Notes for a Course in Development Economics

Debraj Ray

Version 3.3, 2009.

CHAPTER 1

Introduction

Open a book — any book —on the economics of developing countries, and it will begin with the usual litany of woes. Developing countries, notwithstanding the enormous strides they have made in the last few decades, display fundamental economic inadequacies in a wide range of indicators. Levels of physical capital per person are small. Nutrition levels are low. Other indicators of human capital such as education — both at the primary and seconday levels — are well below developed-country benchmarks. So are access to sanitation, safe water and housing. Population growth rates are high, and so are infant mortality rates. One could expand this list indefinitely.

Notice that some of these indicators — infant mortality or life expectancy, for instance — may be regarded as *defining* features of underdevelopment, so in this respect the list above may be viewed, not as a statement of correlations, but as a definition of what we mean by development (or the lack of it). But other indicators, such as low quantities of physical capital per capita, or population growth rates, are at least one step removed. These features don't *define* underdevelopment. For instance, it is unclear whether low fertility rates are *intrinsically* a feature of economic welfare or development. Surely, many families in rich countries may take great pleasure in having a large number of offspring. Likewise, large holdings of physical capital may well have an *instrumental* value to play in the development process, but surely the mere existence of such holdings does not constitute a defining characteristic of economic welfare.

And indeed, that is how it should be. We do not make a list of the features that go hand in hand with underdevelopment simply to define the term. We do so because — implictly or explicitly — we are looking for explanations. Why are underdeveloped countries underdeveloped?[1] It is easy enough to point to these inadequacies in terms of physical and human capital, but the extra step to branding these as *causes* of underdevelopment is perilously close, and we should avoid taking that step. Low levels of capital, or low levels of education, are just as much symptoms of development as causes, and to the extent that

[1]Perhaps the word "underdeveloped" does not constitute politically correct usuage, so that several publications — those by well-known international organizations chief among them — use the somewhat more hopeful and placatory present continuous "developing". I won't be using such niceties in this article, because it should be clear — or at least it is clear in my mind — that *economic* underdevelopment pins no derogatory social label on those who live in, or come from, such societies.

they intertwine with and accompany the development process (or the lack of it), we cannot rely on these observations as explanations.

That doesn't stop economists from offering such explanations, however. More than one influential study has regressed growth rates (alternatively, levels) of per-capita income on variables such as the rate of savings and population growth. There is very little doubt, in fact, that such variables are significantly associated with per-capita income. But nevertheless, we do have to think about the sense in which these studies serve as *explanations* for underdevelopment.

For instance, is it the case that individuals in different parts of the world have some *intrinsic* difference in their willingness — or ability — to save, or to procreate? If this were the case, we could hang our hat on the following sort of theory: such-and-such country is populated by people who habitually save very little. This is why they are underdeveloped.

Somehow, this does not seem right. We would like to have a theory which — while not belittling or downplaying the role of social, cultural and political factors — does not simply stop there. We would like to know, for instance, whether low incomes provoke, in turn, low savings rates so that we have a genuine chicken-and-egg problem. The same is true of demographics — underdevelopment might be a cause of high population growth rates, just as high population growth rates themselves retard the development process.

My goal in these notes is to talk about some of these chicken-and-egg situations, in which underdevelopment is seen not as a failure of some fundamental economic parameters, or socio-cultural values, but as an interacting "equilibrium" that hangs together, perhaps precipitated by inertia or by history. [Indeed, in what follows, I will make a conceptual distinction between equilibria created by inertia and those created by history.]

Why is this view of the development process an important one? There are three reasons why I feel this view should be examined very seriously.

[1] This point of view leads to a theory, or a set of theories, in which economic "convergence" (of incomes, wealth, levels of well-being) across countries is not to be automatically had. Actually, the intelligent layperson reading these words will find this reasoning a bit abstruse: why on earth would one expect convergence in the first place? And why, indeed, should I find a theory interesting on the grounds that it does *not* predict convergence, when I knew that all along? This is not a bad line of reasoning, but to appreciate why it is misguided, it is important to refer to a venerable tradition in economics that has convergence as its very core prediction. The idea is based — roughly — on the argument that countries which are poor will have higher marginal products of capital, and consequently a higher rate of return to capital. This means that a dollar of extra savings will have a higher payoff in poor countries, allowing it grow faster. The prediction: pooere countries will tend to grow faster, so that over time rich and poor countries will come together, or "converge".

This is not the place to examine the convergence hypothesis in detail, as my intention is to cover other views of development.[2] But one should notice that convergence theories in this raw form have rarely been found acceptable (though rarely does not mean never,

[2]See Ray [1998], Chapters 2 and 3.

among some economists), and there are several subtle variants of the theory. Some of these variants still preserve the idea that lots of "other things" being equal, convergence in some conditional sense is still to be had. It's only if we start accepting the possibility that — perhaps — these "other things" *cannot* be kept equal, that the notion of conditional convergence starts losing its relevance and very different views of development, not at all based on the idea of convergence, must be sought.

[2] The second reason why I find these theories important is that they do not reply on "fundamental" differences across peoples or cultures. Thus we may worry about whether Confucianism is better than the Protestant ethic in promoting hard-headed, succesful economic agents, and we might certainly decry Hindu fatalism as deeply inimical to purposeful, economic self-advancement, but we have seen again and again that when it comes down to the economic crunch and circumstances are right, both Confucian and Hindu will make the best of available opportunities — and so will the Catholics and a host of other relgions and cultures besides. Once again, this is not the place to examine in detail fundamentalist explanations based on cultural or religious differences, but I simply don't find them very convincing. This is not to say that culture — like conditional convergence — does not play a role. [In fact, I provide such examples below.] But I also take the view that culture, along with several other economic, social and political institutions, are all part of some broader interactive theory in which "first cause" is to be found — if at all — in historical accident.

[3] The last reason why I wish to focus on these theories is that create a very different role for government policy. Specifically, I will argue that these theories place a much greater weight on one-time, or temporary, interventions than theories that are based on fundamentals. For instance, if it is truly Hindu fatalism that keeps Indian savings rates low, then a policy of encouraging savings (say, through tax breaks) will certainly have an effect on growth rates. But there is no telling when that policy can be taken away, or indeed, if it can be taken away at all. For in the absence of the policy, the theory would tell us that savings would revert to the old Hindu level. In contrast, a theory that is based on an interactive chicken-and-egg approach would promote a policy that attempts to push the chicken-egg cycle into a new equilibrium. Once that happens, *the policy can be removed*. This is not to say that once-and-for-all policies are the correct ones, but only to appreciate that the interactive theories I am going to talk about have very different implications from the traditional ones.

CHAPTER 2

The Calibration Game

The simple model of convergence also has to be put through enormous contortions to fit the most essential development facts regarding per-capita income across countries. This is the point of the current section.

2.1 Some Basic Facts

Low per capita incomes are an important feature of economic underdevelopment—perhaps *the* most important feature—and there is little doubt that the distribution of income across the world's nations is extraordinarily skewed.

The *World Development Report* (see, e.g., World Bank [2003]) contains estimates for all countries, converted to a common currency. By this yardstick, the world produced approximately $32 trillion of output in 2001. A little less than $6 trillion of this — less than 20% — came from low- and middle-income developing countries (around 85% of the world's population). Switzerland, one of the world's richest countries, enjoyed a per capita income close to 400 times that of Ethiopia, one of the world's poorest.

A serious discrepancy arises from the fact that *prices* for many goods in all countries are not appropriately reflected in exchange rates. This is only natural for goods and services that are not internationally traded. The International Comparison Program publishes PPP estimates of income, and under these the differences are still huge, but no longer of the order of 500:1.

Over the period 1960–2000, the richest 5% of the world's nations averaged a per capita income (PPP) that was about twenty-nine times the corresponding figure for the poorest 5%. As Parente and Prescott [2000] quite correctly observed, interstate disparities *within* the United States do not even come close to these international figures. In 2000, the richest state in the United States was Connecticut and the poorest was Mississippi, and the ratio of per capita incomes worked out to slightly less than 2!

Of course, the fact that the richest 5% of countries bear approximately the same ratio of incomes (relative to the poorest 5%) does not suggest that the *entire* world distribution of incomes has remained stationary. Of greatest interest — a recent financial crisis notwithstanding — is the meteoric rise of the East Asian economies: Japan, Korea, Taiwan, Singapore, Hong Kong, Thailand, Malaysia, Indonesia, and (commencing somewhat later) China. Over the period 1965–90, the per capita incomes of the aforementioned eight East Asian economies (excluding China) increased at an annual rate of 5.5%. Over 1990–1999, the pace slowed somewhat, especially in Japan, but averaged well over 3% per year for the remainder.[1]

Impressive as these rates are, they are dwarfed by China's phenomenal performance. Between 1980 and 1990, China's per capita income grew at an annual rate of 8.6%. The corresponding figure for the 1990s is even higher: around 9.6%.

In contrast, much of Latin America languished during the 1980s. After relatively high rates of economic expansion in the two preceding decades, growth slowed to a crawl, and in many cases there was no growth at all. Morley's [1995] study observed that in Latin America, per capita income *fell* by 11% during the 1980s, and only Chile and Colombia had a significantly higher per capita income in 1990 than they did in 1980. It is certainly true that such figures should be treated cautiously, given the extreme problems of accurate GNP measurement in high-inflation countries, but they illustrate the situation well enough. With some notable exceptions (such as Chile, 5.7%, and Argentina, 3.6%), annual per-capita growth in incomes continues to be extremely slow for Latin America in the 1990s, though these rates did turn positive through most of the region.

Similarly, much of Africa stagnated or declined over the 1980s. Countries such as Nigeria and Tanzania experienced substantial declines of per capita income, whereas countries such as Kenya and Uganda barely grew in per capita terms. Notable turnarounds in the 1990s have occurred in both directions, with alarming declines in countries such as the Congo, Rwanda and Burundi, and substantial progress in Uganda.

Looking at the overall picture once more without naming countries, one can get a good sense of the world income distribution by looking at *mobility matrices*, an idea first applied to countries by Danny Quah. I've constructed one such matrix using 132 countries over the period 1980–2000; see Figure 2.1.

Each row and column in this matrix is per-capita income relative to world per-capita income. The rows represent these ratios in 1980; the columns the corresponding ratios in 2000. The cell entries represent percentages of countries in each row-column combination, the rows adding up to 100 each. So, for instance, 88% of the countries that earned less than than a quarter of world per-capita income in 1980 continued to do just that in 2000.

Clearly, while there is no evidence that very poor countries are doomed to eternal poverty, there is some indication that both very low and very high incomes are extremely sticky. Middle-income countries have far greater mobility than either the poorest or the richest countries. For instance, countries in category 1 (between half the world average and the world average) in 1980 moved away to "right" and "left": less than half of them remained where they were in 1980. In stark contrast to this, fully 88% of the poorest countries (category 1/4) in 1980 remained where they were, and none of them went above the world average by 2000. Likewise, another 88% of the richest countries in 1980 stayed right where

[1]To appreciate how high these rates of growth really are, note that for the entire data set of 102 countries studied by Parente and Prescott, per capita growth averaged 1.9% per year over the period 1960–85.

	Quarter or less	Quarter to Half	Half to Average	Average to Twice	Twice or more
Quarter or less	88	8	4	0	0
Quarter to Half	48	43	10	0	0
Half to Average	0	28	56	14	3
Average to Twice	0	0	32	42	26
Twice or more	0	3	3	6	88

The Income Mobility of Countries, 1980--2000.
132 Countries

FIGURE 2.1. THE INCOME MOBILITY OF COUNTRIES, 1980–2000.

they were. This is interesting because it suggests that although everything is possible (in principle), a history of underdevelopment or extreme poverty puts countries at a tremendous disadvantage.

There is actually a bit more to Figure 2.1 than lack of mobility at the extremes. Look at the next-to-poorest category (those with incomes between one-quarter and one-half of the world average in 1962). Almost half of them *dropped* to an even lower category. Thus it is not only the lowest-income countries that might be caught in a very difficult situation. In general, at low levels of income, the overall tendency seems to be movement in the downward direction.

My book contains a corresponding mobility matrix for 1963–1984, with very similar findings.

To summarize, then, we have the following observations.

(1) Over the period 1960–2000, the *relative* distribution of world income appears to have been quite stable. The richest 5% of the world's nations averaged a level of per capita income that was about 29 times the corresponding figure for the poorest 5%. By any standards, this disparity is staggering, and especially so when we remember that we are talking about incomes that have been corrected for purchasing power parity.

(2) The fact that the overall distribution has remained stationary does *not* mean that there has been little movement of countries within the world distribution. Of particular interest in the 1980s is the rise of the East Asian economies and the languishing of other economies, particularly those of sub-Saharan Africa and Latin America. Diverse growth experiences such as these can change the economic composition of the world in the space of a few decades. Nonetheless, a single explanation for this diversity remains elusive.

(3) The observation that several countries have changed relative positions suggests that there are no ultimate traps to development. At the same time, a history of wealth or poverty does seem to partly foretell future developments. The mobility of countries appears to be highest somewhere in the middle of the wealth distribution, whereas a history of underdevelopment or extreme poverty appears to put countries at a disadvantage.

(4) That history matters in this way is an observation that requires a careful explanation. Poor countries do seem to have some advantages. They can use, relatively free of charge, technologies that are developed by their richer counterparts. Scarce capital in these countries should display a higher rate of profit, because of the law of diminishing returns. They can learn from mistakes that their predecessors have made. In this way differences across countries should iron themselves out over the longer run. Thus the observation that history matters in maintaining persistent differences needs more of a justification than might be obvious at first glance.

One can see different attempts to reconcile the failure of convergence with the traditional theory:

(A) *Hardline View.* Don't abandon the traditional aggregative theory but seek reasons for productivity and other controls to be systematically different across countries. Conditional on those controls, attempt to establish convergence.

(B) *Multiplicity View.* Abandon the convergence argument. Argue that the same fundamentals can progress in very different directions depending on initial conditions.

(C) *Interactive View.* Argue that the world is one interactive system and cannot be split up into several growth models (with or without convergence) running side by side.

In these notes we shall spend some time with each of these views.

2.2 The Parente-Prescott and Lucas Calibrations

Normalize labor to 1, and consider the following aggregative production function:

$$y_t = A_t k_t^{\theta},$$

where A_t captures some exogenous growth of TFP. It will be convenient to bestow on A_t the exponential form

$$A_t = A(1+\gamma)^{(1-\theta)t},$$

so that $(1+\gamma)^{(1-\theta)t} - 1$ can be thought of as the growth rate of TFP. We have written things in the slightly ugly form so as to simplify the expressions later.

One way to think about it is to interpret γ as the growth rate of *labor* productivity. That would translate into output productivity by a power of $1-\theta$.

Write the capital accumulation equation:

$$k_{t+1} = (1-\delta)k_t + x_t,$$

where δ is the depreciation rate and x is the flow of fresh investment. By the savings–investment equality and the Solow assumption that savings is proportional to per-capita income, we see that

$$x_t = s y_t,$$

where s is the savings rate. We can combine all these equations and proceed as follows:

Define $k_t^* \equiv k_t/(1+\gamma)^t$; then the capital-accumulation equation implies that

$$(1+\gamma)^{t+1}k_{t+1}^* = (1-\delta)(1+\gamma)^t k_t^* + \frac{x_t}{k_t^*}k_t^*,$$

and dividing through by $(1+\gamma)^t$, we may conclude that

$$(1+\gamma)k_{t+1}^* = \left[(1-\delta) + s\frac{y_t}{k_t}\right]k_t^*.$$

Now y_t/k_t simply equals $A/k_t^{*1-\theta}$, so that k_t^* simply converges to k^*, where k^* is given by

$$(1-\delta) + s\frac{A}{k^{*1-\theta}} = 1+\gamma,$$

or

$$k^* = \left(\frac{sA}{\gamma+\delta}\right)^{1/(1-\theta)}.$$

It follows that k_t converges to the path

$$(1+\gamma)^t\left(\frac{sA}{\gamma+\delta}\right)^{1/(1-\theta)},$$

so that y_t converges to the path

$$A(1+\gamma)^t\left(\frac{sA}{\gamma+\delta}\right)^{\theta/(1-\theta)}.$$

We may therefore conclude that in steady state, the effect of varying the savings rate only depends on the ratio $\theta/(1-\theta)$.

But this has small effects. We know that in a Cobb-Douglas world of perfect competition, θ is a good proxy for the share of capital in national income. Lucas (1990) uses an estimate for that share of around 0.4, so that the ratio in question is around 2/3. This means that a doubling of the savings rate — a huge increase — will only raise steady state per-capita incomes by a factor of $2^{2/3}$, which is around a 60% increase. This comes nowhere close to the inequalities we see around us.

Parente and Prescott (2000) impute around 70% to labor income and 5% to land, which leaves them with a capital share of 25%. With that our required ratio is even lower: $\theta/(1-\theta) = 1/3$. That means that a doubling of the savings rate only translates into a 25% variation in per-capita income.

And moreover, the savings rates in the richest countries are nowhere close to double that of their poor counterparts. In 1993, the industrialized countries averaged a savings rate of 19.4%. The LDCs actually had a higher savings rate during that period — 23.3% — while even Africa had a savings rate of 18.8%.

Notice that TFP differentials in productivity give us a better chance to explain differentials: whereas across two countries 1 and 2,

$$\frac{y_1}{y_2} = \left(\frac{s_1}{s_2}\right)^{\theta/(1-\theta)},$$

for technical levels the difference is more amplified:

$$\frac{y_1}{y_2} = \left(\frac{A_1}{A_2}\right)^{1/(1-\theta)}.$$

When θ is 1/3, the savings difference translate into income differences as the square root, while for technology differences the ratio is taken to the power 1.5. So, for instance, a doubling of the technology difference "explains" a difference of close to three times in per-capita output. This is a bit closer to what we see, and it is small wonder, then, that those who are wedded to Solow-type convergence models have been inclined to focus on technological differences.

Actually, the above calibrations, which are due to Parente and Prescott (2000) (see also Mankiw, Romer and Weil (1992)) can be given a sharper and more immediate expresson by simply using the production function and no more. This is the route taken in Lucas (1990). Once again begin with the Cobb-Douglas production function

$$y = Ak^{\theta}$$

(note, no time subscripts, I won't even need to do any growth theory). Use the competitive condition to assert that the rate of return to capital is given by

$$r = A\theta k^{\theta-1},$$

or equivalently

$$r = \theta A^{1/\theta} y^{(\theta-1)/\theta}.$$

Once again take the share of capital to be 1/3; then $\theta = 1/3$, so that across two countries 1 and 2,

$$\frac{r_1}{r_2} = \left(\frac{y_2}{y_1}\right)^2.$$

This yields absurd numbers. If the per-capita income in the US is 15 times larger than that of India, the rate of return on capital in India should be over 200 times higher! Even if the share of capital in production is taken to be 0.4 (used by Lucas), the ratio in the rates of return should be close to 60, also plainly absurd.

There are alternative routes out of this dilemma.

2.2.1 Differences in Human Capital. Of course, we should get the immediate alternative out of the way first, which is that labor qualities in the two countries are not the same. So differences in per-capita income are not the same as differences in income per "effective capita". For instance, Anne Krueger (1968) attempts to compare US and Indian workers by looking at information on each country's mix of workers (by age, education and sector) and combining this with (US-based) estimates of how these factors affect productivity (as measured by relative earnings).

Krueger obtains an overall ratio of one US worker = approx. 5 Indian workers. This means that the ratio income per effective capita is 3, but this too generates a rate of return differential between 5 (if capital's share is 40%) and 9 (if that share is set lower at 1/3). This difference is also "too large", and there is still a lot left to explain.

[Update on Krueger and per-capita income differentials: Heston, Summers and Aten (2002) argue that in 1990 the PPP income differences were perhaps 11:1. Banerjee and Duflo (2004) adjust the Krueger estimates of relative worker productivity to about 10:3. This leaves us effectively in the same place: the adjusted ratio is then about 3.2, which creates the same differentials in the rates of return as in the previous paragraph.]

2.2.2 Differences in TFP. Now let's turn our attention to technological differences. Let us look at the implicit TFP ratios needed if we were to equalize rates of return in the two countries *and* maintain the requirement that per-(effective) capita income ratios are around 3. Use subscripts I for India and U for the US; then the equality of the two rates of return to capital demands that

$$A_I y_I^{\theta-1} = A_U y_U^{\theta-1},$$

so that

$$\frac{y_U}{y_I} = \left(\frac{A_U}{A_I}\right)^{1/(1-\theta)} \simeq \left(\frac{A_U}{A_I}\right)^{1.5},$$

provided that the share of capital is around a third. So this means that

$$\frac{A_U}{A_I} \simeq 3^{2/3} = 2.08.$$

It is hard to get a feel for whether this is a large difference or a small one. One way of looking at it: if the US and India put in the same amounts of capital and quality-corrected labor into production, the US will produce twice as much as India. This may be a tall order.

Nother way: Lucas's view is that this difference is attributable to an *externality* created by human capital. Suppose that the externality is proportional to h^a, where a is some coefficient and h is the human capital endowment per capita. Then

$$\frac{A_U}{A_I} = \left(\frac{h_U}{h_I}\right)^a.$$

Lucas estimates a at around 0.36, using Denision's productivity comparisons within the United States over 1909 and 1958, and combining them with human capital endowments over the same period. Because $5^{0.36} \simeq 1.8$, this takes care of the problem as far as Lucas is concerned.

2.2.3 Misallocation of Capital. Another way to think about it is to generate the productivity differences from the misallocation of capital in a disaggregated model. Banerjee and Duflo (2004) adopt this approach, but there is an interesting tension here. To generate serious misallocation problems, one must presume that the marginal product of capital is substantially different across small and large firms. But this means that capital has high curvature in production, so that one must choose correspondingly smaller values of θ. Assuming that capital is misallocated cannot provide a ready fix on this problem.

That said, credit constraints and consequent misallocation of resources may well be important.

2.2.4 The Share of Capital. One way out is to somehow enlarge the share of capital, and in this way the value of θ. Parente and Prescott (2000, p. 44–55) discuss this route in some detail, by considering intangible forms of capital and the possibility that physical capital is grossly mismeasured, but these adjustments are just not enough.

2.2.5 Government Failure. One view is that governments might expropriate new investors, while existing investors (who may be unproductive) are overprotected. This is a view in which incumbent elites are not necessarily the best business hands, yet they are in a position to control the entrance of others more efficient than they are. This is related to political-economy arguments made by Engerman and Sokoloff and Acemoglu-Johnson-Robinson that we will discuss later in the course.

Parente and Prescott consider a variant of this point of view, in which they regard the government as intervening excessively and thus lowering productivity.

Another sort of government failure may arise from the *lack* of intervention, such as intervention to protect property rights. Certain types of long-run investment may then not be made (see Besley, Bandiera, or Goldstein-Udry). Or there may be various free-rider problems in joint production, as also overexploitation of the commons.

2.3 Summing Up

Convergence relies on diminishing returns to "capital". If this is our assumed starting point, the share of capital in national income does give us rough estimates of the concavity of production in capital. The problem is that the resulting concavity understates observed variation in cross-country income by orders of magnitude. Huge variations in the savings rate do not change world income by much. For instance, doubling the savings rate leads to a change in steady state income by a factor of 1.25, which is inadequate to explain an observed range of around 20:1 (PPP). Indeed, as Lucas (1990) observes, the discrepancy actually appears in a more primitive way, at the level of the production function (even without the attendant steady state theory). For the same simple production function to fit the data on per-capita income differences, a poor country would have to have enormously higher rates of return to capital; say, 60 times higher if it is one-fifteenth as rich. This is implausible. And so begins the hunt for other factors that might explain the difference. What did we not control for, but should have?

This is the kind of mindset that you will take on board if you get on the convergence boat. The Solow benchmark of convergence must be tested against the empirical evidence on world income distributions, savings rates, or rates of return to capital. The two will usually fail to agree. Then we look for the missing variables that will bridge Solow (or some close variant thereof) to the data. Thus it is not uncommon to find economists "explaining" inter-country variation by stating that one country is more corrupt than another, or more democratic, or is imbued with some particularly hardworking cultural ethic.

With careful economists such as the ones I have cited here, the argument is conducted far more responsibly. "Human capital" is often used as a first port of call: might differences here account for observed cross-country variation? The rest is usually attributed to that familiar black box: "technological differences". As one might imagine, that slot can be filled in a variety of ways: externalities arising from human capital, incomplete diffusion of technology, excessive government intervention, within-country misallocation of resources, take your pick. All of these — and more — are interesting candidates, but by now we have wandered far from the original convergence model, and if at all that model still continues to illuminate, it is by way of occasional return to the recalibration exercise, after choosing plausible specifications for each of these potential explanations.

The Solow model and its immediate variants don't do a bad job. In the right hands, the model serves as a quick and ready fix on the world, and it organizes a search for possible explanations. Taken with the right grain of salt, and viewed as a first pass, such an exercise can be immensely useful. At another level, playing this game too seriously reveals a particular world-view. It suggests a fundamental belief that the world economy is ultimately a great leveller, and that if the levelling is not taking place we must search for that explanation in parameters that are somehow structurally rooted in a society. These parameters *cause* economic growth, or the lack of it.

To be sure, the factors identified in these calibration exercises do go hand in hand with underdevelopment. So do bad nutrition, high mortality rates, or lack of access to sanitation, safe water and housing. Yet there is no ultimate causal chain: many of these features go hand in hand with low income in self-reinforcing interplay. By the same token, corruption, culture, procreation and politics are all up for serious cross-examination: just because "cultural factors" (for instance) seems more weighty an "explanation" does not permit us to assign it the status of a truly exogenous variable.

In other words, the convergence predicted by technologically diminishing returns to inputs should not blind us to the possibility of nonconvergent behavior when all variables are treated as they should be — as variables that potentially make for underdevelopment, but also as variables that are profoundly affected by the development process.

This leads to a different way of asking the development question, one that is not grounded in any presumption of convergence. Quite unlike the convergence hypothesis, the starting presumption is distinct: two economies with the same fundamentals can move apart along very different paths. Several factors might lead to such divergence, among them various processes of cumulative causation, or poverty-traps, or initial histories that determined — at least to some degree — the future that followed.

CHAPTER 3

Expectations and Multiple Equilibrium

3.1 Complementarities

Let n be the number of players and $A_1, \ldots, A_n$ be n action sets, one for each player. Suppose that the sets are ordered by "$\geq$". For each player i there is a payoff function $\pi_i : A \to \mathbb{R}$, where A is the product of the actions sets. Say that this game exhibits *complementarities* if whenever $a_{-i} \geq a'_{-i}$, then

$$\arg\max_{a_i} \pi(a_i, a_{-i}) \geq \arg\max_{a_i} \pi(a_i, a'_{-i}).$$

It will suffice for the purpose of these notes to provide a simplified and more special description. Suppose that a set of individuals all have access to some set of actions A, taken to be a subset of the real line. Denote by a a generic action, a_i the action taken by individual i, and by m_i the average of all actions other than the one taken by i.

Assume that the payoff function is given by $\pi_i(a, m)$ for each individual i, where a denotes his action and m denotes the average action taken by everybody else. Then it is easy to see that there are complementarities in our more general sense if for all i,

$$\pi_i(a, m) - \pi_i(a', m) \text{ is increasing in } m \tag{3.1}$$

whenever $a > a'$ are two actions in the set A.

Notice the difference between complementarities and positive externalities. The former change the *marginal* gain to taking an action while the latter affects payoff *levels*. Changes in the marginal gain are compatible with payoff levels going in either direction.

As we shall see, Pareto-ordered outcomes are typical of these situations (though they won't *necessarily* happen).

3.2 Some Examples

3.2.1 Qwerty. There are two technologies; call them [Q]werty and [D]vorak. There are many individuals, each of whom employs a single Q-trained secretary or a single D-trained secretary. The cost of installing each technology is the same, but the cost expended on a secretary is a decreasing function of the number of other people using the same secretary type. [More secretarial schools exist for that type.] This is a situation of complementarities.

The same goes for technologies such as PCs and Macs, in which the benefits from adopting the technology depend positively on the number of other users (networking).

3.2.2 Infrastructure.
A railroad is used for transporting products from the interior to the ports. People are indexed on $[0, 1]$, and person i gets a benefit $B(i)$ from being able to use the railroad. The cost of railroad use is declining in the number of users: $c(n)$, where n is the number of users and $c'(n) < 0$. This is a situation of complementarities.

3.2.3 Finance.
A thicker financial market caused by lots of people putting their money in financial assets can create the possibilities of greater diversification. So at the margin, it becomes easier for an individual investor to invest.

3.2.4 Capital Deepening.
Greater roundaboutness in production increases the productivity of capital, the scale of aggregate production, and in this way the final demand for individual machine varieties. This may in turn justify the greater roundaboutness of production.

3.2.5 Social Capital.
High rural-urban migration can destroy social capital back in rural areas. In turn, that destruction can increase the pace of rural-urban migration.

3.2.6 Discrimination.
Individuals discriminated against may not invest in human capital, perpetuating that discrimination.

3.2.7 Currency Crises.
Apart from the fundamentals of holding or selling a currency, there is a strong incentive to sell if other individuals are selling. This forms the basis of a class of currency-crisis theories based on complementarities.

