THE EFFECT OF UNDERCLASS SOCIAL ISOLATION ON SCHOOLING CHOICE

PETER STREUFERT
University of Western Ontario

Abstract
This paper models underclass social isolation as the loss of high-income role models and then studies the plausible conjecture that this isolation depresses the level of schooling chosen by underclass youth. It is found that although this conjecture stands in roughly calibrated simulations, it is not a theoretical necessity. The paper also shows that the introduction of more representative role models during university can polarize underclass youth and that income support programs depress schooling choice through a distinct and reinforcing channel.

1. Introduction

This paper studies a central thesis of Wilson’s (1987) book The Truly Disadvantaged, namely, the thesis that the social isolation of the underclass causes underclass youth to underestimate the effect of schooling on income and consequently, to choose too little schooling.

When he states this thesis most precisely (pp. 56–58), Wilson argues that “a perceptive ghetto youngster in a neighborhood that includes a good number of working and professional families . . . can see a connection between education and meaningful employment.” Yet, “in a neighborhood with a paucity of regularly employed families . . . the relationship between schooling and postschool employment takes on a different meaning.” This results in “a shockingly high degree of educational retardation.”
We begin by modeling how role models shape a young person’s perception of the relationship between schooling and income. A role model is taken to be a single observation of the schooling and income of an adult worker, and many such observations enable a young person to (non-parametrically) regress income on schooling. Thus a young person’s information-gathering process resembles a labor economist’s estimation of the earnings function. If the role models that a young person observes are representative of the labor force, she will be well informed when making her schooling decision. Section 2 defines schooling choice in this ideal “benchmark” model, provides a useful incremental characterization of this choice (Theorem 1), and roughly calibrates a “Benchmark Simulation” to fit Mincer’s earnings function and the schooling distribution of the U.S. labor force.

Underclass social isolation is then modeled as the elimination of high-income observations at each level of schooling. That is, we assume that an underclass youth observes a sample of role models that is truncated from above. Wilson’s thesis suggests that the resulting bias in the regression depresses schooling.

Section 3’s “Isolation Simulation” accords with this thesis. In it, a young person who completed high school in the Benchmark Simulation only completes grade eight. This precipitous drop in her schooling results from the lack of any role models earning more than $15,000/year. An economist would say that the representative role models she lacks define a public good that is lacking in an underclass neighborhood. Alternatively, a sociologist might be pleased to hear an economist argue that a “rational” individual’s schooling decision depends critically on that individual’s social environment.

However, Section 3’s “Theoretical Counterexample” shows that social isolation need not depress schooling. If Wilson’s conjecture stands, it stands because social isolation reduces the perceived increment to income that would result from further schooling. But this need not be the case, as demonstrated by the counterexample. Furthermore, social isolation inevitably causes a young person to underestimate the income she forgoes while attending school, and this factor can only serve to make school more attractive. Thus, social isolation depresses schooling only if (1) it leads to a decrease rather than an increase in perceived incremental income and (2) the magnitude of this decrease is sufficiently great to overcome the underestimation of forgone income. (In statistical terminology, truncation depresses schooling only if (1) it decreases rather than increases the slope of the regression and (2) the magnitude of this decrease outweighs the drop in the level of the regression.)

Section 4 extends the model by allowing a person’s collection of role models to change as she proceeds through school. For example, the “Polarization Simulation” assumes that a person’s social isolation is lifted if and when she enters university. Such changing role models split underclass youth into two disparate groups: one whose schooling is stunted by
underclass isolation and another that is ultimately unaffected by its under-
class origins. No one in the first group starts university, while everyone in
the second group goes on to graduate school.

Section 5 alters the model to study the effects of income support
programs. In contrast to social isolation, income support unambiguously
reduces incremental income (i.e., the slope of the regression) and increases
forgone income (i.e., the level of the regression). Thus, income support
inevitably depresses schooling (Theorem 2). The “Support Simulation”
shows that this effect can be quite pronounced in the lower tail of the
population.

The “Combination Simulation” of Section 6 shows that social isolation
and income support are likely to depress schooling through distinct and
reinforcing channels. One squeezes the regression from above, and the
other squeezes from below.

Section 7 concludes.

2. Benchmark Model

2.1 Specification

Imagine that a young person must choose a level of schooling $s$ from the
13-element set \{7, 8, ..., 19\}. For example, the choice $s = 18$ describes
someone who chooses to enter the labor force after completing 18 grades
(i.e., high school, four years of university, and two years of postgraduate
training), at about 24 years of age. The choice $s = 7$ is less straightforward
because few children literally enter the labor force after completing seven
grades, at about 13 years of age. Rather, the choice $s = 7$ is better inter-
preted as someone who becomes unmotivated during junior high school,
fails two grades, and then quits altogether as soon as she is able to do so
(in the United States, this becomes legal on her 16th birthday).

One key factor in a young person’s decision is the effect that $s$ will
have upon her future annual income $y$ (for simplicity, assume $y$ is constant
throughout her career and ignore nonwage income). The actual relation-
ship between $s$ and $y$ is specified by the set of cumulative distribution
functions\(^1\) \(\{F_s \}_{s \in \{7, 8, ..., 19\}}\). For every schooling $s$ and income $y$, $F_s(y)$ is the
probability that a person with schooling $s$ will earn an income of less than
or equal to $y$.

