϕ determined by the technology or cost function (13). For each δ , the parent's dynastic utility expressed as a function of ϕ , or equivalently the $g(\phi)$ defined in (14), is single-peaked. Therefore majority voting will result in implementation of the most preferred ϕ of the individual with the median δ . If this is below the threshold of about $36k$, no pro-social education will be undertaken. For higher median δ , it will be undertaken at a level determined by (15).

Theoretically more interesting than heterogeneity within the population is the possibility that δ depends on the parent generation's income y . Very poor parents may have to focus too much of their attention and resources on survival to think about making their children more pro-social. Or poor societies may have lower life-expectancies, reducing the expected value of a child in the calculation of the parent.⁴ These possibilities would make δ a decreasing function of y . It is also possible that poor parents attach greater importance to improving the well-being of their children, and therefore have higher δ . I now consider the intergenerational dynamics of pro-socialness and income that results from such dependence of δ on y .

To continue the numerical example in the simplest way, take $k = 0.01$. Then, corresponding to the critical value $\theta^* = 0.305$, we have the critical value of δ , given by

$$\delta^* = k / (\theta^*)^3 = 0.352.$$

If $\delta < \delta^*$, there is no expenditure on pro-social schooling; if $\delta > \delta^*$, there is positive expenditure of this kind. Figure 2 shows the resulting income per capita y as a function of δ :

Now consider an overlapping generation structure where each generation lives for two periods. It is born and educated in the first of these periods, and is economically active in the second. Denote each generation by the period in which it is economically active. Consider two successive periods t and $(t + 1)$. The income per capita, y_t , of the t -generation is determined by the pro-socialness that was instilled into it by the schooling it received in period $(t - 1)$. Its dynastic concern parameter for its children is denoted by δ_t . Let this be a function of the t -generation's income per capita: $\delta_t = h(y_t)$.

First consider the possibility that $h(y)$ is an increasing function. Let y^* be defined by $h(y^*) = \delta^* \approx 0.352$.⁵

⁴ In his study of the family in 16-18th century England, Stone (1977) gives a more powerful motive: "to preserve their mental stability, parents were obliged to limit the degree of their psychological involvement with their infant children. Even when children were genuinely wanted and not regarded as economically crippling nuisances, it was very rash for parents to get too emotionally concerned about creatures whose expectation of life was so very low" (p. 70) and "to maintain a psychological distance between parents and children, . . . [t]he methods included infantile abandonment to a wet-nurse" (p. 194).

⁵ I consider the "general case" where $0 < y^* < 4$; the extreme cases where $h(0) > \delta^*$ or $h(4) < \delta^*$ have trivial analyses.

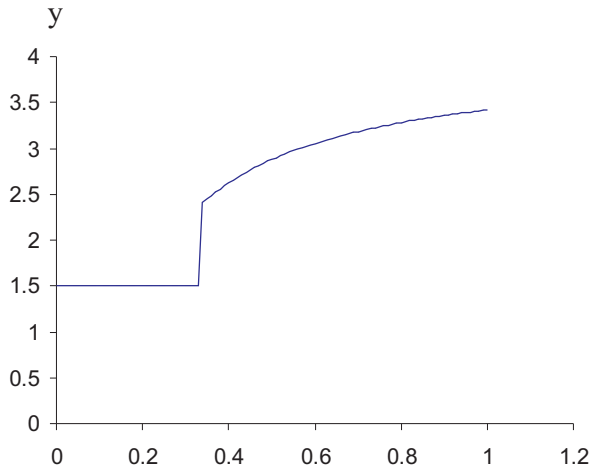


Figure 2: Income per capita y as function of δ

If $y_t < y^*$, then $\delta_t < \delta^*$. Generation t spends nothing on pro-social schooling for generation $(t + 1)$. Therefore $y_{t+1} = 1.5$, in the purely selfish Nash equilibrium of (4).

If $y_t > y^*$, then $\delta_t > \delta^*$. Generation t spends enough on pro-social schooling for generation $(t + 1)$ to raise its income according to the partially cooperative outcome of (9). Therefore y_{t+1} is an increasing function of y_t in this range.

Depending on the exact shape of the function $h(y)$, two cases now arise. If $y^* < 1.5$, then the dependence of y_{t+1} on y_t is as shown in Figure 3. This function intersects the 45-degree line at only one point to the right of 1.5, and its slope there is < 1 . Therefore we have a unique stable steady state with some pro-social education.

If $y^* > 1.5$, we have two stable steady states as shown in Figure 4: one without pro-social education and low income, the other with pro-social education and higher income. Thus we can have a poverty trap, where the parents are too poor to spend on pro-social education that would raise their children above similar poverty. This happens if the initial y_0 is below y^* . If the initial $y_0 > y^*$, then the dynamics converge to the upper steady state.

If $\delta_t = h(y_t)$ is a decreasing function, then y_{t+1} is also a decreasing function of y_t . This can cut the 45-degree line only once; therefore we have a unique steady state. But it need not be stable. As Figure 5 shows, there can be a limit cycle, where the even-period generations are rich and unconcerned about their children, reducing the odd-period generations to poverty, who in turn are determined to lift the following (even-period) generations out of poverty. Locally, the steady state can be stable or unstable depending on whether the slope of the curve showing y_{t+1} as a function of y_t is numerically less than or greater than 1. Figure 6 shows the possibilities.