3.2.8 Endogenous Growth.
Economy-wide investment raises the return to individual investment, thus potentially generating a sustainable growth path.

3.2.9 Social Norms.
Sometimes, social norms can change a Prisoner's Dilemma to a coordination game. Examples: spitting in public, throwing garbage on the streets, or engaging in tariff wars. Sometimes repeated interactions can imitate the same outcome (though with many agents this is hard).

3.3 Complementarities and Development

The ordinary view of capitalist development is that it inflicts negative externalities: pollution, greed and so on. This is certainly true. But there is an important sense in which the capitalist investment process creates severe complementarities (whether the underlying externalities are negative or positive; they could be either).

For instance, a firm that prides itself on quality and fair dealing will induce its competitors to take the same actions simply to maintain business competitiveness, and could spark off a quality race (the same applies to research and innovation, or indeed, low prices). Note that the underlying externalities are *negative* but that we have a case of complementarities in the appropriate action space.

In another context, the combined actions of several firms can (a) lower infrastructural costs, (b) create demand for each others' products, both directly and (c) by creating higher incomes; and can (d) enable the creation of new products or the startup of some other productive activity by making inputs available. These are complementarities, too, in the sense that these actions of "investment" increase the incentive for other firms to 'invest" as well. This time the externalities are positive.

So complementarities can exist both with positive and with negative externalities. Indeed, the medieval guild system was designed to avoid some of the complementaries which had negative externalities attached to them — e.g., a new improvement had to be sanctioned by existing guild members. As you can imagine, such an outcome cannot be stable unless severe punishments were available for offenders, and such punishments themselves became weaker as the guild system died away, provoking others to leave the guilds. [Thus the guild story is also one of complementarities!]

It is very important to understand that while the distinction between positive and negative externalities is important in understanding the *normative* properties of a particular equilibrium, the distinction between complementarities and what might be called *anti-complementarities* (reverse the movement of (3.1) in others' actions) is essential in understanding the possible variety in the development experience. Complementarities create the possibility of multiple equilibria, so that we might argue that countries — or societies — are in different equilibria though there is nothing intrinsically different between them.

3.4 Multiple Equilibrium

3.4.1 Complementarities through Demand. How do complementarities manifest themselves in multiple equilibria? One way to begin exploring this is through the standard general equilibrium model. Understand, first, that some amount of limited competition is necessary. If the price at which an output is being sold (or an input being bought) is fixed, then the agent can inflict no externality on another — he internalizes these externalities through the price. But if the price drops as you sell more of a product (or the input price raises), you are creating a gain for another agent that you fail to internalize. Thus perfect competition may be at odds with the multiple equilibrium story that creates a set of Pareto-dominating equilibria (of course, multiple equilibria that are Pareto-undominated are clearly possible by classical considerations).

The route envisaged by Rosenstein-Rodan was two-fold. Investments everywhere create a climate for more investment: (a) directly, via intersectoral changes in price, and (b) indirectly via the generation of incomes. The parable of the shoe-factory emphasized (b). This is the Rosenstein-Rodan story as extended by Murphy, Shleifer and Vishny [1989].

3.4.2 Pecuniary Externalities. Before I get into the Murphy-Shleifer-Vishny model it is important to observe that the Rosenstein-Rodan view is one of *pecuniary* externalities, in which one sort of change (investment) provokes another (investment somewhere else) via a change in prices. Such externalities are to be contrasted with what one might call *technological* externalities, in which there is a direct effect that has nothing to do with prices. (Using a PC when lots of other people use a PC is a networking effect that is a technological externality. Of course, one could also tell a pecuniary story in which a big market for PCs drives down their prices and therefore makes it more attractive to buy one.)

Pecuniary externalities are natural. Tibor Scitovsky (1954), in a famous article, argued that they are orders of magnitude more compelling than technological externalities. Indeed, because of Rosenstein-Rodan and Hirschman, pecuniary externalities are viewed as the fundamental process underlying development (or the lack of it).

But percuniary externalities are hard to model using competitive markets. Basically the first fundamental theorem of welfare economics stands in the way. That theorem rules out Pareto-ranked equilibria whenever there are no technological externalities. Let us very quickly review the theorem in the context of a CRS production model.

Suppose that each person i has a utility function u_i and a vector of endowments $x(i)$. The total endowment vector is $x \equiv \sum_i x(i)$. Denote by c the vector of final consumption goods: it is allocated among the population of all individuals with person i getting $c(i)$.

There is a production technology T to convert x into c. Thus T is a set of feasible (c, x) output-input pairs. We don't make any particular assumption on T.

Note that production may generate positive profits so we have to distribute them among the agents. Let $\theta(i)$ be the share of aggregate profits π that accrue to agent i.

Finally, there is a price vector p for the final goods and w for the endowments. Say that $(p^*, w^*, c*(i))$ is a *competitive equilibrium* if, defining $c^* \equiv \sum_i c^*(i)$, we have profit maximization:

$$\pi \equiv p^*c^* - w^*x \geq p^*c' - w^*x' \text{ for all } (c', x') \in T,$$

and utility maximization subject to budget constraints:

$$c^*(i) \text{ maximizes } u(c(i)) \text{ on } \{c(i) | p^*c(i) \leq w^*x(i) + \theta(i)\pi\}.$$

Proposition 3.1. *A competitive equilibrium is Pareto optimal, so in particular there can be no Pareto-ranked equilibria.*

Proof. Suppose not. Then there exists an alternative vector of outputs c (feasible, so that $(c, x) \in T$) and an allocation $c(i)$ of it such that $u(c(j)) \geq u(c^*(j))$ for all j, with strict inequality for some j. But then by utility maximization, we must have

$$p^*c(j) \geq p^*c^*(j) \text{ for all } j, \text{ with strict inequality for some } j,$$

and adding over all j, we must conclude that

$$p^*c > p^*c^*.$$

Substracting the common term w^*x from both sides of this inequality, we contradict the profit maximization property. ∎

This is why we employ some imperfect competition in the models of pecuniary externalities that follow.

3.4.3 Model 1. The Profit Externality.

There is a continuum of sectors indexed by $q \in [0, 1]$. The utility function for a consumer is

$$\int_0^1 \ln x(q)dq.$$

With this utility function, when consumer income is given by y, an amount of y is spent on every good q.

Now, normalize wage rate to unity; then $y = \pi + L$, where π is profits and L is the labor endowment of a typical agent.

Now suppose that each sector has two technologies, a *cottage technology* which is freely available without any setup cost, and an *industrialized technology*, which requires a setup cost. In the former, assume that one unit of labor produces one unit of output. In the latter, assume one unit of labor produces α units of output, where $\alpha > 1$. But there is a setup cost, which we denote by $F > 0$.

Now the cottage sector is competitive (while demand for each good is unitary elastic), so it follows that if there is industrialization in some sector the price will be set at the limit price, equal to one. Thus the profit from industrialization is given by

$$y - \frac{y}{\alpha} - F = \frac{\alpha - 1}{\alpha} y - F \equiv ay - F \tag{3.2}$$

The point, therefore, is that a larger y is more conducive to industrialization. To complete the circle, more industrialization is also conducive to a larger value of y. To see this, notice that if a fraction n of the sectors do industrialize, then profits per firm are

$$\pi(n) = ay - F$$

so that aggregate income $y(n)$ is given by

$$y(n) = n\pi(n) + L = n[ay(n) - F] + L,$$

or equivalently

$$y(n) = \frac{L - nF}{1 - an}. \tag{3.3}$$

Notice that

$$y'(n) = \frac{(aL - F)/(1 - an)}{1 - an}$$

and that

$$\pi(n) = ay(n) - F = \frac{aL - F}{1 - an} \tag{3.4}$$

so that combining these two equations we get

$$y'(n) = \frac{\pi(n)}{1 - an} \tag{3.5}$$

Note that (3.5) exhibits a multiplier-like quality because of the externality. Extra profits create more than their own weight in income, while higher income in turn can spur industrialization. This is a classic case of complementarities.

But it isn't so classic in one respect. Oddly enough, despite the complementarity, there can only be a single equilibrium in this model. If $F < aL$, then this means that at $y(0) = L$ it is worth industrializing, so *everyone* will. On the other hand, if $F > aL$, then from equation (3.4) it is not worth industrializing *even when* income is all the way at $y(1)$ (set $n = 1$ in (3.4) and evaluate $\pi(1)$).

[To be sure, when $F = aL$ there are, in fact, a continuum of equilibria but they are all equivalent in that they generate the same level of national income.]

One possible "explanation" for uniqueness is that the model is too simple: there isn't enough heterogeneity among the firms. Let's satisfy ourselves that this has nothing to do with it. Suppose that the fixed costs vary smoothly across sectors, all the way from zero to infinity. Order the sectors so that $F(0) = 0$, $F(1) = \infty$, and $F(i)$ is smoothly increasing. Then if n sectors invest, it must be the interval of firms $[0, n]$ that's doing the investing, and the following zero-profit condition must hold:

$$ay(n) - F(n) = 0, \tag{3.6}$$

where

$$y(n) = \int_0^n \pi(i)di + L, \tag{3.7}$$

and

$$\pi(i) = ay(n) - F(i) \tag{3.8}$$

for each i. Let $A(n)$ denote the average value of all the fixed costs on $[0, n]$; then combining (3.7) and (3.8) we see that

$$y(n) = any(n) - nA(n) + L$$

or

$$y(n) = \frac{L - nA(n)}{1 - an},$$

and using this information in (3.6), and moving terms around, we see that

$$[1 - an]F(n) + anA(n) = aL \tag{3.9}$$

is the fully reduced-form zero-profit condition. Can there be more than one solution in n to this condition? Differentiating the LHS of (3.9) shows that the derivative is $[1 - an]F'(n)$, which is positive. There cannot be more than one solution.

Why does the complementarity have no effect? It does not because the externality is generated by the payoff of the firm. At all points for which the marginal payoff is positive (and so is the externality), the firm pushes ahead and produces more. It does not internalize the externality but it does not need to — the privately profitable and the socially profitable moves *coincide*. Likewise for the case in which profits are negative. There is a cutback which enhances both private and social outcomes at the same time. So even though there is an externality, the actions fully internalize the externality, as it were.

One lesson from all this is that the Rosenstein-Rodan intuition needs to be examined more carefully. The source of the complementarity must be something other than private profit alone.

The main idea in what follows — and this is perhaps the lesson of the Murphy-Shleifer-Vishny exercise — is that (to obtain multiplicity) one needs an externality source which isn't effectively internalized by the externality provider. For instance, if there are taxes on output (with proceeds given, say, lump-sum to the economy), then the profit-maximum for a firm will not correspond to the point at which the marginal externality washes out. Likewise for the example that we discuss below.

3.4.4 Model 2. The Wage Externality. Suppose that industrialization not only benefits the industrializing firm, it also benefits workers in that firm in the form of higher wages. These higher wages need not mean higher utility: they could be compensation for a higher disutility of labor. Or they may be Shapiro-Stiglitz type wages designed to prevent shirking in a hard-to-monitor activity. Or there may be political and economic pressures to keep wages above some stipulated minimum in the organized industrial sector.

Let the (additive) wage premium be v, so that the wage to be paid is $1 + v$. Now the monopolists's profit from industrializing in any particular sector is

$$\pi = y - \frac{1+v}{\alpha}y - F(1+v) \tag{3.10}$$

when the total demand for that sector is given by y.

This specification can generate multiple equilibria. If there is no industrialization, then $y = L$. This is self-justifying if profits, evaluated at this level of income, are nonpositive. That is, using (3.10), we have the condition

$$L\left(1 - \frac{1+v}{\alpha}\right) - F(1+v) \leq 0. \tag{3.11}$$

On the other hand, to see if one can have an equilibrium in which all industrialize, one must have the condition that profits are nonnegative when aggregate income (and hence demand) is evaluated at the full-industrialization point. Aggregate income is given by

$$y(1) = \Pi + L(1+v) = \left(1 - \frac{1+v}{\alpha}\right)y(1) - F(1+v) + L(1+v),$$

or

$$y(1) = \alpha(L - F). \tag{3.12}$$

[This is an intuitive expression, since the RHS corresponds to the total output produced.] So profits are nonnegative at the full industrialization point if

$$\left(1 - \frac{1+v}{\alpha}\right)\alpha(L - F) - F(1+v) \leq 0,$$

and rewriting this, we obtain the condition

$$L\left(1 - \frac{1+v}{\alpha}\right) - F \geq 0. \tag{3.13}$$

Now (3.11) and (3.13) can easily be compared to see why multiplicity is possible.

A similar exercise can be carried out using the heterogeneous-cost variant. Make the same assumptions on the function $F(i)$ as before. Then the equilibrium condition is given by

$$y(n)\left[1 - \frac{1+v}{\alpha}\right] = F(n)(1+v). \tag{3.14}$$

Notice that

$$y(n) = \int_0^n \pi(i)di + L + v\int_0^n F(i)di + \frac{vny(n)}{\alpha},$$

where

$$\pi(i) = y(n) - \frac{1+v}{\alpha}y(n) - (1+v)F(i),$$

and combining these last two equations we see that

$$\begin{aligned} y(n) &= ny(n)\left[1 - \frac{1+v}{\alpha}\right] + L - (1+v)\int_0^n F(i)di + v\int_0^n F(i)di + \frac{vny(n)}{\alpha} \\ &= any(n) + L - \int_0^n F(i)di \qquad (3.15) \end{aligned}$$

Employing (3.15) in (3.14), we see that

$$\frac{L - \int_0^n F(i)di}{1 - an}\left[1 - \frac{1+v}{\alpha}\right] = F(n)(1+v),$$

or

$$L - \int_0^n F(i)di = F(n) \cdot \frac{(1+v)(1-an)}{[1 + (1+v)/\alpha]} \tag{3.16}$$

[Notice that (3.16) reduces to (3.9) when $v = 0$, so that there is no possibility of multiple equilibrium when $v = 0$.]

You can differentiate both sides of this condition, just as we did before, to provide restrictions for which multiple equilibria are possible.

Model 3. Separation of Costs and Benefits

Another example of the general principle discussed earlier, but we don't do this in class; of minor importance.

If the setup costs and the later revenues are separated in time, then the latter will drive demand, while industrializers care about revenues *net* of setup costs. This will be sufficient to drive a wedge between private and socially optimal decision-making.

There is a problem though. For the economy to go through this deferred industrialization process, the interest rate must rise to encourage consumption postponement. For the industrializers, this brings *up* the opportunity cost of investing in industrialization.

If no one industrializes, then income is (L, L) and $\beta = \frac{1}{1+r}$ solves out for the interest rate. So the condition for no industrialization is simply

$$\beta a L - F \leq 0. \tag{3.17}$$

On the other hand, if full industrialzation occurs, current output is $L - F$ and in the next period it is αL. This must be rationalized by the Euler equation for aggregate consumption. If u denotes the indirect utility of current consumption of the composite, then solve

$$\max u(c_1) + \beta u(c_2),$$

subject to

$$c_1 + \frac{1}{1+r} c_2 = \text{ present value of income},$$

so that

$$\frac{1}{1+r} = \beta \frac{u'(c_2)}{u'(c_1)} = \beta \frac{u'(\alpha L)}{u'(L-F)}.$$

It follows that the condition for full industrialization is

$$\beta a \alpha L \frac{u'(\alpha L)}{u'(L-F)} - F \geq 0. \tag{3.18}$$

Are (3.17) and (3.18) compatible? Yes, if the interest rate does not rise by too much. [Actually, with the logarithmic specification employed by Murphy, Shleifer and Vishny [1989], this does not work, but if the indirect utility function of composite consumption has higher elasticity than logarithmic, (3.17) and (3.18) can be compatible.]

3.4.5 Complementarity through the Production Process. "The division of labor is limited by the size of the market." Thus spake Adam Smith, and he spake well. A large market size encourages investments that may be complicated and costly, but have immensely high productivity. The deeper insight of Allyn Young was to note that the converse is also true: the "division of labor", or production roundaboutness, *also determines the size of the market*. In this sense, roundaboutness begets roundaboutness.

The example that we consider has to do with the provision of intermediate inputs that are required in the production of final output in the economy. One feature of economic development is the creation and use of increasingly sophisticated methods of production, often characterized by their "roundaboutness." Almost any productive activity can serve as an example. Let's take construction. In developing countries, construction is a pretty labor-intensive activity. The area is cleared by hand, rubble is removed in small baskets carried by hand, cement is often mixed at the site and carried by hand, and walls are put up brick by brick. In industrialized economies, each of these tasks has been automated: cranes are used for clearing and prefabricated walls are erected at the site. Each automated instrument, in turn, is produced through a complicated activity: think about how cranes and prefabricated walls are themselves produced. Thus the final production of a house is reduced to a large series of automated steps, each of a high degree of sophistication and requiring the provision of many intermediate inputs.

These sophisticated inputs can be extremely costly to produce if they going to be sold in tiny markets. The manufacture and sale of cranes requires that there be a fairly large demand for cranes in construction, and so it is with prefabricated walls. Otherwise it is simply not worth setting up separate plants to manufacture these items. In other words, intermediate inputs are often produced under conditions of increasing returns to scale.

At the same time, the provision of intermediate inputs, and the consequent roundaboutness of production, can have very productive consequences, because production not only benefits from *scale*, it also benefits from the *variety* of inputs that are employed. To see this in concrete terms, suppose that output is produced using a constant returns to scale technology that includes as inputs intermediate goods as well as labor. If the quantity of labor as well as of all *existing* varieties of intermediate inputs is doubled, then output doubles: this is just a feature of constant returns to scale. This notion suggests that if the production budget is doubled, output doubles too. Not so, simply because a doubling of all inputs is only *one option* under a doubling of the budget. It is also possible to expand the *variety* of intermediate inputs that are used in production. The option to expand variety leads to a situation where output more than doubles: with input variety, increasing returns to scale is built in provided that the underlying production function exhibits constant returns. It follows that the productivity of the economy depends on the scale and richness of operations of the intermediate goods sector.

To formalize this, suppose that an intermediate good — which we loosely call capital —- is "produced" by means of several intermediate inputs (machines). Machines are indexed by i: they appear on a continuum $[0, \infty)$. The more machines are used, the more variety there is in the production process, but all machines are used ina symmetric way. the easiest way to capture this is to suppose that

$$X = \left[\int_0^n x(i)^{1-\frac{1}{\sigma}} di \right]^{\sigma/(\sigma-1)} \tag{3.19}$$

where n is an index of the variety of machines or the roundaboutness of production, and $\sigma > 1$ proxies the degree of substitution across machines. This is like a CES production function where the elasticity of substitution exceeds 1 (so that no one machine is necessary in the production process).

Suppose that capital is produced by means of a "budget" K which can be used — at a normalized price of one — to produce intermediate inputs. Then by strict concavity of the production function in (3.19) and the symmetry of the problem, the budget would be equally divided among available machines, so that $K = xn$ would solve for x, the quantity of each machine used in the production process. Consequently,

$$X = \{n\,(K/n)^{(\sigma-1)/\sigma}\}^{\sigma/(\sigma-1)} = n^{1/(\sigma-1)}K. \tag{3.20}$$

It is in this sense that variety can be equated with total factor productivity.

We are now in a position to see how this leads to multiple equilibria. Suppose that the economy is "poor" and exhibits a low demand for the final product. This situation means that intermediate production cannot occur at an economically viable scale, which means that the prices of intermediate goods are high. Consequently, firms substitute away from intermediate goods to the use of raw labor. This lowers productivity because of the argument in the previous paragraph and generates low income in the economy. Low income in turn generates a low demand for the final good, and the vicious cycle is complete. The other side of the coin is a virtuous circle. High demand for the final consumption good increases the demand for intermediates, and because these intermediates are produced under conditions of increasing returns to scale, prices of intermediates fall. Falling prices encourage a further substitution away from labor to intermediates, which raises the productivity of the economy. Incomes rise as a consequence and so does demand, completing the virtuous circle.

A formalization follows, based on Ciccone and Matsuyama [1996].

The idea of this exercise is to endogenize the degree of roundaboutness in production (notice in passing that the same kind of analysis can be carried out using variety in consumption). To this end, we suppose that each of the machine sectors must be set up (with a fixed cost of S, denominated in terms of labor) before production can commence. Thereafter production of machines takes place under constant returns to scale (see description below). Each machine sector is run by a monopolist, but since different machines can — in principle — be very close substitutes in production, there are limitations on his pricing behavior. This is what we begin by exploring.

Imagine that varieties $[0, n]$ are in force. Then producers of the final good will demand a quantity $x(i)$ of machine i, chosen to

$$\max F\left(\{\int_0^n x(i)^{(\sigma-1)/\sigma} di\}^{\sigma/(\sigma-1)}, L\right) - wL - \int_0^n p(i)x(i)di.$$

Of course, the input i is of measure 0 but the following calculation can easily be justified by thinking of this as an approximation for a large but finite number of input varieties. The necessary and sufficient first-order condition is

$$F_X \frac{\sigma}{\sigma-1}\{\int_0^n x(i)^{(\sigma-1)/\sigma} di\}^{1/(\sigma-1)} \frac{\sigma-1}{\sigma} x(i)^{-1/\sigma} = p(i),$$

or

$$F_X X^{1/\sigma} x(i)^{-1/\sigma} = p(i),$$

or

(3.21)
$$x(i) = \frac{F_X^\sigma X}{p(i)^\sigma}.$$

[Note: $p(i)x(i) = \frac{F_X^\sigma X}{p(i)^{\sigma-1}}$, so demand is elastic.]

Now turn attention to machine producers. Suppose that machines are produced using labor alone, and that a units of labor are needed to produce one unit of a machine. Then the marginal cost of producing y units of a machine is given by way, where w is the wage rate.

Using (3.21),we can conclude that producers of the intermediate good will choose $p(i)$ to

$$\max \frac{F_X^\sigma X}{p(i)^{\sigma-1}} - wa\frac{F_X^\sigma X}{p(i)^\sigma},$$

and (because each machine is of measure 0), this is equivalent to the problem

$$\max_{p(i)} \frac{1}{p(i)^{\sigma-1}} - \frac{wa}{p(i)^{\sigma}}.$$

You should check that the first-order conditions characterizing this problem are necessary and sufficient. They are

$$(1-\sigma)p(i)^{-\sigma} + wa\sigma p(i)^{-\sigma-1} = 0,$$

or

$$p(i) = \frac{wa\sigma}{\sigma - 1}.$$

To make things easier to write, choose units of labor so that $a = 1 - (1/\sigma)$; then

$$p(i) = w \text{ for all } i. \tag{3.22}$$

Before we go on to determine varieties, think of what it will now cost to buy one unit of the composite capital X, given that this price determination process is in place. Recall that the composite is

$$\begin{aligned} X &= \left[\int_0^n x(i)^{(\sigma-1)/\sigma} di\right]^{\sigma/(\sigma-1)} \\ &= n^{\sigma/(\sigma-1)} x, \end{aligned}$$

because the same amount of every machine is bought given the pricing rule (3.22). So the effective price of X — call it P — is given by

$$P = \frac{\text{cost of buying } X}{X} = \frac{pnx}{n^{\sigma/(\sigma-1)}x} = \frac{w}{n^{1/(\sigma-1)}}.$$

Thus P/w — the *effective* relative factor price — is equal to $n^{1/(1-\sigma)}$, which is declining in n.

So if F is CRS and we denote by $\alpha \equiv \frac{F_X(X,L)X}{F(X,L)}$ the factor share of X in production, then

$$\alpha = \alpha(n^{1/(1-\sigma)}) \equiv A(n). \tag{3.23}$$

If the elasticity of substitution between X and L exceeds unity, $A(n)$ is an increasing function (it is flat if F is Cobb-Douglas).

Now we return to the problem of determining equilibrium variety. Denoting the value of final output by Y, the operating profit for a producer of intermediates is given by

$$\pi = (p - aw)x = px(1-a) = \frac{px}{\sigma} = \frac{\alpha Y}{\sigma n},$$

where the second equality follows from the pricing rule $p = w$, the third equality from the choice of labor units, and the last inequality from the fact that $\alpha Y = npx$ by definition of factor share. Thus

$$\pi = \frac{A(n)}{n}\frac{Y}{\sigma}. \tag{3.24}$$

Equation (3.24) tells us that an increase in variety has three effects on the operating profit of a typical producer of intermediates:

[1] A larger n increases variety and this decreases the share to each variety (this is the $1/n$ term).

[2] A larger n affects the factor share of intermediates, generally in a positive direction (this is the $A(n)$ term).

[3] A larger n affects final output (and therefore national income).

Now let us go ahead and finish the endogenization of n using the free-entry condition:

$$\pi = \frac{A(n)}{n}\frac{Y}{\sigma} = \text{ startup costs } = wS,$$

or

$$\frac{n\sigma}{A(n)} = \frac{Y}{wS} = \frac{wL + pnx}{wS} = \frac{w(L + anx) + n\pi}{wS}.$$

Now observe that $L + anx = (T - nS)$, where T is the total labor endowment in the economy, and π is simply equal to wS. Using these in the equation above, we conclude that

$$\frac{n\sigma}{A(n)} = \frac{T}{S}. \tag{3.25}$$

This expression tells us that the potential for multiplicity is intimately linked to the behavior of $n/A(n)$ — to its possible nonmonotonicity in n, to be more accurate. For instance, in the Cobb-Douglas case, $A(n)$ is a constant as we have already seen, so that there is a unique level of product variety in the economy. On the other hand, if $A(n)$ increases sharply with n (at least over some range), then the complementarity is strong and multiple equilibria are indeed possible.

3.4.6 Complementarity and Finance. The focus in this section is not so much the idea of multiple equilibria as the notion of how externalities might permeate the development process through different channels. In this section we study the financial sector.

It is well known that financial deepening is one of the characteristics of the development process. The introduction of money into a subsistence economy opens up opportunities for trade that never existed before. Similarly, the expansion of the credit system opens up new opportunities for investment, and this is taken one step further when a stock market comes into being. Now, the investment-enhancing effects of financial deepening are well worth studying, but we concentrate here on a slightly different set of questions: how is the *availability* of finance (from end-savers) tied to the extent of financial deepening?

The answer that we would like to explore is that financial deepening offers opportunities for diversification, and this encourages a greater flow of savings from low-risk (but low-return) activities to the higher-risk (but higher-return) sectors. To be sure, the greater flow permits, in its turn, greater deepening of the financial market. This is the phenomenon explored in this section (we draw on Acemoglu and Zilibotti [1997]).

Notice that if each sector costs nothing to set up, then large amounts of diversification can be achieved even with small amounts of finance by simply spreading the available finance arbitrarily thinly over the existing sectors. Thus the idea of setup costs (or more generally, a nonconvexity) must enter the picture again.

Begin by looking at individual choices. Suppose that financial securities for different production sectors are indexed by $j \in [0,1]$. Sector j pays off a return of R if state j occurs. States are realized with uniform probability on $[0,1]$.

Notice that our goal is not to examine the determinants of productivity in each sector (as we did earlier), nor do we ascribe different rates of return to each production sector. The only goal here is to examine diversification possibilities.

Suppose that a measure n of sectors is "open" at any particular date. Rearrange sectors so that we can think of these as sectors in $[0,n]$, where $0 \leq n \leq 1$. Now an agent must assign his current assets to a savings portfolio. Denote assets by A. The sectors available are all these (of measure n), and a riskless sector which pays a return factor of r. That is, if an agent assigns a fraction α of his savings to the risky portfolios, and owns a "density" $f(j)$ of asset j, his return in state j is

$$\begin{aligned} c(j) &\equiv (1-\alpha)Ar + \alpha ARf(j) \text{ if } j \in [0,n], \\ &\equiv (1-\alpha)Ar \text{ if } j \notin [0,n]. \end{aligned} \tag{3.26}$$

Assume for now that these returns are all consumed (later we outline a model that allows for this as well as for other possibilities). Assume that the agent is risk-averse and that his utility is given by the familiar functional:

$$\int_0^1 \ln(c(j))dj.$$

Substituting in the value of $c(j)$ from (3.26), we see that the agent's objective is to choose α and the density $f(j)$ to maximize

$$\int_0^n \ln\left([1-\alpha]Ar + \alpha ARf(j)\right)dj + (1-n)\ln\left([1-\alpha]Ar\right), \tag{3.27}$$

subject to the constraints that $\alpha \in [0,1]$, $f(j) = 0$ for $j > n$, and $\int_0^n f(j)dj = 1$.

The first thing to note about this maximization problem is that *given* any value of α, the density $f(j)$ must be uniform over $[0,n]$. This equalizes agent returns over all states in $[0,n]$, and by second-order stochastic dominance the agent must prefer this to any other allocation over the risky portfolio. Consequently, our maximization problem reduces to the task of finding $\alpha \in [0,1]$ to maximize

$$n\ln\left([1-\alpha]Ar + \frac{\alpha AR}{n}\right) + (1-n)\ln\left([1-\alpha]Ar\right). \tag{3.28}$$

Write down the first-order conditions for this to solve the problem (and verify the second-order conditions yourself):

$$\frac{R - rn}{(1-\alpha)r + \frac{\alpha R}{n}} = \frac{1-n}{1-\alpha}.$$

Rearrange this to get

$$\alpha = \frac{(R-r)n}{R-rn} \tag{3.29}$$

Notice that α is increasing in n (simply differentiate (3.29)). This summarizes our intuition that the greater the amount of financial deepening (as proxied by n), the the greater the amount of savings coming into the (risky) financial sector.