Imagine that a young person learns about \(\{F_s\}\) by observing role
models from the labor force. Formally, a role model is a single observa-
tion of schooling $s$ and income $y$. As elucidated by Manski (1993), this
mathematical concept of role models accords with the sociological con-
cept of role models, and a collection (i.e., sample) of such role models
accords with the sociological concept of a reference group.

\(^1\)Each $F_s: \mathbb{R} \rightarrow [0,1]$ is assumed to have a finite mean.
For simplicity, we assume asymptotic sampling. That is, we assume that for every \( s \), a young person learns the entire cumulative distribution function rather than needing to draw statistical inferences from a finite sample. Thus, at every \( s \), the sample mean income and the population mean income coincide at \( E[F_s] \). In other words, \( \langle E[F_s] \rangle \), can be regarded as a perfect nonparametric regression of income \( y \) on schooling \( s \).

Let \( \delta \in (0,1) \) be a person’s discount factor and let \( \theta \in \mathbb{R} \) specify the annual disutility a person experiences as the result of attending school. She then chooses her schooling \( s \) so as to maximize the difference between the income of her work years and the disutility of her school years. Formally, a person with disutility parameter \( \theta \) uses the regression \( \langle E[F_s] \rangle \) to select the schooling

\[
S^\theta(\langle E[F_s] \rangle) = \max \left\{ \arg \max \left\{ \sum_{a = \sigma + 1}^{a = 59} \delta^{a - 1} E[F_a] - \sum_{a = 1}^{a = \sigma} \delta^{a - 1} \theta \right| \sigma \in \{7,8,\ldots,19\} \right\}.
\]

Note that \( \sigma \) rather than \( s \) is used as the choice variable so as to avoid confusion with the \( s \) in \( \langle E[F_s] \rangle \), that \( a = 59 \) coincides with retirement at 65 years from birth, and that the outer max operator simply means that if several schoolings maximize utility then the tie is resolved in favor of the highest schooling. This choice problem closely resembles Rosen (1977, pp. 9–13), and numerous variations appear elsewhere in the labor economics literature (e.g., Heckman 1976; and Ryder, Stafford, and Stefan 1976, to name but two). Notice that schooling unambiguously declines with the disutility parameter \( \theta \) but that the effects of the discount factor \( \delta \) and the regression \( \langle E[F_s] \rangle \), are ambiguous.

### 2.2 Incremental Characterization of Schooling

Given the regression \( \langle E[F_s] \rangle \), we can derive the incremental benefit of attending grade \( \sigma \in \{8,9,\ldots,19\} \) as

\[
B_\sigma(\langle E[F_s] \rangle) = \delta E[F_{\sigma + 1}] - E[F_{\sigma - 1}](1 - \delta^{59 - \sigma})/(1 - \delta) - E[F_{\sigma - 1}].
\]

The first term is the present discounted value of the increase in future income that a person will enjoy as a result of investing in grade \( \sigma \) as opposed to grade \( \sigma - 1 \). The second term is the income a person will forgo while attending grade \( \sigma \). The following theorem states that a person will continue to attend school until the benefit of an additional year of schooling falls below her disutility.

**THEOREM 1:** Suppose that \( B_\sigma(\langle E[F_s] \rangle) \) is weakly decreasing in \( \sigma \). Then

\[
S^\theta(\langle E[F_s] \rangle) = \max\{7\} \cup \{\sigma \geq 8 \mid B_\sigma(\langle E[F_s] \rangle) \geq \theta\}.
\]

**Proof:** See the Appendix.

The incremental net benefit \( B_\sigma(\langle E[F_s] \rangle) \) is weakly decreasing in \( \sigma \) whenever two conditions are met: (1) that the increment to future income
is declining with each additional year of schooling, in other words, that \( \langle E[F]\rangle_i \) is concave:

\[
(\forall s \in [9,10,\ldots,19]) E[F_s] - E[F_{s-1}] \leq E[F_{s-1}] - E[F_{s-2}]
\]

and (2) that forgone income is increasing with each additional year of schooling, in other words, that \( \langle E[F]\rangle_i \) is increasing.

However, \( B_\sigma(\langle E[F]\rangle_i) \) can be weakly decreasing in \( \sigma \) even when the increment to future income is not decreasing (as, e.g., in the Benchmark Simulation). This occurs when the increase in the (discounted) increment to future income is outweighed by the increase in forgone income (which is more likely to happen when the discount factor is low). This discussion is important because \( \langle E[F]\rangle_i \) is convex in the large empirical literature that linearly regresses the logarithm of income on schooling.

### 2.3 Benchmark Simulation

**Parameters.** Each person’s optimization problem has three parameters: the nonparametric regression \( \langle E[F]\rangle_i \), the discount factor \( \delta \), and the disutility parameter \( \theta \). All people share the same \( \langle E[F]\rangle_i \) and \( \delta \), while each has her own unique \( \theta \).