These are interesting theoretical possibilities. Unfortunately it is not yet possible to judge their relevance, because I have not been able to find any empirical research examining whether and how dynastic preference varies with parental income or wealth. This can be a useful topic for future research, because its potential implications go beyond the context of education to instill pro-social preferences. More generally, parents invest in their children's

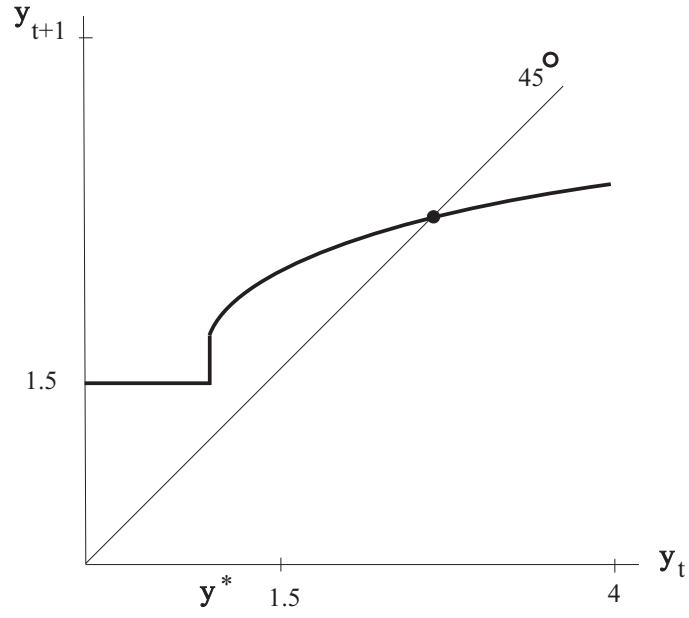


Figure 3: Unique steady state with pro-social education

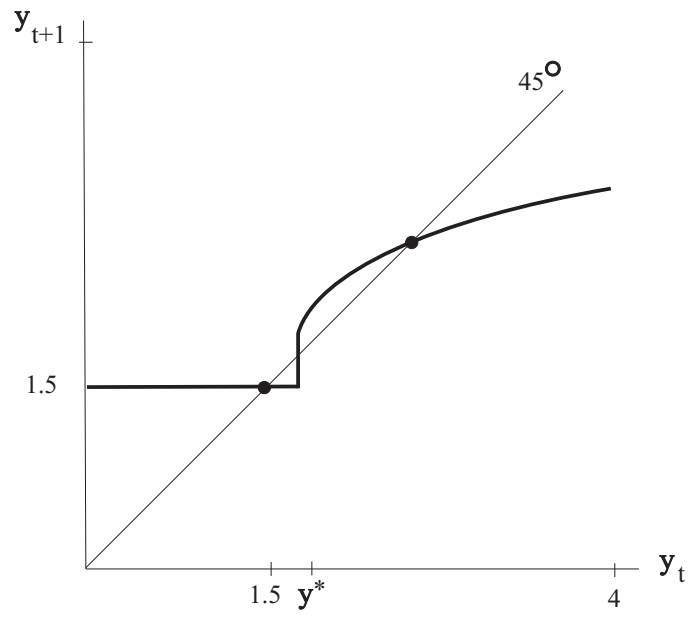


Figure 4: Multiple steady states and a poverty trap

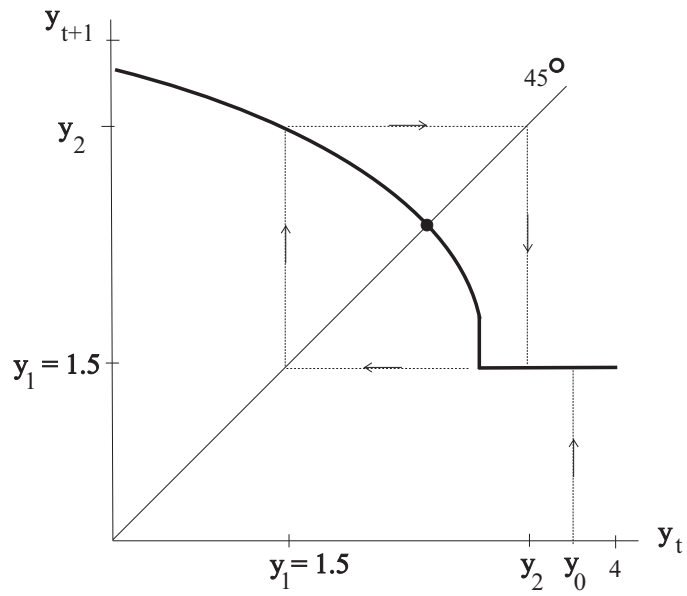


Figure 5: Limit cycle when δ is a decreasing function of y

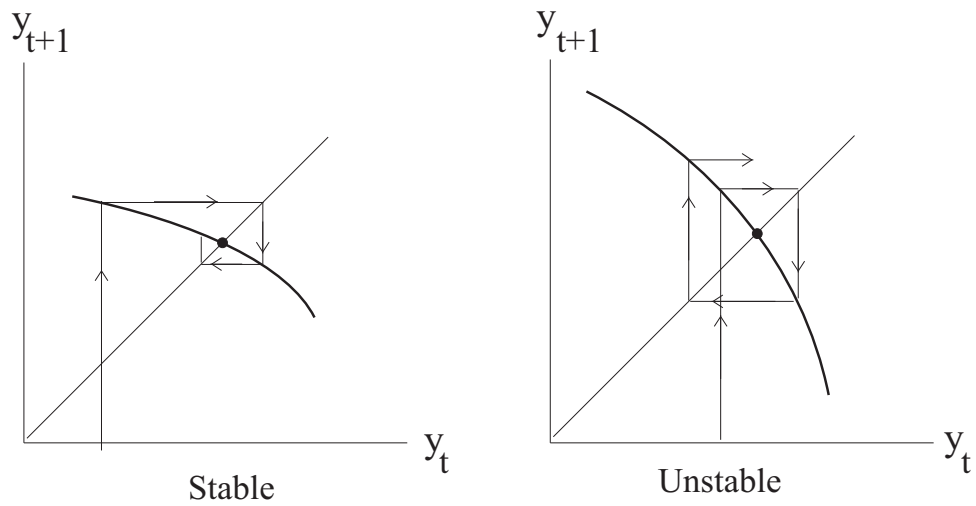


Figure 6: Local stability and instability when δ is a decreasing function of y

education to increase the children’s human capital. The resulting dynamics of income over generations can similarly exhibit multiple or unstable steady states, corresponding to poverty traps or to perpetuation of initial inequalities of income or wealth, or the phenomenon of “rags to rags in three generations.”

