School Attendance and Child Labor – A Model of Collective Behavior

Holger Strulik*

First Version: November 2009. Final Version: April 2011

Abstract. This article investigates how community attitudes affect school attendance and child labor and how aggregate behavior of the community feeds back on the formation of schooling attitudes. The theory takes aggregate and idiosyncratic poverty into account as an important driver of absence from school and provides an explanation for why equally poor villages or regions can display very different attitudes towards schooling. Distinguishing between three modes of child time allocation, school attendance, work, and leisure, the article shows how child labor productivity and the time costs of schooling contribute to the existence of a locally stable anti-schooling norm and how policy can exploit social dynamics and help a community to escape permanently from low attendance at school and child labor.

Keywords: School Attendance, Child Labor, Social Norms, Targeted Transfers

JEL: I20 I29 J13 O12

*University of Hannover, Wirtschaftswissenschaftliche Fakultaet, 30167 Hannover, Germany; email: strulik@vwl.uni-hannover.de. I would like to thank Arnab Basu, Arjun Bedi, Karl Dietrich, Katja Michaelowa, Jean-Philippe Platteau, and participants at the ESPE conference, London, at the VfS Conference on Development Economics and Policy, Zurich, and at the EUDN Conference, Paris, and two anonymous referees for their useful comments.

1. Introduction

About 69 million school-age children are not in school, according to the most recent UN statistics, and the Millennium development goal that by 2015 children everywhere, boys and girls alike, will be able to complete a full course of primary schooling is "unlikely to be met" (UN, 2010). At the same time about 152 million children ages 5 to 14 are engaged in child labor (ILO, 2010). It is tempting to employ the child time budget constraint and argue that the children are not in school because they have to work in order to supplement family income and this is indeed the main mechanism explored so far in the economics literature on child labor and schooling.

If working and attending school were the only activities of children and if child labor were merely a side-effect of poverty than it would be sufficient to address policy to poverty alleviation and solve along-side the child labor and education problem. There would be no need for a separate development goal for education. Consequently, anti child labor interventions have been motivated in the economics literature mainly by the existence of multiple equilibria that occur if children compete with adults for wage work. When there are at the same time two equilibria available, one with and one without child labor, a ban on child labor or a boycott of child labor products can be desirable as a coordination device as shown in Basu's (1998) seminal article. Subsequent research, however, has argued that when international trade, price discrimination, skilled and unskilled labor, or political economy aspects are taken into account, bans and boycotts are less desirable than originally thought and probably detrimental to the fight against child labor (see e.g. Basu and Zarghamee, 2009, Doepke and Zilibotti, 2010).

The present article uses the existence of multiple equilibria as well to motivate policy intervention but it is based on a novel mechanism: the influence of community approval of schooling on the division of children's time on work, education, and leisure and play. The new framework helps to explain several phenomena that have been neglected by the conventional literature, for example, why working children are not attending school although this would be compatible with their weekly work load, why school attending children continue to work, why school attendance varies largely between regions of similar poverty and child labor regulations, and why supply-side policies, like the provision of a local school, are successful in drawing children to school.[1]

In order to address these questions the present article sets up a dynamic model of a community

[1]Across countries, the majority of working children works less than 24 hours per week (UCW, 2010), which should, in principle, be compatible with attendance at school (see Edmonds, 2007). For differences across similar regions, see Ray (2000) on Peru and Pakistan or Kingdon (2005) on India where 98 percent of the girls are enrolled in middle school in Kerala but only 36 percent in Rajasthan, a state of similar poverty. For the impact of local schools on attendance, see for example, Sawada and Lokshin (2009) and Burde and Linden (2010). See Basu and Tzannatos (2003), Edmonds (2007) and UCW (2010) for surveys on child labor issues, child time allocation, and the role of norms.

in which the decision of parents about the allocation of their children's time is – besides the usual economic determinants – also dependent on the community's approval of schooling. The realized attendance at school in turn determines the evolution of the schooling attitude in the community. Households are heterogenous with respect to poverty and with respect to susceptibility to community approval. As a result it remains to be true that poverty is an important driver of absence from school. But the link between poverty and child labor is less clear cut than in the conventional model. Children of some relatively poor families attend school while other children of relatively rich parents are absent. Depending on the environment (model parameters) school attending children may continue or stop working. More importantly the twofold heterogeneity of households creates social dynamics according to which similarly poor communities can end up at different locally stable equilibria: at one equilibrium all children are going to school and the community sustains a pro-schooling norm whereas at the other equilibrium many children are not going to school and the community sustains an anti-schooling norm.

Empirically, norms, traditions, and beliefs are hard to measure quantitatively and econometric exercises trying to provide evidence for true community effects are frequently plagued by the self-selection problem (Manski, 1993). In support of the proposed theory there exists, however, ample narrative evidence. One general conclusion is that the perceived anti-schooling norm is strongest with respect girls. For example, the UNICEF (2008) report on schooling in Sudan concludes:

Cultural barriers to schooling might present the biggest challenge especially in trying to achieve gender parity in education, but yet they have to be addressed if progress is to be achieved in this crucial area. Most cultural barriers manifest themselves as attitudes and practices that have gained acceptance over long periods of repeated practice and approval.

That gender-biased anti-schooling attitudes can be shared by a community at large is also impressively demonstrated in Davison and Kanyuka's (1992) study on education in rural Malawi where 68 percent of fathers and 70 percent of mothers believed that boys are more intelligent than girls, 60 percent of the teachers stated that girls are lazy, and 85 percent of the girls themselves thought that boys are more intelligent.[2]

But community dynamics work also in the other direction, towards higher attendance at school. This has been shown by two recent studies of PROGRESA. Both Lalive and Cattaneo (2009) and Bobonis and Finan (2009) find that schooling-contingent subsidies did not only increase school attendance of the children of targeted households but also that of children form ineligible households

[2]On a more general level, neighborhood peer effects on schooling have been shown by Zelizer (1985), Crane (1991), Case and Katz (1991), and Kling et al. (2007), and with focus on developing countries by Jensen and Nielsen (1997), Binder (1999), Dreze and Kingdon (2001), and Chamarbagwala (2009).

when the program was introduced in their community. Besides demonstrating the existence of such a "social multiplier" (Glaeser et al., 2003) the studies provide a couple of other interesting results, which will be addressed by the theory developed below. They found that poorer families are more strongly affected by the behavior of peers, that the social interaction runs through the parents (rather than the children) and that the strength of the response of ineligible families is increasing in the share of eligible families.

Generally, schooling attitudes and norms could not only be influenced by peers and community behavior, as assumed here, but also intergenerationally by the effort that parents make in shaping their children's preferences (vertical socialization). Modeling vertical socialization is in particular relevant when the subject of investigation is cultural heterogeneity. As Bisin and Verdier (2001) have shown the stable coexistence of a majority trait and a minority trait requires perpetual engagement in vertical socialization. In a homogenous society, in contrast, vertical socialization is unnecessary. In this sense the present article considers a sufficiently homogenous community where parents share the same preferences about child work, leisure, and education. Parents are thus not interested in shaping their children's *preferences* but in adjusting their *behavior* according to the social norm (Granovetter, 1978, Bernheim, 1994).

The present article continues a small but growing literature on the role of norms in economics. Problems addressed so far include the growing welfare state (Lindbeck, et al., 1999), out-of-wedlock childbearing (Nechyba, 2001), family size (Palivos, 2001), women's labor force participation (Hazan and Maoz, 2002), corruption (Hauk and Saez-Marti, 2002), occupational choice (Mani and Mullin, 2004), contraceptive use (Munshi and Myaux, 2006), work effort (Lindbeck and Nyberg, 2006), patience (Doepke and Zilibotti, 2008), cooperation in prisoner's dilemmas (Tabellini, 2008), and doping in sports (Strulik, 2011). The studies of Lopez-Calva (2002) and Kirchsteiger and Sebald (2010) are particularly related to the present work.

Lopez-Calva (2002) introduces a social norm in Basu and Van's (1998) framework, in which child labor is supplied to firms for a salary. The social norm is attached to child labor, which is considered to be "inherently bad" although the individual disutility that a parent experiences from sending a child to work decreases with aggregate child labor supply. Parents are assumed to be homogenous and education is not considered. In a static framework it is shown that two equilibria exist: one equilibrium without child labor and high stigma and another equilibrium in which all children supply market labor and stigma is low. For a given set of parameters both equilibria are equally likely such that the child labor decision is reduced to a coordination problem. In contrast to Basu and Van's finding, when social norms matter, multiple equilibria occur independently from the shape of the labor demand curve, even in the case when child labor does not depress adult wages.

Kirchsteiger and Sebald (2010) investigate a setup of vertical socialization, in which better educated parents experience higher utility from educating their children. Child labor is not considered. They show that multiple equilibria exist for a society consisting of dynasties of either high or low innate talent. In particular it is possible that either the dynasty of low talent or both dynasties are stuck in an equilibrium without education. With respect to policy they conclude that compulsory schooling is necessary to escape from illiterateness. A subsidy alone could not initiate a transition towards education of all groups.

The present article extends the above literature by investigating how schooling attitudes affect behavior when all three possible modes of child time allocation – (domestic) work, school attendance, and leisure and play – are taken into account. Since targeted subsidies have been successful in getting children to school and have initiated social dynamics through peers, it seems appropriate to assume that schooling attitudes are affected by the community (rather than by vertical socialization as in Kirchsteiger and Sebald, 2010). Since the assumption that children are either working fulltime or not at all will be abandoned, it seems appropriate to attach the social norm to schooling (rather than to child labor as in Lopez-Calva, 2002). A schooling norm can be more easily developed and sustained because schooling is a binary decision (sent the girl to school or not) whereas child labor is a continuous variable to which it is harder to attach a specific normative value. Furthermore, the schooling decision is very visible within the community. Children show up at school, wear uniforms and satchels, wait with fellow pupils at the bus stop etc. Whether and how long children work cannot so easily be monitored by others, in particular if they work at home in the household or on the family farm.

Indirectly, however, schooling attitudes affect child labor supply. In particular, it will be shown that depending on the environment (the specification of model parameters) schooling affects child labor either on the intensive margin or on the extensive margin. On the intensive margin school attending children continue to supply labor and spend less time working. On the extensive margin school attending children stop working because their parents want to leave them some hours of leisure and play after school. Distinguishing the impact of policy on the intensive and extensive margin leads to some non-obvious policy conclusions. For example, increasing the length of the school day is helpful to reduce child labor on the intensive margin. On the extensive margin, however, if schooling becomes too time-intensive, parents may prefer to withdraw their children from school, in particular when child labor is valuable and schooling gets little approval from the community.

The setup of a dynamic community of heterogenous households allows us to investigate issues that have not been addressed by the so far available literature. The dynamic setup is useful to distinguish between attitudes (current beliefs) and norms (persistent attitudes). The two-fold heterogeneity of households is useful to explain the social multiplier observed in conjunction with the evaluation of PROGRESA, the local stability of an anti-schooling norm, and how its breakdown can be initiated by education supply policies or income subsidies. Section 4 and 5 provide two examples for how the social multiplier can be exploited. Section 4 shows how a two-step education supply policy can reverse the schooling norm and eradicate child labor. Section 5 explores the design and costs of conditional transfers. Derivations of equations and non-obvious proofs of lemmata and propositions can be found in the Appendix.