For every \( y \), specify \( F_s \) as the lognormal cumulative distribution function with mean \( E[F_s] \) satisfying

\[
\sum_{a=s+1}^{50} \delta^{-a(s+1)} E[F_s] = \sum_{a=s+1}^{50} \delta^{-a(s+1)} y(s, a - s),
\]

where

\[
y(s, x) = 4480(4.87 + 0.255s - 0.0029s^2 - 0.0043sx + 0.148x - 0.0018s^2),
\]

and with standard deviation equal to half of this mean. The first equation states that \( E[F_s] \) is the present discounted value of lifetime earnings, expressed at an annual rate. The second equation calculates each year’s earnings as a function of schooling \( s \) and experience \( x = a - s \). This second equation is the earnings function estimated by Mincer (1974), discussed by Willis (1986, Table 10.5), and crudely adjusted for inflation (so that income is measured in $/year). Note that \( \langle E[F]\rangle_i \) is depicted in Figure 1a as the solid curve.

Specify the discount factor as \( \delta = 0.85 \). This discount rate can be regarded as modeling not only time preference but also financial constraints. A discount factor of 0.85 corresponds to a discount rate of 17%. As the rate of return to schooling is about 17% at \( s = 8 \) (Willis 1986, p. 546), significantly lower discount rates result in no schooling choices near \( s = 8 \).

The simulation concerns 100 individuals that differ only in their disutility parameter \( \theta \). The 100 values chosen can be seen as the vertical
coordinates of the 100 stars in Figure 1c. These 100 $\theta$s are “log-beta-ly” distributed: the base-2 logarithms of a linear transformation of the 100 $\theta$'s are percentiles from a symmetric beta distribution (the remarks in the procedure maketv of Streufert 1999 give full details). We chose this distribution of $\theta$s so that the resulting distribution of schooling would roughly
match the actual distribution of schooling (U.S. Bureau of the Census 1990, Table 632). Most individuals have been assigned negative disutilities: this should be interpreted to mean that most individuals find school less onerous than work.

**Results.** Each of the 100 individuals is depicted by a star in Figure 1c. The star’s vertical coordinate gives her disutility parameter \( \theta \), and its horizontal coordinate shows her chosen schooling \( S^\theta(\langle E[F_s] \rangle) \). For example, persons \( \theta = \$402/\text{year} \) and \( \theta = \$301/\text{year} \) have been assigned the greatest disutility parameters and only complete grade seven. Meanwhile person \( \theta = \$229/\text{year} \) has been assigned the third-highest disutility parameter and completes grade eight. In accord with Theorem 1, their decisions are completely explained by the fact that the incremental net benefit to attending grade eight, namely, \( B_8(\langle E[F_s] \rangle) \), lies below \$301 and above \$229. In general, the incremental net benefit \( B_s(\langle E[F_s] \rangle) \) lies just slightly above the top star at schooling \( s \). Note that \( B_s(\langle E[F_s] \rangle) \) is decreasing in \( s \), as assumed by Theorem 1.

As a consequence of their chosen schooling \( s \), each person receives an income \( y \) drawn from the distribution \( F_s \). This is depicted by the 100 stars in Figure 1a. For example, persons \( \theta = -\$9,951/\text{year} \) and \( \theta = -\$11,801/\text{year} \) (the two persons with the lowest disutility parameters) chose 19 years of schooling, and their two incomes are depicted by the two stars at \( s = 19 \) in Figure 1a. In particular, these stars are at the 25th and 75th percentiles of the distribution \( F_{19} \) (they are located asymmetrically around the mean \( E[F_{19}] \) because \( F_{19} \) is a lognormal distribution).

Figure 1b shows the density of the aggregate distribution of income that results from the schooling decisions shown in Figure 1c. Crudely speaking, this density is found by sweeping all of the stars in Figure 1a to the vertical axis and letting them pile up.

### 3. Social Isolation

#### 3.1 Theory

We model underclass social isolation by assuming that an underclass youth observes, for each schooling \( s \), a distribution of incomes that is truncated from above by \( \alpha \). This truncation models selective out-migration from an underclass neighborhood: everyone with an income above \( \alpha \) leaves the ghetto, while everyone with an income at or below \( \alpha \) remains in the ghetto and provides a role model for underclass youth.

Formally, let \( E[F_s^\alpha] \) denote the expectation of the distribution \( F_s^\alpha \) that is obtained by truncating the distribution \( F_s \) from above at \( \alpha \geq 0 \). A person with disutility parameter \( \theta \) observing the truncated regression \( \langle E[F_s^\alpha] \rangle \), will choose schooling \( S^\theta(\langle E[F_s^\alpha] \rangle) \). Wilson’s conjecture is that

\[ F_s^\alpha(y) = \begin{cases} 1 & \text{if } y > \alpha \\ F_s(y)/F_s(\alpha) & \text{if } y \leq \alpha \end{cases} \]

\[ F_s^\alpha(y) = \begin{cases} 1 & \text{if } y > \alpha \\ F_s(y)/F_s(\alpha) & \text{if } y \leq \alpha \end{cases} \]

That is, define \( F_s^\alpha(y) = 1 \) if \( y > \alpha \) and \( F_s^\alpha(y) = F_s(y)/F_s(\alpha) \) if \( y \leq \alpha \).
truncation discourages schooling, in other words, that a decrease in $\alpha$ will cause a decrease in $S^\alpha(\mathbb{E}[F_\alpha])$. This does occur in the Isolation Simulation of Section 3.2. However, this is not a logical necessity, as demonstrated by the Theoretical Counterexample of Section 3.3.