5 Heterogeneous γ

In this section I consider heterogeneity of the pro-socialness parameter γ in the population. To keep the calculations simple I assume that there are just two types L and H . A fraction θ_L of the population is L -type, and the fraction θ_H is H -type. The total population n is large. Each L -type has parameter γ_L , and each H -type has parameter γ_H .

$$\theta_L + \theta_H = 1 \quad \text{and} \quad 0 \leq \gamma_L < \gamma_H \leq 1.$$

Even within this simple dichotomy, to avoid tedious taxonomy I focus on only one kind of equilibrium, the one that seems conceptually the most interesting in this context. I stipulate that each L -type makes only private effort, so $x_L > 0$ and $z_L = 0$, while each H -type makes both types of effort, so $x_H > 0$ and $z_H > 0$.⁶

Once again the details of the calculation are in the appendix. It is found that for an equilibrium of this kind we need

$$\theta_H > \frac{3 \gamma_H - 2}{3 \gamma_H}, \quad \text{or} \quad \gamma_H \theta_L < 2/3. \quad (17)$$

The equilibrium effort levels are

$$x_L = \frac{3 \gamma_H}{2 - \gamma_H}, \quad (18)$$

$$x_H = \frac{1}{2 - \gamma_H} \left[3 \gamma_H - \frac{1}{\theta_H} (3 \gamma_H - 2) \right], \quad (19)$$

$$z_H = \frac{1}{\theta_H} \frac{3 \gamma_H - 2}{2 - \gamma_H}. \quad (20)$$

The resulting public effort averaged over the whole population is

$$\bar{z} = \theta_H z_H = \frac{3 \gamma_H - 2}{2 - \gamma_H}. \quad (21)$$

⁶ In the equilibria that arise if condition (17) is violated, the H -types are saints who make only public effort ($x_H = 0$). Totally forgoing private benefit seems unrealistic, except perhaps for very few people of a super- H -type. However, here is a brief statement about other equilibria; details are in the appendix. If $\gamma_H \theta_L > 2/3 > \gamma_L \theta_L$, the L -types make only private effort and H -types only public effort; if $\gamma_L \theta_L > 2/3$, the L -types make both types of effort and the H -types only public effort. Equilibria where both types make positive efforts of both kinds are not possible, for reasons broadly similar to those that yield corner solutions in contribution equilibria of pure public goods (e.g. Bergstrom, Blume and Varian 1986, p. 41).

The following features of the equilibrium should be noted:

[1] For $z_H > 0$ in (20), we need $\gamma_H > 2/3$, a condition we already know from Section 3.

[2] Equations (18)-(20) imply $x_L = x_H + z_H$. The two types make the same total effort; they just divide it differently between private and public components.

[3] The population average public effort \bar{z} is independent of θ_H . If there are fewer H -types, each of them makes proportionately more public effort to keep the population average unchanged. Combining this with points [1] and [2] immediately above shows why the lower bound (17) on θ_H is needed: if there are too few H -types, each of them has to make so much public effort that the private effort x_H of the H -types, given by (A.17), cannot stay non-negative.

[4] The equilibrium quantities are independent of γ_L (of course so long as $\gamma_L < \gamma_H$; otherwise the types are reversed). This is obvious; if the L -types are not making any public effort, the exact extent of their pro-socialness is immaterial.

The resulting utility levels of the two types are

$$u_L = 3 \left(\frac{\gamma_H}{2 - \gamma_H} \right)^2 \quad (22)$$

$$u_H = 3 \left(\frac{\gamma_H}{2 - \gamma_H} \right)^2 - \frac{1}{\theta_H} \frac{2\gamma_H(3\gamma_H - 2)}{(2 - \gamma_H)^2} \quad (23)$$

The H -types get a lower level of selfish utility because they are making less private effort. Of course they get higher v_H taking into account the pro-social component. However, u_H is still relevant as this is what concerns a parent deciding whether to vote for financing education that may make his/her child an H -type.

The L -types benefit from the pro-socialness of the H -types: u_L is an increasing function of γ_H . At the lower end of the range, $\gamma_H = 2/3$, we have $u_L = 3/4$, the same as in the purely selfish Nash equilibrium (4). At the upper end, $\gamma_H = 1$, we have $u_L = 3$, even higher than in the optimum (5) for a homogeneous population. This is because, as (21) shows, $\bar{z} = 1$ when $\gamma_H = 1$, so the L -types get the full benefit of the same average public effort as in (5) without themselves making any contribution to public effort.

Turning to the H -types, u_H is an increasing function of θ_H , because the same population-average public effort is shared by more people. The effect of γ_H is more complicated. It can be shown that

$$\frac{\partial u_H}{\partial \gamma_H} > 0 \quad \text{if and only if} \quad \theta_H > \frac{5\gamma_H - 2}{3\gamma_H}. \quad (24)$$

If there are too few H -types, each is making a high level of public effort. Then an increase in the pro-socialness parameter raises the burden even further at an increasing marginal cost. But if there are many H -types, then the burden of the public effort is mild, and the effect of an increase in it is more than offset by the higher productivity of private effort that results from greater pro-socialness. Figure 7 shows regions in (γ_H, θ_H) space corresponding to the two cases, and the separating curve between the two cases defined by (24). Note that we need $\gamma_H > 2/3$ and (17) for the equilibrium, so the shaded areas are ruled out.