Now we move away from the individual level to understand some more about the process of opening new financial sectors. Specifically, we wish to caputure the possibility that some sectors require certain minimum investment sizes in order to open, and that these can vary from sector to sector. the easiest way to do this is to suppose that there is a function $S(i)$ defined on $[0, 1]$ that captures the cost of opening a new sector i. We order sectors (this will not be at variance with our previous ordering) so that $S(i)$ is increasing — assume it is smooth, with $S(0) = 0$.

The macroeconomic equilibrium at any one point of time can now be described.

Assuming that n sectors are open, investment per sector — as we have seen — is given by the amount

$$\frac{\alpha A}{n} = \frac{R - r}{R - rn} A,$$

using (3.29) It follows that the equilibrium financial depth is given by

$$\frac{R - r}{R - rn} A = S(n). \tag{3.30}$$

To be sure, there may be multiple solutions to this equation for the usual reason that there are multiple equilibria, but we are not going to pay this too much attention. The point that we wish to emphasize is this — *as long as there is no equilibrium that involves maximal depth* ($n = 1$), *every equilibrium is Pareto-inefficient*. There are externalities in the process that will not go away, even if coordination failures are resolved.

To understand why, notice that the opening of a fresh sector confers positive benefits on all agents, simply because of the diversification possibilities that are involved. However, individual agents wish to equalize their holdings across sectors. This tendency towards equalization means that sectors with large setup requirements cannot be accomodated. At the margin, *some* accomodation must be beneficial. This is the source of the inefficiency.

To make this point more precisely, imagine a situation in which a social planner chooses not only α and the portfolio for the representative agent in the economy, *but also the value of* n. In other words, such a planner would actually maximize

$$\int_0^n \ln\left([1-\alpha]Ar + \alpha A R f(j)\right) dj + (1-n)\ln\left([1-\alpha]Ar\right), \tag{3.31}$$

subject to all the constraints described after (3.27), *and* a suitable choice of n, under the additional constraint that $\alpha A f(j) \geq S(j)$ for all $j \in [0, n]$.

The solution to this problem is a bit more complicated because of this last constraint. In particular, equal division of the portfolio cannot be assured. But what we *can* say is this: if

$$\alpha A f(j) > S(j)$$

over any interval of securities, then over that interval we must have equal division of the portfolio. For if not, a second-order dominance move can be created without violating any of the constraints. It follows that along some interval $[0, n]$, we have equal division, while *possibly* along an additional interval $[n, n^*]$, we have portfolio holdings that exactly compensate for the fixed cost of those sectors. So our maximization problem may be rewritten as: choose n, n^* (not less than n) and $\alpha \in [0, 1]$ to maximize

$$n \ln([1-\alpha]Ar + S(n)R) + \int_n^{n^*} \ln([1-\alpha]Ar + S(i)R)\, di + (1-n^*) \ln([1-\alpha]Ar), \tag{3.32}$$

subject to the "budget constraint"

$$S(n)n + \int_n^{n^*} S(i)di = \alpha A. \tag{3.33}$$

Set up the Lagrangean

$$\begin{aligned} & n \ln([1-\alpha]Ar + S(n)R) + \int_n^{n^*} \ln([1-\alpha]Ar + S(i)R)\, di + (1-n^*) \ln([1-\alpha]Ar) \\ & +\lambda \left[\alpha A - S(n)n - \int_n^{n^*} S(i)di\right], \end{aligned}$$

and write down the Kuhn-Tucker conditions (we'll only need to do so for the derivatives with respect to n^* and n):

$$\ln([1-\alpha]Ar + S(n^*)R) - \ln([1-\alpha]Ar) - \lambda S(n^*) \leq 0, \text{ with equality if } n^* > n, \tag{3.34}$$

and

$$\frac{nS'(n)R}{[1-\alpha]Ar + S(n)R} - \lambda n S'(n) \leq 0, \text{ with equality if } n > 0. \tag{3.35}$$

Now the question is: is it ever possible for an equilibrium to satisfy the Kuhn-Tucker conditions, which, as we know, are *necessary* for an optimum? [Never mind whether they are sufficient.] The answer is no. To see this, notice that in an equilibrium, $n > 0$ (because $S(0) = 0$). Consequently, (3.35) holds with equality, so that

$$\lambda = \frac{R}{[1-\alpha]Ar + S(n)R}.$$

Therefore, if $n = n^*$, (3.34) implies that

$$\ln([1-\alpha]Ar + S(n)R) - \ln([1-\alpha]Ar) \leq \frac{RS(n)}{[1-\alpha]Ar + S(n)R}. \tag{3.36}$$

But this contradicts the strict concavity of the logarithmic function. We know that $\ln(y) - \ln(x) > (y-x)/y$. Applying this to $x = [1-\alpha]Ar$ and $y = [1-\alpha]Ar + S(n)R$, we may conclude that

$$\ln([1-\alpha]Ar + S(n)R) - \ln([1-\alpha]Ar) > \frac{RS(n)}{[1-\alpha]Ar + S(n)R},$$

which contradicts (3.36).

It follows that we must always have $n^* > n$ in the social planning problem; that is, some sectors must be "cross-subsidized". But no agent wants to take up the burden of cross-subsidization! *Given* that other agents are holding unequal portfolios in order to keep a sector open, a given agent will always want to allocate his portfolio equally among the sectors. To be sure, if all agents do so, some sectors will have to close down, perpetuating the inefficiency.

Observe that the financial deepening problem has implications for the pace of development, as well as its uncertainty. In countries with small asset levels (proxied by A), aggregate development will be beset by frequent crises, as full insurance (diversification) is not possible. Extend the model a little bit to recursively define A in each period in one of two ways:

[1] A is the product of savings decisions made by young agents who only earn a wage; the wage is the marginal product of labor evaluated at the going capital stock of the economy. Total returns to the financial sector constitute the amount of capital, which is ploughed into the deterministic production of a final good, under an Inada-style production function. Agents live two periods. This is the approach taken by Acemoglu-Zilibotti and this model can be easily embedded into sucha dynamic model.

[2] A is the product of savings decisions made by long-lived agents, in which the log utility function postulated here must be viewed as the value function of a dynamic program with underlying utility functions also logarithmic. Income is then the sum of wages and capital returns at any date. This is a bit more complicated as a dynamic program needs to be solved.

With either of these extensions, A will fluctuate for long periods of time at low levels (if one starts low) until a certain asset level is reached that permits full or near-full diversification. At this point the economy settles down to a steady rate of growth. Thus, while the ergodic behavior of all economies is the same, the transitions can be very different: for a poor economy, the steady state can be a long time coming.

Notice in conclusion that this is one view of history-dependence. Another has to do with the possibility that *long run* behavior may also be different depending on initial conditions. This is what we take up in coming sections of the course.

CHAPTER 4

History Versus Expectations

So far, we've discussed the existence of a virtuous circle of expectations that might lead communities to Pareto-superior states among multiple potential equilibria. As we have already noted, it is possible to think of these as generalized coordination games. Once the game-theoretic analogy is pressed into service, however, certain troubling issues arise.

Rosenstein-Rodan and Hirschman were concerned with multiple equilibria, no question, but this was only the starting point of their analysis. They took it as self-evident that *history* somehow plays a role in pinning down the "starting equilibrium", and that the role of policy would be to construct ways of "moving" society from one equilibrium to another.

It is of interest to note that the theory of pure coordination games makes it very difficult, if not impossible, to ask this question in a well-defined way. Imagine a sequence of such situations. Suppose that a bad equilibrium outcome has been in force for the previous 100 plays of the game. In what sense does that make it more likely that in the 101st iteration, the bad equilibrium will occur again? Formulated in this way, the answer must be: in no sense at all. For as is well-known, all sorts of complicated switches between the good and bad equilibria can actually constitute a perfect equilibrium of the repeated game. There is no role for history in such a formulation.

Thus, by and large, this literature ignores a central question raised by Rosenstein-Rodan and Hirschman: how does an economy "move from a bad to a good equilibrium"? We place this entire phrase in quotes because it is imprecise: so called transitions from one "equilibrium" to another must themselves be viewed as the equilibria of some encompassing intertemporal process.

It appears (unless some convincing psychological model can be found that weds people to the historical status quo) that some sort of state variable is needed to make sense of the Rosenstein-Rodan/Hirschman argument. [This will also serve as a tentative introduction to the next part of the course in which such state variables, such as initial inequalities, are strong enough to pin down a unique equilibrium.] But the state variable must be introduced in some "economical way", that does not rule out — by fiat or by assumption — the possibility of moving to a new equilibrium.

In what follows, we look at some different approaches to the question of equilibrium-picking when there is potential coordnation failure.

4.1 State Variables and Common Knowledge

The first avenue that we explore is due to Morris and Shin (1998) (see also Carlsson and Van Damme (1993)). The basic idea is very simple and very powerful. Suppose that a coordination game is played between several agents — we shall write a concrete one down in a moment — and that the realization of a state variable affects the payoffs. Suppose that for certain values of the state variable the play of one action or the other is a dominant strategy, but for other values of the state there is a coordination problem. If the realization of the state variable is common knowledge, there is no bite at all in introducing the state variable: each realization of the state precipitates a different game, and each game is played separately.

But suppose that there isn't common knowledge of the realizations. Suppose that we all see the realized value with a bit of noise. Then I know that you have seen something, but you may have seen something a bit different from me. Now there isn't a string of separate games for each realization of the state variable. The implications can be surprising, as the following model of financial crises reveals.

Consider a country which has pegged its exchange rate at some value e. (For concreteness, think of e as the number of dollars required to buy one unit of the domestic currency.) We shall assume that this exchange rate is unambiguously overvalued, in the following sense: suppose that there is some random variable θ (the state) on $[0, 1]$ which determines the "true" exchange rate $f(\theta)$ were the currency to be allowed to float at θ. Then e always exceeds $f(\theta)$ for all $\theta \in [0, 1]$.

But θ also influences the exchange rate: which is to say that $f(\theta)$ varies with θ. Arrange so that $f(\theta)$ is strictly increasing in θ. So the idea is that θ is some "fundamental" which influences the country's capacity to export or import, or to attract investment; the higher being θ, the more favorable the climate.

Now there is a bunch of speculators (of total measure 1), each of whom can sell one unit of the local currency. If they do, they pay a transactions cost t. If the government holds the peg, the exchange rate stays where it is, and this is the payoff to selling: $-t$. If the government abandons the peg, then the speculators make a profit of $e - f(\theta)$, so their net payoff is $e - f(\theta) - t$.

What about the government's decisions? It has only one decision to make: whether to abandon or to retain the peg. We assume that it will abandon the peg if the measure of speculators exceeds $a(\theta)$, where a is increasing in θ (that is, if the basic health of the economy is better, the government is more reluctant to abandon[1]).

[1]See Morris and Shin (1998) for a very simple account of how to derive $a(\theta)$ from a somewhat more basic starting point.

We will assume that there is some positive value of θ, call it $\underline{\theta}$, such that below $\underline{\theta}$ the situation is so bad that the government will abandon the peg anyway. In other words we are assuming that $a(\theta) = 0$ for $\theta \in [0, \underline{\theta}]$. Then it rises but always stays less than one by assumption.

Consider, now, a second threshold for θ which we'll call $\bar{\theta}$: this is the point above which no one wants to sell the currency even though she feels that the government will abandon the peg for sure. In other words, $\bar{\theta}$ solves the equation

$$e - f(\bar{\theta}) - t = 0. \tag{4.1}$$

We will assume that such a $\bar{\theta}$, strictly less than one, indeed exists. But we also suppose that there is a gap between $\underline{\theta}$ and $\bar{\theta}$: that $\underline{\theta} < \bar{\theta}$.

[If there were no such gap, there wouldn't be a coordination problem to start with.]

Now we are ready to begin our discussion of this model. First assume that the realization of θ is perfectly observed by all agents, and that this information is common knowledge. Then there are obviously three cases to consider.

Case 1. $\theta \leq \underline{\theta}$. In this case, the government will abandon the peg for sure. The economy is not viable, all speculators must sell, and a currency crisis occurs.

Case 2. $\theta \geq \bar{\theta}$. In this case no speculator will attack the currency, and the peg will hold for sure.

Case 3. $\underline{\theta} < \theta < \bar{\theta}$. Obviously this is the interesting case, in which multiple equilibria obtain. There is an equilibrium in which no one attacks, and the government maintains the peg. There is another equilibrium in which everyone attacks and the government abandons the peg. This is a prototype of the so-called second-generation financial crises models, in which expectations — over and above fundamentals — play an important role (see Obstfeld (1994, 1996)).

So much for this standard model, variants of which we have already seen in excruciating detail. Now we drop the common knowledge of realizations (but we will maintain the assumption of common knowledge of the information *structure* that I am going to write down).

Suppose that θ is distributed uniformly on $[0, 1]$: its value will be known perfectly at the time the government decides whether or not to hold the peg or to abandon it. Initially, however, the realization of θ is noisy in the following sense: each individuals sees a signal x which is distributed uniformly on $[\theta - \epsilon, \theta + \epsilon]$, for some tiny $\epsilon > 0$ (where θ is the true realization). Conditional on the realization of θ, this additional noise is iid across agents.

Proposition 4.1. *There is a unique value of the signal x such that an agent attacks the currency if $x < x^*$ and does not attack if $x > x^*$.*

This is an extraordinary result in the sense that a tiny amount of noise refines the equilibrium map considerably. Notice that as $\epsilon \to 0$, we are practically at the common knowledge limit (or are we? the question of what sort of convergence is taking place is delicate and important here), yet there is no "zone" of multiple equilibria! The equilibrium is unique.

What is central to the argument is the "infection" created by the lack of common knowledge (of realizations). To see this, we work through a proof of Proposition 4.1, with some informal discussion.

Start by looking at the point $\underline{\theta} - \epsilon$. Suppose that someone receives a signal x of this value or less. What is she to conclude? She doesn't know what everyone else has seen, but she *does* know that the signal is distributed around the truth with support of size 2ϵ. This means that the true realization *cannot* exceed $\underline{\theta}$, so that the government will abandon the peg for sure. So she will sell. That is, we've shown that for all

$$x \leq x_0 \equiv \underline{\theta} - \epsilon,$$

it is dominant to sell.

Now pick someone who has a signal *just* bigger than x_0. What does he conclude? Suppose, for now, he makes the assumption that *only* those with signals less than x_0 are selling; no one else is. Now what is the chance — given his own signal x — that someone else has received a signal not exceeding x_0? To see this, first note that the true θ must lie in $[x - \epsilon, x + \epsilon]$. For each such θ the chances that the signal for someone else is below x_0 is $(1/2\epsilon)[x_0 - (\theta - \epsilon)]$, so the overall chances are just these probabilities integrated over all conceivable values of θ, which yields $(1/2\epsilon)[x_0 - (x - \epsilon)]$. So the "infection" spreads: if x is close to x_0, these chances are close to 1/2. In this region, moreover, it is well known that the government's threshold is very low: close to zero sellers (and certainly well less than half the population) will cause an abandonment of the peg. Knowing this, such an x must sell. *Knowing that all with signals less than x_0 must sell, we have deduced something stronger: that some signals above x_0 must create sales as well.*

So let us proceed recursively: Suppose we have satisfied ourselves that for some index n, everyone sells if the signal is no bigger than x_n (we already know this for x_0). We define x_{n+1} as the *largest* value of the signal for which people will want to sell, knowing that all below x_n are selling.

This is a simple matter to work out. Fix $x \geq x_n$, and imagine any $\theta \in [x - \epsilon, x + \epsilon]$. For such θ, everybody with a signal between $\theta - \epsilon$ and x_n (such an interval may be empty, of course) will attack, by the recursive assumption. Because these are the *only* attackers (also by the recursive assumption), the government will yield iff

$$\frac{1}{2\epsilon}[x_n - (\theta - \epsilon)] \geq \alpha(\theta),$$

or

$$\theta + 2\epsilon a(\theta) \leq x_n + \epsilon$$

So we can define an implicit function $h(x, \epsilon)$ such that the above inequality translates into

$$\theta \leq h(x_n, \epsilon).$$

Put another way, the implicit function $h(x, \epsilon)$ solves the equation

$$h(x, \epsilon) + 2\epsilon a(h(x, \epsilon)) = x + \epsilon \tag{4.2}$$

It follows that if our person with signal x were to attack, her expected payoff would be given by

$$\frac{1}{2\epsilon}\int_{x-\epsilon}^{h(x_n,\epsilon)}[e-f(\theta)]d\theta - t. \tag{4.3}$$

Now retrace the recursion starting all the way from $n = 0$: we have $x_0 = \underline{\theta} - \epsilon$. Then (remembering that $a(\theta) = 0$ for all $\theta \le \underline{\theta}$) it is easy to see that (4.3) reduces to

$$\frac{1}{2\epsilon}\int_{x-\epsilon}^{\underline{\theta}}[e-f(\theta)]d\theta - t.$$

For $x \simeq x_0$, this is just

$$\frac{1}{2\epsilon}\int_{\underline{\theta}-2\epsilon}^{\underline{\theta}}[e-f(\theta)]d\theta - t,$$

which is certainly strictly positive. So x_1 is well-defined, and $x_1 > x_0$.

Now put x_1 in place of x_0, and repeat the process. Notice that h is increasing in x, so if we replace x_0 by x_1 in (4.3), then, evaluated at $x = x_1$, the payoff must turn strictly positive.[2] So the new x_2, which is the maximal signal for which people will sell under the belief that everyone less than x_1 sells, will be still higher than x_1. And so on: the recursion creates a strictly increasing sequence $\{x_n\}$, which converges from below to x^*, where x^* solves

$$\frac{1}{2\epsilon}\int_{x^*-\epsilon}^{h(x^*,\epsilon)}[e-f(\theta)]d\theta - t = 0. \tag{4.4}$$

It is very easy to see that there is a unique solution to x^* defined in this way. In fact, something stronger can be established:

Claim. If x^* is some solution to (4.4), and $x' > x^*$, then

$$\frac{1}{2\epsilon}\int_{x'-\epsilon}^{h(x',\epsilon)}[e-f(\theta)]d\theta - t < 0. \tag{4.5}$$

To prove this, consider any $x' > x^*$. Then two things happen: first, it is easy to see that

$$h(x',\epsilon) - x' < h(x^*,\epsilon) - x^*,$$

so that the support over which integration takes place in (4.4) is narrowed. Moreover, the stuff *inside* the integral is also smaller when we move from x^* to x', because $f(\theta)$ is increasing. So the LHS of (4.4) *unambiguously* falls when we move from x^* to x', and we are done with the Claim.

To learn a bit more about x^*, use (4.2) to see that $h(x,\epsilon) - x + \epsilon = 2\epsilon[1 - a(h(x,\epsilon))]$, so that

$$0 = \frac{1}{2\epsilon}\int_{x^*-\epsilon}^{h(x^*,\epsilon)}[e-f(\theta)]d\theta - t = [1-a(h(x,\epsilon))]e - \frac{1}{2\epsilon}\int_{x^*-\epsilon}^{h(x^*,\epsilon)}f(\theta)d\theta - t,$$

or

$$e - \frac{1}{2\epsilon}\int_{x^*-\epsilon}^{h(x^*,\epsilon)}f(\theta)d\theta - t = a(h(x,\epsilon))e$$

[2] This is on the assumption that the sequence $\{x_n\}$ stays bounded below $\bar{\theta}$. This will certainly be the case, see below, so it's not really an assumption at all.

A comnparison of this equation with (4.1) categorically shows that x^* is bounded below $\bar{\theta}$ for small ϵ.

So there is a unique solution to x^* and it is below $\bar{\theta}$, which justifies the previous recursive analysis (see in particular, footnote 2). Notice also that our analysis shows that *every* equilibrium must involve attack for signals less than x^*.

To complete the proof, we must show that no signal above x^* can ever attack. Suppose, on the contrary, that in some equilibrium some signal above x^* finds it profitable to attack. Take the supremum of all signals under which it is weakly profitable to attack: call this x'. Then at x' it is weakly profitable to attack. Suppose we now entertain a change in belief by supposing that everybody below x' attacks *for sure*; then this cannot change the weak profitability of attack at x'. But the profit is

$$\frac{1}{2\epsilon}\int_{x'-\epsilon}^{h(x',\epsilon)}[e - f(\theta)]d\theta - t,$$

which is nonnegative as we've just argued. But this contradicts the Claim.

So we have proved that there is a unique equilibrium to the "perturbed" game, in which a speculative attack is carried out by an individual if and only if $x \le x^*$. As $\epsilon \to 0$, this has an effect of refining the equilibrium correspondence dramatically. To describe this, calculate the threshold x^* as $\epsilon \to 0$. The easiest way to do this is the "sandwich" inequality:

$$[e - f(h(x^*,\epsilon))][1 - a(h(x^*,\epsilon)] \le \frac{1}{2\epsilon}\int_{x^*-\epsilon}^{h(x^*,\epsilon)}[e - f(\theta)]d\theta \le [e - f(h(x^*,\epsilon))][1 - a(h(x^*,\epsilon)],$$

which is obtained by noting that $f(x^* - \epsilon) \le f(\theta) \le f(h(x^*,\epsilon))$ for all $\theta \in [x^* - \epsilon, h(x^*,\epsilon)]$. Both sides of the sandwich go to the same limit, because x^* and $h(x^*,\epsilon)$ — as well as the realization of the state — all go to a common limit, call it θ^*. This limit solves the condition

$$[e - f(\theta^*)][1 - a(\theta^*)] = t. \tag{4.6}$$

It is obvious that there is a unique solution to (4.6).

Note: At this point be careful when reading Morris-Shin. There is an error in Theorem 2. See Heinemann (AER 2000) for a correction of this error which agrees with the calculations provided here.

4.2 Lagged Externalities

The use of a state variable which may change the fortunes of one sector versus another (irrespective of sectoral membership) *at any one point of time* is a useful way of thinking about (relatively) short-run episodes such as financial crises. It may not be as useful when thinking about a country or region which is stuck in some low-investment trap for long periods of time. It is much harder to conceive of a state variable that will signal positive profitability for the high-investment outcome (regardless of the actions of other investors) in any one period. In what follows, we take two approaches; one that dispenses with this sort of state variables altogether (Adserà and Ray [1998]; this section) and one that resurrects state variables, but

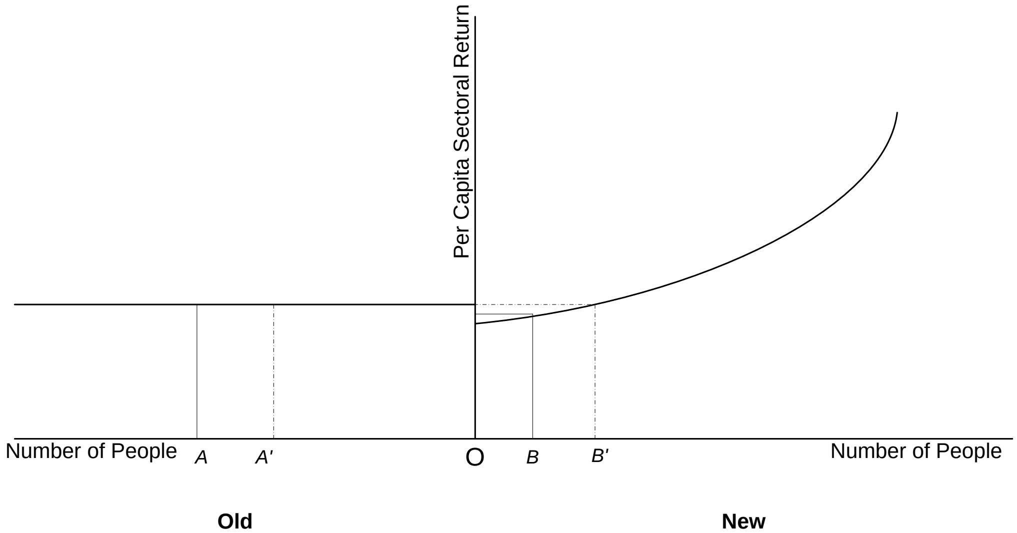


FIGURE 4.1. A DEPICTION OF SECTORS A AND B.

has them follow a Markov process (possibly with small local support) over time (Frankel and Pauzner [2001]; next section).

Suppose that an economy has two regions, A and B. A total capital (or labor) endowment of $\bar{K}$ is split at date 0 between the two regions. Denote by K the capital in region B, so that $\bar{K} - K$ is the capital stock in region A. Capital invested in Region A yields a fixed rate of return, normalized to zero. Region B's rate of return r is taken to depend positively on its capital endowment:

$$r = f(K) \tag{4.7}$$

where f is continuous, strictly increasing, and $f(0) < 0 < f(\bar{K})$.

Imagine that there is a continuum of agents, and each agent owns a single unit of capital. Capital is free to move between regions but each relocation entails a nonnegative cost.

This model may be conveniently summarized by using a simple diagram.

Figure 4.1 shows us the two sectors. In A the rate of return is flat at the (normalized) level zero. In B, the return to any one individual to participating in a sector depends positively on the number of individuals already active in that sector (starting below zero and ending above). To complete the description, we locate the initial allocation of individuals across the two sectors (given by history). This is given by OA people in A and OB people in B. You could think of the line segment AB as the total number of people in the economy: as the allocation of people changes, the only thing that alters is the *position* of AB but not its length.

Now, the initial allocation of people has been chosen to deliberately illustrate a point. Even though B is, in principle, "better" than A (if everyone were in B the rate of return would exceed that of A), at the starting location, the *actual* rate of return in A exceeds the actual rate of return in B.

Note that our examples discussed so far fit in quite neatly into this framework. For instance, the return in B could be interpreted as the wages paid to people who have decided to acquire a certain set of skills, *as a function of the number of individuals who have already acquired such skills*. In this interpretation, if you were to think of the alternative occupation as conferring no externalities, you might imagine that the rate-of-return line in A as nearly flat, while the one in B slopes upwards. Or you could think of B as typewriters with the Dvorak system, or for that matter as alternative chip designs, while A contains the QWERTY typewriters, or Intel chips. The rate of return is then to be interpreted as the total amount of satisfaction accruing to a consumer or user, net of the cost of purchase or usage. In all these examples, it might be useful to think of OA as a large segment, while OB is very small, perhaps zero.

Now we are ready to make this model run. Imagine that, as indicated in Figure 4.1, the rate of return in A exceeds the corresponding rate in B. Then, as time passes, individuals will gravitate from B to A. This describes the failure of of an exciting new sector when there is enough critical mass to keep the momentum going. Matters will end with everybody in A and nobody in B.

Figure 4.1 also reveals that if there had been sufficient critical mass, matters could have been entirely different. For instance, suppose that initial history initially put us at the allocation OA' and OB', where the rates of returns in the two sectors are exactly equal. Then the slightest additional tilt towards B can spark off an accelerating tempo of beneficial change, as people switch over to the new technology (or the new product, or a new way of life).

Now here is a different story that works entirely on the basis of expectations. Begin again in the situation where OA individuals are in A, while the rest, OB, are in B. Now imagine that for some reason, *everybody* believes that *everybody else* will be in B tomorrow. Never mind where this belief came from. Simply note that *if* this belief is genuinely held by someone, he must also believe that B is the sector to be in, because the return there is higher. Consequently, he will gravitate to B tomorrow. But if everybody thinks the same way, everybody will be in B, and the seemingly absurd belief is completely justified (provided that the one-time cost of moving is not too high, of course). Thus, it would seem from this argument that history plays no role at all. *Irrespective* of initial conditions, there are only two self-justifying outcomes that are possible, everyone in A or everyone in B, and that *both* these outcomes are always possible, depending only on expectations.[3] How do we square this story with the one that we described earlier?

To analyze this, we enrich both the switching cost specification as well as the intertemporal structure of the externality.

In all that follows, regard K as the amount of capital in Sector B, so that $\bar{K} - K$ is the amount of capital in Sector A. Assume that the cost of moving from B to A is given by a function $\hat{c}_A(K)$ and that the cost of moving from A to B is given by some function $\hat{c}_B(K)$. Say that either of these cost functions exhibits *congestion* if it increases, at least over some interval, in

[3]There is also a third outcome in which the rates of return in the two sector are exactly equalized, but it is possible to rule this out on grounds of “stability”.

the capital in the relevant sector. That is, $\hat{c}_A$ exhibits congestion if it is decreasing over some range, while $\hat{c}_B$ does so if it is increasing over some range.

Before we describe agent behavior, let us track the relevant prices. Let $\gamma \equiv \{r(t), c_A(t), c_B(t)\}_{t=0}^{\infty}$ be some point expectation about the path (measurable in time) of returns and relocation costs in each region. Future returns are discounted in the standard way, using a discount rate ρ. Denote by $V(\gamma, i, t)$ the *optimal value* to an agent in region i, $i = A, B$, beginning at time t, when the commonly anticipated path of returns is γ. Then by standard dynamic programming arguments, an agent in region i will switch sectors at time t if $V(\gamma, i, t) < V(\gamma, j, t) - c_j(t)$, will stay if the opposite inequality holds, and will be indifferent if equality holds.