3.2 Isolation Simulation

Parameters. The Isolation Simulation alters the Benchmark Simulation by truncating every distribution $F_i$ from above at $\alpha = 15,000$/year. This means that no young person ever observes anyone earning more than $15,000$/year. (This cutoff happens to be above both the mean and the median of true income distribution $F_{12}$.)

Results. Figure 2a depicts the $15,000$/year truncation (solid line) which shifts the regression down from $\mathbb{E}[F_i]$ to $\mathbb{E}[F_i^{15000}]$, i.e., from the dotted curve to the solid curve.

Figure 2c shows that this misperception shifts schooling back (from the small stars to the large stars). For example, consider the person with the third highest disutility parameter, that is, person $u = 229$/year. In the Benchmark Simulation (small stars in Figure 2c), she chose eight grades. Here in the Isolation Simulation (large stars in Figure 2c), she chooses seven grades. She does this because the perceived incremental benefit of grade eight has fallen from something above $229$ to something below $229$.

The perceived incremental benefit of grade eight can be determined more precisely. In the Benchmark Simulation, it was between $301$ and $229$ because person $\theta = 301$/year chose grade seven and person $\theta = 229$/year chose grade eight. Here, it lies between $-1,120$/year and $-1,165$/year because person $\theta = -1,120$/year chooses grade seven and person $\theta = -1,165$/year chooses grade eight.

Something similar is occurring at all grade levels. The incremental benefit of grade $s$ is approximately equal to the highest star at grade $s$. Isolation has caused this incremental benefit to shift down. By Theorem 1, this is equivalent to shifting back schooling. All this can be seen in Figure 2c: the downward shift of the highest stars explains the backward shift of every person’s schooling choice.

The consequences of this backward shift in schooling are drawn out in the remaining diagrams. Figure 2d shows the (lifetime) utility loss that each person suffers, and Figure 2b shows the downward shift in the income distribution. The stars in Figure 2a depict the incomes that the 100 persons receive. Twelve stars lie above the truncation line. This suggests that 12 persons will escape the ghetto, while the remaining 88 will stay behind and provide a truncated sample of role models for the next generation (this sentence is not formally consistent with the model since the income coordinates of Figure 2a’s stars are percentiles rather than random variables and since the next generation will draw an asymptotic rather than finite sample).
3.3 Theoretical Counterexample

As suggested by the Isolation Simulation (and as formally proven in Theorem 1 under the assumption that the incremental benefit of grade $\sigma$...
decreases with $\sigma$), truncation will depress schooling precisely when it depresses the perceived incremental benefit

$$B_{\sigma}(E[F_{\sigma}]) = \delta(E[F_{\sigma}^\sigma] - E[F_{\sigma-1}^\sigma](1 - \delta^{s_{\sigma}})/(1 - \delta) - E[F_{\sigma-1}^\sigma].$$

Truncation need not depress the incremental benefit: the effect of truncation on the first term is ambiguous, and its effect on the second term is unambiguously in the opposite direction.

Begin with the second term. $E[F_{\sigma-1}]$ is the level of the regression at $\sigma - 1$, and it is the income one forgoes while attending grade $\sigma$ rather than starting work after grade $\sigma - 1$. Truncation must depress one’s perception of forgone income, and this can only serve to increase the perceived incremental benefit of grade $\sigma$.

The first term concerns the slope of the regression line between $s = \sigma - 1$ and $s = \sigma$, namely, $E[F_{\sigma}^\sigma] - E[F_{\sigma-1}^\sigma]$. This slope is the incremental income one receives in each working year as a result of attending grade $\sigma$ rather than grade $\sigma - 1$. Although it is plausible that truncation will flatten the regression line as it did in the Isolation Simulation, this is not a theoretical necessity.

This conceptual possibility is illustrated by the Theoretical Counterexample. In this simulation, each $F_s$ is defined to place one-third of the probability at each of the three triangles shown at $s$ in Figure 3a. At every $s$, one triangle is located at $\$50,000/year, a second is located at $\$30,000/year, and the third is located at $s$ times $\$1,000/year. Truncation at $\$20,000/year (Figure 3a’s flat solid line) lops off the top two triangles from every $F_s$, and this serves to accentuate the differences between the distributions. As a result, the truncated regression (solid upward-sloping line) is steeper than the true regression (dotted upward-sloping line).

Thus, in the Theoretical Counterexample, truncation leads one to underestimate the income forgone by another year of schooling and to overestimate the incremental income that will result. Both factors lead one to overestimate the incremental benefit of schooling.