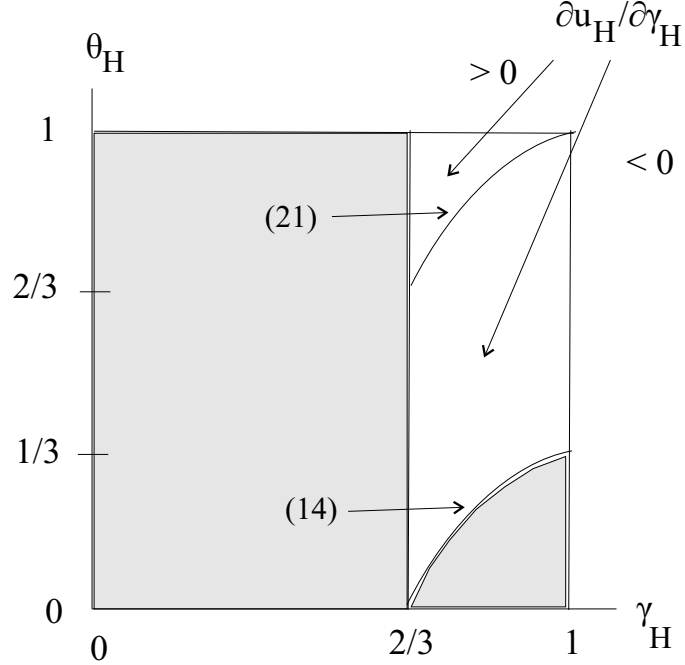


Figure 7: Effect of higher γ_H on u_H

The derivation above is for an equilibrium with a given structure of pro-social preferences. Now consider how this fits into the story of parents voting on educating the next generation. Many different specifications of the process are conceivable; to avoid tedious taxonomy I will outline only one. Suppose each parent has one child. In the absence of any socializing education, the child inherits the parent's pro-socialness parameter, γ_L or γ_H as the case may be. Education can increase the values of these parameters, or it can convert some L -types into H -types, that is, increase θ_H . For the parent, regardless of his/her own γ , the next generation's pattern of pro-socialness serves only the instrumental purpose of achieving a better equilibrium, so the parent cares only about his/her child's selfish utility.

The two types of parents now differ in their preferences about educating the next generation:

L -type parents want a higher γ_H . They are indifferent to an increase in θ_H if their own child stays L -type. But they are against education that creates a probability of their own child turning into an H -type, as that would decrease its selfish utility from (22) to (23).

H -type parents want the next generation to have a higher proportion of H -types, that is, they favor education that raises θ_H . They favor education that raises γ_H if θ_H is high enough to satisfy (24), and dislike it otherwise.

Both types of parents are indifferent about an increase in γ_L (so long as it stays $< \gamma_H$). Therefore neither type wants to spend on education that will increase the L -type children's pro-socialness only a little. The two will differ about a bigger increase that reverses the types. The H -type parents still regard their children's pro-socialness as instrumental, and will be glad to relieve them of the burden. The L -type parents will not want their children to become the new H -type. In any case, reversal of types seems unlikely to result from education unless

it can be specifically targeted to the L -type children, and that seems unrealistic.

All of this pertains to changes that preserve the type of equilibrium that is the focus of analysis in the text. If γ_H goes so high as to violate (17), then the H -types' children will become saints who make only public effort. Again regarding the pro-socialness as instrumental, the H -type parents will dislike this, and the L -type parents will like it.

These statements pertain to the benefit side in the parents' calculation. Their voting behavior will of course take into account the cost of education. But without doing this explicitly, it is easy to see how the votes will differ. The outcome of majority voting will depend on whether the parent generation's θ_H is greater than or less than $\frac{1}{2}$.

6 Concluding Comments

This has been an exploratory analysis of the formation of pro-social preferences through a collectively determined and costly activity of education in schools. I hope it has yielded some interesting results and has shown that the subject is amenable to modeling and potentially important. But the simple model should not be overinterpreted; many of its results may fail to survive generalization.

The 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 do not think that will produce any major new insights. Here are two suggestions for more substantive generalization.

[1] The literature on developmental psychology brings out the substantial uncertainty in the process by which children learn to recognize and respect others' perspectives and claims. Such uncertainty should be brought into the model. For example, there may be 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.

[2] More generally and most importantly, the model specifies the “technology” by which educational expenditures raise pro-socialness as a “black box” in (13). The model should be made richer and more realistic by closer integration with social and developmental psychology to improve the specification of this technology.

References

- Bar-Gill, Oren and Chaim Fershtman. 2005. "Public policy with endogenous preferences." *Journal of Public Economic Theory* 7(5), 841–8577
- Becker, Gary S. and Kevin M. Murphy. 1988. "A theory of rational addiction." *Journal of Political Economy* 96(4), August, 675–700.
- Bénabou, Roland and Jean Tirole. 2002. "Self-confidence and personal motivation." *Quarterly Journal of Economics* 117(3), August, 871–915.
- Bénabou, Roland and Jean Tirole. 2003. "Intrinsic and extrinsic motivation." *Review of Economic Studies* 70(3), July, 489–520.
- Bénabou, Roland and Jean Tirole. 2006. "Belief in a just world and redistributive politics." *Quarterly Journal of Economics* 121(2), May, 699–746.
- Bereby-Meyer, Yoella and Shelly Fisk. 2009. "Is homo economicus a five year old?" Ben Gurion University, working paper, available at <http://ssrn.com/abstract=1460482>.
- Bergstrom, Theodore, Lawrence Blume and Hal Varian. 1986. "On the private provision of public goods." *Journal of Public Economics* 29(1), February, 25–49.
- Bisin, Alberto and Thierry Verdier. 2001. "The economics of cultural transmission and the dynamics of preferences." *Journal of Economic Theory* 97(2), April, 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.
- 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.
- Gerber, Elisabeth R. and John E. Jackson. 1993. "Endogenous preferences and the study of institutions." *American Political Science Review* 87(3), September, 639–656.
- Gul, Faruk and Wolfgang Pesendorfer. 2001. "Temptation and self-control" *Econometrica* 69(6), November, 1403–1435.
- Hauk, Esther and Maria Saez-Marti. 2002. "On the cultural transmission of corruption." *Journal of Economic Theory* 10(2), December, 311–335.
- Hoffman, Martin L. 2000. *Empathy and Moral Development*. Cambridge, UK: Cambridge University Press.
- Kohn, Meir G. and Steven Shavell. 1974. "The theory of search." *Journal of Economic Theory* 9(2), October, 93–123.

- 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.
- Palacios-Huerta, Ignacio and Tano J. Santos. 2004. "A theory of markets, institutions, and endogenous preferences." *Journal of Public Economics* 88(3–4), March, 601–627.
- Rabin, Matthew. 1993. "Incorporating fairness into game theory and economics." *American Economic Review* 83(5), December, 1281–1302.
- Schelling, Thomas. 1984. *Choice and Consequence*. Cambridge, MA: Harvard University Press.
- Stone, Lawrence. 1977. *The Family, Sex and Marriage in England 1500–1800*. New York: Harper & Row.
- Tabellini, Guido. 2008. "The scope of cooperation: Values and incentives." *Quarterly Journal of Economics* 123(3), August, 905–950.
- Weibull, Jörgen and Marcus Salomonsson. 2006. "Natural selection and social preferences." *Journal of Theoretical Biology* 239(1), March, 79–92.