A path γ is an *equilibrium* if it is generated by the optimal decisions of (almost) all agents in response to γ.

To discuss the generation of γ, consider now some exogenously given measurable path $\{K(t)\}_{t=0}^{\infty}$ (recall the interpretation that $K(t)$ is the amount of capital in sector B).[4] We assume that there is *some* lag (however small) in the speed at which external effects induced by incoming/outgoing factors affect the going rates of return. From this point of view, we regard the return function $f(K)$ as representing a "long run level" of the rate of return, once the economy has settled at a certain level of capital K. We assume that at date 0, $r(0)$ is precisely $f(K(0))$ (see (4.7)). Thereafter, we introduce an increasing function g, with $g(0) = 0$, such that

$$\dot{r}(t) = g(f(K(t)) - r(t)). \tag{4.8}$$

Thus, the rate of return at any date "chases" the "appropriate" rate of return corresponding to the division of the capital endowment at that date. The specific functional form of $g(.)$ determines the speed at which returns adjust. In any case, capital owners will get paid $r(t)$ at date t.

Thus a path of capital allocations $\{K(t)\}$ generates a path of returns $\{r(t)\}$ using (4.8), and a path $\{c_A(t), c_B(t)\}$ using the relationships $c_B(t) = \hat{c}_B(K(t))$ and $c_A(t) = \hat{c}_A(\bar{K} - K(t))$ for all t.

Several economic situations conform quite naturally to this specification. In models of search or matching, the productivity of some fixed amount of capital may depend on the ability of that capital to find partners (with more capital), say, because of minimum scale requirements in production. This ability, in turn, will depend on the total amount of capital in the economy (see, e.g., Diamond [1982]). Note that a discontinuous jump in the capital stock will lead to a smooth intertemporal increase in productivity as long as the process of "matching partners" takes place in continuous time. Likewise, if one replaces "capital" by "population" and "rate of return" by "utility", the concentration of population in a particular geographical region may provoke large amounts of productive activity and a variety of goods and services, attracting still more people because of the greater utility to be had (Krugman [1991b]). Again, the degree of productive activity might react smoothly to

[4] Note that we do not a priori restrict $\{K(t)\}$ to be a continuous path, so that self-fulfilling "jumps" are, in principle, permitted.

a sudden influx of population (perhaps because the information regarding a larger market needs time to permeate to all the producers).

Say that an intertemporal equilibrium is *exclusively history dependent* if the long-run outcome either equals the initial allocation, or entails migration *only* to the sector that is initially profitable. Note that myopic tatonnement has the same properties, though obviously the exact path may be different. The main similarity is that no room is left for farsighted expectations.

Proposition 4.2. *Assume $f(K(0)) \neq 0$. Unless the cost of relocation exhibits congestion, every equilibrium must be exclusively history dependent, irrespective of the discount rate.*

This observation is independent of the magnitude of discounting, and of the degree of responsiveness of returns (as long as it is not instantaneous). Thus by a minor and reasonable weakening of one of the assumptions in the literature, we obtain a class of models where expectations are dwarfed by history, where initial conditions determine the final equilibrium. Of course, if rates of return adjust *instantaneously*, then expectations-driven equilibria are possible.

We reiterate: our claim is not that ahistorical equilibria are impossible. But that, in this class of models, in addition to the intersectoral agglomeration externalities, the "migration technology", is crucial to understand the sources of such equilibria. The only way in which such outcomes can occur is by introducing a cost to postponement; i.e., by making future relocation costs increase in the stock of settlers in our case.

Proof of Proposition 4.2. Consider the case in which $f(K(0)) < 0$. The case $f(K(0)) > 0$ can be settled by a parallel argument.

Fix any equilibrium γ. We claim that $K(t) \leq K(s)$ for all $t \geq s$, which settles exclusive history-dependence.

Suppose this is false. Then, indeed, there is some t and s with $t > s$, and exhibiting the following features:

a. $K(t) > K(s)$.

b. $r(\tau) < 0$ for all $\tau \in [s, t]$.

c. Some agent moves to sector B at date s.

d. The moving agent does not return to A until after date t.

Part (a) is simply the negation of our proposition. Part (b) follows from the fact that returns adjust continuously, and that $r(0) < 0$. Part (c) follows from the fact that K cannot begin to climb unless there is movement from A to B. And Part (d) follows from Part (b): there is no point in someone going to B and coming back while the returns there are strictly negative throughout.

Now consider our moving agent at date s and have her move at date t instead, following thereafter her original optimal strategy (by (d), this is possible). Then her deviation return, discounted by ρ to date s, is given by

$$e^{-\rho(t-s)}[V(\gamma, B, t) - c_B(K(t))], \tag{4.9}$$

while along the presumed optimum it is (once again discounted to s)

$$\begin{aligned} V(\gamma, A, s) &= \int_s^t r(\tau)d\tau + e^{-\rho t-s}V(\gamma, B, t) - c_B(K(s)) \\ &< e^{-\rho(t-s)}V(\gamma, B, t) - c_B(K(s)) \\ &\leq e^{-\rho(t-s)}[V(\gamma, B, t) - c_B(K(s))] \\ &\leq e^{-\rho(t-s)}[V(\gamma, B, t) - c_B(K(t))], \end{aligned} \tag{4.10}$$

where the first inequality follows from Part (b), the third from discounting, and the last from the no-congestion assumption.

But together, (4.9) and (4.10) yield a contradiction. ■

4.3 State Variables and Time

We now resurrect state variables but in a way that concentrates on the possibly long-term nature of the coordination problem (as in Adserà-Ray and in contrast to Morris-Shin). (The material here is based on Frankel and Pauzner [2001].) As in the previous sections, we assume that in every period there is a coordination game to be played. However, we will take the return in Sector B to depend, not just on the amount of K in that sector, but also on some exogenous state variable z_t. Thus we write $f(K, z)$ for the return in that sector, while the return in sector A is normalized to zero just as before. Without loss of any generality assume that f is increasing in z (the realization of the random variable z_t at date t).

The idea is that z_t is some random variable which is changing over time and is exogenous to the model at hand. For example, it could be the price of oil, which makes the industrial sector B relatively less or more attractive than the agricultural sector A.

The first main assumption on z is that it follows some nondegenerate stochastic process with no trend. So the increments to z all look the same, no matter what the starting level. The second assumption is a joint one on f and z: there are values of z (in its overall support, which you can think of as the entire real line) so that it pays for everyone to be in Sector B today no matter what happens in the future, and no matter how many people are in Sector B today. Likewise, there are values of z so low that the opposite is true: it is worth being in Sector A today no matter what happens in the future.

Like the previous model, this one is easier to do in continuous time. Let ρ be the discount factor of the agents. The former assumption can be formally expressed as

$$\mathbb{E}\left[\int_{t=0}^{\infty} e^{-\rho t} f(0, z_t) | z_0 = \bar{z}\right] \geq 0 \tag{4.11}$$

for some $\bar{z}$ in the support of the random variable. The latter assumption can be expressed as

$$\mathbb{E}\left[\int_{t=0}^{\infty} e^{-\rho t} f(\bar{K}, z_t) | z_0 = \underline{z}\right] \leq 0 \tag{4.12}$$

for some $\underline{z}$ in the support of the random variable.

Notice that this model *emphatically does not rule out* the multiple equilibrium problem. The assumptions embodied in (4.11) and (4.12) only rule out the problem for extreme and possibly very low-probability values that z_t *might* take on at some later date. For "most" intermediate values of z, there should be a genuine coordination problem as before: my behavior will depend, in principle, on what I anticipate others to do, now and in the future.

To describe behavior a bit more clearly, let us specify the model a bit further. We shall suppose that there is no cost of moving, but that everybody gets the chance to move at some point according to a Poisson process with independent arrivals. Thus each person is arbitrarily and randomly given the chance to move (she may not exercise this right, however). The remainder are stuck in whichever sector they may happen to be in that instant. All moves are made under the asumption of rational expectations and full intertemporal utility maximization, just as in the previous model.

Proposition 4.3. *In the infinite-horizon moving game, there is a function $\hat{k}(z)$ such that everyone who has a chance to move at any date, chooses sector B when $k > \hat{k}(z)$, chooses Sector A when $k < \hat{k}(z)$, and is indifferent when $k = \hat{k}(z)$. The specific nature of the function $\hat{k}$ will depend on the data of the problem.*

The proposition states that despite the potential scope for multiple equilibria, history once again fully pins down the outcome (but in a completely different way from the previous exercise). The extreme values that z can conceivably assume somehow serve to pin down behavior in the intermediate (more likely) zone.

Understanding this proposition teaches us something subtle about the role of beliefs in equilibrium models, so it is worthwhile to go through the details. To emphasize the nontriviality of this proposition, it may be worth considering — just for a moment — the case in which z does not change at all over time (see Figure 4.2). This sort of diagram will be repeated more than once so let us get used to it. On the horizontal axis are various values of z; in particular, the threshold values of $\bar{z}$ and $\underline{z}$ are clearly marked. On the vertical axis are different values of the capital stock in Sector B. The maximum such value is obviously $\bar{K}$; this is marked as well.

You are to interpret both these values as *current* values. [Of course, this injunction is irrelevant for z at the moment because its value is fixed, but it will become relevant later.]

The diagram also contains two downward-sloping lines. The first of these lines — call it line I — depicts the combinations of (K, z) values such that if Sector B starts with K, and then gains capital *whenever* people get a chance to move, the present discounted value of sector B is zero. Formally, this first line is the collection of all (K, z) pairs such that

$$\mathbb{E}\left[\int_{t=0}^{\infty} e^{-\rho t} f(K_t^1, z) | K_0 = K\right] = 0$$

where $\{K_t^1\}$ is the process in which people move to sector B whenever they have a chance to move, starting from K.

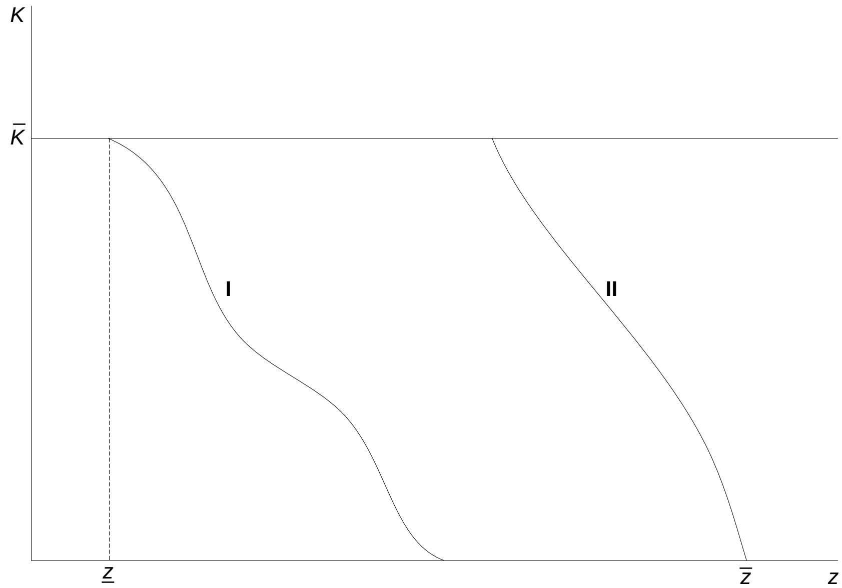


FIGURE 4.2. BOUNDARIES FOR MULTIPLE EQUILIBRIA WHEN z IS FIXED.

Notice that as z goes up, this makes returns better for Sector B, so that the threshold value of K needed to sstisfy the condition goes down. This is why the line is downward-sloping.

The second line — call it line II — describes the combinations of (K, z) values such that if Sector B starts with K, and then *loses* capital whenever people get a chance to move, the present discounted value of sector B is zero. Formally, this first line is the collection of all (K, z) pairs such that

$$\mathbb{E}\left[\int_{t=0}^{\infty} e^{-\rho t} f(K_t^2, z) | K_0 = K, z_0\right] = 0$$

where $\{K_t^2\}$ is the process in which people move to sector A whenever they have a chance to move, starting from K.

Line II is downward-sloping for the same reason that line I is. In addition, it should be obvious that line II must lie to the right of line I, as it imposes zero profits for Sector B on the basis of more pessimistic expectations about that sector.

Now notice that lines I and II mark out demarcations for unique and multiple equilibria. If the system finds itself to the right of line II, everybody *must* move to B when they get the chance. Likewise, if the system finds itself to the left of line I, everyone must move to A when they get the chance. But between lines I and II the situation is up for grabs. Because such a configuration is to the left of line II, it is possible to speculate that if everyone flees for Sector A, then a particular agent must too (when she gets the chance to move). This is self-fulfilling. Similarly, because the configuration is to the right of line I, one can sustain the opposite movement in this zone. Multiple equilibria are not ubiquitous (for extreme values of z there is only one equilibrium), but they haven't vanished, not by a long shot.

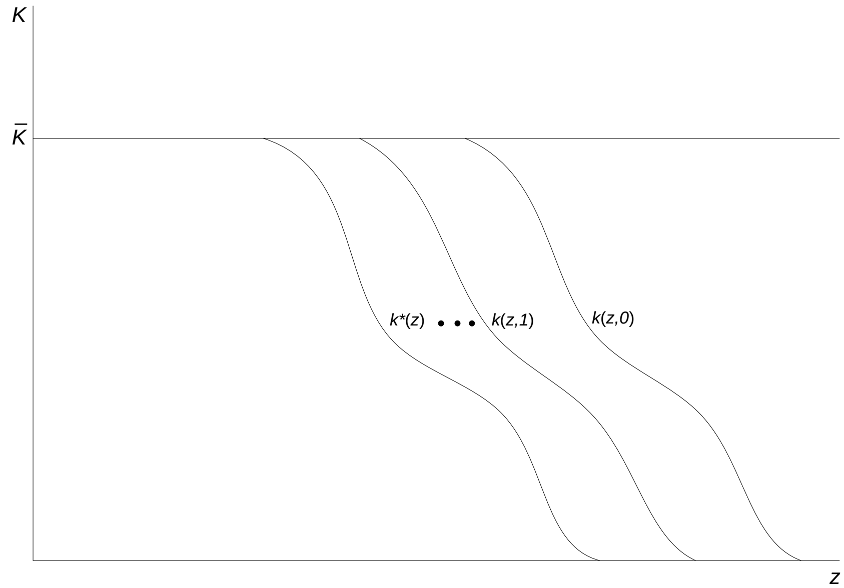


FIGURE 4.3. RECURSIVE CONSTRUCTION OF THE LOCUS $k^*(z)$.

So much for the case in which the value of z is exogenously given. Now let's suppose that z follows a Markov process as described above. Figure 4.3 starts this analysis by drawing the analogue of line II: the one with pessimistic expectations from the point of view of Sector B. I say "analogue" because the earlier description took z to be given; now we have to take expectations over changes in z as well. Formally, we look at all pairs of points (k, z) — forming the locus $k(z, 0)$ — such that

$$\mathbb{E}\left[\int_{t=0}^{\infty} e^{-\rho t} f(K_t, z_t) | K_0 = K, z_0 = z\right] = 0 \tag{4.13}$$

where K_t is the "pessimistic process" in which individuals move to A whenever they get the chance, and z_t is the exogenous stochastic process starting from $z_0 = z$.

The locus $k(z, 0)$ is the analogue of line II, and it is downward-sloping for exactly the same reason. But now matters get interesting just to the left of this locus. The pessimistic expectation: can it be validated? It can if z is fixed, but now it is possible that tomorrow, z might increase, taking the whole system into the area right of $k(z, 0)$. In that case we know that whoever gets a chance to move will *not* move to A. So expectations that are *fully pessimistic* are inconsistent with what we already know about the model, at least with some probability. Let us modify, then, our process K to say that it pessimistically moves down (as before) *except* under the above contingency. But now this makes Sector B a bit more attractive than it was when constructing the locus $k(z, 0)$. It follows that the *new* zero-payoff locus associated with these somewhat brighter expectations — call it $k(z, 1)$ — must be to the left of $k(z, 0)$ (see Figure 4.3).

This starts an iteration. Even $k(z, 1)$ must be too pessimistic, because the prospects of moving to B are brighter than those described by $k(z, 0)$. In general, define $k(z, n + 1)$ to be the locus

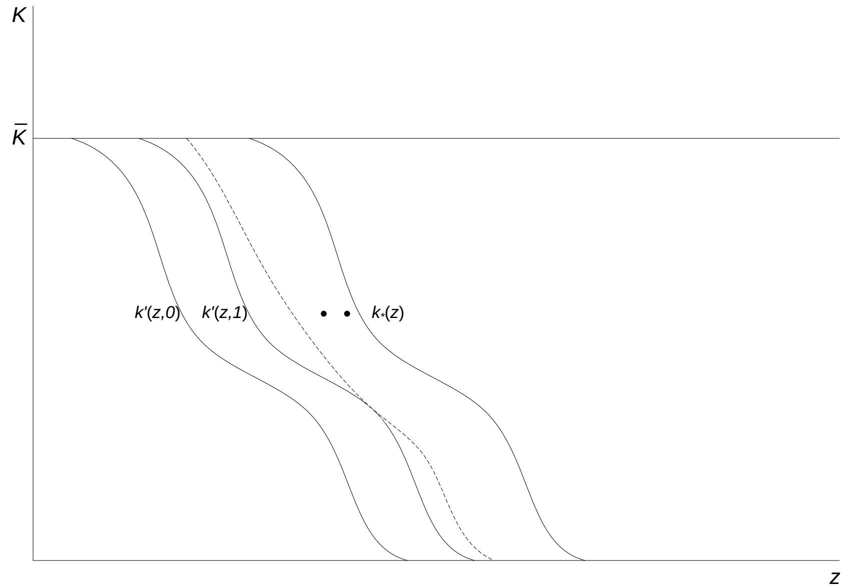


FIGURE 4.4. RECURSIVE CONSTRUCTION OF THE LOCUS $k_*(z)$.

of all (k, z) points which make people indifferent between Sectors A and B, *assuming that whenever the system is to the right of* $k(z, n)$ *people will move to B, and will move to A otherwise.* [Because z has a density, we don't have to worry about what happens if we land right on $k(z, n)$.] All these loci are downward-sloping, and they converge to some limit which we shall call $k^*(z)$. This limit locus has the following properties:

[1] If the system is to the right of this locus, everyone who has a chance to move at that point *must* move to Sector B.

[2] If the system is to the left of this locus, matters are ambiguous. Certainly, if everyone is *expected* to go to A whenever the system is to the left of $k^*(z)$, then they *will* do so. But there is no ruling out multiple equilibria yet. It may be that points to the left of $k^*(z)$ are *also* consistent with a move to B.

So we need to continue the analysis, and to do so we start an iteration from the left this time. The iteration is not going to be entirely symmetric, so pay attention to what follows.

First, we translate the locus $k(z, 0)$ way over to the left. Figure 4.4 illustrates. By our assumption on extreme values, I can find, in fact, a translate that is so far over that for every (k, z) on that (translated) locus,

$$\mathbb{E}\left[\int_{t=0}^{\infty} e^{-\rho t} f(K_t, z_t) | K_0 = K, z_0 = z\right] \leq 0 \tag{4.14}$$

where K_t is now the "optimistic process" in which individuals move to B whenever they get the chance, and z_t is the exogenous stochastic process starting from $z_0 = z$. That is, these values of z are so bad for sector B that despite the harbouring of optimistic expectations about future moves to Sector B, it is best to stay in Sector A. In this way, we have identified a *sufficient* condition for moves to Sector A to always constitute an equilibrium: that (k, z) be to the left of this translated locus; call it $k'(z, 0)$.

But now, once again, we know that these optimistic predictions are too optimistic. People won't always move to B; in fact, as we have just seen, if the state moves into the extreme left zone that we have just identified, they will move to A (or stay there, as the case may be). Thus we can define a fresh collection of (k, z) with the property that starting from each of these points and entertaining the optimistic expectation *except* when the state wanders to the left of $k'(z, 0)$, state A is weakly better. Figure 4.4 shows the boundary of this set as a dotted line to the right of $k'(z, 0)$.

Now iterate as follows: translate the locus $k'(z, 0)$ — to the right this time — as far as possible, subject to the condition that no part of it wander out of the dotted line (see Figure 4.4 to follow this fully). This translation we will call $k'(z, 1)$. More generally, continue the iteration as follows: given $k'(z, n)$, consider the collection of all (k, z) points which make people weakly prefer Sector A to B, *assuming that whenever the system is to the right of* $k'(z, n)$ *people will move to B, and will move to A otherwise*. Having done so, define $k'(z, n + 1)$ to be the furthest translate of $k'(z, n)$ subject to the constraint that it fully lies within the set of points defined in the previous sentence. Let $k_*(z)$ denote the limit of this recursion.

Appreciate again that this iteration is different from the first one. It has a similar flavor, but we only restrict ourselves to translates. Nevertheless, we can certainly say that for all points to the left of $k_*(z)$, it is a dominant strategy to move to sector A (after all, we have been — if anything — "conservative" towards A when performing this iteration).

Now we need to look at points *on* the locus a bit more carefully. Notice that the first limit — $k^*(z)$ — certainly has the property that *on* $k^*(z)$, one is indifferent between A and B , provided we assume everyone moves to B on the right of this locus, and towards A on the left of this locus. But because we have thrown away some points in doing the second iteration, we cannot say the same of $k_*(z)$. All we can say is that if we assume everyone moves to B on the right of $k_*(z)$, and towards A on the left of $k_*(z)$, then *on* the locus people weakly prefer A to B.

But we can go one step further: *there must be some point on* $k_*(z)$ *at which individuals must be indifferent between A and B*, provided we assume everyone moves to B on the right of this locus, and towards A on the left of this locus. The reason is the nature of our iteration: at each stage we chose the *maximal* translation to the right that is possible. If strict preference for A held at all points on $k_*(z)$, the iteration could not have ended at that stage.[5]

Call this point of indifference α. Now go back to the locus $k^*(z)$ and pick its "twin point" β (see Figure 4.5). That there is a twin point, at eactly the same vertical height, follows from the fact that the loci $k^*(z)$ and $k_*(z)$ are translates of each other.

Our final claim is that α and β must coincide. To see this, recap the following two points:

[5] This is a crucial technical point in the proof and establishing it formally needs a bit of care, though the intuition is very clear.

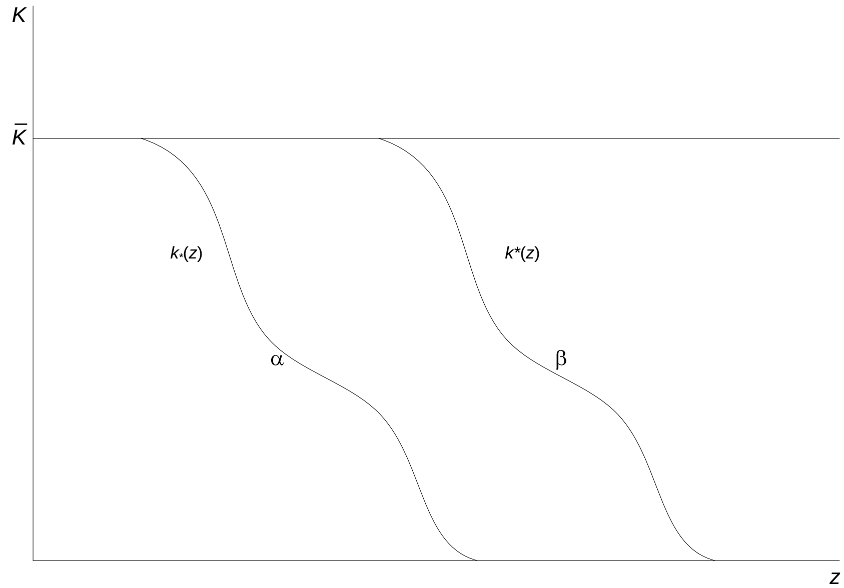


FIGURE 4.5. SHOWING THAT THE LOCI $k^*(z)$ AND $k^*(z)$ MUST COINCIDE.

[i] α is a (k, z) combination at which all movers are indifferent between staying at A and B, assuming that everybody (who can) will choose A to the left of the locus $k_*(z)$ and everybody will choose B to the right of that locus.

[ii] β is a (k, z) combination at which all movers are indifferent between staying at A and B, assuming that everybody (who can) will choose A to the left of the locus $k^*(z)$ and everybody will choose B to the right of that locus.

Now recall that z has no trend. But then statements [i] and [ii] must be *incompatible* if we insist that α and β are distinct! To see this, simply couple together the events at α and β, and thereafter. For every path starting from α, there is a twin path starting from β, forever separated by the same distance in z. And the two share exactly the same probability measure! But the expectation of returns over the first set of paths is zero. Therefore the expectation of returns over the second set of paths must be positive, because f is increasing in z. This proves that α and β cannot be distinct.

But then the *entire* functions $k^*(z)$ and $k_*(z)$ must coincide! Call this common function $\hat{k}(z)$. We have proved the proposition. It must be that to the left of $\hat{k}(z)$, everyone moves to A, while to its right everyone moves to B. Multiple equilibria have been eliminated.

CHAPTER 5

History-Dependence: An Introduction

5.1 An Overview

To summarize what we have studied so far: we've looked at multiple equilibrium models in which there are "good" and "bad" equilibria, often Pareto-ranked. The presence of such multiplicities has often been put forward as an explanation for underdevelopment. Ot at least, even if it cannot serve as an *explanation*, it does succeed in pointing out why entirely different levels of economic outcomes are consistent with the same underlying fundamentals.

As we have seen, the potential difficulty with the multiple-equilibrium explanation is that we have no theory of how one equilibrium or the other comes about. Indeed, if we run such models in real time, there is no particular reason why the economic system cannot "jump" from one equilibrium to another, such jumps helped along by acts of coordination among the population. At the same time we know that such jumps make no intuitive sense. An economy locked into many years of one bad equilibrium is not likely to change overnight (or at any rate, overnight on any one particular night). We have looked at three different theoretical frameworks that attempt to come to grips with this. The first takes seriously the common-knowledge presumptions that underlie these models and shows that a departure from such presumptions can often pin down unique equilibria, typically mediated by some public signal. The second approach studies lagged external effects. The third approach returns to public signals in a dynamic context to precipitate a unique equilibrium.

What's nice about these models is that they will typically generate rapid transitions from one equilibrium to another. What these models say that we will observe are long periods spent in a particular equilibrium, and then a series of events that suddenly hurl the economy into an entirely different configuration. Many social and economic transformations do have this "logistic" feature.

In what follows, we take a very different — and complementary — path. Now the state variable will get center stage and we will push multiple equilibrium into the background. Indeed, we will consider models in which at every date (and for each going value of the state variable), there will be a *unique* outcome. But the dynamics of the system will allow for *multiple steady states*. Multiple steady states are therefore not to be confused with multiple equilibrium. There will typically be just one for every initial condition. But different initial conditions will often map into different steady states.

There are many different state variables that we can study. Each of them is important enough to have a book written on them. First begin with economics:

1. The capital stock is, of course, the most common example of a "state" in development economics. It is the central variable of growth theory. Section 5.3 below tells you a little bit about the Romer model, in which the initial capital stock can have long-term effects on economic growth (unlike in the convergence models).

2. Infrastructure represents a special case of item [1], but an extremely important special case that has yet to be studied carefully in the context of history-dependence.

3. Likewise, it would be of great use to have economic theories of culture and corruption, which then feed back into economic progress (or the lack thereof). Not that models of economics and culture don't exist, but once again there is not that much that fits the history-dependence framework that I emphasize here.

4. Poor societies have legal systems that are limited, if not in terms of laws on the books, then in terms of actual enforcement of those laws. The same is true of contract enforcement. Limited enforcement of contracts and laws then feed back on the economic system.

5. Economic inequality: this will be the one topic that we discuss in great detail below. But stay tuned for future editions in which I plan to take up all the other themes one by one.

Then there is politics:

1. Elites that deny power and therefore cripple subsequent growth.

2. Historical inequalities in (say) the distribution of land that distorts and limits political participation to demands for redistribution rather than for growth-enhacing investments.

3. the poverty-conflict-poverty cycle, about which we shall have something to say in these notes.

5.2 Introductory Notes on Economic Inequality

We now turn to a particular — but important — state variable: the disparity in the historical distribution of assets, or *economic inequality*. To be sure, inequality is a concept that is of intrinsic interest, and there is a large literature that has attempted to axiomatize measures of inequality so that one can keep track of changing income or wealth distributions from a purely ethical perspective. But there is also the functional aspect of inequality: the *interaction* between inequality and other variables of significance in the economy, such as aggregate output, efficiency, unemployment, and so on. These interactions may take place "within the period", so that they are analyzable using a static model, or they be dynamic, in which case we must use intertemporal models that keep track — among other things — of accumulation decisions.

Why does inequality matter (not intrinsically, which is a question of ethics, but functionally)? There are several answers to this question:

[1] *Inequality matters because it affects productive potential.* The cleanest parable of this has come to be known as the nutrition-based efficiency wage hypothesis. In poor societies, wages affect food consumption, and food consumption affects the ability to work. This will lead to an equilibrium in which the distribution of asset ownership will have economy-wide effects.