Figure 3b’s small stars show that all persons would choose seven grades when perceiving the true regression. This implies that the true incremental benefit of grade eight is less than the lowest disutility parameter of $\theta = -\$11,801/year. Figure 3b’s large stars show that everyone with disutility parameter less than $\theta = -\$1,300/year would choose more schooling if they misperceived the regression because of truncation. This implies that the incremental benefit of grade eight has risen to at least $-\$1,300/year because of truncation.

4. Polarization

4.1 Theory

This section modifies the model of the previous section by letting the truncation cutoff $\alpha$ vary as an underclass youth moves through school.
Her schooling decision then becomes a sequential problem: she continues on in school if and only if the role models that she has observed up until that time suggest that some further schooling is better than stopping immediately.

It is reasonable to assume that $\alpha$ increases over time, for this would model the notion that, as an underclass youth moves from grade school to
high school to university and finally to graduate school, the social isolation of her underclass origins is gradually lifted. Unfortunately, early social isolation might so depress her schooling choice that she drops out before observing the new role models.

Formally, let $\langle \alpha_t \rangle_{t \in \{7, 8, \ldots, 19\}}$ be a sequence of cutoffs. We assume that a person with disutility parameter $\theta$ and truncation cutoffs $\langle \alpha_t \rangle_{t \in \{7, 8, \ldots, 19\}}$ will choose schooling

$$\min\{t \geq 7 | S_t^\theta((E[F_t^{\alpha_t}]),) = t\},$$

where

$$S_t^\theta((E[F_t^{\alpha_t}]),)$$

$$= \max \left\{ \arg \max \left\{ \sum_{\alpha = \sigma+1}^{19} \delta^{\alpha-1} E[F_{\sigma}^{\alpha}] - \sum_{\sigma = 1}^{\alpha} \delta^{\alpha-1} \theta | \alpha \in \{t, t+1, \ldots, 19\} \right\} \right\}.$$

The first expression says that a person will stop attending school in the first grade $t$ such that the role models she is viewing under the truncation cutoff $\alpha_t$ lead her to choose $t$ grades of schooling. The second expression defines the symbol $S_t$ in a manner that is identical to $S$ except for the constraint that a person cannot choose a level of schooling lower than $t$ (thus if she doesn’t choose $t$ grades she is choosing to continue rather than regress).

Note that a person does not anticipate that her role models might change and therefore, she has no incentive to continue in school for the sake of gathering information (Manski and Wise 1983, p. 10).

### 4.2 Polarization Simulation

**Parameters.** This simulation differs from the Isolation Simulation in that a person’s isolation ends abruptly when she enters university. Formally, $\alpha_t = -$15,000/year for $t \leq 12$, and $\alpha_t = +\infty$/year for $t > 13$.

**Results.** See Figure 4. This simulation is essentially an amalgamation of the Isolation and Benchmark Simulations: isolation causes an underclass youth to choose too little schooling (as in the Isolation Simulation) unless she chooses to begin university in spite of her isolation, in which case she chooses an efficient level of schooling (as in the Benchmark Simulation) on the basis of role models observed after entering university.

This discontinuous change in information divides underclass youth into two classes: school-loving youth (with disutility parameters $\theta < -$7,000/year and postgraduate training) who are ultimately unaffected by the isolation of their ghetto origins and their comparatively school-averse peers (with disutility parameters $\theta > -$7,000/year and no more than a high-school education) who never overcome their isolation. This tendency to polarization accords with empirical findings that the income distribution of black men has become polarized (Murray 1984, p. 92 (citing Kilson 1981); and Wilson 1987, p. 45 (citing Levy 1986)).
5. Income Support

5.1 Theory

Return to the benchmark model. Then modify it by introducing a stylized income support program in which everyone who has left school is guar-
anteed an income of at least $\beta$. This implies that an underclass youth will observe no role models with an income below $\beta$ and will observe an appreciable number of role models earning exactly $\beta$.

Formally, let $E[F_s, \beta]$ denote the mean of the distribution $F_s, \beta$ derived from $F_s$ by censoring from below at $\beta \geq 0$. Then suppose that a person with disutility parameter $\theta$ uses the regression $\langle E[F_s, \beta] \rangle_s$ to choose the schooling $S^\theta(\langle E[F_s, \beta] \rangle_s)$, where $S$ is defined as in the benchmark model.

The following theorem shows that support must depress schooling. The gist of its proof concerns the effect of support on the incremental benefit of grade $s$, namely,

$$B_s(\langle E[F_s, \beta] \rangle_s) = \delta(\langle F_s, \beta \rangle - \langle F_{s-1}, \beta \rangle)(1 - \delta^{59-s})/(1 - \delta) - \langle F_{s-1}, \beta \rangle.$$  

Support must diminish incremental future income (the first term, which is the slope of the regression) and must increase forgone income (the magnitude of the second term, which is the level of the regression). Both effects decrease the incremental benefit of grade $s$ and thereby depress schooling.

**THEOREM 2:** Suppose that $(\forall s \in \{8, 9, \ldots, 19\}) F_s$ first-order stochastically dominates $F_{s-1}$. Then $\beta' \geq \beta$ implies $S^\theta(\langle E[F_s, \beta'] \rangle_s) \leq S^\theta(\langle E[F_s, \beta] \rangle_s)$.