Appendix – Mathematical Derivations

Baseline Model

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

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

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

with complementary slackness in each equation. Note that in (A.2) I have used $\partial \bar{z} / \partial z_i = 1/n$. 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

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

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 $\bar{z} = 0$ also, and the conditions (A.1) and (A.2) are

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

or

$$x_i = \frac{3}{2}, \quad \frac{1}{n} \leq \frac{2}{3}.$$

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 $\bar{z} = z$, and

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

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

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

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:

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

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

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

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

or

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

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

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

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.

Purely Instrumental Pro-socialness

The policy choice was shown in the text to be the problem of choosing ϕ to maximize the $f(\phi)$ defined in (14). Define $\theta = (k/\delta)^{1/3}$, so $k = \delta \theta^3$, and write the formula defining the function for $\phi \geq \frac{2}{3}$ as

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

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

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

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

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

If this exceeds $\frac{3}{4}\delta$, then ϕ^* maximizes $f(\phi)$ in (14); otherwise 0 is 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

$$h(\theta) = \theta^3 - 3\theta^2 + 1 - \frac{3}{4} = \theta^3 - 3\theta^2 + \frac{1}{4}.$$

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 (15).

Comprehensive pro-socialness

Recall that the objective function is now given by (23), namely

$$g(\phi^c) = \begin{cases} n \left[\frac{3}{4} \delta \phi^c - \frac{k \phi^a}{1 - \phi^c} \right] & \text{if } \phi^c < \frac{2}{3} \\ n \left[\delta \frac{(\phi^c)^2 (4 - 3\phi^c)}{(2 - \phi^c)^2} - \frac{k \phi^a}{1 - \phi^c} \right] & \text{if } \phi^c \geq \frac{2}{3} \end{cases}$$

For $\phi^c < \frac{2}{3}$, we have

$$g'(\phi^c) = n \left[\frac{3}{4} \delta - \frac{k \phi^a}{(1 - \phi^c)^2} \right].$$

This is increasing over its range if

$$\frac{3}{4} \delta > 9k \phi^a, \quad \text{or} \quad \delta > 12k \phi^a.$$

This will be the case over the range of parameters of interest. However, for $\phi^c \geq \frac{2}{3}$,

$$g'(\phi^c) = n \left[\delta \phi^c \frac{16 - 18\phi^c + 3(\phi^c)^2}{(2 - \phi^c)^2} - \frac{k \phi^a}{(1 - \phi^c)^2} \right].$$

At $\phi^c = \frac{2}{3}$, this becomes

$$n [2\delta - 9k \phi^c].$$

The condition for $g(\phi^c)$ to be increasing as ϕ^c approaches $\frac{2}{3}$ from the left is then sufficient to ensure that $g'(\phi^c)$ is positive to the right of $\frac{2}{3}$. Therefore pro-socialness will be pushed farther. However, the first-order condition for the optimum, namely $g'(\phi^c) = 0$, is not analytically solvable. It is easy to see that $g'(\phi^c)$ is a decreasing function of ϕ^a , and therefore the optimal ϕ^c is a decreasing function of ϕ^a . Some numerical solutions are reported in the text.

Heterogeneous γ

With two types L, H and $\gamma_L < \gamma_H$, there are three types of equilibria to consider where the types show distinct behavior:

$$\begin{aligned} \text{Equilibrium type A:} & \quad x_L > 0, z_L = 0; & \quad x_H > 0, z_H > 0 \\ \text{Equilibrium type B:} & \quad x_L > 0, z_L = 0; & \quad x_H = 0, z_H > 0 \\ \text{Equilibrium type C:} & \quad x_L > 0, z_L > 0; & \quad x_H = 0, z_H > 0 \end{aligned}$$

Type A was the focus in the text in Section 5. Here I outline the derivation of all three types. There are also three logical possibilities where the two types exhibit the same behavior. The one where both types are purely selfish was the starting point in Section 3. The one where both types exert both types of effort will be seen to be impossible. The one where both types exert purely public effort is obviously impossible: even highly pro-social saints get no utility if no one else gets any private output.

In all three types of equilibria A, B, C, individual i chooses (x_i, z_i) to maximize social utility

$$v_i = (1 + \bar{z})x_i - \frac{1}{3}(x_i + z_i)^2 + \gamma_i \sum_{j \neq i} (1 + \bar{z})x_j.$$

Then

$$\frac{\partial v_i}{\partial x_i} = (1 + \bar{z}) - \frac{2}{3}(x_i + z_i), \quad (\text{A.7})$$

$$\begin{aligned} \frac{\partial v_i}{\partial z_i} &= \left[x_i + \gamma_i \sum_{j \neq i} x_j \right] \frac{1}{n} - \frac{2}{3}(x_i + z_i) \\ &= \frac{1}{n} [(1 - \gamma_i)x_i + \gamma_i X] - \frac{2}{3}(x_i + z_i), \end{aligned} \quad (\text{A.8})$$

where $X = \sum_j x_j$ including x_i , and I have used $\partial \bar{z} / \partial z_i = 1/n$.

Equilibrium of Type A:

Here each L -type individual chooses $x_L > 0, z_L = 0$, and each H -type chooses $x_H > 0, z_H > 0$. Using (A.7) and (A.8), the Kuhn-Tucker conditions are

$$(1 + \bar{z}) - \frac{2}{3}x_L = 0, \quad (\text{A.9})$$

$$\frac{1}{n} [(1 - \gamma_L)x_L + \gamma_L X] - \frac{2}{3}x_L \leq 0, \quad (\text{A.10})$$

$$(1 + \bar{z}) - \frac{2}{3}(x_H + z_H) = 0, \quad (\text{A.11})$$

$$\frac{1}{n} [(1 - \gamma_H)x_H + \gamma_H X] - \frac{2}{3}(x_H + z_H) = 0. \quad (\text{A.12})$$