2. The Model

2.1. **Setup of Society and Family Constraints.** Consider a community populated by a continuum of families indexed by i. Suppose that each family consists of one parent and one child. The daily time budget of children is normalized to one and parents decide how it is spent. At any day t child time may be spent on attending school $h_t(i)$, work $\ell_t(i)$, or leisure and play. Attending school can be modeled conveniently as a binary choice, yes ($h_t(i) = 1$) or no ($h_t(i) = 0$). Child labor, in contrast, is better conceptualized as a continuous variable. Parents decide not only whether at all but also how long their children have to work. Let the time cost of attending school including walking or driving to school be given by τ. The implied child time budget constraint for family i at day t is then given by (1).

$$1 - \ell_t(i) - \tau h_t(i) \geq 0. \tag{1}$$

Suppose that adults allocate all of their time – net of household duties – to work. Without child labor household i earns income $w(i)$. How child labor enters the budget constraint depends on whether the child earns a market wage or whether it is occupied with unpaid family work, i.e. helping on the family farm, the family business, or doing household chores. In order to avoid distracting case differentiation this article focusses on unpaid family work, which is empirically the most prevalent form of child labor.[3]

Suppose that a child working ℓ_t time units sets free $\gamma\ell_t$ units of extra time of the parent, $0 \leq \gamma \leq 1$, and that the extra time is used to supply additional adult labor and earn income. Income is used

[3]Of the child laborers world-wide 64 percent of boys and 73 percent of girls are unpaid family workers. 23 percent of boys and 19 percent of girls are in paid employment (ILO, 2010). A webpage appendix demonstrates robustness of results with respect to wage work. It is available at `http://kaldor.vwl.uni-hannover.de/holger/research/index.php`.

to finance adult consumption $c_t(i)$ so that family i's budget constraint is given by (2).[4]

$$w(i) + \gamma w(i)\ell_t(i) = c_t(i). \tag{2}$$

2.2. **Preferences.** Utility of any adult i consists of a "normal" or autonomous component $u_t^a(i)$ and a socially-dependent component $u_t^s(i)$. In solving the time allocation problem parents consider the trade-off between the empathetically perceived utility from child leisure and the children's contribution to family income. They take also into account the empathetically perceived utility from letting their children attend school and the community's evaluation of such a decision.

Suppose the autonomous part of utility can be captured by the following CES function.

$$u_t^a(i) = c_t(i)^{1-\alpha} \cdot [1 - \ell_t(i) - \tau h_t(i)]^{\alpha} \cdot [1 + a h_t(i)]\,. \tag{3}$$

Here the parameter $\alpha \in (0,1)$ measures empathy, i.e. the relative weight that parents put on child leisure relative to own consumption. For $\alpha \to 0$ totally insensitive parents care only about consumption (and the evaluation of their behavior by the community). As α rises parents increasingly suffer from letting their children work. The parameter a measures how parents evaluate schooling versus leisure time of their children, $a > 0$. Since school attendance h_t is either 0 or 1, the associated exponent in the utility function is normalized to unity without loss of generality. The preference for child leisure distorts the usually assumed mutual exclusiveness of schooling and working and allows us to distinguish between effects on the intensive and extensive margin. On the intensive margin, if the time cost of schooling (τ) increases, it is partly "financed" by less leisure and play. But since parents want to leave their children some leisure time, $1 - \ell_t - \tau h_t > 0$. This implies that, on the extensive margin, a certain time cost of schooling below $1 - \ell_t$ is already sufficiently high such that children are not sent to school.

The individual schooling decision is evaluated by the community. Although the model is more general and gender is nowhere explicitly stated, community attitudes can be introduced probably best with a special focus on schooling of girls. Let ϕ denote the *maximum* strength of an anti-schooling norm ("girls don't belong in school") which results when indeed there are no girls in the community attending school, $0 \leq \phi \leq 1$. Thus, ϕ is the stigma cost imposed on the first family of the community whose daughter is allowed to attend school. Let S_t denote the strength of community support for a

[4]An alternative assumption providing similar results with more notational clutter would be that income is spent on consumption of adults and children and that adults take utility from child consumption (imperfectly) into account. Also, we could alternatively assume that adults spend the time saved through helping children on (utility enhancing) leisure or – with a focus on daughters – that parents have one or several sons if preference parameters are such that sons are attending school.

pro-schooling decision in period t such that the actual evaluation of a pro-schooling decision is given by $(S_t - \phi)$. When S_t exceeds ϕ disapproval turns into approval. Over time, community approval of schooling depends on the actual attendance at school. In this section, however, we focus on the static decision problem for predetermined S_t.

The model does not provide a deeper explanation for why an anti-schooling norm existed in the first place. Formally, ϕ is exogenously given by history. For example, in the past there was no school within daytime reach ($\tau > 1$), or there was no school for girls, or there was no separate toilet for girls at school. The fact that no girl attended school in the initial state then "explains" the initial strength of the community's attitude that girls do not belong in school.[5]

People are to different degrees susceptible to the evaluation of their behavior by others. We capture this fact by assuming that susceptibility to approval is uniformly and independently distributed within the unit interval. A parent i has susceptibility to approval $\sigma(i) \in [0,1]$. The most self-assured parent is assigned with value 0 and the most indeterminate one with value 1. Summarizing, the social component of utility experienced by parent i for a "pro-schooling" decision is given by (4).

$$u_t^s(i) = \sigma(i) \cdot (S_t - \phi) \cdot h_t(i). \tag{4}$$

Note that the "character trait" of susceptibility to approval is heterogenous across the community but individually given while the social evaluation of an individual action is the same for all community members but endogenously explained and time-variant.

2.3. **Child Labor and Schooling.** Putting the autonomous and social elements of utility together and assuming separability, i.e. $u_t(i) = u_t^a(i) + u_t^s(i)$, parent i at day t solves the problem

$$\max_{c_t(i),\ell_t(i),h_t(i)} u_t(i) = c_t(i)^{1-\alpha}(1 - \ell_t(i) - \tau \cdot h_t(i))^{\alpha}(1 + a \cdot h_t(i)) + \sigma(i) \cdot (S_t - \phi) \cdot h_t(i). \tag{5}$$

subject to the budget constraint (2). The solution with respect to child labor is given by (6).

$$\ell_t(i) = \max\left\{0,\ \ (1-\alpha)(1 - \tau h_t(i)) - \frac{\alpha}{\gamma}\right\}. \tag{6}$$

Child labor depends positively on child productivity γ and negatively on preference for child leisure α. At the interior solution, schooling *partly* substitutes child labor. Note that child labor is always smaller than one. Only totally insensitive parents would let their children work all of the time ($\ell \to 1$ for $\alpha \to 0$ and $h_t = 0$). Empathetic parents leave their children some spare time. For

[5]In the spirit of North (1990) it can be imagined that the "institution of daughters not attending school" was once functional in the history of a traditional society, at a time when the skill premium was small and daughters were seen as a household duties performing asset, sold at marriage; see Dreze and Sen (1995) and Kingdon (2009).

children to work at all, child labor productivity must be sufficiently high in comparison to preference for child leisure. Some of a child's daily time is used for work, if

$$\gamma(1-\tau h_t) > \frac{\alpha}{1-\alpha}. \tag{7}$$

Since the condition may hold for $h_t = 0$ but not for $h_t = 1$, we have to distinguish three cases:

case WW**:** for $\gamma(1-\tau) > \alpha/(1-\alpha)$: school attending children continue to work.

case NW**:** for $\gamma \geq \alpha/(1-\alpha) \geq \gamma(1-\tau)$: school attending children stop working.

case NN**:** for $\alpha/(1-\alpha) > \gamma$: children are not working regardless of the schooling decision.

High child labor productivity γ increases the parameter domain of case WW and decreases the domain of case NN whereas low time cost of schooling τ increases the parameter domain of case NW at the expense of case WW.

The cases WW and NW are perhaps the most interesting ones from the developing countries' viewpoint. But there are also incidences reported supporting the NN case of completely idle children (Biggeri et al., 2003, Bacolod and Ranjan, 2008). Besides the natural explanation of low productivity, possibly due to insufficient health, captured by low γ in (7), the present article provides another explanation for the "puzzle of idle children": a prevailing anti-schooling attitude in the community.

Insert (6) into (2) to get consumption and then insert $\ell_t(i)$ and $c_t(i)$ into (5) to get indirect utility $\tilde{u}_t(i)$ for a given schooling decision $h_t(i)$. The child of family i is attending school if $\tilde{u}_t(i)|h_t(i) = 1 > \tilde{u}_t(i)|h_t(i) = 0$. Evaluate and compare utilities for the three cases from above to find that the child of family i is attending school if

$$E_j \cdot w(i)^{1-\alpha} \geq \sigma(i)\cdot(\phi - S_t), \qquad j \in \{WW, NW, NN\}, \tag{8}$$

$$E_{WW} = \{[1+\gamma(1-\tau)]\, a - \gamma\tau\}\,(1-\alpha)^{1-\alpha}\cdot\left(\frac{\alpha}{\gamma}\right)^{\alpha}$$

$$E_{NW} = (1-\tau)^{\alpha}(1+a) - (1-\alpha)^{1-\alpha}(1+\gamma)\cdot\left(\frac{\alpha}{\gamma}\right)^{\alpha}$$

$$E_{NN} = (1-\tau)^{\alpha}(1+a) - 1.$$

The left hand side of (8) reflects the utility gain experienced from a pro-schooling decision while the right hand side reflects the utility loss experienced from community disapproval of schooling. Ceteris paribus the utility gain from schooling is lower for poorer families. The constant schooling propensity E_j, summarizes all decision-relevant parameters. It provides the utility gain experienced from a pro-schooling decision that an additional unit of weighted income $w(i)^{1-\alpha}$ brings about. Taking the first derivative of the E_j's with respect to a and τ provides the following result.

Lemma 1. *$\partial E_j/\partial a > 0$, $\partial E_j/\partial \tau < 0$, $\qquad j \in \{WW, NW, NN\}$.*

This means that the schooling propensity E_j is high (and for given income and community approval the child is more likely attending school) if the individual evaluation of schooling is high and if schooling is not time intensive (τ is low). An immediate corollary is that making schooling more time consuming in order to eradicate or reduce child labor, could be a bad idea. Some parents may indeed let their children work less, as indicated by (6). Other parents, however, react with withdrawing their children from school, as indicated by (8) and Lemma 1.

The comparative statics with respect to child labor productivity reveals a non-monotony.

Lemma 2. *$\partial E_{NW}/\partial\gamma < 0$ and, if $\gamma < \bar{\gamma} \equiv \alpha a/[(1-\tau)(1+a)-1]$, $\partial E_{WW}/\partial\gamma < 0$. If $\gamma > \bar{\gamma}$, $\partial E_{WW}/\partial\gamma > 0$.*

Naturally, if school attending children stop working (case NW), higher child labor productivity increases the opportunity costs of schooling and reduces the schooling propensity. If child labor productivity is sufficiently small, this "normal" reaction is also evoked in the WW case, in which school attending children continue to work. On the other hand, if children are very productive, $\gamma > \bar{\gamma}$, increasing child labor productivity further *raises* the schooling propensity.

The result with respect to E_{WW} is explained by two countervailing forces of child labor productivity on schooling. First, higher child labor productivity increases the opportunity cost of schooling, which leads for itself to a lower schooling propensity. This effect is reflected by the term $(\alpha/\gamma)^\alpha$ of E_{WW} in (8). Second, higher child labor productivity increases family income and consumption, which in turn reduces the need for child labor and increases the schooling propensity. This effect is reflected by the first term of E_{WW} in (8). Ceteris paribus, the positive effect through the first term dominates if child labor productivity is high.

While discriminating child labor productivity (at the extreme through an effective ban) increases the schooling propensity in some cases, the model points also to a case for which just the opposite is true. Discriminating child labor productivity reduces the schooling propensity if school attending children continue to work and if their productivity is sufficiently high (their contribution to family income is sufficiently important).

The strength of the schooling propensity differs, depending on the mode of child time allocation:

Lemma 3. *$E_{NW} \leq E_{WW}$. $E_{NW} < E_{NN}$.*

This means that the schooling propensity is lowest when school attending children stop working (case NW). The schooling propensity is higher if parameters are such that school attending children continue to contribute to family income (case WW) or if children do not work regardless of the schooling decision (case NN).