**Proof:** See the Appendix.

Furthermore, an income support program harms society as a whole in the sense of the following theorem’s inequality. Its left-hand side is the utility gained by a person as a consequence of income support $\beta$. Its right-hand side is the government’s cost of the support provided to this person. The two are comparable because a person’s utility is measured in units of income.

**THEOREM 3:** Let $s^* = S^\theta(\langle E[F_s] \rangle_s)$ be schooling in the benchmark model and let $s = S^\theta(\langle E[F_s, \beta] \rangle_s)$ be schooling given income support $\beta$. Then

$$\left( \sum_{a=s+1}^{59} \delta^{a-1}E[F_{s, \beta}] - \sum_{a=1}^{s} \delta^{a-1} \theta \right) - \left( \sum_{a=s^*+1}^{59} \delta^{a-1}E[F_{s^*}] - \sum_{a=1}^{s^*} \delta^{a-1} \theta \right) \leq \sum_{a=s+1}^{59} \delta^{a-1}(E[F_{s, \beta}] - E[F_s]).$$  

**Proof:** See the Appendix.

### 5.2 Support Simulation

**Parameters.** The Support Simulation alters the Benchmark Simulation by censoring every distribution $F_s$ from below at $\beta = $ $5,000$/year. This means

\[F_{s, \beta}(y) = 0 \text{ if } y < \beta \text{ and } F_{s, \beta}(y) = F_s(y) \text{ if } y \geq \beta.\]
that every role model who would have earned less than $5,000/year instead making $5,000/year as a consequence of the income support program. This guaranteed income happens to be below both the mean and the median of the true income distribution $F$. 

Figure 5: Support simulation. Figure 5a depicts a $5,000/year income support (solid line) that shifts the regression up (from dotted curve to solid curve). Figure 5c shows that this causes 34 persons to complete only grade seven. Consequently, they earn the (unsupported) incomes depicted at $s = 7$ in Figure 5a and cause Figure 5b's aggregate income distribution to shift down (from dotted curve to solid curve).
Results. Theorem 2 states that support must depress schooling. This simulation shows that the effect can be rather severe under a modest support program that ensures that everyone makes at least $5,000/year. Figure 5c’s large stars show that 34% of the simulation’s hundred persons complete only seven grades. In marked contrast, Figure 5c’s small stars show that only two persons made this choice in the Benchmark Simulation.

While the Isolation Simulation showed that social isolation depresses almost everyone’s schooling choice, this simulation suggests that support affects mainly the lower portion of the population. The most severely affected person has disutility parameter $\theta = -996/\text{year}$. She completes only seven grades even though she would have chosen 11 grades in the absence of income support (Figure 5c). (There is no figure depicting utility loss. Support benefits every individual (if they don’t have to pay for it) and harms society as a whole (in the sense of Theorem 3).)

6. Combination Simulation

Parameters. This simulation combines the Polarization and Support Simulations. In particular, it models social isolation with $\alpha_t = 15,000/\text{year}$ for $t \leq 12$ and $\alpha_t = +20/\text{year}$ for $t \geq 13$, and it models an income support program with $\beta = 5,000/\text{year}$.

Results. See Figure 6. Because social isolation and income support squeeze the regression line from opposite sides, one might anticipate that the two depress schooling through distinct and reinforcing channels. This simulation accords with that intuition. No person chooses more schooling here than they did in either the Polarization Simulation or the Support Simulation, and 43 persons choose less than they did in either. For example, person $\theta = -2,752/\text{year}$ chose 14 grades in the Benchmark Simulation, nine grades in the Isolation and Polarization Simulations, 14 grades in the Support Simulation, and seven grades here in the Combination Simulation.

7. Conclusions

Although the effect of social isolation on schooling choice is ambiguous from a theoretical perspective, our simulations suggest that it can markedly depress schooling and that it can divide underclass youth into two polarized classes. Social isolation would be reduced if geographical neighborhoods were made more representative of the entire society. But, it may be more efficacious to focus on other sorts of “neighborhoods,” such as networks of family or friends, or organizations concerned with cultural or recreational pursuits. It seems that disembodied information about the true relationship between schooling and income is not easily transmitted.

Income support inevitably depresses schooling, and our simulations suggest that it affects primarily the lower portion of the population. This
is not a matter of misinformation. It is an unintended consequence of income support.

Finally, it is hoped that this paper and its computer programs will inform empirical studies of the relationship between social isolation and schooling choice. The fundamental factor is more complex than mean neighborhood income. It is the representativeness of the adults with which young people can identify.