Also,

$$X = n_L x_L + n_H x_H = n (\theta_L x_L + \theta_H x_H) = n \bar{x}, \quad (\text{A.13})$$

where \bar{x} denotes the average of the x 's in the population, and

$$\bar{z} = n_H z_H / n = \theta_H z_H. \quad (\text{A.14})$$

From (A.9) and (A.11) we have

$$x_L = x_H + z_H = \frac{3}{2} (1 + \bar{z}). \quad (\text{A.15})$$

Therefore

$$n_L X_L + n_H (x_H + z_H) = \frac{3}{2} (1 + \bar{z}) (n_L + n_H),$$

or

$$n \bar{x} + n \bar{z} = \frac{3}{2} n (1 + \bar{z}),$$

or

$$\bar{x} = \frac{3 + \bar{z}}{2}, \quad \text{or} \quad \bar{z} = 2\bar{x} - 3. \quad (\text{A.16})$$

Next, (A.11) and (A.14) imply

$$x_H + \frac{1}{\theta_H} \bar{z} = \frac{3}{2} (1 + \bar{z}), \quad \text{or} \quad x_H = \frac{3}{2} + \left(\frac{3}{2} - \frac{1}{\theta_H} \right) \bar{z}.$$

Combine (A.11) and (A.12) to write

$$\frac{1 - \gamma_H}{n} x_H + \gamma_H \bar{x} = 1 + \bar{z},$$

and use (A.16) to get

$$\frac{1 - \gamma_H}{n} x_H = 1 + \bar{z} - \gamma_H \frac{3 + \bar{z}}{2} = \left(1 - \frac{1}{2} \gamma_H\right) \bar{z} - \left(\frac{3}{2} \gamma_H - 1\right).$$

Substituting from the previous expression for x_H in terms of \bar{z} gives

$$\frac{1 - \gamma_H}{n} \left[\frac{3}{2} + \left(\frac{3}{2} - \frac{1}{\theta_H} \right) \bar{z} \right] = \left(1 - \frac{1}{2} \gamma_H\right) \bar{z} - \left(\frac{3}{2} \gamma_H - 1\right).$$

This yields a solution for \bar{z} in terms of exogenous parameters alone:

$$\bar{z} = \frac{\left(\frac{3}{2} \gamma_H - 1\right) + \frac{1}{n} \frac{3}{2} (1 - \gamma_H)}{\left(1 - \frac{1}{2} \gamma_H\right) - \frac{1 - \gamma_H}{n} \left(\frac{3}{2} - \frac{1}{\theta_H}\right)}.$$

If n is large, which is the case of particular interest in the context of public goods and is the assumption made in the text, this is approximated by

$$\bar{z} = \frac{\frac{3}{2} \gamma_H - 1}{1 - \frac{1}{2} \gamma_H} = \frac{3 \gamma_H - 2}{2 - \gamma_H}, \quad (\text{A.17})$$

and is guaranteed to be positive because $1 > \gamma_H > 2/3$.

Using the solution for \bar{z} we can successively find:

$$1 + \bar{z} = \frac{2\gamma_H}{2 - \gamma_H};$$

and from (A.16),

$$\bar{x} = \frac{3}{2} + \frac{1}{2} \frac{3\gamma_H - 2}{2 - \gamma_H} = \frac{2}{2 - \gamma_H}.$$

Next, from (A.9),

$$x_L = \frac{3\gamma_H}{2 - \gamma_H};$$

from (A.14),

$$z_H = \frac{1}{\theta_H} \frac{3\gamma_H - 2}{2 - \gamma_H};$$

and from (A.15),

$$x_H = \frac{1}{2 - \gamma_H} \left[3\gamma_H - \frac{1}{\theta_H} \frac{3\gamma_H - 2}{2 - \gamma_H} \right].$$

This is the solution (18)–(20) in the text. (Recall that for large n , we have $\phi \approx \gamma$.)

For $x_H > 0$ we need $\theta_H > (3\gamma_H - 2)/(3\gamma_H)$; this is the restriction on the parameter range for Type A equilibrium, and was stated as (17) in the text.

It remains to verify the Kuhn-Tucker inequality condition (A.10) for $z_L = 0$ to be optimal. Write it as

$$\frac{1 - \gamma_L}{n} x_L + \gamma_L \bar{x} - \frac{2}{3} x_L \leq 0.$$

For large n , this becomes

$$\gamma_L \bar{x} \leq \frac{2}{3} x_L,$$

or

$$\gamma_L \frac{2}{2 - \gamma_H} \leq \frac{2}{3} \frac{3\gamma_H}{2 - \gamma_H}$$

or simply $\gamma_L \leq \gamma_H$, which is true.

Now the expressions (22) and (23) in the text for utilities u_L and u_H can be found by substitution, and differentiation followed by tedious algebra gives

$$\frac{\partial u_H}{\partial \gamma_H} = \frac{4}{(2 - \gamma_H)^3} \left[3\gamma_H - \frac{1}{\theta_H} (5\gamma_H - 2) \right],$$

yielding the condition (24) in the text.

Equilibrium of Type B:

Here each L -type individual chooses $x_L > 0$, $z_L = 0$, and each H -type chooses $x_H = 0$, $z_H > 0$. Using (A.7) and (A.8), the Kuhn-Tucker conditions are

$$(1 + \bar{z}) - \frac{2}{3}x_L = 0, \quad (\text{A.18})$$

$$\frac{1}{n} [(1 - \gamma_L)x_L + \gamma_L X] - \frac{2}{3}x_L \leq 0, \quad (\text{A.19})$$

$$(1 + \bar{z}) - \frac{2}{3}z_H \leq 0, \quad (\text{A.20})$$

$$\frac{1}{n} \gamma_H X - \frac{2}{3}z_H = 0. \quad (\text{A.21})$$

Also,

$$X = n_L x_L = n \theta_L x_L, \quad \bar{x} = \theta_L x_L, \quad (\text{A.22})$$

and

$$\bar{z} = n_H z_H / n = \theta_H z_H. \quad (\text{A.23})$$

From (A.18) and (A.23), we have

$$1 + \theta_H z_H = \frac{2}{3}x_L,$$

and from (A.21) and (A.22),

$$\gamma_H \theta_L x_L = \frac{2}{3}z_H. \quad (\text{A.24})$$

Therefore

$$1 + \frac{3}{2}\gamma_H \theta_H \theta_L x_L = \frac{2}{3}x_L,$$

or

$$x_L = \frac{1}{\frac{2}{3} - \frac{3}{2}\gamma_H \theta_H \theta_L}.$$

The denominator of this is positive, because

$$\theta_H \theta_L \leq \theta_H (1 - \theta_H) \leq \frac{1}{4},$$

so

$$\frac{3}{2}\gamma_H \theta_H \theta_L \leq \frac{3}{2} \times 1 \times \frac{1}{4} = \frac{3}{8} < \frac{2}{3}.$$

Therefore the solution has $x_H > 0$.