In order to avoid uninteresting case differentiation we assume that parameters are such that the schooling propensity is always positive, $E_j > 0$ for all j. This implies that we know that, if a child is not attending school, the cause is a prevailing anti-schooling norm in the community (the right hand side of (8) is positive and larger than the left hand side). In the Appendix it shown that the following assumption guarantees a positive schooling propensity, $E_j > 0$ for all j.

Assumption 1 (Schooling Feasibility). $a > \bar{a} \equiv \alpha\tau/(1-\tau)$.

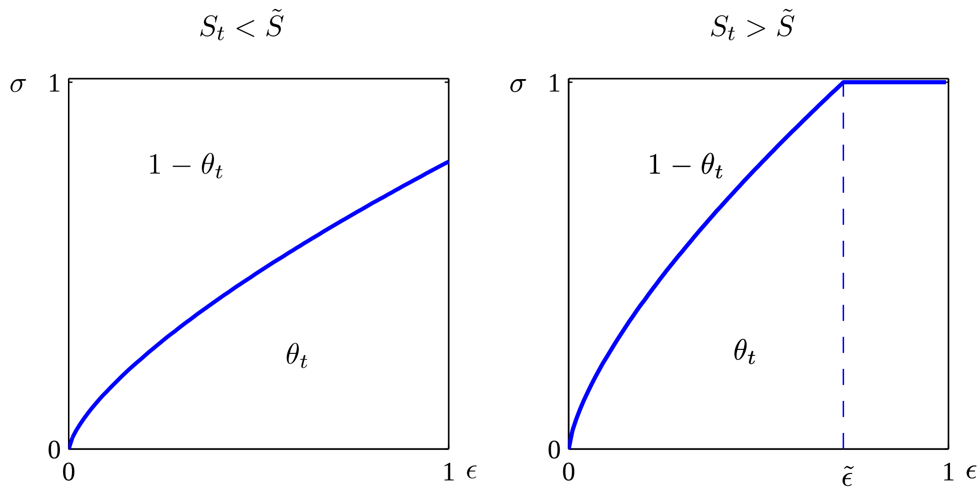
The condition requires that parents experience sufficient utility from an educated child. Naturally, for $E_j > 0$ the weight of education in utility (a) has to be larger if parents put a lot of weight on child leisure (α is large) or if schooling takes a lot of time (τ is large). Schooling feasibility is assumed to hold henceforth. Applying it in (8) produces the familiar result that, ceteris paribus, children of poor families (of low $w(i)$) are less likely attending school. The mechanism at work, however, is different from the poverty channel established in the conventional literature. Here, poor parents, given their low income, achieve lower autonomous utility such that the social evaluation of their schooling decision plays a relatively larger role in their decision making. The result gets support from empirical studies on the PROGRESA program which find that the schooling decision of poorer parents is more easily influenced by their peers (Bobonis and Finan, 2009).

2.4. **How Many and Whose Children Are Going to School?** In the following we distinguish between aggregate productivity (of the whole community or region) and idiosyncratic productivity (of a specific family i) and assume that wages are given by $w(i) = \epsilon(i)A$. In order to allow for an analytical solution we assume that the income index $\epsilon(i)$ is uniformly and independently distributed within the unit interval, $\epsilon(i) \in (0,1)$. Employing the new notation we can restate condition (8). The child of family i is attending school if $A^{1-\alpha}E_j/(\phi - S_t)\epsilon(i)^{1-\alpha} > \sigma(i)$. Diagrammatically, a child is attending school if the family's (ϵ, σ) tuple lies below the $\sigma(\epsilon)$ threshold (9).

$$\sigma(\epsilon) = \frac{A^{1-\alpha}E_j}{\phi - S_t} \cdot \epsilon^{1-\alpha}, \quad j \in \{WW, NW, NN\}. \tag{9}$$

Because the right hand side of (9) is an increasing function of aggregate and idiosyncratic productivity, the model supports, in principle, the conventional result that higher income affects education positively. Besides income, however, the schooling decision depends also on the strength of community approval and the individual susceptibility to the evaluation by others. This fact generates a less clear-cut association between income, education, and child labor. In particular, some relatively rich but highly approval-dependent parents are not sending their children to school if the community evaluates this negatively. On the other hand, some relatively poor yet independently-minded parents let their children attend school because they do not care much about the prevalent anti-schooling attitude in their community.

Figure 1: School Attendance Threshold



Families are distinguished by income index $\epsilon \in [0,1]$, with $\epsilon = 1$ identifying the richest parent and by susceptibility to community approval $\sigma \in [0,1]$, with $\sigma = 1$ reflecting the highest susceptibility. The share of children attending school is θ_t. If community approval gets sufficiently high, all parents above a critical income level $\tilde{\epsilon}$ let their children attend school. $S_t < \phi$ in both panels.

Figure 1 visualizes these results by showing the threshold in an $\epsilon-$ σ–diagram. The area below the threshold gives the share of children who are attending school at day t, denoted by θ_t. The figure illustrates that for a *given* value of σ it is always the children of poor parents who are not participating in school. But because σ is distributed independently from poverty, there may be also some rich parents who leave their children at home. If community approval is sufficiently high, all sufficiently rich parents let their children attend school. Formally, there exists an $\tilde{\epsilon}$ for which $\sigma(\tilde{\epsilon}) = 1$, i.e. $\tilde{\epsilon} = \left[(\phi - S_t)/E_j\right]^{1/(1-\alpha)}/A$. For $\tilde{\epsilon}$ to be in the relevant range, i.e. $\tilde{\epsilon} < 1$, community approval of schooling has to be sufficiently high, $S_t > \tilde{S} \equiv \phi - A^{1-\alpha}E_j$, see the right panel of Figure 1. If S_t is even higher such that $S_t > \phi$ all children in the community are attending school. Diagrammatically, the slope of the threshold becomes infinite for $S_t \to \phi$.

The simple structure of the model allows us to compute analytically the area below the threshold. Taking the piece-wise definition of the threshold into account we arrive at the following attendance rate.

$$\theta_t = \theta(S_t) = \begin{cases} \frac{1}{2-\alpha}\frac{A^{1-\alpha}E_j}{\phi - S_t} & \text{for } S_t \leq \tilde{S} \equiv \phi - A^{1-\alpha}E_j \\ 1 - \frac{1-\alpha}{2-\alpha}\frac{1}{A}\left(\frac{\phi - S_t}{E_j}\right)^{\frac{1}{1-\alpha}} & \text{for } \phi > S_t \geq \tilde{S} \\ 1 & \text{for } S_t \geq \phi. \end{cases} \tag{10}$$

Inspecting (10) and using Lemma 1 and 2 proves the following comparative statics.

PROPOSITION 1. *$\partial\theta_t/\partial S_t > 0$, $\partial\theta_t/\partial A > 0$, and $\partial\theta_t/\partial\tau < 0$. For case NW and if $\gamma < \bar{\gamma}$ for case WW, $\partial\theta_t/\partial\gamma < 0$. For case WW, if $\gamma > \bar{\gamma}$, $\partial\theta_t/\partial\gamma > 0$.*

The attendance rate is increasing in community support S_t and in average productivity (income) of the community A and it is decreasing in the time-cost of schooling τ. If parameters are such that attending children stop working, the attendance rate is decreasing in child labor productivity. If attending children continue to work, attendance is decreasing in child labor productivity if productivity is low and increasing otherwise.

Using Lemma 3 we verify that the slope of the threshold is flattest for the *NW* case, which proves the following result.

PROPOSITION 2. $\theta(S_t, \alpha, \phi, A, E_{NW}) < \theta(S_t, \alpha, \phi, A, E_j), \qquad j \in \{WW, NN\}$.

This means that, ceteris paribus, the attendance rate is lowest if parameter values of child productivity γ and schooling intensity τ are such that school attending children stop working. To understand this result note how the indivisible nature of the schooling activity, i.e. participating or not, affects child labor at the extensive margin. The higher the time cost of schooling the larger the parameter domain supporting the *NW* case. If schooling is sufficiently time consuming, for example, because the next school is far away, sufficiently empathetic parents would not demand any work from their children after school and leave them some time for leisure and play instead. However, in this case parents are also, ceteris paribus, more reluctant to let their children attend school in the first place since they would forego any child labor contribution to family income. As a consequence the schooling propensity is relatively low and thus a relatively large share of families in the community prefers that their children are not attending school.

3. The Evolution and Persistence of Community Norms

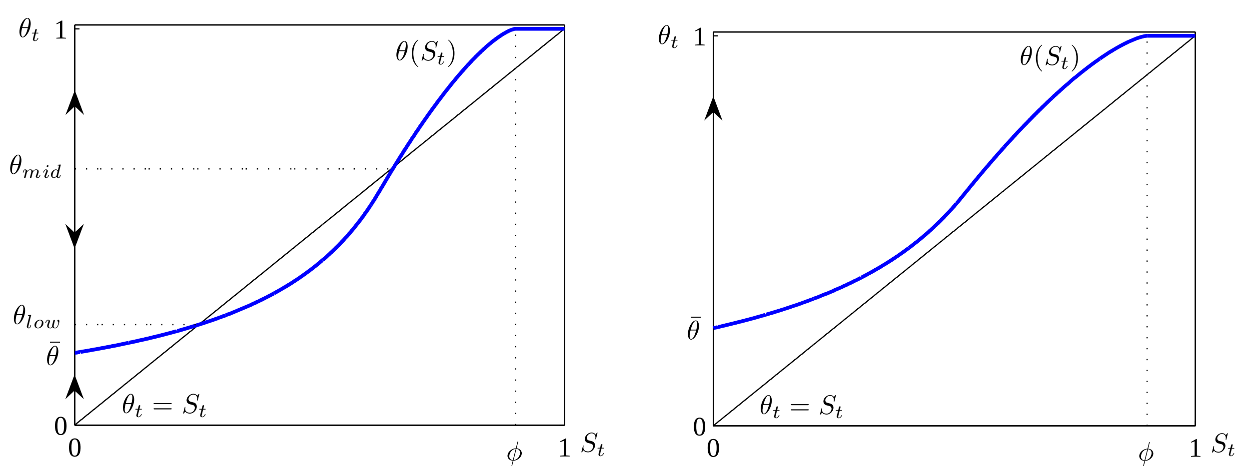
3.1. **The Evolution of Community Approval.** A community's evaluation of behavior is not a given constant but evolves as a lagged endogenous variable depending on the observation of actual community behavior. The results obtained so far were just providing a snapshot of the community at a given time t. In order to proceed we assume that the strength of community support S_t depends positively on the attendance rate in the past. Let δ denote the time preference rate or rate of oblivion by which past observations are depreciated in mind so that approval is given by $S_t = (1-\delta)\sum_{i=0}^{\infty}\delta^i\theta_{t-1-i}$. Alternatively, this can be written in period-by-period notation (11).

$$S_t = (1-\delta)\cdot\theta_{t-1} + \delta\cdot S_{t-1}. \tag{11}$$

A social equilibrium is obtained where attendance stays constant and accords with the community attitude to schooling, i.e. $\theta_t = S_t$. To find these fixed points, inspect (10) to see that $\theta(S_t)$ originates at $\bar{\theta} > 0$, is continuous in $(0,1)$ and increasing. It is convex for small $S_t \leq \tilde{S}$ (i.e. for $\tilde{\epsilon} = 1$), concave for large $S_t \geq \tilde{S}$ (i.e. for $\tilde{\epsilon} < 1$), and equal to unity if S_t exceeds ϕ. These features imply that in an $S_t - \theta_t$ diagram, the $\theta(S_t)$ curve intersects the identity line either twice (as shown in the left panel of Figure 2) or not at all (as shown in the right panel).

If the curve does not intersect the identity line, the only long-run equilibrium is at $\theta_t = S_t = 1$. A possibly observable anti-schooling attitude in the community is not persistent. Even if the community starts out at a situation of high disapproval of schooling, the attendance rate is always greater than the one needed to sustain the currently prevailing anti-schooling attitude. Diagrammatically $\theta(S_t)$ is always above S_t. Thus over time, as the community travels up the $\theta(S_t)$ curve, more families jump on the schooling bandwagon, and disapproval vanishes and turns eventually into approval.