[2] *Inequality matters because it affects the ability of individuals to gain access to productive resources.* As long as the distribution of productive ideas or new projects is loosely correlated with the distribution of wealth, there will be people who have ideas (and abilities), but do not have access to the wealth needed to put those ideas into real outcomes. This sort of argument hinges on a missing or imperfect market for capital.

[3] *Inequality matters because it affects incentives.* There are various aspects of this problem and we shall address some of them. For instance, what is the connection between an egalitarian society and incentives? What is the connection between egalitarianism and the extent of collective action?

[4] *Inequality matters for "political-economy" reasons.* Under this catch-all are a number of effects that may be discussed. Inequality may affect voting behavior, leading to different degrees of redistribution and consequently affecting the incentives to accumulate capital. Inequality may create social conflict and in this way be related to low output. Finally, inequality (within an organization) may create efficiency losses, because individuals at different wealth levels may effectively have different objectives.

[5] *Inequality matters because it prevents correct resource allocation on the part of the government.* That is, high inequality may affect the ability of individuals to transmit information accurately to a policy-maker.

We begin with dynamic models that track the evolution of inequality over generations. The emphasis will be on items [1] and [2] discussed above.

5.3 A Special Note on Capital Stocks

We now begin our transition to a more serious assessment of initial conditions. Unlike the more abstract role for history postulated in repeated coordination games, these considerations arise from full-fledged dynamic models in which state variables play a fundamental role.

I provide a preliminary illustration of this using a variant of a growth model due to Romer [1986]. In this model, multiplicity of equilibrium paths can exist from the same initial condition, but the initial conditions play a fundamental role as well.

Suppose that output is produced at each date t according to the equation

$$y_t = F\left(k_t, \hat{k}_t\right), \tag{5.1}$$

where $\hat{k}$ is the average economy-wide stock of capital, introduced to express the social externalities from capital (or the knowledge embodied in capital). The idea is that greater capital in the society at large has positive spillovers for any one producer. In the sequel we will take $k_t = \hat{k}_t$ for all t, but no one individual internalizes this connection in his decision-making.

Produced output is divided among consumption and investment as follows:

$$y_t = c_t + a_t k_t, \tag{5.2}$$

where c represents consumption and ak represents the amount of net investment in the production of new capital. The fact that we express new investment as a fraction a of the existing capital stock is simply a matter of convenience.

Investment creates growth in the capital stock. We postulate that

$$\frac{k_{t+1} - k_t}{k_t} = G(a_t), \tag{5.3}$$

so that the rate of growth of the capital stock depends on the "intensity" of investment: on the investment-capital ratio a.

The objective function takes the familiar form:

$$\max \sum_{t=0}^{\infty} \beta^t u(c_t), \tag{5.4}$$

where β, the discount factor lies between 0 and 1, and u is a one-period utility function.

The following assumptions will be in force:

(U.1) The utility function u is increasing, continuous and strictly concave on $\mathbb{R}_+$.

(G.1) The function $G(a)$ is continuous, increasing and concave, with $G(0) = 0$ and $\sup G(a) \equiv B < \infty$.

(F.1) The production function F is continuous, nondecreasing, and increasing and concave in its first argument. However, when private and social capital move *together*, we assume that the (private) marginal product of capital is nondecreasing; that is, $F_1(k, k)$ is nondecreasing in k.

Also restrict the discount factor β to ensure that the infinite sum of utilities is always well-defined. I omit the details here.

Fix some initial stock k and a path of "social capital" stocks $\hat{\mathbf{z}} \equiv \{\hat{k}_t\}$, with $\hat{k}_0 = k$. Notice that a program (for some individual) can simply be identified with the (private) sequence $\mathbf{z} \equiv \{k_t\}$ (everything else can be recovered from it).Say that such a program $\mathbf{z}$ is *optimal from k*

under $\hat{\mathbf{z}}$ if it maximizes utility (see (5.4)) in the class of all feasible programs. Finally, $\mathbf{z}$ is an *equilibrium program* if it is *optimal from k under* $\mathbf{z}$ itself.

Using fixed point arguments, it is possible to prove that an equilibrium exists. But of course, there may be many of them, and this comes from exactly the same considerations that we have seen earlier in these notes. The fact that an increase in social capital increases private marginal product means that a complementarity is at work here — higher investments in the economy as a whole may be self-fulfilling. The point of this section, however, is not to emphasize the multiplicity but to see how an initial state (proxied in this instance by the capital stock) — or history — can determine the set of equilibrium paths.

To this end, define an equilibrium program from k to be *stationary* if $k_t = k$ for all t. We may now state

Proposition 5.1. *There exists an equilibrium program from k which is stationary if and only if*

$$F_1(k,k)G'(0) \leq \beta^{-1} - 1. \tag{5.5}$$

Proof. *Necessity.* Here is an outline which can easily be made fully rigorous. The idea is that if an equilibrium program is stationary, then a tiny sacrifice of consumption today in return for a larger stationary stream of output in the future is not worth it. Let's suppose that the tiny sacrifice of consumption is ϵ. Then the current loss is *approximately* $u'(c)\epsilon$. What is the future gain? The extra capital stock produced is *approximately* $G'(0)\epsilon$, and consequently the extra stream of output available from tomorrow is *approximately* $F_1(k,k)G'(0)\epsilon$. If this is all consumed, then the extra stream of utility from tomorrow is *approximately* $u'(c)F_1(k,k)G'(0)\epsilon$. Calculate the present value of this extra stream of gains. Our program is an equilibrium (and therefore an optimum from k under itself), so this deviation should not be worth it. This yields (5.5).

Sufficiency. Suppose that expectations of social capital are stationary at $\hat{k}_t = k$ for all t. We want to show that the stationary program is optimal under these expectations, starting from k. let c be the consumption under the stationary program and let $\{k_t\}$ be any other program (feasible from k under stationary expectations) with associated consumption sequence $\{c_t\}$.

Then for any date T,

$$\begin{aligned}
\sum_{t=0}^{T}\beta^t[u(c)-u(c_t)] &\geq u'(c)\sum_{t=0}^{T}\beta^t[c-c_t] \\
&= u'(c)\sum_{t=0}^{T}\beta^t[F(k,k)-F(k_t,k)+a_tk_t] \\
&\geq u'(c)\sum_{t=0}^{T}\beta^t[F_1(k,k)(k-k_t)+a_tk_t] \\
&\geq u'(c)\sum_{t=0}^{T}\beta^t[F_1(k,k)(k-k_t)+\frac{1}{G'(0)}(k_{t+1}-k_t)] \\
&\geq u'(c)\sum_{t=0}^{T}\beta^t[\frac{1-\beta}{\beta G'(0)}(k-k_t)+\frac{1}{G'(0)}(k_{t+1}-k_t)] \\
&= \frac{u'(c)}{G'(0)}\sum_{t=0}^{T}\{(\beta^{t-1}-\beta^t)(k-k_t)+\beta^t(k_{t+1}-k_t)\} \\
&= \frac{u'(c)}{G'(0)}\{(\beta^{-1}-\beta^T)k-\sum_{t=0}^{T}(\beta^{t-1}k_t-\beta^tk_{t+1})\} \\
&= \frac{u'(c)}{G'(0)}\{\beta^T(k_{T+1}-k)\}\geq 0,
\end{aligned}$$

which completes the proof. ∎

Proposition 5.1 shows us that stationary equilibria exist under certain conditions, and moreover, that they *cannot* exist unless those conditions are met. Notice that the condition (5.5) combines technological parameters, the discount factor, *and the initial capital stock* in a specific way. If you combine this restriction with our assumption that $F_1(k,k)\geq 0$, we see that at certain low levels of the capital stock, it is possible to stagnate forever.

The role of the initial capital stock is made even sharper in the next proposition, which tells us that if (5.5) is not met — that is, if the initial capital stock is large enough — then equilibrium programs must exhibit sustained growth.

Proposition 5.2. *If*

$$F_1(k,k)G'(0) > \beta^{-1}-1, \tag{5.6}$$

then every equilibrium program must exhibit $k_t\to\infty$.

Proof. Notice that $\{k_t\}$ must be a nondecreasing sequence. Suppose, contrary to our claim, that $k_t\to\bar{k}<\infty$ as $t\to\infty$. Consider the sequence of programs $\{k_t^n\}$ given by $k_t^n=k(t+n)$ for each n and t. Note that the nth sequence has as its starting stock k_n, and that for each n, the program is an equilibrium from k_n. By a standard maximum-theorem-style argument that we do not include here, the pointwise limit of these programs also represents an equilibrium. But the limit is the stationary sequence $\{\bar{k}_t\}$ given by $\bar{k}_t = \bar{k}$ for all t. By Proposition 5.1 ,

$$F_1(\bar{k}, \bar{k})G'(0) \leq \beta^{-1} - 1.$$

But $\bar{k} \geq k$, the original initial stock, and given that k fails (9) and $F_1(k, k)$ is nondecreasing, we have a contradiction. ■

Thus in the Romer model, if the initial capital stock is large enough, equilibrium programs *must* grow. No self-fulfilling expectational argument can create stagnation. However, if the initial capital stock is low, an expectations trap is possible. In this way history and expectations interact.

In case you have not noticed this already, notice that Propositions 5.1 and 5.2 are different in one significant respect. Proposition 5.2 rules out stagnation, while Proposition 5.1 says that stagnation is one possible outcome, but does not rule out growth. So there is a role here for expectations, despite the influence of history. With more work, one can show that there is an even lower (positive) value of the capital stock from which *no* growth can ever occur, and in the remaining "intermediate" zone there are multiple equilibria involving both stagnation (as we've already seen) and sustained growth, depending on expectations.

CHAPTER 6

The Dynamics of Inequality

6.1 A Introductory Framework

Every dynamic model of growth and distribution will exhibit some form of the following equations at the level of the individual unit (person, household, region, country). Think of time as being divided into discrete periods and adopt the view that each period is the length of an entire generation. So individuals indexed by different dates belong to different generations. The first equation is just budget balance:

$$y_t = c_t + x_t, \tag{6.1}$$

where y is the lifetime income or wealth of an individual in generation t, c is her consumption and x is the bequest she leaves to the next generation.

The next equation describes how bequests translate into lifetime income for the next generation:

$$y_{t+1} = f(x_t), \tag{6.2}$$

where f is some production function. Such production functions have different interpretations depending on the context:

1. It may just be a standard production function as used in the theory of growth, say of the Cobb-Douglas form.

2. If the individual is part of a competitive economy everyone could earn a wage rate w and then get a return of $1 + r$ on bequests, so that

$$f(x) = w + (1 + r)x.$$

3. The "production function" may simply represent returns to different "occupations". For instance, suppose that it costs nothing to keep your child unskilled but costs $\bar{x}$ to turn her into a skilled laborer; then

$$\begin{aligned} f(x) &= \underline{w} \text{ for } x < \bar{x} \\ &= \bar{w} \text{ for } x > \bar{x}. \end{aligned}$$

in this case the production function will be exogenous to an individual dynasty but may well be endogenous from the point of view of the economy as a whole (because the wages for skilled and unskilled labor will be determined in equilibrium). These represent the sort of interactive models that will occupy a lot of our attention.

The last ingredient has to do with preferences. There are four notions that appear in the literature.

WG. *Warm Glow*. Generation t has a utility function given by $U(c_t, x_t)$. See, e.g., Banerjee and Newman (1993).

CB. *Consumption-Based*. Generation t has a utility function given by $U(c_t, c_{t+1})$. See, e.g., Arrow (1973), Bernheim and Ray (1986).

IB. *Income-Based*. Generation t has a utility function given by $U(c_t, y_{t+1})$. See, e.g., Becker and Tomes (1979, 1981).

NP. *NonPaternalistic*. Generation t has a utility function given by $U(c_t, V_{t+1})$, where V_{t+1} is the lifetime utility of generation $t + 1$. See, e.g., Barro (1978) and Loury (1981).

Notice that WG starts to look really problematic in those cases in which the production function is endogenous, as in the case of occupational choice. It is hard to cling to some exogenous utility function that is based on bequests when what those bequests are accomplishing is endogenous to the system. Thus, while WG and IP are equivalent when f is exogenous, WG is not really something we will be looking at seriously.

What about CB versus IB? I prefer the latter for two reasons.

First, I should be getting a utility out of my child's "capability", measured here by lifetime income or wealth. Just *what* she does with it — how much she consumes, how much she gives away — should not really be my concern. That said, we might sometimes see paternalistic parents who care more about the child's consumption rather than the resources that she has access to. I'm not saying this is a logical rebuttal of CB.

Second, a model based on CB is typically hard. There are all sorts of game-theoretic subtleties involved here (see Kohlberg (1976) and Bernheim and Ray (1986)). It does not seem sensible to bring in all those subtleties and still try to address the different questions of inequality and development that we are more interested in here.

That said, however, why not go the whole hog and choose NP over IB? After all, NB is *completely* nonpaternalistic: parents only care about the utility of their children: not the bequests they leave them, or how much they consume, or indeed how much they earn. There is some sense in this view and we will try and incorporate it in what follows. But it has to be realized that the value-function approach can be conceptually problematic. It is well known that NP is equivalent to the maximization of a utility function defined on an infinite generational stream of consumptions. Do we really internalize *that* much? Do parents really calculate future utilities, or do they use the "capability measure" — income/wealth — as a convenienet shorthand?

There is another reason — but more on this later — why NP may not be the correct specification.

In any case, even at this level of abstraction, we can glean a few general principles. The most important of these — at least for the aggregative models that we consider — is the principle of *monotonicity*. The easiest way to see this is to suppose that a generation has the utility function

$$\mathbb{E}_\alpha U(c, \Psi(k, \alpha)),$$

where c is its consumption, k is its investment, α is some random shock, and $\Psi(k, \alpha)$ is some mapping. If it is the identity mapping, then $\Psi(k, \alpha) = k$ and this is like warm-glow. But Ψ could also represent next-period's consumption, or income, or utility. For now let us not worry about the endogeneity of these mappings; we will do that when we get to the equilibrium analysis.

Assume that

[U] The utility function U is differentiable and strictly concave in c, and exhibits complementarities between c and Ψ: $U_{12}(c, \Psi) \geq 0$.

Renark. Obviously U is not an ordinal property and all of the above can be written as a suitable supermodularity condition. However, condition [U] is fine enough for what we want to do.

Proposition 6.1. *Assume* [U] *and suppose that Ψ is nondecreasing in k. Let h be the policy correspondence that describes the optimal choice of k for each y, subject to the constraint that $c = y - k$. Then if $y > y'$, $k \in h(y)$, and $k' \in h(y')$, it must be the case that $k \geq k'$.*

Proof. Suppose this assertion is false for some (y, y', k, k') as described in the statement of the proposition. Then $k' > k$. Notice that k' is feasible for y (because it is feasible under y', which is smaller), while k is feasible under y' (because k', which is bigger, is feasible under y'). It follows from optimality that

$$\mathbb{E}U(y - k, \Psi(k, \alpha)) \geq \mathbb{E}U(y - k', \Psi(k', \alpha)),$$

while

$$\mathbb{E}U(y' - k', \Psi(k', \alpha)) \geq \mathbb{E}U(y' - k, \Psi(k, \alpha)).$$

Adding these two inequalities and transposing terms, we see that

$$\mathbb{E}U(y - k, \Psi(k, \alpha)) - \mathbb{E}U(y' - k, \Psi(k, \alpha)) \geq \mathbb{E}U(y - k', \Psi(k', \alpha)) - \mathbb{E}U(y' - k', \Psi(k', \alpha)).$$

Now use the assumption that Ψ is nondecreasing in k, plus the complementarities condition, to conclude that

$$\mathbb{E}U(y - k', \Psi(k', \alpha)) - \mathbb{E}U(y' - k', \Psi(k', \alpha)) \geq \mathbb{E}U(y - k', \Psi(k, \alpha)) - \mathbb{E}U(y' - k', \Psi(k, \alpha)).$$

Combine these last two inequalities to obtain that

$$\mathbb{E}U(y - k, \Psi(k, \alpha)) - \mathbb{E}U(y' - k, \Psi(k, \alpha)) \geq \mathbb{E}U(y - k', \Psi(k, \alpha)) - \mathbb{E}U(y' - k', \Psi(k, \alpha)),$$

and now draw yourself a diagram to see that the strict concavity of U in c is violated by this last inequality. ∎

This proposition yields the following corollary when there is no uncertainty:

Proposition 6.2. *Suppose that there is no uncertainty, that the production function is increasing in k, and that the optimization problem above is solved repeatedly by successive generations starting from some initial y. Then the resulting paths of y_t and k_t must be monotonic in time.*

Proof. Suppose not. Then two cases are possible.

Case 1. There exist dates t and s with $s \geq t+1$ such that $y_t < y_{t+1}$, $y_{t+1} = \cdots = y_s$, and $y_{s+1} < y_s$.

Case 2. There exist dates t and s with $s \geq t+1$ such that $y_t > y_{t+1}$, $y_{t+1} = \cdots = y_s$, and $y_{s+1} > y_s$.

It is easy to see that one of these two cases *must* occur if the proposition is false. We now obtain a contradiction in Case 1; the same argument holds for Case 2.

In Case 1, we have $y_{t+1} > y_{s+1}$, it follows from f increasing that $k_t > k_s$. However $y_t < y_s$. This contradicts Proposition 6.1. ■

6.2 Inequality and Capital Markets: Noninteraction

There are two approaches to study of evolving inequality. One is to look at ongoing random shocks. The other is to argue that there are intrinsic market forces which tend to separate individuals. The first approach typically consists in looking at the ergodic behavior of a *single* dynasty, and then equating this to the cross-sectional distribution of the system as a whole. This is the viewpoint epitomized in several papers: we discuss a variant of Loury [1981] in these notes.

On the other hand, it should be obvious that the second type of analysis — inequality arising from market-generated separations across individuals — cannot be conducted by simply studying the intertemporal behavior of a single agent. *Interactions* are fundamental. This is the topic of the next section. Here we study the single-dynasty model.

There is one agent alive at every date, and each agent has a single child, who becomes an adult in the next "period" (or generation). An agent at date t has access to total resources y_t, which we may sometimes loosely call income, but is really to be interpreted as the sum of income and any starting wealth. The agent divides this into consumption (c_t) and bequests for the next generation (k_t):

$$y_t = c_t + k_t.$$

Bequests create starting wealth y_{t+1} for the next generation. More than one interpretation is possible: these may be financial bequests, or upfront educational investments (the preferred interpretation by Loury). In any case, y_{t+1} is not fully pinned down by k_t. A random shock $\alpha_{t+1} \in [0, 1]$ is also assumed to play a role. One might call this the *ability* of the agent at date $t+1$. Thus we have

$$y_{t+1} = f(k_t, \alpha_{t+1})$$

where it is obviously reasonable to assume that f is increasing and smooth in the first argument and that we simply label abilities so that f is also increasing (and continuous) in the second argument. We assume that abilities are iid on $[0, 1]$, with a continuous density that is strictly positive on $(0, 1)$.

We will make further assumptions on f, but at present we only need to make one: that even for the highest ability, the production function ultimately "flattens out": $\lim_{k\to\infty} f'(k, 1) < 1$ (where f' is the derivative with respect to the first argument). This can be justified at a couple of levels, but the main reason we do this is convenience: we want everything to be bounded so that we can analyze it simply. [Why does this assumption guarantee that income and capital must be bounded? Draw a diagram of the "best-case" production function under this assumption and satisfy yourself.]

Note that — unlike Loury — we do *not* assume that f is concave. This allows us to put other models in perspective (more on this below).

Now turn to preferences. We shall assume that each generation is a dynastic utility maximizer; that is, it seeks to maximize

$$u(c) + \delta \mathbb{E} V,$$

where u is some smooth strictly concave utility function defined on current consumption, $\delta \in (0, 1)$ is a discount factor, and V is the expected lifetime utility to be experienced by the child. The expectation is taken before the V because it is assumed that the parent does not know how the child's ability will turn around at the time of making the educational bequest.

That is, given some y, a parent tries to maximize, choosing k,

$$u(y - k) + \delta \mathbb{E}_\alpha V\left(f(k, \alpha)\right) \tag{6.3}$$

subject to the constraint that $0 \leq k \leq y$. Notice that the function V is fundamentally endogenous and that it must solve the well-known functional equation:

$$V(y) = \max_{0 \leq k \leq y} \left[u(y - k) + \delta \mathbb{E}_\alpha V\left(f(k, \alpha)\right)\right].$$

That such a consistent V exists can be proved using standard arguments. which we omit here. It can also be shown that V is continuous in y.[1]

The beauty of this sort of argument is that it essentially reduces an infinite-dimensional problem to something very simple: the one-variable maximization problem expressed in equation (6.3). From this problem we get what is called an (optimal) *policy correspondence* $\Gamma(y)$. This is the set of all k's which solve (6.3), given the starting value y. Notice that we have made no assumptions on the curvature of f so in principle this correspondence could have arbitrary shape. It turns out that the following monotonicity argument must be true:

Proposition 6.3. *Suppose that $y \leq y'$ and that $k \in \Gamma(y)$ and $k' \in \Gamma(y')$. Then it must be that $k' \geq k$.*

Proof. Verify that all the conditions of Proposition 6.1 are satisfied. ■

[1]The argument used to establish the existence, uniqueness and continuity of V is the famous Banach contraction mapping theorem. See Stokey and Lucas [1989] for how this works with value functions such as those in this model.

Recall from Proposition 6.2 that this would immediately yield a convergence argument if there is no uncertainty. Of course, the current model does have uncertainty so this simple sequence argument cannot be applied. [Try it and see where it fails.] But nevertheless, the proposition continues to be helpful in this and in other cases, as we shall see in a bit.

Let's return to the convergence argument in the face of uncertainty. Is there some way in which we can approach this problem? Here I follow Loury in making the following additional assumptions on the production function:

Poor Geniuses Exist. $f(0, 1) > 0$.

Rich Fools Exist. $f(k, 0) < k$ for all $k > 0$.

I've labeled the assumptions so that they are self-explanatory. Just in case: the first assumption states that even if you don't invest anything into a high-ability child, she will end up making *some* money. The second assumption states that low-ability children drain resources: whatever you put in, less comes out.

Under these two assumptions, the following important proposition is true:

Proposition 6.4. *Let μ_0 be an initial probability measure on the set of initial incomes of generation zero, and denote by μ_t the distribution of income for generation t induced by the equilibrium behavior described above. Then there exists a unique measure μ^* such that μ_t converges to μ^* as $t \to \infty$, and this measure μ^* is independent of μ_0.*

It is important to understand why this result is true, as it leads to a significant insight into how uncertainty works in these models. To this end, consult Figure 6.1. He we show diagrammatically how income evolves from generation to generation within a dynasty, starting with income y_0 at date 0. An investment of k_0 is made, and this leads to an uncertain income at the next generation (the little hill on the $t = 1$ line marks the density of such incomes). A typical income realization is the level y_1, at which the investment k_1 is made, and the whole process repeats itself. The density conditional on an investment k_1 is shown by the little hill on the $t = 1$ line, while the thick density with larger support shows *all* possible values of y_2 two periods hence (that is, conditioning on y_0 but *not* on the specific realization of y_1).

Now, in this diagram I have marked out a special interval of incomes I which has the property that *no matter where you start from*, there will be a (common) date T such that these densities (conditional on the starting point) "envelop" the interval I (with probability uniformly positive. This fact follows from the two assumptions we've made above regarding "poor fools" and "rich geniuses" (though, as we shall see in a bit, the assumptions are *not* necessary). The point is that with positive probability and at some uniformly chosen date, the system will wander into I no matter what the initial conditions were.

A formal proof of this assertion is provided in Section 6.2.2. But intuitively, what this means is that the system must lose its memory, its history, at some point of time (for whatever happens with positive probability must happen for sure). That suggests that limit behavior is independent of history, which indeed is the broad substance of Proposition 6.4.

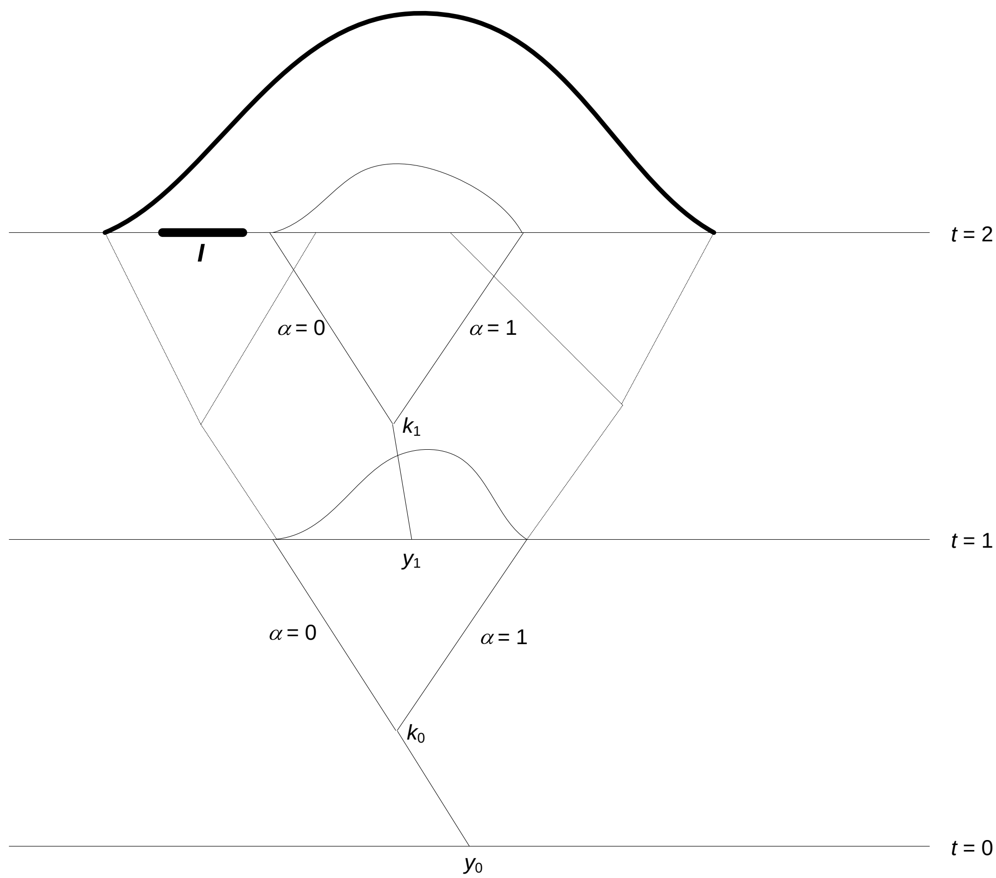


FIGURE 6.1. A DYNASTIC INCOME PATH

If you want to know more about the technicalities of this ergodic theorem, look at the general result (Proposition 6.5) given in Section 6.2.2. Essentialy, the ability of the process to "communicate" or "mix" no matter what the initial conditions were, is responsible for the history-independent limit behavior of the system.

Another way to graphically examine the mixing property is to look at the policy functions generated as a result of dynastic optimization. While these map from current y to the current choice of k, we shall do this slightly different: the result is shown in Figure 6.2. The diagram places current income on the horizontal axis, and tomorrow's income on the vertical axis. Of course, tomorrow's income is uncertain, and may therefore be represented by a *band* of possible incomes, which boils down to an interval of incomes for each value of today's income.

Now, in part (a) of this figure, notice that there are two clear zones. In Zone I, income *must* be bounded above by the value Y_I, which is the biggest intersection (in this region) of the stochastic correspondence with the 45^0 line. By monotonicity of the policy correspondence, it follows that if $y \leq Y_I$, y can *never* cross the value Y_I, even under the most optimistic conjectures regarding the realization of ability shocks. Likewise, there is a lower bound Y_{II}

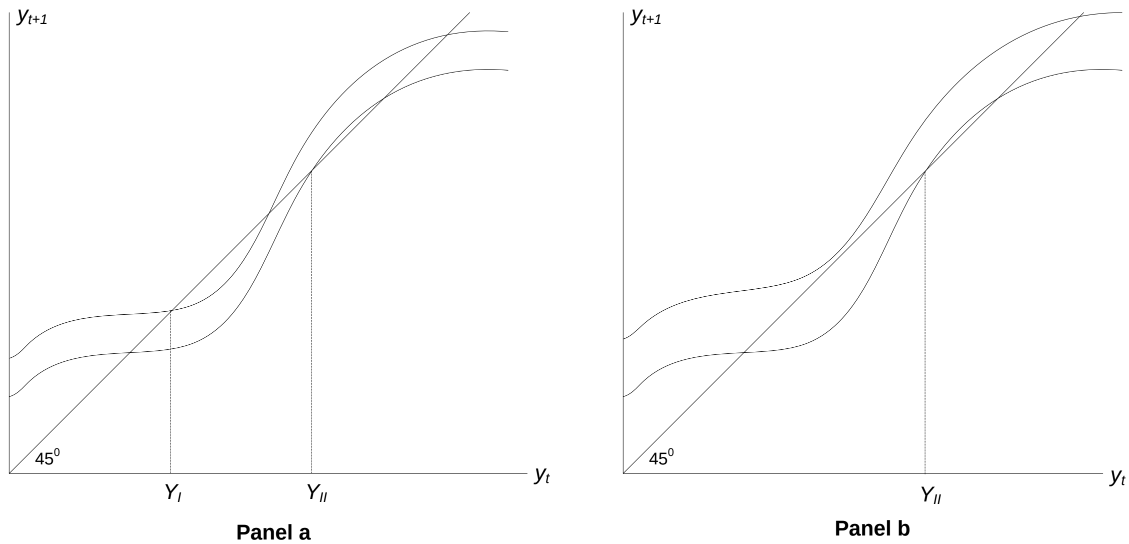


FIGURE 6.2. MIXING

in a different region of the policy correspondence. By the same logic, if $y \geq Y_{II}$, y can *never* fall below the value Y_{II}, even under the most pessimistic realization of ability shocks. Notice that in this diagram, Y_{II} exceeds Y_I, so that there two regions are effectively segregated. Proposition 6.4 can thus never hold for this scenario.