Then using (A.24), we have

$$z_H = \frac{\frac{3}{2}\gamma_H \theta_L}{\frac{2}{3} - \frac{3}{2}\gamma_H \theta_H \theta_L}.$$

This is also positive.

It remains to verify the complementary slackness inequalities (A.19) and (A.20). Using (A.22), write (A.19) as

$$\left[\frac{1 - \gamma_L}{n} + \gamma_L \theta_L - \frac{2}{3} \right] x_L \leq 0.$$

For large n , this is simply

$$\gamma_L \theta_L \leq \frac{2}{3}.$$

Next, using (A.23), write (A.20) as

$$1 + \theta_H \frac{\frac{3}{2} \gamma_H \theta_L}{\frac{2}{3} - \frac{3}{2} \gamma_H \theta_H \theta_L} \leq \frac{2}{3} \frac{\frac{3}{2} \gamma_H \theta_L}{\frac{2}{3} - \frac{3}{2} \gamma_H \theta_H \theta_L}.$$

Simplifying the left hand side and canceling the (positive) denominator from both sides, we have

$$\frac{2}{3} \leq \gamma_H \theta_L.$$

Thus the conditions for the Type B equilibrium are

$$\gamma_L \theta_L \leq \frac{2}{3} \leq \gamma_H \theta_L.$$

These are as stated in the text Footnote 5.

Equilibrium of Type C:

Here $x_L > 0$, $z_L > 0$ and $x_H = 0$, $z_H > 0$. Using (A.7) and (A.8), the Kuhn-Tucker conditions are

$$(1 + \bar{z}) - \frac{2}{3}(x_L + z_L) = 0, \quad (\text{A.25})$$

$$\frac{1}{n} [(1 - \gamma_L) x_L + \gamma_L X] - \frac{2}{3}(x_L + z_L) = 0, \quad (\text{A.26})$$

$$(1 + \bar{z}) - \frac{2}{3} z_H \leq 0, \quad (\text{A.27})$$

$$\frac{1}{n} \gamma_H X - \frac{2}{3} z_H = 0. \quad (\text{A.28})$$

Also,

$$X = n_L x_L = n \theta_L x_L, \quad \bar{x} = \theta_L x_L, \quad (\text{A.29})$$

and

$$\bar{z} = (n_L z_L + n_H z_H) / n = \theta_L z_L + \theta_H z_H. \quad (\text{A.30})$$

Using (A.29), we have (A.26) as

$$\left[\frac{1 - \gamma_L}{n} + \gamma_L \theta_L - \frac{2}{3} \right] x_L = \frac{2}{3} z_L, \quad (\text{A.31})$$

while (A.28) and (A.29) yield

$$\gamma_H \theta_L x_L = \frac{2}{3} z_H. \quad (\text{A.32})$$

Now use (A.30) in (A.25), and substitute for z_L and z_H using the above two equations, to express everything in terms of x_L :

$$\begin{aligned} 0 &= 1 + \theta_L z_L + \theta_H z_H - \frac{2}{3}(x_L + z_L) \\ &= 1 + \left(\theta_L - \frac{2}{3}\right) z_L + \theta_H z_H - \frac{2}{3} x_L \\ &= 1 + \left(\theta_L - \frac{2}{3}\right) \frac{3}{2} \left[\frac{1 - \gamma_L}{n} + \gamma_L \theta_L - \frac{2}{3} \right] x_L + \theta_H \frac{3}{2} \gamma_H \theta_L x_L - \frac{2}{3} x_L \\ &= 1 + \left\{ \left(\theta_L - \frac{2}{3}\right) \left[\frac{3}{2} \frac{1 - \gamma_L}{n} + \frac{3}{2} \gamma_L \theta_L - 1 \right] + \frac{3}{2} \gamma_H \theta_L \theta_H - \frac{2}{3} \right\} x_L. \end{aligned}$$

This lets us solve for x_L in terms of the exogenous parameters. For large n , the long expression in the brackets multiplying x_L on the right hand side simplifies to

$$\begin{aligned} \left(\theta_L - \frac{2}{3}\right) \left[\frac{3}{2}\gamma_L\theta_L - 1\right] + \frac{3}{2}\gamma_H\theta_L\theta_H - \frac{2}{3} &= \frac{3}{2}\gamma_L(\theta_L)^2 - \theta_L - \gamma_L\theta_L + \frac{2}{3} + \frac{3}{2}\gamma_H\theta_L\theta_H - \frac{2}{3} \\ &= -\left[1 + \gamma_L - \frac{3}{2}(\gamma_L\theta_L + \gamma_H\theta_H)\right] \theta_L \end{aligned}$$

Then

$$x_L = \frac{1}{\left[1 + \gamma_L - \frac{3}{2}(\gamma_L\theta_L + \gamma_H\theta_H)\right] \theta_L}. \quad (\text{A.33})$$

For a meaningful solution, the denominator of this should be positive. To verify this, note that for large n , (A.31) becomes

$$\frac{2}{3}z_L = \left[\gamma_L\theta_L - \frac{2}{3}\right] x_L.$$

For both z_L and x_L to be positive, we need $\gamma_L\theta_L > 2/3$. Therefore this is going to be a condition for the Type C equilibrium. (And it fits exactly with the parameter ranges we have obtained so far for equilibria of Types A and B.) Now $\theta_L + \theta_H = 1$, so

$$\gamma_L\theta_L + \gamma_H\theta_H < \max(\gamma_L, \gamma_H) = \gamma_H \leq 1.$$

Therefore

$$1 + \gamma_L - \frac{3}{2}(\gamma_L\theta_L + \gamma_H\theta_H) \geq 1 + \gamma_L - \frac{3}{2} = \gamma_L - \frac{1}{2}.$$

And

$$\gamma_L\theta_L > \frac{2}{3} \quad \text{implies} \quad \gamma_L > \frac{2}{3} \frac{1}{\theta_L} > \frac{2}{3} > \frac{1}{2}.$$

This proves the desired result.