Figure 2: Community Dynamics of School Attendance



The attendance rate at day t is given by $\theta(S_t)$ according to (10). The left-hand panel identifies three equilibria, the outer ones being locally stable and separated by θ_{mid}. The right hand side shows the case where $\theta = 1$, i.e. complete attendance, is the only long-run equilibrium.

If the $\theta(S_t)$ curve intersects the identity line twice, it identifies three social equilibria. The equilibrium in the middle, θ_{mid}, is unstable. If attendance at school is slightly larger, then the associated

$\theta(S_t)$ value lies above the identity line implying that the current schooling attitude in the community triggers further attendance, which raises community support of schooling further etc. Likewise for slightly lower attendance, i.e. for θ_t slightly below θ_{mid}, social dynamics (11) drive θ down until only a few children (mainly from the richest families) are attending school. Employing analogous arguments, the equilibria $\theta_{high} = 1$ and $\theta_{low} > 0$ are identified as locally stable.

3.2. **Persistence of an Anti-Schooling Norm.** The equilibrium of low attendance at school θ_{low} is located at the lower intersection of the convex branch with the identity line. To identify it insert $S_t = \theta_t$ into the convex branch of the $\theta(S_t)$-curve (10) and solve for the smallest root.

$$\theta = \theta_{low} \equiv \frac{\phi}{2} - \sqrt{\frac{\phi^2}{4} - \frac{E_j A^{1-\alpha}}{(2-\alpha)}}. \tag{12}$$

Inspect (12) to conclude that θ_{low}, if it exists, is smaller than $\phi/2 < 1/2$, implying that only a minority of children is going to school. For the equilibrium to exist the radicand in (12) has to be positive implying that θ_{low} exists if

$$E_j A^{1-\alpha} < (2-\alpha)\phi^2/4, \qquad j \in \{WW, NW, NN\}. \tag{13}$$

Note that existence of θ_{low} is not sufficient to establish an anti-schooling norm. Additionally, the actual attendance rate has to be below θ_{mid}. If "somehow" school attendance became sufficiently high ($\theta_t > \theta_{mid}$) social dynamics lead the community towards $\theta_{high} = 1$ irrespective of the fact that poverty and preferences would, in principle, also support the θ_{low} equilibrium. This feature of *local* stability will be exploited in the policy evaluation of Section 4.

The dynamic nature of the model also helps to clarify, linguistically and formally, the difference between a current *attitude* and a persistent *norm.* For example, if the community starts out at low S_t in the right panel of Figure 2, we would say that there exists an anti-schooling attitude. The attitude however does not classify as a norm because it is not persistent and transforms endogenously into social approval of schooling. If, however, the community starts out at low S_t in the left panel of Figure 2, it converges towards an equilibrium of low school attendance and persistent community disapproval of schooling. The feature of stability of the attitude justifies to characterize it as a norm.

Inspecting condition (13) and employing Lemma 1 provides the following result.

PROPOSITION 3. *A social equilibrium supporting an anti-schooling norm exists if aggregate productivity of the community is sufficiently low (A is sufficiently low), if parents' autonomous evaluation of education is sufficiently low (a is sufficiently low), and if schooling is sufficiently time intensive*

(τ is sufficiently high).

Applying Lemma 3 on condition (13) verifies the following result.

PROPOSITION 4. *The set of parameters values for $\{a, A, \phi\}$ supporting an equilibrium θ_{low} is largest if school attending children stop working, i.e. if γ and τ fulfill*

$$\gamma \geq \alpha/(1-\alpha) \geq \gamma(1-\tau).$$

As explained above, the indivisible nature of the schooling activity creates a set of values for $\{\gamma, \tau\}$ for which parents prefer their children to work if they are not going to school and to stop working if they are going to school. The resulting decision about child labor at the extensive margin and the implied forgone child labor if a child is going to school implies that the schooling propensity is particularly low (E_{NW} is smaller than E_{WW} and E_{NN}). This means $\theta(S_t)$ is lowest in the E_{NW} case for any given level of S_t and any given level of preference for schooling (a), aggregate productivity (A) and maximum disapproval (ϕ). As a consequence the set of parameter values $\{a, A, \phi\}$ for which an intersection θ_{low} exists is largest. In other words, if γ is so low that children are not working no matter whether they attend school or not, or if γ is so high that children are anyway working, the fact that they are going to school does not distort child labor at the extensive margin and an anti-schooling norm is harder to sustain in the community.

Using Lemma 2 we prove the non-monotonous effect of productivity on norm prevalence.

PROPOSITION 5. *Assume an anti-schooling norm prevails. (a) A sufficiently large decrease of child labor productivity γ initiates a transition towards θ_{high} if school attending children stop working or if they continue to work and are sufficiently unproductive, $\gamma < \bar{\gamma}$. (b) If, however, $\gamma > \bar{\gamma}$, and school attending children continue to work, decreasing child labor productivity deteriorates school attendance.*

Diagrammatically, in case (a) lower child productivity γ increases the schooling propensity E_j and shifts the $\theta(S_t)$–curve upwards, whereas in case (b) it reduces the schooling propensity and shifts the $\theta(S_t)$–curve further down. Economically, income from child labor is more important for the family in case (b) in the sense that the negative effect of decreasing γ on lower family income dominates the positive effect on lower opportunity cost of schooling. The policy implication is that if child labor productivity is not too high, a discrimination of child work, for example, a (partly) successful child labor ban that effectively drives down child labor productivity, is helpful to increase attendance at school. However, if child labor productivity is high and school attending children continue to work, reducing the children's contribution to family income is detrimental to schooling via the poverty channel. It further deteriorates attendance at school and increases aggregate child labor supply.[6]

More generally, anything that shifts the $\theta(S_t)$–curve sufficiently far up such that the intersection θ_{low} ceases to exist, initiates social dynamics towards high attendance at school. This could, in principle, be brought about by an increase of average productivity of the community (A), for example, by widespread introduction of fertilizer or irrigation, or by lowering the time cost of schooling (τ), for example, by providing bus service to school or by establishing a local school, Yet, if such a policy shock is not sufficiently strong, the effect on school attendance will be only marginal since θ_{low} continues to exist and the force of social dynamics are not positively exploited.

4. A Two Step Policy to Eliminate Child Labor

We next investigate how a two-step policy can exploit the force of social dynamics and achieve both high attendance at school and eradication of child labor. The mechanism of the policy uses the fact that once a pro-schooling norm has been established by a sufficiently strong upward shift of the $\theta(S_t)$–curve (from left to right hand side in Figure 2) and the entailed convergence towards θ_{high}, a downward shift of the $\theta(S_t)$–curve does *not* trigger the return to θ_{low}.

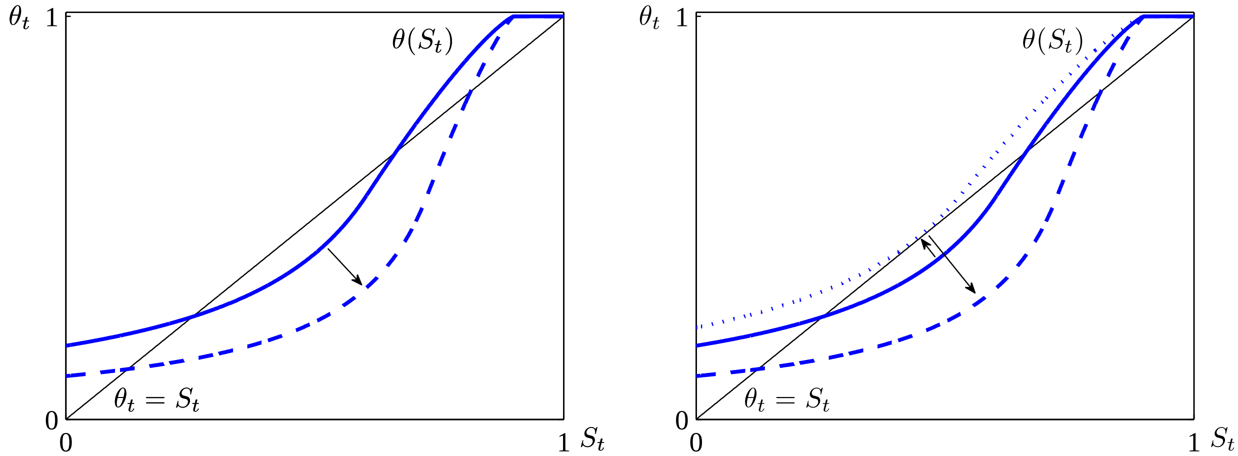
Proposition 6. *The social equilibrium θ_{high} is robust. Once it is attained, the community stays there irrespective of a change of the model's parameter values.*

In contrast to "stability", which characterizes the impact of a change of the endogenous variable θ_t on the equilibrium, "robustness" refers to the impact of a non-marginal and possibly large parameter change on the attained equilibrium. Note that the equilibrium at θ_{low} is not robust. Robustness of θ_{high} can be exploited to eradicate child labor in a two-step way, reminiscent of the Chinese proverb "To reach your goal take a detour". So far we have established the result that increasing the time requirement for schooling τ reduces child labor and eventually eliminates it if children are attending school. It could thus be tempting for policymakers to increase τ in order to reduce child labor. The "if", however, is rather big, since we have also established the result that increasing τ reduces the schooling propensity (Lemma 1) and that it reduces through this channel the attendance rate (Proposition 1).

Figure 3 and 4 illustrate these arguments using a parameterized model. Initially, schooling takes up a a quarter of a child's day, $\tau = 0.25$. If the non-sleeping time per day is 16 hours this implies

[6]That reducing child labor income through boycotts and bans can have the detrimental effect of increasing child work has also been established in the conventional child labor literature, see Basu and Zarghamee (2009) and in the political economy theory of child labor, see Doepke and Zilibotti (2010).

Figure 3: One-Step and Two-Step Attempts to Eliminate Child Labor through Schooling



Parameters: $A = 1$, $\alpha = 0.2$, $\gamma = 0.5$, $a = 0.4$, $\phi = 0.9$. Initially $\tau = 0.25$ (solid lines). Left panel: unsuccessful attempt to eliminate child labor by implementing $\tau = 0.5$. Right panel: successful policy implements $\tau = 0.1$ first (dotted line) and when $\theta \approx \theta_{high}$ it implements $\tau = 0.5$ (dashed line).

that the schooling activity including homework and traveling to school and back takes about 4 hours, which is about the typical time cost of primary schooling in many developed countries (Edmonds, 2007). Without loss of generality we consider an anti-schooling norm with respect to girls and set parameter values such that the initial attendance rate is very low. We set parameters $\gamma = 0.5$, $\alpha = 0.2$, $a = 0.4$, and $\phi = 0.9$ such that at the equilibrium $\theta_{low} = 0.25$.

The chosen parameter values also imply that girls who are not attending school work 40 percent of the day (i.e. about 6.4 hours) and girls who are attending school work 20 percent of the day (3.2 hours). This means that the setup supports the WW-case: attending children continue to work, though much less than non-attending girls. The model outcome fits nicely with Edmonds' (2007) estimate that – across countries – school attendance is compatible with work up to about 29 hours per week from which on attendance declines dramatically. The implied average labor supply of girls is $L_{low} = 0.2\theta_{low} + 0.4(1 - \theta_{low}) = 0.35$. This figure in turn implies that children contribute on average $0.35\gamma/(1 + 0.35\gamma) = 15$ percent to family income.