Now study panel (b) of Figure 6.2. Here, there is an upper trap Y_{II} just as we had before, but there is no "lower trap": incomes starting below can always wander into the zone to the right of Y_{II}, and must stay there therafter. Thus there is full mixing and the ergodicity result of Proposition 6.4 is restored.

But are the two assumptions concerning fools and geniuses met? Not really? We have, reflected in panel (b), a variant of these assumptions, which does just as well. In fact, convince yourself by drawing an imaginary panel (c), that if Y_I is well-defined but there is no Y_{II}, that this would be pretty much in line with the assumptions we do have. But now you also see that the specific form of our assumptions is not really necessary: what we are after is really the existence of a mixing zone, as described in Figure 6.2's panel (b), for instance.

6.2.1 Uncertainty, Inequality, and Ergodicity. What, then, is the precise role played by uncertainty in these models? To understand this, it will be useful to first strip the uncertainty away altogether, and simply look at a deterministic version.

First of all, suppose that the model is fully convex: that is, in addition to u being strictly concave, we shall also assume that f is strictly concave (in addition to assuming the endpoint conditions $f'(0) = \infty$ and $f'(\infty) < 1$). Now the optimal policy correspondence is really a *function*, and by proposition 6.3 it must be a nondecreasing function. It is therefore easy enough to see — by iteration — that the sequence of dynastic incomes $\{y_0, y_1, y_2 \ldots\}$ must converge to some steady state y^*. It is also easy to see (under our assumptions) that y^* must be strictly positive.

At this stage we don't know what y^* is. We do not *even* know whether y^* is unique or not. Of course, it is unique given y_0, but nothing in what we've said so far rules out the possibility that it might change as y_0 changes. But in fact, this cannot happen, and the easiest way to see this is to use the Euler equation to the solution of the optimization problem, which shows us that

$$u'(c_t) = \delta u'(c_{t+1}) f'(k_t) \tag{6.4}$$

for all $t \geq 0$. In the limit, $y_t \to y^*$, so it is trivial to see that k_t and c_t must converge as well, say to k^* and c^* respectively. Passing to this limit in the Euler equation (6.4), we may conclude that

$$\delta f'(k^*) = 1, \tag{6.5}$$

which, by the way, is the famous formula for the "modified golden rule" in optimal growth problems. When f is strictly concave, the value of k^* is uniquely pinned down. There can be, therefore, no more than one value for limit income, no matter where one starts from. Convergence occurs to perfect equality (though this may take time).

This brings us to the first role of uncertainty: by creating ongoing shocks (in this case to ability), it keeps individuals away from perfect equality. With each shock, one might think of the convergence problem as beginning again, but it is regularly perturbed by ongoing, further shocks. Thus uncertainty acts as a tool to create inequality in a world of convergence. Several economists who rely on convergence-based models generally invoke uncertainty to "explain" inequality. [In fact, Loury's model assumes that f is strictly concave and is therefore an example of this.]

Now suppose that we drop the concavity of f. Now (6.4) is still valid as an interior first-order condition, though we must be careful about checking second-order conditions. The main point, however, is that (6.5) will now admit several solutions in general. And indeed, there are now several steady states, depending on initial conditions. With decreasing returns dispensed with, history is perfectly capable of creating a lock-in effect. If there is inequality to start with, it may not go away: individuals may remain stuck in different steady states. This multiplicity in the face of nonconvexities has been known for some time (see, e.g., Majumdar and Mitra [1982] and more recently, the first part of Galor and Zeira [1993]).

But the possibility of such historical lock-in (arising from the convex f) brings us to a second role for uncertainty, which is its ability to *remove* such lock-ins. After all, the analysis in the previous section did not rely at all on the concavity of f. It is quite possible that the "mixing" condition discussed there holds even when f has the "wrong" curvature. This means that uncertainty can actually remove lock-ins and restore equality (or at least equality of long-run opportunity) when in the presence of perfect certainty, such equality would be missing.

This sort of discussion suggests that a study of uncertainty in this context may be misleading at the same time that it may be illuminating. It may be misleading because even very small mixing probabiluities lead us back to ergodicity (for instance, assume that everyone has a small but positive probability of winning the state lottery; then the mixing condition would be satisfied). But it may be a long time coming. For this reason, an uncertainty-based theory may hide certain structural features of the model. (For instance, we we've seen, there is a deep difference between the case of concave and convex f, but somehow the presence of uncertainty permitted us to provide a *mathematically* similar treatment.)

There is a final point about the role of uncertainty that might be worth making. It is peculiar to the noninteractive models that we have been discussing. Notice that with or without a mixing condition, it is possible to derive limit theorems regarding the distribution of income, in the following sense: starting from an initial y_0, the resulting distributions of income converge. The mixing condition simply adds that there will be a common limit to which the convergence occurs. Without the mixing condition — as in panel (a) of Figure 6.2 — there will be several steady states, *but two different limit distributions must have no incomes in common*. [If they did, the mixing condition would be satisfied, and we would not have two different limits in the first place.]

This is a peculiar and interesting characteristic of the noninteractive model, and also points to its inadequacy. If one wants to use such models to explain varying different distributions in, say, different countries with the same underlying fundamentals, one would have to contend with the uncomfortable prediction that the poorest person in one country must be richer than the richest person in the other country! The more realistic (but more complicated) interactive models that study later will set us free of this difficulty.

6.2.2 Technical Aside on Markov Processes and Existence of Mixing Interval .

This section included for completeness only. Not for a development course.

Let P be a transition probability on some state space X. It will be useful to have notation for the m-step transition probability generated by P. This is, intuitively, the probability of the system being in the subset A after m periods, starting from some given state x "today". Clearly, this is given by the measure μ_m, starting from the case where μ_0 assigns probability one to x. This measure we will denote in transition probability form as $P^m(x, A)$, for any measurable subset A of X.

The fundamental condition to be investigated is

Condition M (Stokey and Lucas [1989]) There exist $\epsilon > 0$ and an integer $M \geq 1$ such that for any event A, either (i) $P^M(x, A) \geq \epsilon$ for all $x \in X$, or (ii) $P^M(x, A^C) \geq \epsilon$ for all $x \in X$.

To appreciate condition M, let's look at a case when it is *not* satisfied. Consider the familiar two-state Markov chain in which $\pi_{ij} = 1$ if and only if $i \neq j$, for $i, j = 1, 2$. Pick $A = \{1\}$. Then for any positive integer M, $P^M(x, A) = 1$ either if M is even and $x = 1$, or if M is odd and $x = 2$. Otherwise, $P^M(x, A) = 0$. This means that condition M fails. We see therefore, that the real bite of condition M is in the postulated *uniformity* with which all states hit particular events.

Before we state the main result of this section, let us also relate Condition M to convergence in the Loury model. Let Y be the solution to $f(Y, 1) = Y$: it is the maximum possible output level. We first prove a formalization of the claim made in the main text regarding the "mixing interval" I:

Claim. There exists a date M, an interval I of incomes, and $\epsilon' > 0$ such that for every measurable subset I_λ of I of measure λ,

$$\text{Prob}\,\{y_M \in I_\lambda | y_0\} \geq \lambda\epsilon'$$

independently of $y_0 \in [0, Y]$.

Proof. By the assumptions on f and α, and using the "rich genius" condition, there exists a compact interval $I = [A, B]$, with $0 < A < B$, and a number a with $0 < a < A$, such that the random variable $f(k, \alpha)$ has strictly positive density on I for every $k \in [0, a]$. Let ϵ' be the minimum value of this density. It is easy to see that under our assumptions, $\epsilon' > 0$.

It follows that for every measurable subset I_λ of I of measure λ,

$$\text{Prob}\,\{f(k, \alpha) \in I_\lambda | k\} \geq \lambda\epsilon' \tag{6.6}$$

independently of $k \in [0, a]$.

For any y_0 and $t \geq 1$, let Y_t denote the random variable describing output at date t if all output is systematically invested up to date t, and none consumed. By the rich fools assumption, there is some date $M \geq 1$ and some probability $\eta > 0$ such that

$$\text{Prob}\,\{Y_{M-1} \in [0, a] | y_0\} \geq \eta$$

independently of $y_0 \in [0, Y]$.

Now turn to the equilibrium policy. Look at the equilibrium value of k_M conditional on any $y_0 \in [0, Y]$. Because $k_M \leq y_M \leq Y_{M-1}$, we see that

$$\text{Prob}\,\{k_M \in [0, a] | y_0\} \geq \eta \tag{6.7}$$

independently of $y_0 \in [0, Y]$.

Combining (6.6) and (6.7), we must conclude that for every measurable subset I_λ of I of measure λ,

$$\text{Prob}\,\{y_M \in I_\lambda | y_0\} \geq \lambda\epsilon'$$

independently of $y_0 \in [0, Y]$, and the claim is proved. ■

With the Claim in hand, it is easy to verify Condition M for the Loury model. Pick I, M and ϵ' as in the Claim, let ι be the measure of I, and define $\epsilon \equiv \iota\epsilon'/2$. Now for any event A, either $A \cap I$ has measure at least $\iota/2$ (in which case define $I_{\iota/2} \equiv A \cap I$) or $I - A$ has measure at least $\iota/2$ (in which case define $I_{\iota/2} \equiv I - A$). Now apply the Claim to the set $I_{\iota/2}$ to verify Condition M.

The main result of this section concerns the implication of Condition M:

Proposition 6.5. *Under condition M, there exists a unique invariant probability measure μ^* such that for any initial μ_0 on X, the generated sequence $\{\mu_t\}$ converges strongly to μ^*.*

Proof. We will follow the finite horizon case exactly. That is, we will show that

(1) $\mathcal{M}$ — the set of all probability measures on X — equipped with the total variation metric is a complete metric space.

(2) The operator $T^M : \mathcal{M} \to \mathcal{M}$ given by

$$T^M(\mu)(A) \equiv \int_X P^M(x, A)\mu(dx),$$

for all A, is a contraction.

(3) Thus T^M has a unique fixed point μ^* and the M-step iterates of any initial probability measure must converge to μ^*.

(4) The convergence of the entire sequence of measures, and not just this particular subsequence, can then be established by a subsequence argument identical to that used in the finite horizon case.

All the new stuff is in the first two items. To these we now proceed.

First we establish the completeness of $\mathcal{M}$. To this end, suppose that $\{\mu^n\}$ is a Cauchy sequence in $\mathcal{M}$. then from the definition of the total variation metric, it follows that for each event A, $\mu^n(A)$ is a Cauchy sequence of *numbers*. By the completeness of the real line, $\mu^n(A)$ converges to some $\mu(A)$ for each A. We will show that μ is a probability measure and that μ^n converges strongly to μ.

It is obvious that $\mu(A) \in [0, 1]$ for all A, that $\mu(X) = 1$, and that $\mu(\emptyset) = 0$. It remains to prove countable additivity to establish that μ is indeed a probability measure. To this end, let $\{A_i\}$ be a countable collection of disjoint events in X. Then

$$\mu(\cup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \mu^n(\cup_{i=1}^{\infty} A_i) = \lim_{n\to\infty} \sum_{i=1}^{\infty} \mu^n(A_i) = \sum_{i=1}^{\infty} \lim_{n\to\infty} \mu^n(A_i) = sum_{i=1}^{\infty} \mu(A_i)$$

where the second-last equality follows from the dominated convergence theorem. This proves that μ is indeed a bonafide probability measure. What's left to do is to show that $\|\mu^n, \mu\| \to 0$ as $n \to \infty$. Note that because $\{\mu^n\}$ is Cauchy, for all $\epsilon > 0$, there is N such that if $n, m \geq N$,

$$|\mu^n(A) - \mu^m(A)| \leq \epsilon$$

for all sets A. Taking limits in m, it follows that

$$|\mu^n(A) - \mu(A)| \leq \epsilon$$

for all $n \geq N$, and for all A, which implies that $\|\mu^n, \mu\| \to 0$ as $n \to \infty$.

This establishes the completeness of $\mathcal{M}$.

Our next task is to show that T^M is a contraction of modulus $1 - \epsilon$. To this end, pick μ and μ' in $\mathcal{M}$. Then there is a "common" part γ and "idiosyncratic" parts μ^1 and μ^2, so that $\mu = \mu^1 + \gamma$, $\mu' = \mu^2 + \gamma$, and μ^1 and μ^2 have disjoint support.

Digression. The technical details of this assertion rely on the Radon-Nikodym Theorem (see Stokey and Lucas [1989, Lemma 7.12]). But one can informally illustrate how the common part is obtained in the case where μ and μ' have densities f and f' on the real line. In this case, simply define $g(x) \equiv \min\{f(x), f'(x)\}$ for each $x \in \mathbb{R}$, and integrate this as you would to get a cdf, to arrive at the common part of the measure γ. This measure γ is, of course, not a probability measure. The idiosyncratic residuals μ^1 and μ^2 are then defined by inserting the rest of the probability, event by event, to bring total probability up to μ and μ'. In this way you see that the measures μ^1 and μ^2 are not probability measures either. Now study the values $\mu^1(X)$ and $\mu^2(X)$, which are the values assumed by the measures on the entire state space. Note, first, that these must be equal to each other. Note, also, that their common value must be equal to the supremum difference between the probability measures μ and μ' over any event. To see this, note that for any event A

$$\begin{aligned} |\mu(A) - \mu'(A)| &= |\mu^1(A) - \mu^2(A)| \\ &\leq \max\{\mu^1(A), \mu^2(A)\} \\ &\leq \max\{\mu^1(X), \mu^2(X)\}. \end{aligned}$$

This ends our digression.

Returning to the main argument, we see that

$$\begin{aligned} \|T^M\mu, T^M\mu'\| &= \sup_A \left| \int P^M(x,A)\mu(dx) - \int P^M(x,A)\mu'(dx) \right| \\ &= \sup_A \left| \int P^M(x,A)\mu^1(dx) - \int P^M(x,A)\mu^2(dx) \right|. \end{aligned}$$

Now consider any event A and its complement A^C. Without loss of generality suppose that $P^M(x,A) \geq \epsilon$ for all $x \in X$. If K denotes the common value of $\mu^1(X)$ and $\mu^2(X)$ (see digression above), then it must be the case that

$$\left| \int P^M(x,A)\mu^1(dx) - \int P^M(x,A)\mu^2(dx) \right| \leq (1-\epsilon)K.$$

Combining these last two observations, and the observation in the digression, we see that

$$\|T^M\mu, T^M\mu'\| \leq (1-\epsilon)\|\mu, \mu'\|,$$

which completes the proof that T^M is a contraction. ■

6.3 Interactive Inequality

It is now time to unpack the nature of the household production function. In many important situations, it is determined by relative prices. This nring us to models of interactive inequality, in which no household or dynasty can be treated as an isolated unit.

6.3.1 Relative Prices and Efficiency Units. Economists have traditionally employed a simple shorthand for the study of occupational diversity, which is to reduce different qualifications and skills to aggregate quantities of "human capital". In other words, all human capital is — *even before we write down the definition of equilibrium for the society in question* — commonly expressible in some common efficiency unit.[2] This approach is summarized by Becker and Tomes in their 1986 paper:

[2]So, for instance the investment choice in Loury's model is interpreted as a choice of "how much" education to acquire: there is no formal difference between human and physical capital. The so-called endogenous growth models (see, e.g., Lucas (1988)) continue, by and large, to retain this shorthand.

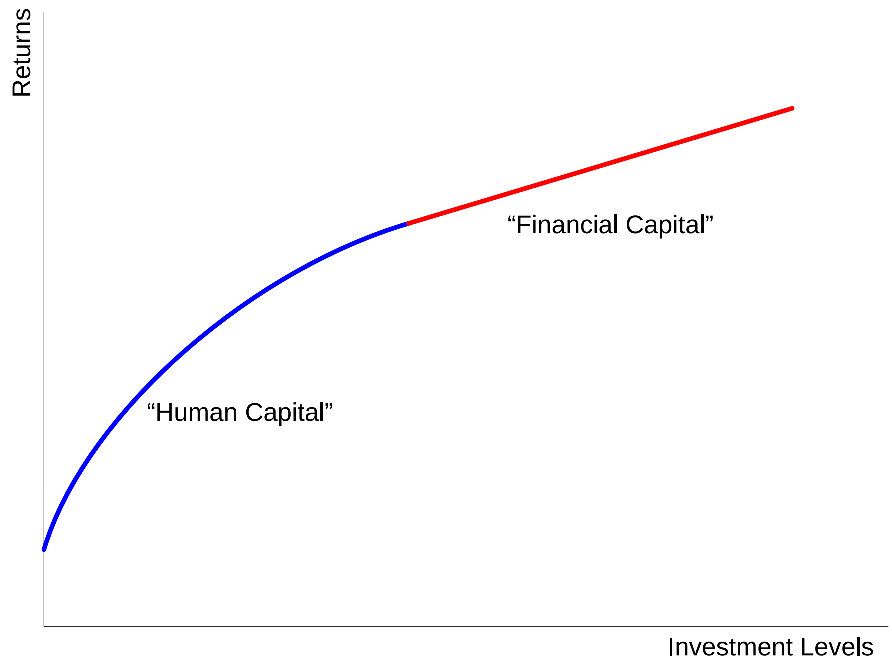


FIGURE 6.3. EFFICIENCY UNITS

> "Although human capital takes many forms, including skills and abilities, personality, appearance, reputation and appropriate credentials, we further *simplify* by assuming that it is homogeneous and the same "stuff" in different families." (Becker-Tomes (1986, p.56), emphasis ours)

The crucial assumption is that the relative returns to different occupations are exogenous, so that the reduction to efficiency units can be carried out separately from the behavioral decisions made in the population. It is, in fact, common to specify that the returns to human capital are concave while the return to financial investment is linear (see Figure 6.3).

It is unclear what different "levels" of human capital mean independent of the relative market returns, which are typically endogenous. We might all agree that skilled labor embodies more human capital than unskilled labor. But the assumption that skilled and unskilled can be reduced to a common and determinate yardstick of efficiency units presumes much more than this ordinal comparison. The implications of such an assumption can be quite significant, as we show in the next section.

6.3.2 Endogenous Relative Prices: Persistent Inequality With Two Skills. The following analysis is based on Ray (1990). Elements of this model appear in Ljungqvist (1993), Freeman (1996) and Mookherjee and Ray (2003). The specific exposition follows Ray (2006).

Suppose that aggregate production is a CRS function of just two inputs: skilled and unskilled tasks, satisfying Inada conditions in each. There are two occupations: skilled and unskilled labor. The latter can only do the unskilled tasks. Wages in each occupation equal their respective marginal products, and so depend on relative supplies of workers in the two occupations. Skill acquisition requires a fixed parental investment in education. Assume that this is the only way a parent can transfer wealth to their children, i.e., there are no financial bequests. In a later section we drop this assumption.

Now observe that even if all parents in the economy have identical wealth and preferences, *they cannot all leave the same bequests*. The reason is simple. If every parent keeps their child unskilled, there will be no skilled people in the next generation, raising the return to skilled labor enough that investment in skill will be the optimal response. Conversely, if all children are skilled, the return to skill will vanish, killing off the investment motive.[3] Hence even if all families start equal in generation 0, some will invest and others will not; in the next generation their fortunes must separate.

Just who goes in one direction and who in another is entirely accidental, but such accidents will cast long shadows on dynastic welfare. Indeed, the two directions are utility-equivalent for generation 0, but not for generation 1! Furthermore, in succeeding generations wealthier parents will have a greater incentive to train their children, so that the "primitive inequality" that sets in at the first generation will be reinforced: children of skilled parents will be more likely to acquire skills themselves. The logic of "symmetry-breaking" implies that every steady state in this example must involve persistent inequality. The endogeneity of occupational returns is central to this argument.[4]

In contrast to this example, observe that the same argument does not apply to activities in which each unit is a perfect substitute for another. For instance, if shares in physical capital can be divisibly held, everyone can derive the very same rate of return on each unit. But a single individual cannot hold an arbitrarily fine portfolio of different occupations.[5]

In the subsections that follow, I develop this argument in more detail.

6.3.2.1 Preliminaries Time is discrete, running $t = 0, 1, 2 \ldots$. A *dynasty* is represented by an infinite sequence of individuals, each individual living for a single period. There is a continuum of dynasties so that a unit mass of atomless individuals belongs to a *generation* at each date.

There are two skill categories, "high" and "low", which are combined via a production function f to produce a single final output, which we take to be the numeraire. An individual in the high-skill category (or a "high individual" for short) earns a wage $\bar{w}_t$ at date t.

[3] If skilled labor cannot perform unskilled tasks, the unskilled wage will become very high by the Inada conditions. But even if they can perform unskilled tasks, this will equalize the two wage rates. Either interpretation has the same outcome.

[4] See e.g., Katz and Murphy (1992) for the responsiveness of US skill premia to relative supply of skilled workers.

[5] This discussion suggests, then, that the correct dividing line is not between "physical" and "human" bequests, but rather bequests that result in endowments that are alienable (e.g. money) and endowments that are not (e.g. occupations). The latter may include transfers of physical assets such as a family business which is not incorporated — perhaps for reasons of moral hazard or simply the lack of development of a stock market. These transfers are no different from human bequests in their implications for disequalization, and should be included in the category of "occupational bequests".

Likewise, a low individual earns $\underline{w}_t$. Whether or not an individual is high or low depends on the investment made by her parent. Being low at any date requires no investment by the parent; being high requires an exogenous investment of x.

Earned income is partly consumed and partly used in educating the individual's offspring. Depending on the education level of the child, the child receives an income next period, and the entire process repeats itself without end.

Now I turn to the determination of wages. Assume that the production function f for final output is smooth, CRS in its two inputs, strictly concave in each input and satisfies the Inada end-point conditions. Given a unit mass of individuals, if a fraction λ of them is high at some date, then the high wage is given by

$$\bar{w}(\lambda) \equiv f_1(\lambda, 1-\lambda),$$

while the low wage is given by

$$\underline{w}(\lambda) \equiv f_2(\lambda, 1-\lambda).$$

where these subscripts represent partial derivatives. We will call these wages the *wages associated with* λ.

It is easy to see that $\bar{w}(\lambda)$ is decreasing and continuous in λ, with $\bar{w}(\lambda) \to \infty$ as $\lambda \to 0$. Likewise, $\underline{w}(\lambda)$ is increasing and continuous in λ, with $\underline{w}(\lambda) \to \infty$ as $\lambda \to 1$. These observations imply, in particular, that there exists a threshold $\tilde{\lambda}$ such that $\bar{w}(\tilde{\lambda}) = \underline{w}(\tilde{\lambda})$.

To complete the description of the model, we presume that each generation t maximizes an additive function of the one-period utility u_t from its own consumption, and the lifetime utility (V_{t+1}) felt by generation $t+1$, discounted by $\delta \in (0,1)$. The utility function u will be assumed to be increasing, smooth and strictly concave in consumption, and defined at least on $[-x, \infty)$. This last requirement is innocuous but serves to simplify notation and exposition. Moreover, the idea that consumption can go negative captures the idea that the borrowing constraint is never absolute, but that the investment of x at lower wealth levels entails ever greater utility losses (by strict concavity of u).

6.3.2.2 Equilibrium Suppose, now, that an infinite sequence of wages is given, one for each skill category. We may denote this by the path $\{\bar{w}_t, \underline{w}_t\}_{t=0}^{\infty}$. With such a sequence given, consider the maximization problem of generation t. Denote by $\bar{V}_t$ the lifetime utility for a high member of that generation, and by $\underline{V}_t$ the corresponding lifetime utility for a low member. Standard arguments tell us that the sequence $\{\bar{V}_t, \underline{V}_t\}_{t=0}^{\infty}$ is connected over time in the following way: for each date t,

$$\bar{V}_t = \max u(c_t) + \delta V_{t+1} \tag{6.8}$$

subject to the conditions that

$$c_t + x_t = \bar{w}_t, \tag{6.9}$$

and

$$\begin{aligned} V_{t+1} &= \bar{V}_{t+1} \text{ if } x_t \geq x \\ &= \underline{V}_{t+1} \text{ if } x_t < x. \end{aligned} \tag{6.10}$$

In exactly the same way, $\underline{V}_t = \max u(c_t) + \delta V_{t+1}$, subject to the analogous budget constraint $c_t + x_t = \underline{w}_t$ and (6.10).

These maximization problems describe how education levels change from generation to generation, given some sequence of wage rates. To complete the equilibrium setting, we remind ourselves that the wages are endogenous; in particular, they will depend on the proportion of high individuals at each date.

Formally, for given $\lambda_0 \in (0, 1)$, a *competitive equilibrium* is a sequence $\{\bar{w}_t, \underline{w}_t, \lambda_t\}_{t=0}^{\infty}$ such that

[i] Given λ_0, the path $\{\lambda_t\}$ is generated by the maximization problems described above.

[ii] For each t, $\bar{w}_t$ and $\underline{w}_t$ are the wages associated with λ_t.

Standard fixed-point arguments suffice to show that a competitive equilibrium exists, and we will not pursue such matters here.

Notice that our definition of competitive equilibrium assigns wages on the presumption that skilled labor *must* carry out skilled tasks. Alternatively, we could restate the definition so that if the "natural wages" (as given by marginal product) for a skilled worker falls short of that of an unskilled worker, the former will move into the sector of the latter so that the two wages will be *ex post* equalized. None of this matters much anyway because of the following easy observation, which holds no matter what definition we use:

Observation 6.1. *Recalling that $\tilde{\lambda}$ solves $\bar{w}(\tilde{\lambda}) = \underline{w}(\tilde{\lambda})$, $0 < \lambda_t < \tilde{\lambda}$ for all $t \geq 1$ along any competitive equilibrium.*

The formalities of the (obvious) proof are omitted.[6] From now on I will also presume that $\lambda_0 \in (0, \tilde{\lambda})$ as well. There is no great mystery in this: it saves the expositional trouble of having to qualify several arguments for the initial value of λ.

For later use, I also record a familiar single-crossing observation.

Observation 6.2. *Under a competitive equilibrium, there is no date at which a low person creates a high child while simultaneously, a high person creates a low child.*

The proof of this observation runs exactly parallel to (and in fact can be derived as a special case of) Proposition 6.1, and is therefore omitted.

6.3.2.3 Steady States

A fraction λ is called a *steady state* if there exists a competitive equilibrium $\{\bar{w}_t, \underline{w}_t, \lambda_t\}_{t=0}^{\infty}$ from λ with $(\bar{w}_t, \underline{w}_t, \lambda_t) = (\bar{w}, \underline{w}, \lambda)$ for all t, where $\bar{w}$ and $\underline{w}$ are the wages associated with λ.