Figure 6: Combination simulation. Social isolation and income support depress schooling through distinct and reinforcing channels. When childhood role models earn less than $15,000/year and a support program guarantees an income of $5,000/year, 66 of the 100 persons complete only seven grades (big stars in Figure 6c).
Appendix

Proof of Theorem 1: Since the following statements are equivalent by algebraic manipulation, we conclude that $\sigma$ is weakly preferred to $\sigma - 1$ if and only if $B_{\sigma}(\langle E[F] \rangle_\sigma) \geq \theta$.

\[
\sum_{a=\sigma+1}^{59} \delta^{a-1} E[F]_{\sigma} - \sum_{a=1}^{\sigma} \delta^{a-1} \theta \geq \sum_{a=\sigma}^{59} \delta^{a-1} E[F]_{\sigma-1} - \sum_{a=1}^{\sigma-1} \delta^{a-1} \theta,
\]

\[
\sum_{a=\sigma+1}^{59} \delta^{a-1} E[F]_{\sigma} - \delta^{\sigma-1} \theta \geq \sum_{a=\sigma}^{59} \delta^{a-1} E[F]_{\sigma-1} + \delta^{\sigma-1} E[F]_{\sigma-1}, \]

\[
\sum_{a=\sigma+1}^{59} \delta^{a-1} (E[F]_{\sigma} - E[F]_{\sigma-1]) - \delta^{\sigma-1} E[F]_{\sigma-1} \geq \delta^{\sigma-1} \theta,
\]

\[
\sum_{a=\sigma+1}^{59} \delta^{(a-1)(\sigma-1)} (E[F]_{\sigma} - E[F]_{\sigma-1}) - E[F]_{\sigma-1} \geq \theta,
\]

\[
\sum_{q=1}^{50-\sigma} \delta^{q} (E[F]_{\sigma} - E[F]_{\sigma-1}) - E[F]_{\sigma-1} \geq \theta,
\]

\[
\delta (E[F]_{\sigma} - E[F]_{\sigma-1})(1 - \delta^{50-\sigma})/(1 - \delta) - E[F]_{\sigma-1} \geq \theta,
\]

\[
B_{\sigma}(\langle E[F] \rangle_\sigma) \geq \theta.
\]

For notational ease, define $\sigma^* = S^\theta(\langle E[F] \rangle_\sigma)$. Note that

\[
\sigma^* \in \{7\} \cup \{\sigma \geq 8 | B_{\sigma}(\langle E[F] \rangle_\sigma) \geq \theta\}, \tag{1}
\]

for if $\sigma^* \geq 8$, its optimality implies that it must be weakly preferred to $\sigma^* - 1$ and hence that $B_{\sigma^*}(\langle E[F] \rangle_\sigma) \geq \theta$ by the theorem’s first sentence. On the other hand, since $\sigma^*$ is defined to be the highest schooling that maximizes utility, it must be that $\sigma^* + 1$ is not weakly preferred to $\sigma^*$ and hence that $B_{\sigma^*+1}(\langle E[F] \rangle_\sigma) < \theta$ by the proof’s first sentence. Since $B_{\sigma^*}(\langle E[F] \rangle_\sigma)$ is assumed to be weakly decreasing in $\sigma$, this implies that

\[
(\forall \sigma > \sigma^*) B_{\sigma}(\langle E[F] \rangle_\sigma) < \theta. \tag{2}
\]

Equations (1) and (2) together imply that $\sigma^* = \max\{7\} \cup \{\sigma \geq 8 | B_{\sigma}(\langle E[F] \rangle_\sigma) \geq \theta\}$. ■

Proof of Theorem 2: This paragraph shows that

\[
(\forall \beta^* \geq \beta)(\forall \sigma^+ > \sigma),
\]

\[
E[F_{\sigma^+, \beta^*}] - E[F_{\sigma, \beta^*}] \leq E[F_{\sigma^+, \beta}] - E[F_{\sigma, \beta}],
\]
(intuitively, this means that increased support diminishes incremental income). The inequality is derived from

\[
(\forall \beta' \geq \beta)(\forall \sigma^+ > \sigma)

E[F_{\sigma^+, \beta'}] - E[F_{\sigma^+, \beta}]

= \left\{ \beta' + \int_{\beta'}^{+\infty} (1 - F_{\sigma^+}(y)) \, dy \right\} - \left\{ \beta + \int_{\beta}^{+\infty} (1 - F_{\sigma^+}(y)) \, dy \right\}

= \beta' - \beta - \int_{\beta}^{\beta'} (1 - F_{\sigma^+}(y)) \, dy

\leq \int_{\beta}^{\beta'} F_{\sigma}(y) \, dy

= \beta' - \beta - \int_{\beta}^{\beta'} (1 - F_{\sigma}(y)) \, dy

= \left\{ \beta' + \int_{\beta'}^{+\infty} (1 - F_{\sigma}(y)) \, dy \right\} - \left\{ \beta + \int_{\beta}^{+\infty} (1 - F_{\sigma}(y)) \, dy \right\}

= E[F_{\sigma^+, \beta'}] - E[F_{\sigma, \beta}],
\]

where the inequality follows from the theorem’s assumption that \( F_{\sigma^+} \) first-order stochastically dominates \( F_{\sigma} \).