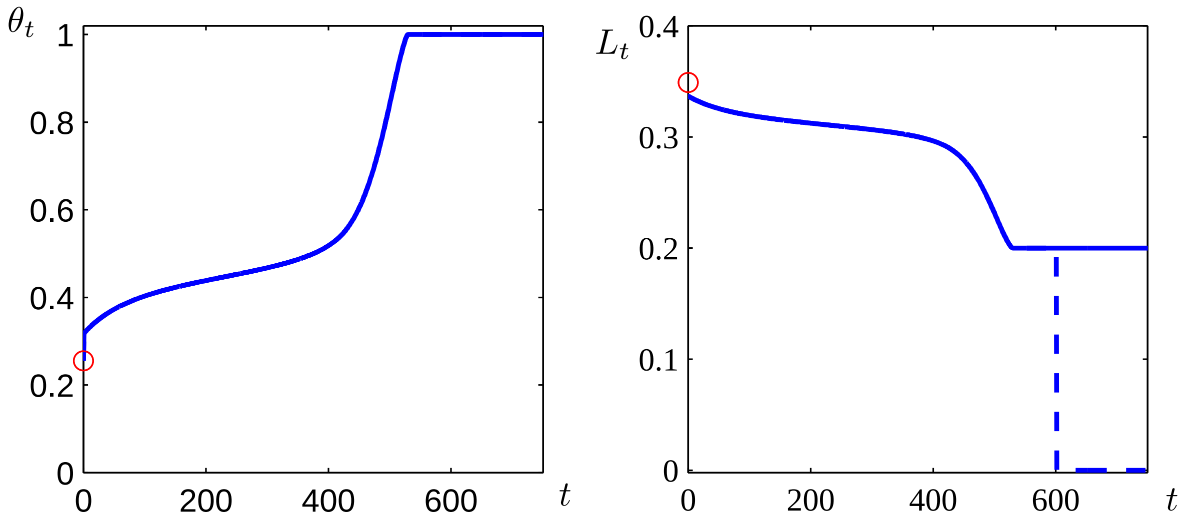
The parameterized model sketches a community that displays a particularly high incidence of child labor and a particularly low attendance rate of girls at school. It roughly corresponds with the situation documented by studies on local school supply in rural Pakistan and Afghanistan. According to Sawada and Lokshin (2009) the probability of a girl entering primary school in rural Pakistan is 22 percent and it rises by 18 percent after the introduction of a local school. Burde and Linden (2010) show that in rural Afghanistan providing a village-based school increases the enrollment rate from 27 to 74 percent and eliminates the gender gap in education. These studies do not provide information about child time use. But, for a similar environment, Dreze and Kingdon (2001) document that in rural Northern India where 44 percent of the girls are not enrolled in school, non-attending children work on average 4.7 hours a day, about two hours more than attending children. Girls work a little longer than boys and they are mainly helping their parents at home and in the fields.

The girls attending school are found below the $\sigma(\epsilon)$–threshold (Figure 1). They come predominantly from families who do not care much about the community evaluation of their actions (low σ families) and from relatively rich families (high ϵ families). At first sight the situation thus looks as if poverty and too much domestic work prevent schooling. For the assumed parameter values, however, the schooling propensity E_{WW} is positive for *all* families. This means that low attendance at school is unambiguously identified as being caused by the prevailing anti-schooling norm.

We next investigate how the time cost of schooling τ affects the equilibrium of low attendance and high child labor. In practice τ can be manipulated in several ways, for example, through travel distance to school, the length of the school day, homework etc. Our first policy experiment makes the school day longer by increasing τ from 0.25 to 0.5 in one step, which could be a move from a half-day school to a full day school. The policy is designed such that school-attending girls stop working. The idea of the policy is that if parents are sufficiently empathetic, they want to leave their children the remainder of the day for leisure and play. In a different context, without a prevailing anti-schooling norm, the idea may work well (e.g. the jornada ampliada in Brazil, see Cardoso and Souza, 2004). Given the prevailing anti-schooling norm, however, the policy is ill-designed. The higher time cost of schooling actually causes some further parents to withdraw their girls from school. The $\theta(S_t)$–curve shifts downwards (shown by dashed lines in Figure 3), eventually leading to lower attendance at school. At the new equilibrium only 12 percent of the girls are attending school. Thus, on the aggregate, lengthening the school day has actually led to *more* child labor.

The right hand side panel of Figure 3 visualizes the two-step policy. In the first step the time cost of schooling is sufficiently strongly *reduced*, for example, by establishing a local primary school. This leads to an upshift of the $\theta(S_t)$ curve (dotted lines in Figure 3). Attending school now takes up only 10 percent of the girls' days, causing some more parents to let their daughters attend school. But what is more important, the equilibrium θ_{low} ceases to exist. The fact that more children are attending school reduces the anti-schooling norm sufficiently strongly so that some further children attend school next period. A bandwagon effect towards complete attendance sets in. The dynamics are shown in Figure 4. The initial situation (at $\tau = 0.25$) is marked with a circle. After reducing the time cost, the share of attending girls jumps up by 7 percent to 32 percent.

Figure 4: Two-Step Policy: Dynamics of School Attendance and Child Labor



School attendance (left) and aggregate child labor (right). Parameters as for Figure 3 and $\delta = 0.96$. The time cost of schooling is reduced from $\tau = 0.25$ (circle) to $\tau = 0.1$. Dashed line: at $t = 600$, $\tau = 0.5$.

An attendance rate of 32 is sufficient to set in motion the bandwagon effect. As time proceeds, more and more girls are sent to school, which further mitigates the anti-schooling attitude in the community. Social dynamics are initially relatively slow since the $\theta(S_t)$–curve is close to the identity line. After about 400 days adjustment dynamics get momentum. Formally, $\tilde{\epsilon}$ gets smaller than unity (see Figure 1) and all richer parents in the community send their girls to school. After about 500 days complete attendance is reached and the community has developed a pro-schooling norm.[7]

The right panel in Figure 4 shows the implied dynamics for aggregate child labor L_t. At the beginning there is only little movement of child labor, since every period only a few more girls are allowed to go to school and to reduce their work load from 40 to 20 percent of their daily time. But eventually, as the anti-schooling attitudes vanishes, child labor is visibly reduced, converging to an aggregate of 0.2. But now, with the pro-schooling norm operative, the $\tau = 0.5$ policy can be safely implemented, for example, by having lessons in the afternoon. Diagrammatically, the $\theta(S_t)$–curve shifts down in Figure 3 (the dashed line again) and θ declines. With the pro-schooling norm at work, the previously ill-designed policy, is now effective. All parents leave their kids in school and stop requiring domestic work (dashed lines in Figure 4). Child labor is effectively eliminated.

[7]The speed of adjustment depends on the numerical specification of δ and is, in absolute terms, arbitrary since we do not know how fast community norms evolve. The shape of the adjustment trajectory, however, is independent from δ and determined by deep parameters, i.e. by the shape of the $\theta(S_t)$–curve and its position with respect to the identity line.

5. Conditional Transfers

Because the proposed model takes heterogeneity of households with respect to poverty and school attendance into account it is well suited for a theoretical analysis of conditional transfer programs, like PROGRESA, Bolsa Escola, or Food for Education. These programs pay a subsidy contingent on attendance at school. They come in two variants. Targeted transfers (like PROGRESA) are addressed to only a subset of a community, usually the poorest families. Non-targeted transfers are paid conditioned on behavior only, i.e. to all children showing up at school (Food for Education). Particularly interesting here is it to provide a theoretical foundation for the success of conditional transfers programs through social norms and collective behavior. We first investigate the optimal design of a targeted transfer scheme (with respect to eligibility and generosity) and then explore the power and costs of non-targeted subsidies.[8]

5.1. **Targeted Transfers.** Suppose that the poorest families of a community are selected to receive an income subsidy of size s contingent on their children going to school. Let p denote the community share of eligible families. This means that, if the program is entirely effective and triggers school attendance of all eligible children, total costs are given by $C = \int_0^p s \mathrm{d}w = ps$. In order to reduce complexity and case differentiation we focus on the NW case, i.e. parameters are such that children who are not going to school are working and school attending children stop working. While the "optimistic " assumption that schooling replaces child work is indeed supported by some studies it is clearly not generally true.[9] The assumption is thus rather justified on theoretical grounds. The major objective of this subsection is to explain the large success of targeted transfer programs through community effects. Since it has been shown (Lemma 3) that the schooling propensity is *lowest* in NW case, a targeted transfer program is – for given expenditure – the least effective in the NW case. We numerically investigate the WW-case later on. For convenience we also normalize average productivity ($A = 1$) such that $w(i) = \epsilon(i)$.

Keeping everything else from the basic model, autonomous utility of a parent whose child is attending school is given by $u_t^a(i) = (\epsilon(i) + s(i))^{1-\alpha}(1-\tau)^{\alpha}(1+a)$. Here $s(i) = s$ if $\epsilon(i) \leq p$ and $s(i) = 0$ otherwise, reflecting the fact that only the poorest families with income below p are eligible to receive transfers. If the child is not attending school, utility is, as before, $u_t^a(i) =$

[8]Empirical studies accounting for the success of conditional transfer programs are, among others, Ravallion and Wodon, 2000, Skoufias and Parker, 2001, Cardoso and Souza, 2004, Schultz, 2004, Bobonis and Finan, 2009, and Lalive and Cattaneo, 2009.

[9]Skoufias and Parker (2001) find for PROGRESA that schooling was largely replacing child work (on the market for boys, in the household for girls). Ravallion and Wodon (2000) found a much smaller impact of schooling on child labor for the Food for Education Program.

$(1-\alpha)^{1-\alpha}(1+\gamma)(\alpha/\gamma)^{\alpha}\epsilon(i)^{1-\alpha}$. Comparing utilities, we find that children of parents with a (ϵ, σ) tuple below the threshold (14) are attending school.

$$\sigma(\epsilon) = \frac{(\epsilon(i) + s(i))^{1-\alpha}(1-\tau)(1+a)^{\alpha} - \lambda\epsilon(i)^{1-\alpha}}{\phi - S_t}, \qquad \lambda \equiv (1-\alpha)^{1-\alpha}(1+\gamma)\left(\frac{\alpha}{\gamma}\right)^{\alpha}. \tag{14}$$

Because eligibility to subsidies is piece-wise defined, the schooling threshold displays a jump discontinuity where $\epsilon = p$, i.e. where eligibility ends. The size of the subsidy s determines the size of the jump at p. Otherwise the model is identical to the simple one from Section 2 and 3. Piece-wise integrating the area below the threshold provides the $\theta(S_t)$–curve.

$$\theta_t = \theta(S_t) = \frac{1}{2-\alpha}\frac{E_S}{\phi - S_t} - (1-\tilde{\epsilon}) \quad \text{where} \tag{15}$$

$$E_S \equiv (1-\tau)^{\alpha}(1+a)\left[(p+s)^{2-\alpha} - s^{2-\alpha} - p^{2-\alpha} + \tilde{\epsilon}^{2-\alpha}\right] - \lambda\tilde{\epsilon}^{2-\alpha}$$

$$\tilde{\epsilon} = \begin{cases} 1 & \text{for } S_t \leq \tilde{S} \equiv \phi + \lambda - (1-\tau)^{\alpha}(1+a) \\ \left(\frac{\phi - S_t}{(1+a)(1-\tau)^{\alpha} - \lambda}\right)^{\frac{1}{1-\alpha}} & \text{for } \phi > S_t \geq \tilde{S} \end{cases}.$$

and $\theta_t = 0$ for $S_t > \phi$. The compound parameter E_S inherits the function of the schooling propensity from the basic model. Differentiating it with respect to the new policy parameters s and p and using the result to compute comparative statics of (15) proves the following insight.

Lemma 4. *$\partial E_S/\partial p > 0$, $\partial E_s/\partial s > 0$, and $\partial\theta_t/\partial p > 0$, $\partial\theta_t/\partial s > 0$.*

The schooling propensity E_S is increasing in the generosity of subsidies s and in the share of targeted recipients p. This implies that for any given schooling attitude S_t the share of school attending children θ_t is increasing in the generosity of subsidies s and in the share of recipients p.