The single-crossing property in the previous section yields a simple characterization of steady states. Let $\bar{w} \equiv \bar{w}(\lambda)$ and $\underline{w} \equiv \underline{w}(\lambda)$ be the associated wages, and let $\bar{V}$ and $\underline{V}$ be the

[6]That $\lambda_t > 0$ for all t follows from the fact that the difference between skilled and unskilled wages would be infinitely high otherwise, so that some educational investment would have taken place prior to that period. [Here we use the assumption that u is defined on $[-x, \infty)$.] On the other hand, λ_t cannot exceed $\tilde{\lambda}$ for any $t \geq 1$, for in that case high and low wages are equalized, and no one in the previous generation would then have invested in high skills.

lifetime utilities associated with being (initially) high and low, respectively. By Observation 6.2, the following two conditions are necessary and sufficient for λ to be a steady state:

$$\bar{V} = u(\bar{w} - x) + \delta\bar{V} \geq u(\bar{w}) + \delta\underline{V},$$

while

$$\underline{V} = u(\underline{w}) + \delta\underline{V} \geq u(\underline{w} - x) + \delta\bar{V}.$$

Combining these two expressions, we may conclude that

$$u(\bar{w}) - u(\bar{w} - x) \leq \delta(\bar{V} - \underline{V}) \leq u(\underline{w}) - u(\underline{w} - x)$$

is a necessary and sufficient condition for λ to be a steady state. Combining this expression for the values of $\bar{V}$ and $\underline{V}$, we have established

Proposition 6.6. *The fraction λ (with associated wages $(\bar{w}, \underline{w})$) is a steady state if and only if*

$$u(\bar{w}) - u(\bar{w} - x) \leq \frac{\delta}{1-\delta}[u(\bar{w} - x) - u(\underline{w})] \leq u(\underline{w}) - u(\underline{w} - x) \tag{6.11}$$

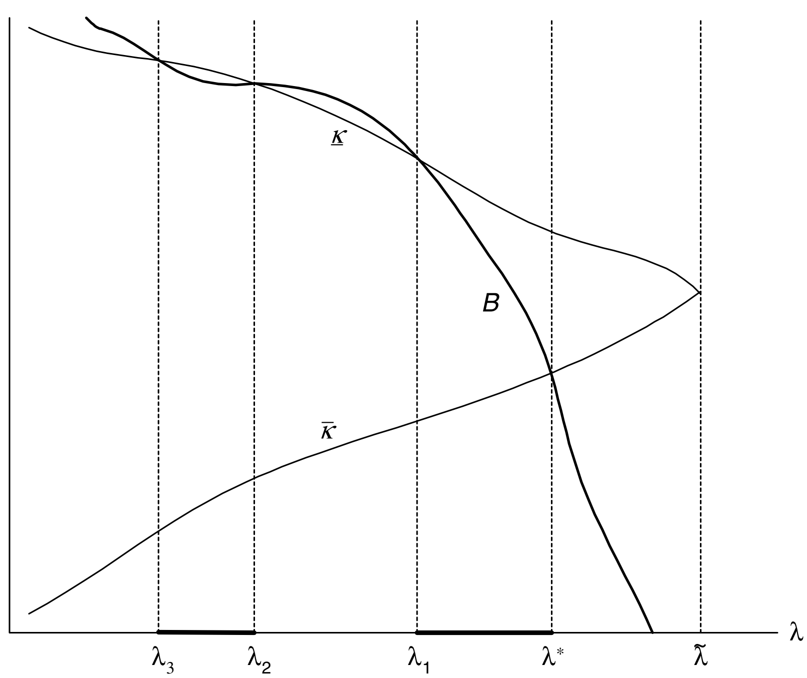


Figure 6.4. The Set of Steady States

Figure 6.5 plots the three terms in (6.11) as a function of λ. The left hand side, denoted by $\bar{\kappa}$, is just the utility cost to a high parent of acquiring skills for her child. This lies uniformly below the right hand side, denoted by $\underline{\kappa}$, which tracks the same utility cost to a low parent. Finally, the middle term, denoted by B, is the present value of benefits to being high rather than low. Of course, the $\bar{\kappa}$ and $\underline{\kappa}$ lines meet at $\tilde{\lambda}$, because wages are equalized there.

Note, moreover, that as $\lambda \to \tilde{\lambda}$, B turns negative while $\bar{\kappa}$ is positive. On the other hand, as $\lambda \to 0$, B grows unboundedly large while $\bar{\kappa}$ is bounded above. Because the changes are monotone, there is a unique $\lambda^* \in (0, \tilde{\lambda})$ such that the first inequality in (6.11) holds with equality. Observe, moreover, that at $\lambda = \lambda^*$, the *second* inequality in (6.11) must hold as well, because of the strict concavity of the utility function. Thus the set of steady states contains some interval to the left of λ^*, and must be a subset of $(0, \lambda^*]$.

Beyond this last observation, the set of steady states may be complicated. In particular, the set need not be connected. For instance, in Figure 6.5, the set of steady states is the union of the two intervals $[\lambda_3, \lambda_2]$ and $[\lambda_1, \lambda^*]$.

Notice that *every steady state involves persistent inequality*, not just in gross incomes but in net utility. In the subsection that follows, I show that the dynamics of an intertemporal competitive equilibrium leads society to one of these steady states.

6.3.2.4 Dynamics Recall from the previous section that λ^* is a steady state, and it is the largest possible steady state. We now state and prove the following

Proposition 6.7. *If $\lambda_0 > \lambda^*$, then there exists a unique competitive equilibrium from λ_0. It goes to a steady state in one period: $\lambda_0 > \lambda_1 = \lambda_t$ for all $t \geq 1$.*

If $\lambda_0 < \lambda^*$, then along a competitive equilibrium λ_t converges monotonically to the smallest steady state no less than λ. Convergence is never attained in finite time unless λ happens to be a steady state to start with, in which case $\lambda_t = \lambda_0$ for all subsequent t.

The proposition provides a full account of the behavior of skill proportions over time, starting from any initial condition. If that initial condition happens to be a steady state, the proposition rules out any equilibrium path other than the steady state path itself. More interesting is the asymmetry of equilibrium behavior under the two remaining kinds of initial conditions. When λ_0 is larger than the largest conceivable steady state, convergence to a steady state occurs in a single unit of time. When λ_0 is such that there are steady states "above" it, convergence is gradual in that the process is never completed in finite time. This asymmetry may have interesting implications for unanticipated technical changes which, once realized, are expected to stay in place thereafter. Changes that call for a reduction in steady state skill proportions take place quickly and dramatically, whereas a climb to a higher steady state is more gradual and drawn out.

Moreover, in the case that convergence is "up" to a steady state, the proposition asserts that it will occur to the *nearest* steady state to the right of the initial conditions. Put another way, only the left-most steady state in each interval of steady states can be an attractor for initial conditions that are distinct from that steady state, and the basin of attraction is precisely the set of initial conditions that lie between it and the next, lower interval of steady states (if any). Thus, despite the multiplicity of steady states, final outcomes can be tagged to initial conditions in a unique way, allowing us in principle to perform comparative dynamics.

To provide some intuition for these results, first study a non-steady-state value of λ that is smaller than λ^*. Why is this not a steady state? Surely, the "no-deviation" condition for the skilled (the first inequality in (6.11)) is satisfied; after all, it was satisfied at λ^*, and now at the smaller value of λ, *both* the wage differential is higher *and* the utility cost of education is lower for the skilled. So, the reason why λ fails to be a steady state is that the "no-deviation" condition for the *unskilled* — the second inequality in (6.11) — is violated; their utility cost of education is high, but not as high as the wage differential. To maintain equilibrium incentives, then, the economy-wide skill ratio must rise, compressing the sequence of wage differentials until the unskilled are exactly indifferent between acquiring and not acquiring skills. [We have already seen in Lemma 6.1 that this compression to indifference is a necessary feature of non-steady-state equilibrium, otherwise no one would stay unskilled.]

Now here is the main point: *the skill ratio achieved in the very next period cannot be a steady state.* For if it were, then the no-deviation condition of the unskilled must be satisfied here, so that the new implied wage differential generates no incentive for them to acquire skills. But the very same wage differential created indifference at date 0, when the utility cost of acquiring education was higher for the unskilled! This is a contradiction. By an obvious recursive argument, it follows that the upward movement in the skill ratio must be gradual and perennial.

Exactly the opposite is true for non-steady-state values of λ that exceed λ^*. For such values, the "no-deviation" condition for the skilled surely fails; the wage differential is too small relative to the utility cost of maintaining skills. So in equilibrium, λ falls. This fall along the equilibrium path raises wage differentials so that the skilled are now indifferent between maintaining and relinquishing skills.

We claim that *the new skill ratio one period later must be a steady state.* Suppose the claim is false. Then the new skill ratio is not a steady state, and this can happen for one of two reasons. First, the no-deviation condition for the unskilled fails — the wage differential at date 1 is too attractive. In that case, we already know that the remaining sequence of wage differentials (counting from date two on) must render the unskilled indifferent at date 1. But this means that the sequence of wage differentials counting from date *one* on is still attractive for the unskilled, and would have been *a fortiori* attractive relative to the skill-acquisition cost for the unskilled at the higher skill ratio prevailing at date 0. But this, in turn, exceeds the cost of skill maintenance for the skilled, which contradicts the fact that the equilibrium creates indifference for the skilled at date 0 (see last sentence of preceding paragraph).

The second reason why the new skill ratio may fail to be a steady state is that the no-deviation condition for the skilled fails again (just as it did at date 0). This means that the skill ratio must fall even further in succeeding periods to create indifference for the skilled at date 1. However, because the wage differential at date 1 is not attractive enough, this means that the entire sequence of wage differentials, *counting* the one at date 1, would not have been attractive enough for the very same skilled individuals, had they been located at date 0, but with the utility costs they possess at date 1. This would be *a fortiori* true if we were to replace the utility costs with the true utility costs of maintaining skills at date 0, which are higher. But now we have a contradiction again (see the last sentence two paragraphs above). This completes our intuitive description.

Proposition 6.7 asserts convergence from every initial condition to some steady state. Moreover, the particular steady state to which convergence occurs can be identified, and this is of interest because there are multiple steady states.

How general are these results on convergence; in particular, do they extend to the case of several occupations? The simple answer is that I do not know. At the same time, as we are going to see in the next section, with endogenous relative prices the *steady state* findings are very general. With endogenous price determination and imperfect credit markets, steady-state inequality among ex-ante identical agents is inevitable under broad conditions. It would, therefore, be of great interest to see if the dynamic counterparts of the results in this section also carry over.

6.4 More on Relative Prices and Persistent Inequality

Recall the two-skill example with which we started the previous section. We argued that even with parents of identical wealth, symmetry-breaking must occur "in equilibrium": after all, it cannot be that all parents skill their children, or that all parents leave their children unskilled. This symmetry-breaking led to persistent inequality. It is natural to ask whether the same results hold when parents can compensate low-skilled children with greater financial bequests. Symmetry-breaking would still occur, but only in "occupational space": there would be no inequality in *overall* wealth, human plus financial. In this section, based on Mookherjee and Ray (2007), I carefully investigate the validity of this argument.

A second assumption that we made is that there are only two occupations. On the face of it, this is is just a simplification, and indeed, most of the literature works off two or three occupations simply for expositional ease. But as we shall see, bringing in many occupations (all with varying relative prices) makes two major conceptual differences. The first is that "many" different levels of human capital can be chosen, along with financial bequests (see previous paragraph). From the household point of view — with all relative prices taken as given — we have then a standard "production function" of the form displayed in Figure 6.3, which is the outer envelope of all human and financial investments. But there is one difference: the *shape* of this "production function" is entirely dependent on endogenous relative prices! Viewed from this perspective, at the household level all returns to investments are driven by relative prices. Whether this gives rise to a concave "technology" or not does not depend on technology but on economics, and on economic equilibrium at that. Yet, we will have something quite definite to say about the curvature of this production function.

The second conceptual difference, as we shall see below, is that steady state inequality might remain but *history-dependence* tends to vanish! Recall in the the two-skill model (see Figure 6.5) that there isn't just inequality in steady state, there is a continuum of steady states. Indeed, the dynamics studied in that model take us to different limit points depending on initial conditions. Thus the two-skill model displays not just persistent inequality, but also persistent dependence on initial conditions, or on "history".

It turns out that with a rich set of occupations, this history-dependence vanishes. This doesn't happen at the level of an individual dynasty, whose fortunes remain profoundly linked to the distant past. Rather, it happens to the economy *as a whole*.

6.4.1 The Model. **Occupations and Training.** There is a compact measurable space $\mathcal{H}$ of *occupations* that will be used in the production of a single, aggregative final good. There is an exogenous *training cost* $x(h)$ for occupation $h \in \mathcal{H}$, denominated in units of final output.[7]

Production. A single aggregate output is produced by physical capital[8] and individuals who hold occupations in $\mathcal{H}$.

Output y (net of the undepreciated capital stock) is produced by a continuous,[9] strictly quasiconcave CRS production function $y = f(k, \boldsymbol{\lambda})$, where k is physical capital and $\boldsymbol{\lambda}$ is an occupational distribution (a finite measure on $\mathcal{H}$). It is helpful to interpret different occupations as corresponding to different kinds of human capital.

Prices and Firms. Firms maximize profits at given prices. Normalize the price of final output to 1. Let $\mathbf{w} \equiv \{w(h)\}$ denote the wage function, and $\mathbf{p} \equiv (r, \mathbf{w})$ the factor price function, where r is the rate of interest. Denote by $c(\mathbf{p})$ the unit cost function.

By constant returns to scale, profit maximization at positive output is possible if and only if $c(\mathbf{p}) = 1$; in that case call $\mathbf{p}$ a *supporting price*. If $(k, \boldsymbol{\lambda})$ is a profit-maximizing choice under the supporting price $\mathbf{p}$, we will refer to it as an *associated* input vector.

In these notes, I assume that the rate of interest r is exogenously given and time-stationary. One simple interpretation is that capital is internationally mobile and that our economy is a price taker on the world market. Under this interpretation we also assume, in effect, that *people* are not internationally mobile: the wage function $\mathbf{w}$ will be determined domestically. See Mookherjee and Ray (2006) for extensions.

Now that r may be treated as a parameter, say that $\mathbf{w}$ is a *supporting wage* if $\mathbf{p} = (r, \mathbf{w})$ is a supporting price.

Families. There is a continuum of families indexed by $i \in [0, 1]$. All families are *ex ante* identical, so we endow $[0, 1]$ with Lebesgue measure ν. Each family i has a single representative at each date or generation, indexed by t. Call this agent (t, i).

Consider a member of generation t. She begins adult life with a financial bequest b and an occupation h, both "selected" by her parent. The latter is obviously shorthand for the assumption that the parent bears the costs of upbringing and education (the *child* can select the particular occupation with no difference to the formal analysis). The overall wealth of our generation-t adult is then $W \equiv b(1 + r) + w_t(h)$, where $w_t(h)$ is the going wage for occupation h at date t.

The agent correctly anticipates factor prices $\mathbf{p}_{t+1} \equiv (r, \mathbf{w}_{t+1})$ for the next generation $t + 1$, and selects her own financial and educational bequests (b', h') to maximize

$$U(W - x(h) - b') + V((1 + r)b' + w_{t+1}(h')) \tag{6.12}$$

[7] As in Mookherjee and Ray (2003), this may be generalized to allow training costs to depend on the pattern of wages. We conjecture that the principal qualitative results of this paper will continue to hold in that setup.

[8] As long as capital goods are alienable and shares in them can be divisibly held, having several capital goods makes no difference to the analysis.

[9] Endow the space of all nonnegative finite measures on $\mathcal{H}$ with the topology of weak convergence. We ask that output be continuous with respect to the product of this topology and the usual topology on k.

subject to the no-intergenerational-debt constraint $b' \geq 0$. We assume that U and V are smooth, increasing and strictly concave, and that U has unbounded steepness at 0.

Now b' and h' become the financial and educational inheritance of her child — generation $t+1$ — and the entire process repeats itself *ad infinitum*.

Clearly, we are using a specification of utility that allows for income-based altruism. More on this specification later.

The condition $b' \geq 0$ is a fundamental restriction stating that children cannot be held responsible for debts incurred by their parents. The capital market is active in all other senses: households can make financial bequests at the going rate r, and firms can freely hire in capital at the very same rate.

Equilibrium. Begin with an initial distribution of financial wealth and occupational choices. A *competitive equilibrium* given these initial conditions is a sequence of wage functions $\mathbf{w}_t$, $t = 0, 1, 2, \ldots$ and occupational distributions $\boldsymbol{\lambda}_t$, $t = 0, 1, 2, \ldots$, as well as occupational and bequest choices for each generation in each family — $\{h_t(i), b_t(i)\}$ — such that for each t and each family i:

(a) person (t, i) chooses $(b_{t+1}(i), h_{t+1}(i))$ to maximize the utility function in (6.12), given that her own starting wealth equals $(1 + r)b_t(i) + w_t(h_t(i))$;

(b) these decisions aggregate to $\boldsymbol{\lambda}_t$ at each t:

$$\lambda_t(H) = \nu\{i \in [0, 1] | h_t(i) \in H\}$$

for every Borel subset H of occupations, and

(c) $\mathbf{w}_t$ is a supporting wage, with associated input vector $(\boldsymbol{\lambda}_t, k_t)$ for some choice of k_t.

Observe that equilibrium conditions place no restrictions on k_t. Because there is international capital mobility, financial holdings by households need bear no relation to capital used in production.[10]

A *steady state* is a competitive equilibrium with stationary prices and distributions; $(\mathbf{w}_t, \boldsymbol{\lambda}_t) = (\mathbf{w}, \boldsymbol{\lambda})$ for all t, and strictly positive output.

The following observation is useful: in steady state, the total wealth of *every* family, not just of the economy as a whole, must be stationary.[11]

Observation 6.3. *The wealth of every family is stationary in any steady state.*

Proof. This follows directly from Proposition 6.2. ■

[10] When there is no international capital mobility, k_t must equal the aggregate of financial holdings, and r must adjust to assure this equalization in equilibrium.

[11] In general, such an assertion is not true of the financial wealth or the occupational choice of a family, which may vary over time.

6.4.2 The Becker-Tomes Benchmark. A special case of this model is the following elementary textbook exercise: *only* financial bequests are possible (earning interest r), and everyone earns a fixed wage w.

Stochastic shocks apart, this is exactly the specification for the Becker-Tomes (1979) model, which assumes a linear rate of return to parental investment in children.[12] We will henceforth refer to this special case as the *Becker-Tomes benchmark*.

In this special case, a parent with wealth W simply selects $b \geq 0$ so as to maximize

$$U(W-b) + V(w + (1+r)b).$$

Let the resulting wealth of the child be denoted $\tilde{W} \equiv w + (1+r)b$. We may write $\tilde{W}$ as a function of W, w and r: $\tilde{W}(W; w, r)$. By our assumptions, $\tilde{W}$ is fully characterized by the first order conditions

$$U'\left(W - \frac{\tilde{W} - w}{1+r}\right) \geq (1+r)V'(\tilde{W}), \tag{6.13}$$

with equality if $\tilde{W} > w$.

It is obvious that $\tilde{W}(W; w, r)$ is nondecreasing and continuous in W. So an iteration of this mapping from any initial condition $W > 0$ will yield long-run wealth starting from W. Call this long-run wealth Ω. Passing to the limit in (8.42), it is trivial to see that if $w \leq \Omega < \infty$,

$$U'\left(\frac{r\Omega + w}{1+r}\right) \geq (1+r)V'(\Omega), \text{ with equality if } \Omega > w. \tag{6.14}$$

Becker and Tomes impose the following restriction on bequest behavior: $\frac{\partial \tilde{W}}{\partial W} \in (0,1)$, justifying it by available empirical evidence. This implies that the wealth of all families will converge to a common limit Ω, *independent of initial wealth*. To ensure that our larger model is consistent with this key equalization property of financial bequests, we impose a similar (though weaker) restriction in this special case:

[LP] *Limited Persistence*. For any $r > -1$ and $w \geq 0$, there is at most one solution in $\Omega \geq w$ to (6.14).

As in Becker-Tomes, [LP] also implies that limit wealth $\Omega(w, r)$ is well defined and independent of starting wealth as long as as that starting wealth strictly exceeds w (that is, as long as starting financial wealth is positive). Given [LP], here is how we define that limit wealth: Set $\Omega(w, r)$ equal to Ω, where Ω solves (6.14), in *all* cases except the one in which

$$U'\left(\frac{r\Omega + w}{1+r}\right) > (1+r)V'(\Omega)$$

[12] A more sophisticated version of their model is Becker and Tomes (1986), in which bequests in the form of human capital are also permitted, but human capital is *a priori* reduced to efficiency units and it is *assumed* that the rate of return to successive units of human capital is declining. In that variant, all families will aim to invest the (same) amount of human capital before turning to linear financial bequests. Indeed, we do not claim that our elementary textbook exercise captures the full implications of the Becker-Tomes models. An important theme in these papers is the interplay between luck and convergence, an issue that is not of relevance here.

for all $\Omega > w$, in which case set $\Omega(w,r) = \infty$.[13] The important point is that $\Omega(w,r)$ is the same no matter what the (positive) level of inital wealth is.

This condition is less restrictive than the original Becker-Tomes assumption on $\frac{\partial W'}{\partial W}$, though the reader is welcome to keep the stronger restriction in mind (as long as it is understood that an implicit restriction on the interest rate is also implied thereby).

Condition [LP] is extremely easy to check. The following observation illustrates this by studying the HARA class of preferences.

Observation 6.4. *Suppose preferences satisfy the following restriction: There exists $\delta > 0$ such that $V = \delta U$, and U belongs to the* HARA *family:*

(H) $\qquad -U''(c)/U'(c) = 1/(\alpha + \beta c),$

where $(\alpha, \beta) \geq 0$ and nonzero. Then the limited persistence property is satisfied, with the single exception in which $\alpha = 0$ and $w = 0$, in which case it is satisfied for all but one value of r.

Provided that preferences satisfy a constant discount rate property, Observation 6.4 states that [LP] holds for utility functions that are iso-elastic or exponential, or belong to the HARA class which nests these as special cases.

It helps to illustrate bequest behavior in the setting with financial bequests alone for the special case of iso-elastic utility with discounting: $U(c) = (c^{1-\sigma} - 1)/(1-\sigma)$ with $\sigma > 0$, and $V \equiv \delta U$, with $\delta \in (0,1)$. Define $\rho \equiv [\delta(1+r)]^{1/\sigma}$. Intergenerational wealth movements in the Becker-Tomes benchmark with stationary (w,r) then takes the form:

$$\tilde{W} = \frac{(1+r)\rho}{1+\rho+r}W + \frac{\rho}{1+\rho+r}w$$

if $W \geq \frac{w}{\rho}$, and $\tilde{W} = w$ otherwise. This allows us to calculate limit wealth:

$$\Omega(w,r) = \begin{cases} w & \text{if } \rho \leq 1, \\ \frac{\rho}{1-r(\rho-1)}w & \text{if } \rho \in (1, 1+\frac{1}{r}), \\ \infty & \text{if } \rho \geq 1+\frac{1}{r}. \end{cases} \tag{6.15}$$

If $\rho \leq 1$, there are no (limiting) financial bequests in steady state in the Becker-Tomes benchmark, and $\Omega(r,w) = w$. Limit wealth $\Omega(w,r)$ is finite only if $\rho \in (1, 1+\frac{1}{r})$.

If r is high enough that this condition is not satisfied, limit wealth is infinite: our version of the limited persistence property is satisfied,[14] whereas the Becker-Tomes version is not. We only point this out to emphasize that the Becker-Tomes version of limited persistence imposes more than a restriction on preferences, but in any case we are also fundamentally interested in the case in which limit wealth is finite.[15]

[13]The only subtlety here is one in which (6.14) holds with equality at $\Omega = w$, while the inequality $>$ holds for all $\Omega > w$. In this case both limits w and ∞ are potential candidates, but the correct limit for starting $W > w$ is easily seen to be the latter. This is a nongeneric case of little import but in any case our definition handles it.

[14]Notice that there are two limit wealths in this case, one at infinity and one at zero. But there is only one limit wealth provided we start with strictly positive wealth, and this is what [LP] requires.

[15]While the case of unbounded limiting wealth is formally a special case, ongoing growth really calls for a different model, ideally one in which training costs are endogenous as well.

A Comment on Nonpaternalism. We end this section by pointing out that preferences are inconsistent with the limited persistence assumption. By the envelope theorem, the derivative of the value function is just the marginal utility of (equilibrium) consumption. At a steady state, then, these marginal utilities drop out from the Euler equation, and the conditions become independent of wealth. In particular, an increase in parental steady state wealth translates into an equal increase in child wealth, so that [LP] fails. Dynastic preferences do not allow us to take the Becker-Tomes postulates fully on board.

Why do we want to incorporate [LP] in the first place? One answer is that it is empirically attractive. That may be so, but we do not impose [LP] for this reason. We do so because we want to show that this assumption generates a form of the steady state wage function that can never exhibit diminishing returns! If at all that wage function is nonlinear, it must be *non*convex at least over a region. This shows that the *assumption* that returns to human capital must be diminishing (Becker and Tomes (1986)) merits careful scrutiny, to say the least. None of these points emerge with dynastic preferences, and this — apart from financial bequests — is another basic difference from the analysis in Mookherjee and Ray (2003).

6.4.3 Two Occupations with No Financial Bequests. In what follows, we study an extremely simply case, one in which there are just two occupations, and financial bequests are nonexistent. You can suppose that we are in a world in which $\Omega(w, r) = w$ for all w.

Call the two occupations "skilled" and "unskilled" labor. For unskilled labor take the training cost to be zero. For skilled labor assume that there is a exogenous training cost X, which is just the number of units of the consumption good used as input into the training process. Let λ denote the fraction of the population at any date that is skilled. If some well-behaved production function f (satisfying the usual curvature and Inada end-point conditions) determines the wage to skill categories, the skilled wage at that date will be given by $\bar{w}(\lambda) \equiv f_1(\lambda, 1-\lambda)$, while the unskilled wage will be given by $\underline{w}(\lambda) \equiv f_2(\lambda, 1-\lambda)$. where subscripts denote appropriate partial derivatives.[16] This yields the following simple characterization: a fraction λ of skilled people is compatible with a steady state if and only if

$$
\begin{aligned}
U(\bar{w}(\lambda)) - u(\bar{w}(\lambda) - X) &\leq V(\bar{w}(\lambda) - x) - V(\underline{w}(\lambda))] \\
&\leq U(\underline{w}(\lambda)) - U(\underline{w}(\lambda) - X)
\end{aligned}
\tag{6.16}
$$

The left hand side of (6.16) represents the utility sacrifice of a skilled parent (hereafter denoted by $\kappa^s(\lambda)$) in educating its child, while the right hand side is the corresponding sacrifice for an unskilled parent (denoted by $\kappa^u(\lambda)$). The term in the middle is the present value benefit of all successive descendants being skilled rather than unskilled (which we shall denote by $b(\lambda)$).

[16]This applies only in the unrealistic event that skilled workers cannot perform unskilled tasks. More generally, if skilled workers can perform unskilled tasks, then the skilled wage cannot ever fall below the unskilled wage. So when the skill intensity λ is large enough that $f_1 < f_2$, wages will not be given by f_1 and f_2, but will be equalized (as a result of skilled workers filling unskilled positions whenever the latter pay higher wages). We omit this minor complication here because a competitive equilibrium with a positive fraction of skilled workers will never give rise to wage differentials that are incompatible with incentives for parents to educate their children.

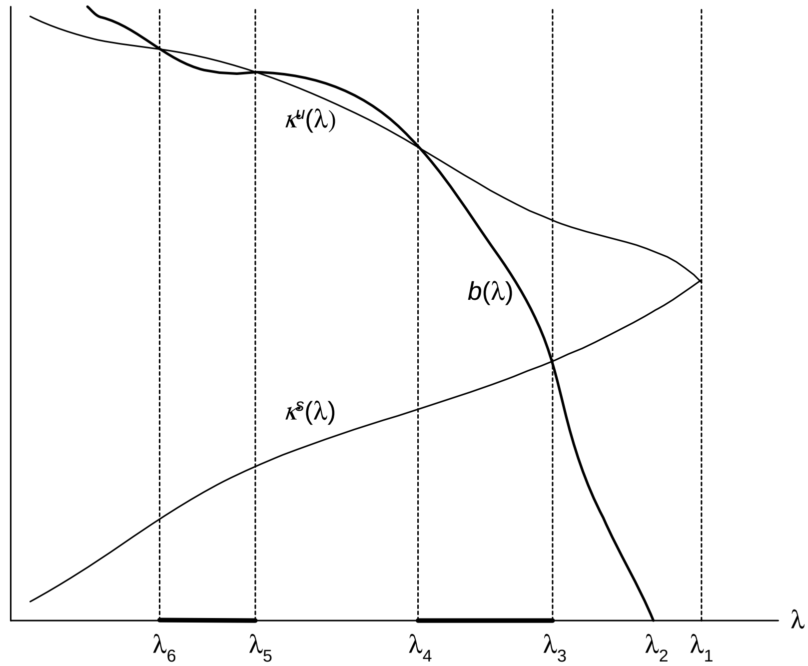


FIGURE 6.5. EDUCATION COSTS AND BENEFITS IN TWO-PROFESSION MODEL

These benefit and sacrifice functions are illustrated in Figure 6.5. $\lambda_1 \in (0, 1)$ denotes the skill intensity of the population at which the skill premium just disappears and the wages of the skilled and unskilled are equal. So κ^s and κ^u intersect there. Likewise, λ_2 is the point at which the wages of the skilled *net* of training equal those of the unskilled. So b drops to zero there. These observations can be used in conjunction with (6.16) to establish:

PROPOSITION 6.8. *There is a continuum of steady states in the two-profession model with exogenous training costs, and both per capita income and consumption rise as the skill proportion in steady state increases.*

Proposition 6.8 tells us that multiplicity — in the sense of a continuum of steady states — is endemic for a small number of professions. While stated only for the two-profession case, it is easy enough to extend the argument to any finite number of distinct professions.

Notice that the *structure* of the set of steady states may be complicated. In particular, the set need not be connected. For instance, in Figure 6.5, the set of steady states is the union of the two intervals (λ_6, λ_5) and (λ_4, λ_3).

The proposition also states that steady states are ordered not only in terms of skill premium but also per capita income: a steady state with a higher λ and lower skill premium corresponds to higher per capita income net of training costs. This does not, however, imply that these steady states are Pareto-ordered. For a detailed discussion of efficiency, see Mookherjee and Ray (2003).

6.4.4 A Rich Set Of Occupations. For a while, now, we shall abandon the idea of just two occupations and go to the opposite extreme, in which we examine what happens with a rich set of occupations. Two assumptions characterize richness:

[R.1] The set of all possible training costs is a compact interval of the form $[0, X]$.