With the solution (A.33) for x_L , we can use (A.31) and (A.32) to complete the solution:

$$z_L = \frac{\frac{3}{2}\gamma_L\theta_L - 1}{\left[1 + \gamma_L - \frac{3}{2}(\gamma_L\theta_L + \gamma_H\theta_H)\right] \theta_L},$$

and

$$z_H = \frac{\frac{3}{2}\gamma_H\theta_L}{\left[1 + \gamma_L - \frac{3}{2}(\gamma_L\theta_L + \gamma_H\theta_H)\right] \theta_L}.$$

Then

$$\bar{z} = \theta_L z_L + \theta_H z_H = \frac{\frac{3}{2}\theta_L(\gamma_L\theta_L + \gamma_H\theta_H) - \theta_L}{\left[1 + \gamma_L - \frac{3}{2}(\gamma_L\theta_L + \gamma_H\theta_H)\right] \theta_L},$$

and

$$1 + \bar{z} = \frac{\gamma_L\theta_L}{\left[1 + \gamma_L - \frac{3}{2}(\gamma_L\theta_L + \gamma_H\theta_H)\right] \theta_L}.$$

It remains to verify (A.27) for the optimality of $x_H = 0$. Comparing the expressions above for $1 + \bar{z}$ and z_H , the inequality becomes

$$\gamma_L\theta_L \leq \frac{2}{3} \frac{3}{2}\gamma_H\theta_L, \quad \text{or} \quad \gamma_L \leq \gamma_H$$

which is true.

Impossibility of Interior Equilibrium:

If all of x_L, z_L, x_H, z_H are positive, the Kuhn-Tucker conditions become simple Lagrangian equations. Using (A.7) and (A.8), we have

$$(1 + \bar{z}) - \frac{2}{3}(x_L + z_L) = 0, \quad (\text{A.34})$$

$$\frac{1}{n} [(1 - \gamma_L)x_L + \gamma_L X] - \frac{2}{3}(x_L + z_L) = 0, \quad (\text{A.35})$$

$$(1 + \bar{z}) - \frac{2}{3}(x_H + z_H) = 0, \quad (\text{A.36})$$

$$\frac{1}{n} [(1 - \gamma_H)x_H + \gamma_H X] - \frac{2}{3}(x_H + z_H) = 0. \quad (\text{A.37})$$

Also,

$$X = n_L x_L + n_H x_H = n(\theta_L x_L + \theta_H x_H), \quad \bar{x} = \theta_L x_L + \theta_H x_H, \quad (\text{A.38})$$

and

$$\bar{z} = (n_L z_L + n_H z_H) / n = \theta_L z_L + \theta_H z_H. \quad (\text{A.39})$$

From (A.34) and (A.36) we have

$$x_L + z_L = x_H + z_H = \frac{3}{2}(1 + \bar{z}).$$

Therefore

$$n_L(x_L + z_L) + n_H(x_H + z_H) = \frac{3}{2}(1 + \bar{z})(n_L + n_H).$$

Using (A.38) and (A.39), this becomes

$$n\bar{x} + n\bar{z} = \frac{3}{2}(1 + \bar{z})n,$$

or

$$\bar{x} = \frac{3}{2} + \frac{1}{2}\bar{z}, \quad \text{or} \quad \bar{z} = 2\bar{x} - 3. \quad (\text{A.40})$$

Combine (A.34) and (A.35), and use (A.40) to write

$$\frac{1 - \gamma_L}{n} x_L + \gamma_L \bar{x} = 1 + \bar{z} = 2\bar{x} - 2.$$

Then

$$\frac{1}{n} x_L = \frac{1}{1 - \gamma_L} [(2 - \gamma_L)\bar{x} - 2]. \quad (\text{A.41})$$

Similarly for x_H . Therefore

$$\begin{aligned} \frac{1}{n} \bar{x} &= \frac{1}{n} (\theta_L x_L + \theta_H x_H) \\ &= \theta_L \frac{1}{1 - \gamma_L} [(2 - \gamma_L)\bar{x} - 2] + \theta_H \frac{1}{1 - \gamma_H} [(2 - \gamma_H)\bar{x} - 2] \\ &= \left[\theta_L \frac{2 - \gamma_L}{1 - \gamma_L} + \theta_H \frac{2 - \gamma_H}{1 - \gamma_H} \right] \bar{x} - \left[\theta_L \frac{2}{1 - \gamma_L} + \theta_H \frac{2}{1 - \gamma_H} \right]. \end{aligned}$$

This yields the solution for \bar{x} in terms of exogenous parameters:

$$\bar{x} = \frac{\theta_L \frac{2}{1-\gamma_L} + \theta_H \frac{2}{1-\gamma_H}}{\theta_L \frac{2-\gamma_L}{1-\gamma_L} + \theta_H \frac{2-\gamma_H}{1-\gamma_H} - \frac{1}{n}}.$$

Substituting back into (A.41) and simplifying,

$$\frac{1-\gamma_L}{n} x_L = \frac{\frac{2\theta_H(\gamma_H-\gamma_L)}{1-\gamma_H} + \frac{2}{n}}{\theta_L \frac{2-\gamma_L}{1-\gamma_L} + \theta_H \frac{2-\gamma_H}{1-\gamma_H} - \frac{1}{n}}.$$

Similarly,

$$\frac{1-\gamma_L}{n} x_H = \frac{\frac{2\theta_L(\gamma_L-\gamma_H)}{1-\gamma_L} + \frac{2}{n}}{\theta_L \frac{2-\gamma_L}{1-\gamma_L} + \theta_H \frac{2-\gamma_H}{1-\gamma_H} - \frac{1}{n}}.$$

This is impossible for large n : since $\gamma_L - \gamma_H < 0$, the numerator on the right hand side becomes negative, and then x_H becomes negative (and large in absolute value).