Inserting the equilibrium condition $S_t = \theta_t$ into (15), noting that θ_{low} is assumed along the convex branch of the $\theta(S_t)$ curve, and solving for θ provides $\theta_{low} \equiv \phi/2 - \sqrt{\phi^2/4 - E_S/(2-\alpha)}$. From this follows that a locally stable anti-schooling norm θ_{low} exists if $E_S < (2-\alpha)\phi^2/4$. Applying Lemma 4 we conclude that larger p or s reduces the set of parameters values for preferences and child contributions to income (a, α, and γ) for which θ_{low} exists, which proves the following result

Proposition 7. *If transfers s are sufficiently high and the share of targeted recipients p is sufficiently large, the equilibrium of low attendance θ_{low} ceases to exist.*

We can now answer the question which combination of p and s moves a socio-economy out of θ_{low} at minimum costs. Formally, we are interested in the solution of the following problem.

$$\min_{p,s} \{C = p \cdot s\} \qquad s.t. \tag{16}$$

$$E_S = (2-\alpha)\phi^2/4. \tag{17}$$

The solution provides a tangency point of the $\theta(S_t)$ curve with the identity line implying that an arbitrarily small further increase of s (or p) initiates a transition towards "education for all". In the Appendix it is shown that the solution of (16)–(17) is (18).

$$p^* = s^* = \left\{ \left[\frac{\phi^2(2-\alpha) + 4\lambda}{4(1-\tau)^\alpha(1+a)} - 1 \right] \cdot \frac{1}{2(2^{1-\alpha}-1)} \right\}^{\frac{1}{2-\alpha}}. \tag{18}$$

Since generosity and eligibility affect the schooling propensity (15) symmetrically, an optimal targeted transfer requires equality of p and s. Taking derivatives of (18) proves the following result on comparative statics.

Proposition 8. *$\partial s^*/\partial\tau < 0$, $\partial p^*/\partial\tau < 0$, and $\partial s^*/\partial\gamma < 0$, $\partial p^*/\partial\gamma < 0$.*

The optimal size of the subsidy (that initiates an escape from θ_{low} at minimum costs) is decreasing in the time costs of schooling τ and in child labor productivity γ. An immediate corollary for policy is that a successful targeted transfer program can be accomplished at lower costs when it is accompanied by a reduction of the time costs of schooling, e.g. by providing bus transfer to school.

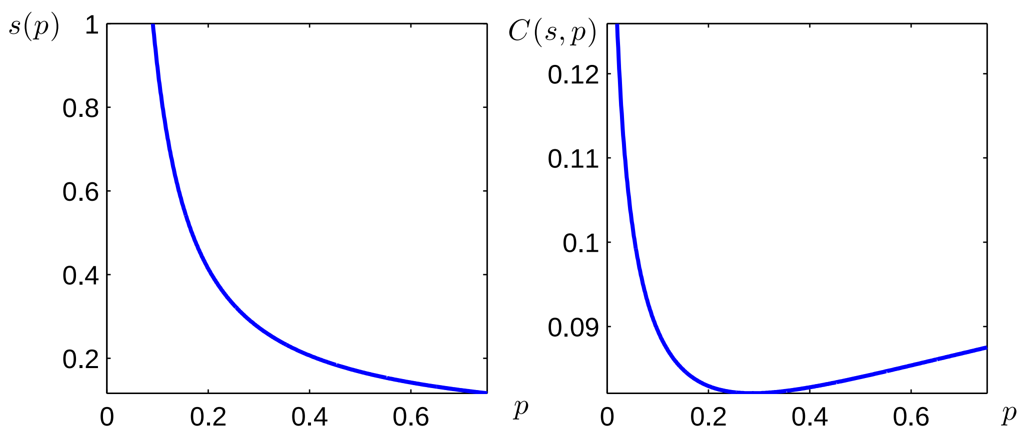
The result that targeted transfers are not only changing the behavior of targeted families but also of ineligible families gets empirical support from recent studies of PROGRESA (Bobonis and Finan, 2009; Lalive and Cattaneo, 2009). Lallive and Cattaneo also provide evidence for $\partial\theta/\partial p > 0$, i.e. that the strength of the response of ineligible families is increasing in the share of eligible families.

We finally investigate the power of targeted transfers for the numerical example from Section 4. Suppose that initially $\tau = 0.5$ such that school attending girls do not work and that the community is initially at an equilibrium θ_{low} where only 12 percent of the girls are attending school. The optimal transfer policy according to (18) is $s^* = p^* = 0.28$, i.e. to address the poorest 28 percent of families with a subsidy of 0.28. Since average adult income in the community is 1/2, the optimal design suggests to address a relatively small share of families with relatively generous subsidies (for comparison, PROGRESA increased income of the beneficiary families by about 22 percent, see Skoufias and Parker, 2001).

The subsidy per family s has to be relatively generous if the addressed group is small because the very poor need relatively large income supplements in order to send their girls to school and initiate the bandwagon effect. If the share of eligible families gets larger, transfers per family can be lower. Since the families at the upper boundary of eligibility are relatively rich, they need smaller subsidies in order to send their girls to school. This behavior decreases community disapproval of schooling and lowers the subsidy needed to convince poorer families to send their girls to school.

Figure 5 shows these results by displaying alternative targeted transfer programs that bring about an escape from θ_{low}. Combinations of subsidy size s and eligibility p on the $s(p)$–curve are efficient in the sense that they fulfil the constraint (17). Combinations below the $s(p)$–curve fail to initiate an escape from low attendance at school and combinations above the curve are inefficient, "too generous" solutions. The $C(s,p)$–curve shows the total costs involved for efficient solutions. The minimum cost strategy solves (16)–(17) and is found where $C(s,p)$ attains its minimum at $p = 0.28$. Interestingly, the cost curve is skewed. It is less costly to deviate from the minimal cost solution by addressing a too large share of families. Again the result is explained with social dynamics. The larger the fraction of eligible families the smaller is the subsidy needed to motivate the richest eligible families to send their children to school and to initiate the social bandwagon.

Figure 5: Successful Targeted Transfers Policies



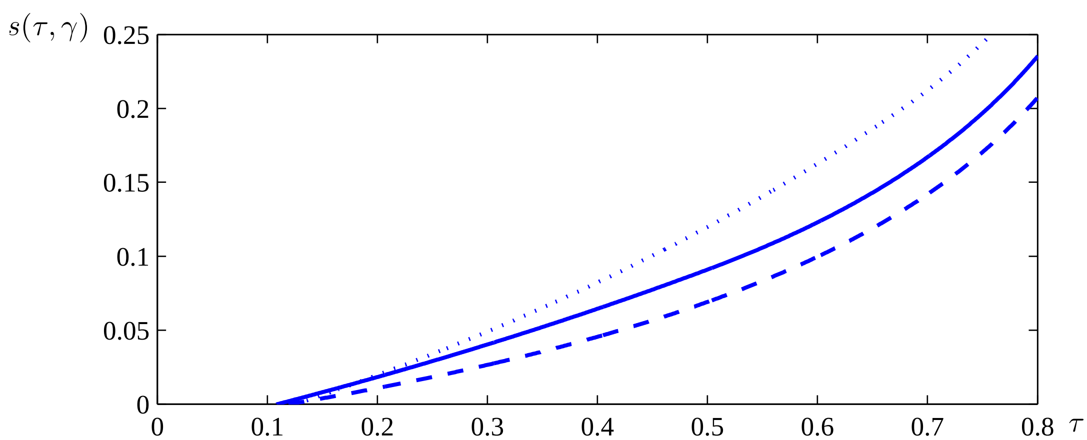
Parameters as for Figure 3 with $\tau = 0.5$. The curve in the left panel shows for alternative eligibility p minimum cost policies that bring about an escape from θ_{low}. The right panel shows the total costs involved.

5.2. **Non-Targeted Conditional Subsidies.** The cost-minimal policy requires to address a relatively small share of families with relatively generous subsidies. If this is not desired, however, Figure 5 shows that the total costs involved when the program targets a larger share of families is insignificantly higher since the subsidy per person can be lower. It is thus instructive to compute the efficient subsidy required when the whole community is eligible. In this case the policy is no longer *targeted* to the poorest families but still *conditioned* on the schooling activity. It could capture, for example, a breakfast or lunch served to all children showing up at school.

Figure 6 shows the efficient subsidy that solves (17) when $p = 1$ for alternative assumptions about child labor productivity γ and time costs of schooling τ. These policies are the cost-minimal policies that get all children into school given that the whole community is eligible. If τ is below 0.5 (for $\gamma = 0.5$, solid line), school attending children continue to work. In that case, Figure 6 shows the numerical solution for the WWcase. The figure reveals that the efficient subsidy is falling in the time cost of schooling (determined, for example by distance to the nearest school) and supports the suggestion from empirical studies to combine demand- and supply side policies (Schultz, 2004).

Since all families of the community are eligible there is no longer a social multiplier at the extensive margin. At the intensive margin community dynamics are, of course, still operative. Their power can be assessed by comparing with the subsidy needed to establish complete attendance on impact, i.e. without exploiting the social bandwagon. It is computed for the basic case ($\gamma = \tau = 0.5$) as 0.56, i.e. more than six times as large as the 0.09 needed with social interaction.

Figure 6: Efficient Non-Targeted Subsidies



The figure shows for alternative time costs of schooling the constrained cost-minimal subsidy needed to escape from θ_{low} given that the whole community is eligible. Child labor productivity: solid line: $\gamma = 0.5$; dashed line: $\gamma = 0.25$; dotted line: $\gamma = 0.75$.

The dashed and dotted lines show efficient subsidies for higher ($\gamma = 0.75$) and lower ($\gamma = 0.25$) child labor productivity. If working children contribute more to family income the efficient subsidy has to be more generous. Given that average income of a family without support from child labor is 0.5 (implied by the normalization $A = 1$) the overall impression is that the efficient subsidy can be surprisingly low. The social multiplier is thus helpful to rationalize the empirical finding that even small subsidies, much below a working child's contribution to family income (like a mid-day meal at school, see e.g. Dreze and Kingdon, 2001) can be very effective in increasing attendance rates.

At the same time, these subsidies are less effective to eliminate child labor. When child labor productivity is 0.5, school attending children continue to work if the time cost of schooling is below 0.5. For the example from Section 4 (with $\tau = 0.25$), school attending children spend 20 percent of the day working and 25 percent with schooling activities. The subsidy is helpful in reducing child labor (from 40 to 20 percent) but it fails to eliminate it. The theory is thus helpful to rationalize the empirical finding that subsidies that are effective in drawing children to school can be relatively ineffective in reducing or eliminating child labor.

6. Final Remarks

This article has proposed a theory of schooling and child labor driven by community attitudes and norms. Maintaining the well-known positive association between income and education, the norm-based theory can resolve some puzzles from the empirical literature, for example, why communities of the same poverty sometimes display very different schooling behavior, and why conditional transfers are so successful in drawing children to school.

The theory has offered a novel rationalization of multiple schooling equilibria and has thus provided further justification of a development goal for education, separately from poverty alleviation. In order to escape from an equilibrium of low schooling sustained by a prevailing anti-schooling norm, policy has not (only) to rely on persuasion. Instead it has been shown how simple policies of education supply (like, for example, the provision of a local school) or demand (conditional transfers) can exploit a social multiplier. But it has also been demonstrated that education policy has to be context-specific. Seemingly well-meant policies to reduce child labor and increase education can work in one environment (for one parameter configuration, in the neighborhood of one equilibrium) and can be ill-designed in another.

The theory is not aiming to replace the poverty channel and other available explanations of child labor and low education. Instead, credit constraints (Baland and Robinson, 2000), bonded child labor (Basu and Chau, 2004), and political economy elements (Doepke and Zilibotti, 2005), constitute conceivable future extensions of the community-norms based model. Another desirable extension would be to consider intergenerational links, taking into account that the current generation's schooling activity is associated with next generation's aggregate productivity. A straightforward extension of the model in this direction, available as Webpage-Appendix C, shows that all the main results are preserved and additionally provides adjustments dynamics: income and productivity rise in line with education if a community manages to escape from the equilibrium of low attendance at school. A more detailed modeling of intergenerational issues, however, should perhaps also take into account the impact of schooling and child labor on fertility (Hazan and Berdugo, 2002; Doepke, 2004) and on child mortality (Strulik, 2004) and possibly integrate the evolution of community attitudes with the intergenerational transmission of schooling norms (Kirchsteiger and Sebald, 2010). There is thus ample room for extending and improving the community-norms driven theory of schooling behavior and child labor.