Now define \( \sigma = S_\delta((E[F_{\sigma, \beta}]), \). Then

\[
(\forall \beta' \geq \beta)(\forall \sigma^+ > \sigma)

\left\{ \sum_{a=\sigma^++1}^{\sigma^+} \delta^{a-1} E[F_{\sigma^+, \beta'}] - \sum_{a=1}^{\sigma^+} \delta^{a-1} \theta \right\} - \left\{ \sum_{a=\sigma^++1}^{\sigma^+} \delta^{a-1} E[F_{\sigma^+, \beta'}] - \sum_{a=1}^{\sigma} \delta^{a-1} \theta \right\}

= \sum_{a=\sigma^++1}^{\sigma^+} \delta^{a-1} \{ E[F_{\sigma^+, \beta'}] - E[F_{\sigma, \beta'}] \} - \sum_{a=1}^{\sigma^+} \delta^{a-1} \theta

- \sum_{a=\sigma^++1}^{\sigma^+} \delta^{a-1} E[F_{\sigma^+, \beta'}] + \sum_{a=1}^{\sigma} \delta^{a-1} \theta
\]
\[
\leq \sum_{a=\sigma^+ + 1}^{59} \delta^{\sigma - 1} \{E[F_{\sigma^+\beta}] - E[F_{\sigma\beta}]\} - \sum_{a=1}^{\sigma^+} \delta^{\sigma - 1} \theta
\]
\[
- \sum_{a=\sigma^+ + 1}^{\sigma^+} \delta^{\sigma - 1} E[F_{\sigma\beta}] + \sum_{a=1}^{\sigma^+} \delta^{\sigma - 1} \theta
\]
\[
\leq \sum_{a=\sigma^+ + 1}^{59} \delta^{\sigma - 1} \{E[F_{\sigma^+\beta}] - E[F_{\sigma\beta}]\} - \sum_{a=1}^{\sigma^+} \delta^{\sigma - 1} \theta
\]
\[
- \sum_{a=\sigma^+ + 1}^{\sigma^+} \delta^{\sigma - 1} E[F_{\sigma\beta}] + \sum_{a=1}^{\sigma^+} \delta^{\sigma - 1} \theta
\]
\[
= \left\{ \sum_{a=\sigma^+ + 1}^{59} \delta^{\sigma - 1} E[F_{\sigma^+\beta}] - \sum_{a=1}^{\sigma^+} \delta^{\sigma - 1} \theta \right\}
\]
\[
- \left\{ \sum_{a=\sigma^+ + 1}^{59} \delta^{\sigma - 1} E[F_{\sigma\beta}] - \sum_{a=1}^{\sigma^+} \delta^{\sigma - 1} \theta \right\}
\]
\[
< 0,
\]

where the first inequality follows from the preceding paragraph (i.e., support diminishes incremental income), the second inequality follows from \(E[F_{\sigma\beta}] \geq E[F_{\sigma\beta}]\) (i.e., support increases forgone income), and the third inequality follows from the fact that \(\sigma = S^\theta((E[F_{\sigma\beta}], \beta)\) (because the penultimate line is the objective at \(\sigma^+\) given \(\beta\) less the objective at \(\sigma\) given \(\beta\)). This last inequality is strict because the definition of \(S\) breaks ties in favor of the highest of the maximizing schooling levels. Since the first line in the entire inequality is the objective at \(\sigma^+\) given \(\beta\) less the objective at \(\sigma\) given \(\beta\), it follows that no \(\sigma^+ > \sigma\) can equal \(S^\theta((E[F_{\sigma\beta}], \beta)\). Hence \(S^\theta((E[F_{\sigma\beta}], \beta) \leq \sigma = S^\theta((E[F_{\sigma\beta}], \beta)\). \(\blacksquare\)

**Proof of Theorem 3:** The theorem’s left-hand side less its right-hand side is nonpositive:

\[
\left( \sum_{a=\sigma + 1}^{59} \delta^{\sigma - 1} E[F_{\sigma\beta}] - \sum_{a=1}^{\sigma} \delta^{\sigma - 1} \theta \right)
\]
\[
- \left( \sum_{a=\sigma^+ + 1}^{59} \delta^{\sigma - 1} E[F_{\sigma^+\beta}] - \sum_{a=1}^{\sigma^+} \delta^{\sigma - 1} \theta \right)
\]
\[
- \sum_{a=\sigma + 1}^{59} \delta^{\sigma - 1} (E[F_{\sigma\beta}] - E[F_{\sigma^+\beta}])
\]
= \left( \sum_{a=\sigma+1}^{59} \delta^{a-1} E[F_{a,\mu}] - \sum_{a=1}^{\sigma} \delta^{a-1} \theta \right)

- \left( \sum_{a=\sigma^*+1}^{59} \delta^{a-1} E[F_{a^*,\mu}] - \sum_{a=1}^{\sigma^*} \delta^{a-1} \theta \right)

- \left( \sum_{a=\sigma+1}^{59} \delta^{a-1} E[F_{a,\mu}] - \sum_{a=1}^{\sigma} \delta^{a-1} \theta \right)

+ \left( \sum_{a=\sigma^*+1}^{59} \delta^{a-1} E[F_{a^*,\mu}] - \sum_{a=1}^{\sigma^*} \delta^{a-1} \theta \right)

\leq 0.

The two equalities are derived by algebra. The final inequality follows from $\sigma^* = S^\theta(E[F_a])$ because the penultimate line is the objective at $\sigma$ less the objective at $\sigma^*$. ■

References