Appendix A: Derivations and Proofs

Derivation of (6) and (8). Insert (2) into (5).

$$u_t(i) = [(1+\gamma\ell_t(i))w_t(i)]^{1-\alpha}(1-\ell_t(i)-\tau\cdot h_t(i))^{\alpha}(1+a\cdot h_t(i)) + h_t(i)\cdot\sigma(i)\cdot(S_t-\phi). \tag{A.1}$$

The first order condition with respect to $\ell_t(i)$ is

$$(1-\alpha)(1-\ell_t(i)-\tau h_t(i))\gamma w_t(i) - \alpha(1+\gamma\ell_t(i))w_t(i) = 0.$$

Solving for $\ell_t(i)$ provides

$$\ell_t(i) = (1-\alpha)(1-\tau h_t(i)) - \frac{\alpha}{\gamma},$$

which is the interior solution in (6). Plugging it into (2) provides consumption

$$c_t(i) = (1-\alpha)\left[1+\gamma(1-\tau h_t(i))\right]w_t(i).$$

Plugging $\ell_t(i)$ and $c_t(i)$ into $u_t^a(i)$ and simplifying provides

$$u_t^a(i) = (1-\alpha)\left(\frac{\alpha}{\gamma}\right)^{\alpha}\left[1+\gamma(1-\tau h_t(i)\right](1+a)w_t(i)^{1-\alpha}.$$

Plugging this into $u_t(i)$ and evaluating $u_t(i)|h_t(i) \geq 1 - u_t(i)|h_t(i) = 0$ provides

$$(1-\alpha)^{1-\alpha}\left(\frac{\alpha}{\gamma}\right)^{\alpha}\left\{\left[1+\gamma(1-\tau)a-\gamma\tau\right]\right\}w_t^{1-\alpha} \geq \sigma(i)(\phi-S_t) \leq 0,$$

which is (8) for $j = WW$. For $\ell_t(i) =$ we get $c_t(i) = w_t(i)$. Plugging this information into $u_t^a(i)$ provides

$$u_t^a(i) = w_t(i)^{1-\alpha}(1-\tau h_t(i))^{\alpha}(1+ah_t(i)).$$

Computing the above utility difference for the case in which $\ell_t(i) = 0$ irrespective of $h_t(i)$ provides

$$\left[(1-\tau)^{\alpha}(1+a)-1\right]w_t(i)^{1-\alpha} \geq \sigma(i)(\phi-S_t) \leq 0$$

which is (8) for $j = NN$. Finally computing the utility difference when $\ell_t(i) > 0$ for $h_t(i) = 0$ and $\ell_t(i) = 0$ for $h_t(i) = 1$ provides (8) for $j = NW$.

Lemma 2. From (8)

$$\frac{\partial E_{NW}}{\partial\gamma} = -(1-\alpha)^{1-\alpha}\left(\frac{\alpha}{\gamma}\right)^{\alpha}\left[1-(1+\gamma)\left(\frac{\alpha}{\gamma}\right)\right]$$

Thus for $E_{NW} < 0$

$$1 > (1+\gamma)\frac{\alpha}{\gamma} \qquad \Rightarrow \qquad \gamma > \frac{\alpha}{1-\gamma}$$

which holds always when the NW-case materialize.

Take the derivative in (8) to see that the sign of $\partial E_{WW}/\partial\gamma$ is equal to the sign of

$$\gamma^{\alpha-1}\left\{-\alpha\left[1+\gamma(1-\tau)\right](1+a)+\alpha(1+\gamma)+\gamma(1-\tau)(1+a)-\gamma\right\}$$

The sign of the above expression equals the sign of $(1-\alpha)\gamma\left[(1-\tau)(1+a)-1\right]-\alpha a$, which is positive if $\gamma > \bar{\gamma}$ and negative if $\gamma < \bar{\gamma}$.

Lemma 3. Begin with showing $E_{NW} < E_{NN}$. Since E_{NW} is monotonously rising in γ (Lemma 2) it is largest where γ is at the upper boundary of the supporting domain i.e. where $\gamma = \alpha/(1-\alpha)$

and thus $\alpha/\gamma = (1-\alpha)$. Substitute these values into E_{NW} to conclude

$$E_{NW} \leq (1-\tau)^{\alpha}(1+a) - (1-\alpha)^{1-\alpha}\left(1+\frac{\alpha}{1-\alpha}\right)(1-\alpha)^{\alpha} = (1-\tau)^{\alpha}(1+a) - (1-\alpha)\left(\frac{1}{1-\alpha}\right)$$
$$E_{NW} \leq (1-\tau)^{\alpha}(1+a) - 1 = E_{NN}.$$

In order to show that $E_{NW} \leq E_{WW}$ let τ_a denote those values of τ that have to hold for E_{WW} to materialize, i.e. according to condition (7)

$$1-\tau_a > \frac{\alpha}{(1-\alpha)\gamma}. \tag{A.2}$$

Likewise, let τ_b denote those values of τ that have to hold for E_{NW} to materialize, i.e. according to condition (7)

$$1-\tau_b < \frac{\alpha}{(1-\alpha)\gamma}. \tag{A.3}$$

Taking the difference of the schooling propensities we see that for $E_{WW} \geq E_{NW}$

$$(1-\alpha)^{1-\alpha}\left(\frac{a}{\gamma}\right)^{\alpha}[1+\gamma(1-\tau_a)] \geq (1-\tau_b)^{\alpha}.$$

Because of (A.3) a sufficient condition for this to be true is

$$(1-\alpha)^{1-\alpha}\left(\frac{a}{\gamma}\right)^{\alpha}[1+\gamma(1-\tau_a)] \geq \left(\frac{\alpha}{\gamma}\right)^{\alpha}(1-\alpha)^{-\alpha} \quad \Rightarrow \quad (1-\alpha)\left[1+\gamma(1-\tau_a)\right] \geq 1.$$

Because of condition (A.2) a sufficient condition for this to be true is

$$(1-\alpha)\left[1+\gamma\frac{\alpha}{(1-\alpha)\gamma}\right] \geq 1 \quad \Rightarrow \quad 1 \geq 1,$$

which is true.

Schooling Feasibility (9). Because of Lemma 1 it is sufficient for $E_j > 0$ to show that

$$E_{NW} = (1-\tau)^{\alpha}(1+a) - (1-\alpha)^{1-\alpha}(1+\gamma)\cdot\left(\frac{\alpha}{\gamma}\right)^{\alpha} > 0.$$

Since $\partial E_{NW}/\partial\gamma < 0$ from Lemma 2, it is sufficient to show that $E_{NW} > 0$ for the largest feasible γ. Since $\alpha/(1-\alpha) \geq \gamma(1-\tau)$ for the NW case to materialize, it is sufficient to show that $E_{NW} > 0$ for $\gamma_{max} \equiv \alpha/[(1-\alpha)(1-\tau)]$.

$$E_{NW}(\gamma_{max}) = (1-\tau)^{\alpha}(1+a) - (1-\alpha)^{1-\alpha}\left(1+\frac{\alpha}{1-\alpha}\frac{1}{1-\tau}\right)\alpha^{\alpha}\alpha^{-\alpha}(1-\alpha)^{\alpha}(1-\tau)^{\alpha} > 0$$

That is

$$(1+a) > (1-\alpha)\left(1+\frac{\alpha}{1-\alpha}\frac{1}{1-\tau}\right) = (1-\alpha) + \frac{\alpha}{1-\tau} \qquad \Rightarrow \qquad a > \frac{\alpha}{1-\tau} - \alpha,$$

which simplifies to (9).

Derivation of (10). Integrating the area below the threshold (9) for $\phi > S_t$.

$$\theta = \int_0^{\tilde{\epsilon}} \epsilon^{1-\alpha}A^{1-\alpha}\frac{E_j}{\phi - S_t}\mathrm{d}\epsilon + (1-\tilde{\epsilon}) = \frac{A^{1-\alpha}}{\phi - S_t}\left|\frac{1}{2-\alpha}\epsilon^{2-\alpha}\right|_0^{\tilde{\epsilon}} + (1-\tilde{\epsilon}).$$

For $S_t \leq \phi - A^{1-\alpha}E_j$, $\tilde{\epsilon} = 1$ and

$$\theta = \frac{A^{1-\alpha}E_j}{(\phi - S_t)(2-\alpha)}.$$

For $S_t > \phi - A^{1-\alpha}E_j$, $\tilde{\epsilon} = [(\phi - S_t)/E_j]^{1/(1-\alpha)}/A$ and

$$\theta = \frac{A^{1-\alpha}E_j}{(\phi - S_t)(2-\alpha)}\frac{1}{A^{2-\alpha}}\left(\frac{\phi - S_t}{E_j}\right)^{\frac{2-\alpha}{1-\alpha}} + 1 - \frac{1}{A}\left(\frac{\phi - S_t}{E_j}\right)^{\frac{1}{1-\alpha}} = 1 - \frac{1-\alpha}{2-\alpha}\cdot\frac{1}{A}\left(\frac{\phi - S_t}{E_j}\right)^{\frac{1}{1-\alpha}}.$$

The last equality sign follows from noting that $1 - (2-\alpha)/(1-\alpha) = -(1-\alpha)$. Altogether this results in (10).

Derivation of (12). Start with noting that the lower (convex) branch of $\theta(S_t)$–curve cuts the identity line at most twice and denote the lower intersection, if it exists, by θ_{low}. Solve for θ_t when $S_t = 0$ to verify that the curve originates at $\bar{\theta} = A^{1-\alpha}E_j/[(2-\alpha)\phi] > 0$. Plug $S_t = \theta_t = \theta$ into the lower branch of (10).

$$\theta = \frac{x}{\phi - \theta}, \quad x \equiv \frac{A^{1-\alpha}E_j}{2-\alpha}.$$

Sorting terms provides $\theta^2 - \theta\phi + x = 0$. Solving for the smallest root of the quadratic provides $\theta_{low} = \phi/2 + \sqrt{\phi^2/4 - x}$., Re-substitute x to get (12) in the text.

Derivation of (15). As for the basic model let $\bar{\epsilon}$denote the income level for which $\sigma(\epsilon)$ intersects the unity line if such an intersection exists. If no intersection exists the upper limit of integration is unity and taking piece-wise continuity into account, the area below the threshold is

$$\begin{aligned}\theta_t &= \int_0^p \frac{1}{\phi - S_t}(\epsilon + s)^{1-\alpha}(1-\tau)^{\alpha}(1+a) - \lambda\epsilon^{1-\alpha}\mathrm{d}\epsilon + \int_p^1 \frac{1}{\phi - S_t}\epsilon^{1-\alpha}\left[(1-\tau)^{\alpha}(1+a) - \lambda\right]\mathrm{d}\epsilon \\ &= \frac{1}{(\phi - S_t)(2-\alpha)}\left\{(1-\tau)^{\alpha}(1+a)\left[(p+s)^{2-\alpha} - p^{2-\alpha} - s^{2-\alpha} + 1\right] - \lambda\right\}\end{aligned}$$

Thus

$$\theta = 1 - \frac{1}{(S_t - \phi)(2-\alpha)}E_S$$

with E_S as defined in (15) for $\tilde{\epsilon} = 1$. Solving (14) for $\sigma = 1$ provides the solution $\tilde{\epsilon} = \left[\frac{\phi - S_t}{(1+a)(1-\tau)^{\alpha} - \lambda}\right]^{\frac{1}{1-\alpha}}$.The threshold in this case is

$$\theta_t = \int_0^p \frac{1}{\phi - S_t}(\epsilon + s)^{1-\alpha}(1-\tau)^{\alpha}(1+a) - \lambda\epsilon^{1-\alpha}\mathrm{d}\epsilon + \int_p^{\tilde{\epsilon}} \frac{1}{\phi - S_t}\epsilon^{1-\alpha}\left[(1-\tau)^{\alpha}(1+a) - \lambda\right]\mathrm{d}\epsilon + (1 - \tilde{\epsilon})$$

i.e.

$$\theta = \frac{1}{(\phi - S_t)(2-\alpha)}\left\{(1-\tau)^{\alpha}(1+a)\left[(p+s)^{2-\alpha} - p^{2-\alpha} - s^{2-\alpha} + \tilde{\epsilon}^{2-\alpha}\right] - \lambda\tilde{\epsilon}^{2-\alpha}\right\} - (1 - \tilde{\epsilon})$$

which is (15) for interior $\tilde{\epsilon}$.

6.1. **Lemma 4.** Take the derivative of E_S with respect to s.

$$\frac{\partial E_S}{\partial s} = c_0 \cdot (2-\alpha)\left[(p+s)^{1-\alpha} - s^{1-\alpha}\right] > 0$$

where c_0 is a positive constant. The derivative is positive because $p + s > s$. Given $\partial E_S/\partial s > 0$, it is obvious from (15) that $\partial\theta/\partial s > 0$. The proof with respect to p obtains

$$\frac{\partial E_S}{\partial p} = c_0 \cdot (2-\alpha)\left[(p+s)^{1-\alpha} - p^{1-\alpha}\right] > 0$$

which is positive because $p + s > p$.

Derivation of (18). Let $L = ps + \mu \left[\phi^2(2-\alpha)/4 - E_S\right]$ be the Lagrangian associated with problem (16)–(17). The first order conditions are

$$\frac{\partial L}{\partial p} = s - \mu \frac{\partial E_S}{\partial p} = 0, \quad \frac{\partial L}{\partial s} = p - \mu \frac{\partial E_S}{\partial s} = 0 \quad \Rightarrow \quad \frac{p}{s} = \frac{\partial E_S/\partial s}{\partial E_S/\partial p} = \frac{(p+s)^{1-\alpha} - s^{1-\alpha}}{(p+s)^{1-\alpha} - p^{1-\alpha}}.$$

The final equality followed from insertion of $\partial E_S/\partial p$ and $\partial E_S/\partial s$ from the proof of Lemma 4. The unique solution (in the $(0,1)$ interval) is $s = p$. Substituting p, E_S simplifies to

$$(1-\tau)^\alpha (1+a) \left[2\left(2^{1-\alpha} - 1\right) s^{2-\alpha} + 1\right] - \lambda.$$

Plugging E_S into (17) and solving for s provides (18).

References

Bacolod, Marigee P. and Priya Ranjan (2008). "Why Children Work, Attend School, or Stay Idle." *Economic Development and Cultural Change*, 56, 791–828.

Baland, Jean-Marie and James A. Robinson (2000). "Is Child Labor Inefficient?" *Journal of Political Economy*, 108, 663–679.

Basu, Arnab K. and Nancy H. Chau (2004). "Exploitation of Child Labor and the Dynamics of Debt Bondage." *Journal of Economic Growth*, 9, 209–238.

Basu, Kaushik and Pham Hoang Van (1998). "The Economics of Child Labor." *American Economic Review*, 88, 412-427.

Basu, Kaushik and Zafiris Tzannatos (2003). "The Global Child Labor Problem: What do We Know and what Can We Do?" *World Bank Economic Review*, 17, 147–173.

Basu, Kaushik and Homa Zarghamee (2009). "Is Product Boycott a Good Idea for Controlling Child Labor? A Theoretical Investigation." *Journal of Development Economics*, 88, 217–220.

Bernheim, Douglas, B. (1994). "A Theory of Conformity." *Journal of Political Economy*, 102, 841–877.

Bhalotra, Sonia (2007). "Is Child Work Necessary?" *Oxford Bulletin of Economics and Statistics*, 69, 29–55.

Biggeri, Mario, Lorenzo Guarcello, Scott Lyon, and Furio Rosati (2003). "The Puzzle of 'Idle' Children: Neither in School nor Performing Economic Activity: Evidence from Six Countries." UCW Working Paper.

Binder, Melissa (1999). "Community Effects and Desired Schooling of Parents and Children in Mexico." *Economics of Education Review*, 18, 311–325.

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

Bobonis, Gustavo J. and Frederica Finan (2009). "Neighborhood Peer Effects in Secondary School Enrollment Decisions." *Review of Economics and Statistics*, 91, 695–716.

Burde, Dana and Leigh L. Linden (2010). "The Effect of Village-based Schools: Evidence from a Randomized Controlled Trial in Afghanistan." Discussion Paper, New York University.

Cardoso, Eliana and Andre P. Souza (2004). "The Impact of Cash Transfers on Child Labor and School Attendance in Brazil." Working Paper 04-W07, Vanderbilt University.

Chamarbagwala, Rubiana (2009). "Social Norms, Spatial Dependence, and Children's Activities: Evidence from India." *Journal of Developing Areas*, 42, 157–178.

Case, Anne C. and Lawrence F. Katz (1991). " The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths." NBER Working Paper 3708.

Davison, Jean and Martin Kanyuka (1992). "Girls Participation in Basic Education in Southern Malawi." *Comparative Education Review*, 36, 446–66.

Doepke, Matthias (2004). “Accounting for Fertility Decline During the Transition to Growth.” *Journal of Economic Growth*, 9, 347–383.

Doepke, Matthias and Fabrizio Zilibotti (2005). “The Macroeconomics of Child Labor Regulation.” *American Economic Review*, 95, 1492–1524.

Doepke, Matthias and Fabrizio, Zilibotti (2008). “Occupational Choice and the Spirit of Capitalism.” *Quarterly Journal of Economics*, 123, 747–793.

Doepke, Matthias and Fabrizio Zilibotti (2010). “Do International Labor Standards Contribute to the Persistence of the Child Labor Problem?” *Journal of Economic Growth*, 15, 1–31.

Dreze, Jean and Amartya Sen (1995). “Basic Education as a Political Issue.” *Journal of Educational Planning and Administration*, 9, 1–26.

Dreze, Jean and Geeta G. Kingdon (2001). “School Participation in Rural India.” it Review of Development Economics, 5, 1–24.

Edmonds, Eric V. (2007). “Child Labor.” In *“Handbook of Development Economics”* Vol. 4., edited by John, A. Strauss and T.Paul Schultz, Elsevier, Amsterdam, 3607-3709.

Glaeser, Edward L., Bruce Sacerdote, and Jose A. Scheinkman (2003). “The Social Multiplier.” *Journal of the European Economic Association*, 1, 345–353.

Granovetter, Mark (1978). “Threshold Models of Collective Behavior.” *American Journal of Sociology*, 83, 1420–1443.

Hazan, Moshe and Binyamin Berdugo (2002). “Child Labor, Fertility, and Economic Growth.” *Economic Journal*, 112, 810–828.

Hazan, Moshe and Yishay D. Maoz (2002). “Womens Labor Force Participation and the Dynamics of Tradition.” *Economics Letters*, 75, 193–198.

Hauk, Esther and Maria Saez-Marti (2002). “On the Cultural Transmission of Corruption.” *Journal of Economic Theory*, 107, 311–335.

ILO (2010). “*Global Child Labour Developments: Measuring Trends from 2004 to 2008*.” International Labor Organization, Geneva.

Jensen, Peter, and Helena S. Nielsen (1997). “Child Labor or School Attendance? Evidence from Zambia.” *Journal of Population Economics*, 10, 407–424.

Kirchsteiger, Georg and Alexander Sebald (2010). “Investments into Education – Doing as the Parents Did.” *European Economic Review*, 54, 501–516.

Kingdon, Geeta G. (2005). “Where Has All the Bias Gone? Detecting Gender Bias in the Household Allocation of Educational Expenditure in India.” *Economic Development and Cultural Change*, 53, 409–451.

Kingdon, Geeta G. (2009). “Education of Females in India: Determinants and Economic Consequences, a Case Study of Uttar Pradesh.” Discussion Paper, Economics Department, Oxford University.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz (2007). "Experimental Analysis of Neighborhood Effects. " *Econometrica*, 75, 83–119.

Lalive, Rafael and Alejandra Cattaneo (2009). "Social Interactions and Schooling Decisions." *Review of Economics and Statistics*, 91, 457–477.

Lindbeck, Assar, Sten Nyberg, and Jörgen W. Weibull (1999). "Social Norms and Economic Incentives in the Welfare State." *Quarterly Journal of Economics*, 114, 1–35.

Lindbeck, Assar and Sten Nyberg (2006). "Raising Children to Work Hard: Altruism, Work Norms, and Social Insurance." *Quarterly Journal of Economics*, 121, 1473–1503.

Lopez-Calva, Luis-Felipe (2002). "A Social Stigma Model of Child Labor." *Estudios Economicos*, 17, 193–217.

Mani, Anandi and Charles H. Mullin (2004). "Choosing the Right Pond: Social Approval and Occupational Choice." *Journal of Labor Economics*, 22, 835–861.

Manski, Charles F. (1993). "Identification of Endogenous Social Effects: The Reflection Problem. " *Review of Economic Studies*, 60, 531–542.

Munshi, Kaivan and Jacques Myaux (2006). "Social Norms and the Fertility Transition." *Journal of Development Economics*, 80, 1–38.

Nechyba, Thomas J. (2001). "Social Approval, Values, and AFDC: A Reexamination of the Illegitimacy Debate." *Journal of Political Economy*, 109, 637–672.

North, Douglas C. (1990). "*Institutions, Institutional Change, and Economic Performance.*" Cambridge University Press, Cambridge.

Palivos, Theodore (2001). "Social Norms, Fertility and Economic Development." *Journal of Economic Dynamics and Control*, 25, 1919–1934.

Psacharopoulos, George (1997). "Child Labor versus Educational Attainment: Some Evidence from Latin America." *Journal of Population Economics*, 10, 377–386.

Ray, Ranjan (2000). "Analysis of Child Labor in Peru and Pakistan: A Comparative Study." *Journal of Population Economics*, 13, 3–19.

Ravallion, Martin and Quentin Wodon (2000). "Does Child Labor Displace Schooling? Evidence on Behavioral Responses to an Enrollment Subsidy." *Economic Journal*, 110, C158–C175.

Sawada, Yasuyuki and Michael Lokshin (2009). "Obstacles to School Progression in Rural Pakistan: An Analysis of Gender and Sibling Rivalry Using Field Survey Data." *Journal of Development Economics*, 88, 335–347.

Schultz, T.Paul (2004). "School Subsidies for the Poor: Evaluating the Mexican PROGRESA Poverty Program. " *Journal of Development Economics*, 74, 199–250.

Skoufias, Emmanuel and Susan W. Parker (2001). "Conditional Cash Transfers and their Impact on Child Work and Schooling: Evidence from the PROGRESA Program in Mexico." *Economica*, Fall 2001, 45–96.

Strulik, Holger (2004). "Child Mortality, Child Labour, and Economic Development." *Economic Journal*, 114, 547–568.

Strulik, Holger (2011). "Riding High – Success in Sports and the Rise of Doping Cultures." *Scandinavian Journal of Economics*, forthcoming.

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

UCW (2010). "*Understanding Children's Work: Child labor: Trends, Challenges, and Policy responses.*" International Labor Organization, Geneva.

UNICEF (2008). "*Socio-economic and Cultural Barriers to Schooling in Southern Sudan.*" United Nations Children's Fund, New York.

United Nations (2010). "Goal 2: Achieve Universal Primary Education." Fact Sheet, UN Department of Public Information, New York.

Zelizer, Viviana A. (1985). "*Pricing the Priceless Child: The Changing Social Value of Children*" Basic Books, New York.