[R.2] For every subset $C \subseteq [0, X]$ of positive Lebesgue measure, if the occupational distribution has zero value over every occupation h with $x(h) \in C$, then no output can be produced.

Thus we don't ask that every occupation be essential, only that (almost) every training cost in $[0, X]$ has an essential occupation attached to it. Observe that conditions [R.1] and [R.2] really go together as a pair: without some restriction like [R.2], [R.1] can always be trivially met by simply inventing useless occupations to fill up the gaps in training costs.

Together, [R.1] and [R.2] imply that whenever positive output is produced, the inhabited range of "equilibrium training costs" is always equal to $[0, X]$.

We justify richness by noting that while there are large differences in training costs between unskilled occupations (such as farm workers or manual jobs) and skilled occupations (such as engineers, doctors and lawyers), there are also many semi-skilled occupations (technicians, nurses and clerks) with intermediate training costs and wages. Besides, there are large differences in the quality of education within any given occupation, which translate into corresponding differences in education costs and wages.

The methodological innovation in the richness asumption is that it allows families to fully fine-tune their investments. Whether or not their investment set is convex depends, then, not on assumed indivisibilities in training costs but in the endogenously determined factor price schedule.

Note that [R.1] includes a bit more than occupational richness. It states that there is an occupation with zero training cost, a restriction that we impose for expositional ease.

6.4.5 Steady States With Rich Occupations. Fix a steady state. Say that an occupation (or training cost) is *inhabited* if some family chooses that occupation (or incurs that training cost). By the richness conditions, we know that every steady state (which has positive output, by definition) must exhibit a full measure of inhabited training costs. Suppose that we can alter the wage function on the small set of uninhabited occupations without changing *any* of the observed features of the steady state. Then we will say that the new steady state wage function (which only differs by specifying different wages for uninhabited occupations) is an *equivalent representation* of the old.

One particular representation is of interest, in which all occupations with the same training cost command the same wage. In that case (though with some abuse of notation), we shall go back and forth between $w(h)$ and the representation $w(x)$.

Proposition 6.9. *Assume* [R.1], [R.2] *and* [LP].

(a) *Every steady state has an equivalent representation with a continuous wage function.*

(b) *In an equal steady state, this equivalent representation may be described as follows: there exists* $w \geq 0$ *such that*

$$w(x) = w + (1+r)x \tag{6.17}$$

for all x. *In that steady state, all families attain a common wealth of* $\Omega(w, r)$, *and*

$$X \leq \frac{\Omega(w,r) - w}{1+r}, \tag{6.18}$$

where X *is the highest training cost across all occupations.*

(b) *In an unequal steady state, the equivalent representation may be described as a two-phase wage function: there exists* $w \geq 0$ *and* $\theta \in [0, X)$ *such that for all occupations* h *with* $x(h) \leq \theta$, (6.17) *holds:*

$$w(h) = w + (1+r)x(h).$$

Families that choose any of these occupations at any date all attain a common wealth of $\Omega(w, r)$ *that is precisely equal to* $w(\theta)$.

On the other hand, for all occupations h *with* $x(h) > \theta$, $\mathbf{w}$ *and* $\mathbf{x}$ *are connected via the following differential equation:*

$$w'(x) = \frac{U'(w(x) - x)}{V'(w(x))}. \tag{6.19}$$

with endpoint constraint that the wage at cost θ *equals* $w(\theta) = w + (1+r)\theta$.

Families in such occupations attain a wealth that is strictly greater than $\Omega(w, r)$, *and the marginal rate of return to these occupations,* $w'(x)$, *strictly exceeds* $1 + r$ *almost everywhere.*

The outline of the argument is as follows. Since (almost) every training cost must have at least one inhabited occupation, the marginal rate of return on training costs must be at least r everywhere. In other words, the presence of human capital allows parents to transfer wealth to their children at (weakly) lower cost than in a Becker-Tomes benchmark where only financial bequests are possible. Hence every family must attain at least the wealth $\Omega(w, r)$ that would have arisen in the latter context (corresponding to the flow wage w that is available to every generation even in the absence of any educational investment). In an equal steady state, the rate of return on human capital at all levels is exactly r, so that the set of investment options is exactly the same as in a Becker-Tomes benchmark with stationary (w, r). Therefore all families attain precisely the wealth $\Omega(w, r)$.

Finally, equation (6.18) must hold in such a steady state, because the occupation with the highest training cost X must have a wage of $w + (1+r)X$, and must be willingly chosen by *some* family.

In an unequal steady state it is trivial to see that the rate of return to *some* occupation must then exceed r. If not, the set of investment options would be just as in a Becker-Tomes benchmark with financial bequests alone, and wealth equality is then the only possible long-run outcome.

What is more subtle is the exact form the wage function must take. The proposition claims that the wage function has two phases. For low-end occupations, financial and human rates of return coincide and the wage function is linear. This phase has an endogenous delineation: any occupation with training cost below the steady state financial bequest in the Becker-Tomes benchmark with (w, r) must generate a rate of return of exactly r. (Note well that w, the lowest wage, is endogenous.)

For occupations with training costs that exceed the Becker-Tomes steady state bequest, the rates of return must be high enough to induce willing settlement. This requirement creates at least a local nonconvexity of the steady state wage function: the marginal rates of return to such occupations will *exceed* r.

A closer inspection of the differential equation (6.19) reveals that the shape of the wage function in the second phase relies entirely on preferences. To be sure (and as we shall see more clearly below), the existence and range of this phase will depend, among other things, on the technology. For a large class of preferences, the wage function exhibits a "global" nonconvexity, in the sense that the marginal rate of return rises monotonically with training costs beyond θ.

Observation 6.5. *Consider either the constant elasticity case with $U(c) = (c^{1-\sigma} - 1)/(1-\sigma)$, $\sigma > 0$, or the case of exponential utility $U(c) = -\exp(-\alpha c)$, $\alpha > 0$, and $V \equiv \delta U$. Then the marginal rate of return on occupations monotonically increases with training cost beyond the boundary θ described in Proposition 6.9.*

In the case of exponential utility, the wage function takes the form (for $x > \theta$):

$$w(x) = \frac{1}{\delta\alpha}\exp(\alpha x) + \Omega(w, r) - \frac{1}{\delta\alpha}\exp(\alpha\theta). \tag{6.20}$$

In the constant elasticity case $w'(x)$ monotonically increases to a finite asymptote that strictly exceeds $1 + r$.

This description stands the traditional theory on its head. That theory presumes — usually by assumption — that the rates of return to human capital must be *declining* in training cost (see, for instance, Loury (1981) and Becker and Tomes (1986)). Therefore the poorer families make all the human capital investment, and once families are rich enough so that the marginal return on human capital falls to the constant rate assumed for financial capital, all other bequests are financial.

In contrast, Proposition 6.9 is stated in a context in which the relative earnings of different occupations are allowed depend on the occupational distribution, for which there is considerable empirical evidence (e.g., Katz and Murphy (1992)). The proposition then asserts that the theory *endogenously* generates rates of return that run counter to the assumptions made in the literature. Financial bequests are made at the low end, while "occupational bequests" carry a higher rate of return and are made by richer families. If a rich set of occupations is essential, the concavity of returns to human capital is *never* an equilibrium outcome in any unequal steady state.

As an aside, we note that we haven't yet provided conditions for a steady state to be unequal, but will soon proceed to an analysis of this question.

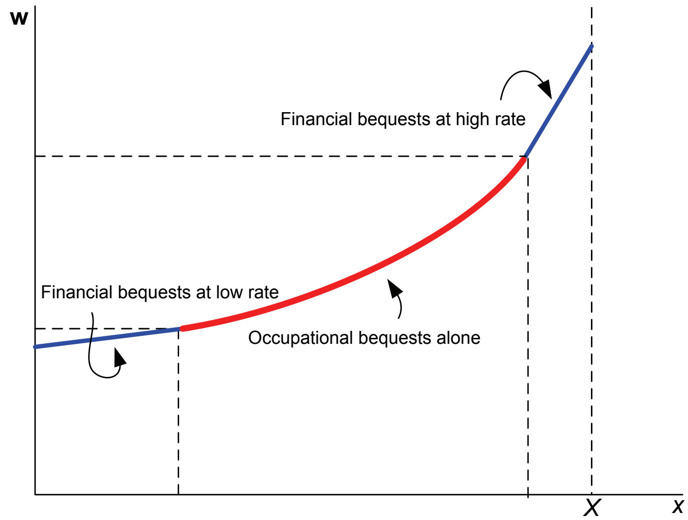


FIGURE 6.6. A EXTENSION TO THREE PHASES

It is important to see that the predicted shape of steady state wage functions depend fundamentally on the limited persistence assumption. If several occupations are essential in steady state, then to sustain the "wealth-spreading" necessitated by those occupations and to counteract convergence, wage functions must display a nonconvexity. A dynastic model, such as the one studied in Mookherjee and Ray (2003), fails [LP] and cannot deliver this property of wage functions in steady state.

Are our predictions counterfactual? We think not. First, it is well known from inequality decomposition studies that earnings inequality accounts for most of overall income inequality. For instance, Fields (2004) summarizes observations from several studies, writing that that "labor income inequality is as important or more important than *all other income sources combined* in explaining total income inequality".

Second, there is evidence that *within* the class of financial bequests, which are admittedly large for rich families, *intentional* bequests are not important. For instance, Gokhale *et al* (2001) argue that most financial bequests in the US economy are unintentional, the result of premature death and imperfect annuitization. In the iso-elastic example, this would correspond to the case with ρ below unity. Our theory then predicts that there are no *intentional* bequests anywhere in the wealth distribution, so human capital differences entirely account for all inequality, perhaps supplemented by unintended financial bequests (which we do not formally model).[17]

Third, it is important to remember that by "occupations", we mean not just human capital but every productive activity that is inalienable. This includes human capital but it is certainly not restricted to it. In particular, it is possible to view large financial bequests observed at the top end of the distribution as a form of occupational investment by parents, in the form of transfer of ownership or control of (partly inalienable) business activities.

[17] For a model of unintended bequests arising from uncertain life span, see, e.g., Fuster (2000).

Finally, it is easy to generate variants of this model which exhibit financial bequests at both upper and lower ends of the wealth distribution. For instance, suppose that financial wealth can be left at two rates of return: one at the existing rate r and another at a higher rate r', only accessible when the level of financial bequests crosses some threshold. Then it can be shows (I do not do this here) that in general, a steady state wage function will have a "three-phase property". Phases I and II look just like they do in this proposition. In phase III there is a return to linearity, this time at the higher rate r'. Financial bequests are left at both ends of the wealth distributrion. Figure 6.6 illustrates.

6.4.6 Existence and Uniqueness With Richness.

Existence. Having gained an understanding of the structure of steady states in this model, we can now provide conditions on the technology and preferences that guarantee the existence of a steady state.

Our definition of a steady state includes the requirement that output must be positive, so that existence is typically nontrivial. Proposition 6.9 informs us that a steady state must assume a particular form. In fact, it is easy to see that *given* some baseline wage w for unskilled labor, that proposition fully pins down the wage function. The only scope for variation lies in w. It comes as no surprise, then, that the existence of a (nondegenerate) steady state depends on the economy being productive enough to sustain positive profit at one of these conceivable wage functions.[18]

Of course, that isn't enough. Proposition 6.9 describes what a steady state necessarily looks like, but is silent on the question of whether such a description is indeed *sufficient* for all the steady state conditions. This, too, will need to be addressed.

We proceed, then, by searching for a steady state using the features described in Proposition 6.9. To this end, we describe the family of all *two-phase wage functions*. Start with a given baseline wage of w, and set $w(x) = w + (1 + r)x$ for all training costs no greater than

$$\theta(w) \equiv \min\{\frac{\Omega(w, r) - w}{1 + r}, X\}.$$

(This corresponds to the old threshold θ used in Proposition 6.9, but now we make the dependence on w explicit.) For occupations with higher training costs — if any — the wage function is set to satisfy the differential equation (6.19):

$$w'(x)) = \frac{U'(w(x) - x)}{V'(w(x))},$$

with the endpoint constraint that the wage for training cost $\theta(w)$ equals $w + (1 + r)\theta(w)$, or equivalently, $\Omega(w, r)$.

[18]There is a family resemblance here to the existence of steady states in the multisectoral optimal growth model, which requires a "productivity condition" (see Khan and Mitra (1986)). Fixed point arguments that neglect such a productivity condition, as in Sutherland (1970), are bound to fail because they do not eliminate the trivial outcome with zero output and wealth all around.

This procedure generates a unique wage function corresponding to any choice of w. Repeat this procedure for every $w \geq 0$: we now have the entire two-phase family, with either one of the phases conceivably degenerate.[19]

Given what we know already, the following "productivity" condition is necessary for the existence of a steady state with positive output:

[P] Unit costs $c(r, \mathbf{w})$ are less than or equal to 1 for some two-phase wage function.

It is easy enough to rewrite [P] as a productivity condition. For instance, if the production function is written as $Af(f, \boldsymbol{\lambda})$, where A is some Hicks-neutral productivity parameter, [P] states simply that A is large enough.

As the following proposition reveals, condition P (in conjunction with the other maintained assumptions) is also sufficient for the existence of a steady state with positive output.[20]

Proposition 6.10. *Under* [R.1], [R.2]*, and* [LP]*, a steady state with positive output exists if and only if* [P] *is satisfied.*

Condition [P] isn't at all difficult to verify, one way or the other. As an example, suppose that each training cost x corresponds to a unique occupation (so name it x as well), and that the production function takes the Cobb-Douglas form

$$\ln y = (1 - \alpha) \ln k + \int_0^X \alpha(x) \ln(\lambda(x)) dx + \ln A,$$

where A is a productivity parameter, $\alpha(x) \geq 0$ and $\int \alpha(x) dx = \alpha \in (0, 1)$. Then it is easy to see that for any wage function $\mathbf{w}$,

$$\ln c(r, \mathbf{w}) = (1 - \alpha)[\ln r - \ln(1 - \alpha)] + \int_0^X \alpha(x)[\ln(w(x)) - \ln(\alpha(x))] dx - \ln A.$$

The verification of [P] therefore simply entails the choice of a wage function that minimizes $\int \alpha(x) w(x)$, and then checking whether the resulting expression above is nonpositive.

Uniqueness. Proposition 6.9 already takes a significant step towards uniqueness by establishing that any steady state wage function, or at least its continuous equivalent representation, must lie in the two-phase class. As described in detail earlier, it is linear with return r over a range of training costs, and then displays a marginal rate of return that strictly exceeds r. In a broad class of cases (see, e.g. Observation 6.5), this marginal return can be shown to be ever-increasing, though it will usually possess a finite asymptote. Finally, the training-cost threshold separating the two phases precisely corresponds to the Becker-Tomes limit bequest with wage equal to the lowest wage along this function.

---

[19] If $\theta(w) = X$, the second phase is degenerate. Remember that at $w = 0$, $\Omega(0, r)$ is the *limit* of $\Omega(w, r)$ as $w \downarrow 0$. It may or may not equal 0, which is always trivially a limit wealth. If it is, the first phase is degenerate.

[20] If [P] fails, it is easy to construct a steady state with zero output: simply construct the two-phase wage function that starts from a baseline wage of zero, and place all individuals in the zero-wage occupation.

To be sure, the same economic fundamentals are consistent — at least in principle — with several wage functions drawn from this two-phase class. But this is what the uniqueness proposition rules out.

Proposition 6.11. *Assume* [R.1], [R.2], *and* [LP]. *Then, apart from equivalent representations which change no observed outcome, there is at most one steady state.*

As already discussed, this uniqueness proposition has far-ranging implications. Apart from Mookherjee and Ray (2003), to be discussed in more detail below, this observation has gone unnoticed in the literature because the literature typically concentrates on a sparse set of occupations (usually two, as in Galor and Zeira (1993) or occasionally three, as in Banerjee and Newman (1993)). In such cases multiplicity is indeed endemic, but once the set of occupations expands such multiplicity must shrink. We reiterate that the expansion of the set of occupations does *not* convexify the set of choices. Indeed, as Proposition 6.9 takes pains to explain, equilibrium *non*convexity is the rule rather than the exception.

Following up on this point, our uniqueness proposition does not rule out the path-dependence of economic fortunes for *individual* families. The identities of those who inhabit the different occupational slots is up for grabs and may — will — depend on historical accident. But their *numbers* cannot.

Proposition 6.11 is a substantial extension of the uniqueness theorem in Mookherjee and Ray (2003) to a context in which financial capital co-exists with human capital. Indeed, given the simplified context of our model,[21] the uniqueness result of Mookherjee and Ray (2003) can be seen very easily and intuitively.

Imagine reworking Proposition 6.9 by imposing the additional constraint that no financial bequests are permitted. One would reasonably suppose, then, that the first phase of the two-phase function would disappear, and that any steady state wage function must be governed by the differential equation (6.19) *throughout*. Now it is easy to see why there can be only one such wage function. If we begin at two different initial conditions and apply (6.19) thereafter, the two wage trajectories cannot cross — a well-known property for this class of differential equations. In short, if there are two steady state wage functions, *one must lie entirely above the other*. But now we have a contradiction, for two wage functions ordered in this way cannot *both* serve as bonafide supporting prices for profit maximization. We obtain uniqueness when there are no financial bequests.

While this serves as some intuition for the result at hand, different considerations emerge when financial bequests are permitted. Now crossings of the two putative steady state wage functions cannot be ruled out by taking recourse to uniqueness theorems for differential equations. After all, the behavior of the wage functions is *not* governed throughout by (6.19); a nontrivial "first phase" makes an appearance. Instead, the formal proof must rely on behavioral arguments, based on household optimization, to rule out such crossings.

[21] Apart from the central difference of financial bequests, there are two differences between our model and that of Mookherjee and Ray (2003). First, they use a nonpaternalistic bequest motive and work with value functions. Second, training costs are endogenously determined in their model. However, these differences are minor and can be readily accommodated.

6.4.7 Conditions for Inequality. We are now in a position a key question: under what conditions does the steady state of the model involve inequality rather than equality? In the language of Proposition 6.9, when must the second phase of the two-phase wage function necessarily be nonempty?

A simple preliminary exercise lays the groundwork for a complete characterization of this question. This exercise concerns the production technology alone and has nothing to do with preferences.

Consider the class of all linear wage functions of the form $w(x) = w + (1 + r)x$ defined on *all* of $[0, X]$, parameterized by $w \geq 0$.

Observation 6.6. *Assume* [P]. *Then there is a unique value of* w *— call it* a *— and a corresponding linear wage function* $\mathbf{w}^*$ *with* $w^*(x) = a + (1 + r)x$ *for all* x *— such that* $c(r, \mathbf{w}^*) = 1$.

Now a isn't an explicit parameter of our model. But for all intents and purposes it is an exogenous primitive. To compute a all one needs is a knowledge of the production function.

As an example, recall the Cobb-Douglas case studied in Section **??**, in which each training cost corresponds to a single occupation: Cobb-Douglas form

$$\ln y = (1 - \alpha) \ln k + \int_0^X \alpha(x) \ln(\lambda(x)) dx.$$

Using the same logic as in that case, it is easy to see that a must solve the equation

$$(1 - \alpha)[\ln r - \ln(1 - \alpha)] + \int_0^X \alpha(x)[\ln(a + [1 + r]x) - \ln(\alpha(x))]dx = 0,$$

provided that condition P holds.

We are now in a position to state our central result concerning persistent inequality.

Proposition 6.12. *Under* [R.1], [R.2], [LP], *and* [P], *the unique steady state is unequal if and only if*

$$\Omega(a, r) < a + (1 + r)X. \tag{6.21}$$

We shall refer to the inequality (6.21) as the *widespan condition*. It is made up of two parts. As we noted above, the parameter a is entirely a feature of the production technology and the range of the occupational structure. On the other hand, the Becker-Tomes limit wealth map Ω is entirely a feature of preferences. The widespan condition states that Becker-Tomes limit wealth — commencing from a — is *not* enough to "span" the entire range of occupations: it (net of a and discounted by the interest rate) is smaller than the span X. Proposition 6.12 declares that in all such cases, the steady state must exhibit persistent inequality.

One should be careful enough to note that (6.21) may not prescribe a unique threshold for the span X. Such an interpretation is suggested by the rewriting of the condition as

$$X > \frac{\Omega(a, r) - a}{1 + r},$$

but observe that both a and X depend on the training cost technology. To make this a more precise, consider economies that are identical in all respects except for their training cost functions, which are drawn from an ordered family (all starting at 0 for some occupation). Such an ordered family may be parameterized by the highest training cost X.

In this class, it is very easy to see that a depends negatively on X. Therefore provided that $\Omega(w, r) - w$ is nondecreasing in w, we do generate a single-threshold restriction on span: "there is X^* such that widespan holds if and only if $X > X^*$". This will be true, for instance, for isoelastic preferences: see (6.22) and the accompanying discussion below. More generally, though, widespan must hold for all training costs large enough (even though (6.21) may not imply a single threshold for X).

Uniqueness plus widespan tells us that there is just one steady state, but it must treat individuals unequally. Conversely, if widespan fails, there is convergence to a common level of wealth for all families. The proposition asserts, therefore, that whether the disequalization or the equalization view of the market is relevant depends on whether or not the widespan condition is satisfied.

Observe that under widespan, the market *must* act to separate identical or near-identical families, as described informally in Section **??**. Long-run equality is not an option. However, this paper does not contain an explicit account of dynamic equilibria from non-steady-state initial conditions.

The discussion following the statement of Proposition 6.11 continues to be relevant here. There is no history-dependence "in the large", as the steady state is unique. But just where an individual family will end up in that distribution profoundly depends on the distant history of that family.

The strength of the characterization result in Proposition 6.9 is such that it renders the proof of this proposition almost trivial. See Appendix.

6.4.8 Applications and Implications.

To illustrate the implications of the span condition in Proposition 6.12, it is useful to invoke the example of an iso-elastic utility function.

Recall, in particular, equation (6.15) and the discussion following it. If $\delta \leq \frac{1}{1+r}$, (6.21) is always satisfied and an equal steady state never exists. There are effectively no financial bequests in the limit, so the model reduces to the disequalization model in which financial bequests are not allowed. If, on the other hand, $\delta \geq 1 + \frac{1}{r}$ then $\Omega(a, r) = \infty$ and (6.21) fails. Financial bequests overwhelm any inequality arising from the need to provide occupational choice incentives, and an unequal steady state cannot exist.

In the intermediate case in which δ is neither too large nor too small, (6.15) tells us that

$$\Omega(a, r) = \frac{\rho}{1 - r(\rho - 1)} a,$$

where $\rho \equiv [\delta(1 + r)]^{1/\sigma}$, so that the widespan condition reduces to

$$\frac{\rho}{1 - r(\rho - 1)} a < a + (1 + r)X.$$

Rearranging, we obtain the following version of widespan:

$$X > a\frac{\rho - 1}{1 - r(\rho - 1)} \tag{6.22}$$

We now describe effects of varying parameters of the model, which are relevant to explaining cross-country differences, or effects of technological change.

(a) *Differences in TFP Levels.* Suppose we compare two countries which differ only in their levels of total factor productivity (TFP). Then for any common value of X, the poorer country has a lower value of a, implying that it is more prone to disequalization. Intuitively, the lower level of wages reduces the intensity of the parental bequest motive: they are less willing to undertake the educational investments for high-end occupations. The resulting shortage of people in high-end occupations causes a rise in the skill premium. This elicits the requires supply into high-end occupations, but makes wealth inequality more acute.

Technologically poorer countries are therefore more prone to disequalization.

Of course, this argument is based partly on the assumption that the range of training costs X is unaffected by wages. However, it is easy to incorporate this extension under the plausible assumption that both human and physical inputs enter into production. Then X lower in the unproductive country, but not by the same factor as a. This argument is obviously reinforced if poorer countries also possess a less productive educational technology.

(b) *Differences in TFP Growth Rates.* While TFP-related differences in poverty are positively associated with disequalization, higher *growth* may be positively related to it as well. For instance, if growth (from Hicks-neutral technical progress) causes all wages and costs to grow at a uniform rate, then — all other things being equal — the level of desired bequests will be dulled, raising the likelihood of disequalization.[22]

To the extent that poorer countries grow faster owing to a "catch-up" phenomenon in technology, the widespan condition is therefore more likely to hold on two counts: higher poverty *and* higher growth. Of course, the net result is ambiguous if subsequent growth isn't positively correlated with initial poverty.

(c) *Changes in Interest Rates.* A change in the rate of return to capital has subtle effects. When r rises, ρ also goes up. Both these effects work against the widespan condition, by raising the rate of return to financial bequests. So a first cut at this issue would suggest that an increase in the global rate of return to physical capital tends to be equality-enhancing. However, there is the possibility that a may be lowered by the increase in r. This effect runs in the opposite direction, and a full analysis is yet to be conducted.

(d) *Reliance on Physical Capital* Now let us compare economies with differing degrees of mechanization, i.e., reliance on physical capital *vis-a-vis* human capital in production. One

[22]We omit a formal demonstration of this assertion, which proceeds by deriving an equivalent of the widespan condition (6.22) in the presence of neutral technical progress.

simple way to do this is to suppose that final output is produced via a nested function

$$y = Ak^{\alpha}m^{1-\alpha},$$

where m is a composite of the occupational inputs: e.g., an intermediate good "produced" by workers. Then greater mechanization corresponds to a rise in α. Setting the marginal product of capital to the interest rate r, we obtain

$$\frac{k}{m} = \left(\frac{A}{r}\right)^{\frac{1}{1-\alpha}},$$

so that the indirect "reduced-form" production function is linear in m:

$$y = Bm,$$

where

$$B = A\left(\frac{A}{r}\right)^{\frac{\alpha}{1-\alpha}}.$$

Notice that B essentially prices the composite in terms of the final output. If B goes down for some reason, then $w(0)$ will decline. So a reduction in B, other things being equal, will contribute to a greater likelihood of disequalization. Whether B goes up or down with α depends on the ratio of A (TFP) to r.[23] In relatively "unproductive" economies in which A is small, an increase in physical capital intensity lowers B, making inequality more likely. The opposite is the case in "productive" economies in which A is large. We thus obtain an interesting answer to a classic question in the theory of distribution: the impact of greater mechanization in production on long-run inequality.

(e) *Wider Product Variety.* Wider occupational spans may be the outcome of introduction of new goods and services, owing to technological change. The production of new goods and services such as information and communication technology creates an entirely new set of occupations. Such occupations are likely to require high levels of education and training, which may be thought of as an increase in the span of occupations and associated training costs. Unlike the parameterization used in Proposition 6.12, such changes involve an increase both in X and in the productivity of the technology. In terms of (6.22), both X and a tend to rise and the net effect depends on the ratio of these two variables.

We have not yet analyzed this application in detail, but it is clear that the effect of wider product variety on inequality is a very important question.

(f) When (6.21) holds, the rising rate of return to occupations (as captured by Observation 6.5) is an important prediction of the theory. We have discussed this already, but it bears repetition that this *derived* property of the model contradicts the *assumed* properties of human capital in Becker and Tomes (1986) and others. Empirical research will take us closer to settling this issue.

[23] The intuitive reason why this is the appropriate comparison is that A becomes the productivity of capital in the limiting case when $\alpha = 1$.

6.4.9 A Summary. We've addressed two central questions in the theory of income distribution:

[I] Do competitive markets equalize or disequalize wealth allocations?

[II] Does history matter? Are the same economic fundamentals consistent with multiple steady states?

We study a model of intergenerational bequests which allows for both financial bequests as well the choice of a rich variety of occupations. A fundamental postulate of the model is that occupational inputs are *imperfect* substitutes, so that factor prices are endogenously determined. At the level of an individual household, occupational investments may be fine-tuned to an arbitrary degree, provided that there is rich variation in occupations and training costs. But the returns to these occupations are endogenous, and the equilibrium of the market will determine whether households face a convex or nonconvex investment technology.

We provide a complete characterization of steady states. We show that if there is inequality in steady state (so that question I has been answered affirmatively), then the steady state wage function must be linear over a section and convex over others, so that each family must face an investment nonconvexity. This finding is at sharp odds with existing literature that simply assumes the opposite: that the rate of retrun to human capital must exhibit diminishing returns.

Our characterization permits us to move on to the two central questions. We prove that with a rich set of occupations, the steady state must be unique. There may (and generally will) be path-dependence at the level of dynastic choices, but the *overall* distribution of outcomes must be independent of history. This provides an interesting negative answer to question II, and in this way challenges a literature that is predicated on steady state multiplicity.

The second principal goal is to address question I: we extend and unify three seemingly different views of market-driven inequality: *equalization*, embodied in traditional theories of income distribution, *disequalization*, central to the recent endogenous inequality literature, and *neutrality*, that either can happen depending on historical conditions. We show that a fundamental condition, which we call the widespan condition, allows us to predict whether the steady state must involve inequality or not.

The widespan condition draws attention to an aspect of technology that has received little attention in the literature: the range or "span" of occupational structure. Whether equality or inequality results from market mechanisms depends on this key parameter representing the extent of occupational diversity, relative to the strength of bequests.

The theory has several potential applications and implications.

We end with a final comment on question II. Our uniqueness result stems from the assumed richness in occupational investments. Whether or not such richness exists therefore deserves to be studied empirically. If the assumption is valid, theories of macroeconomic history dependence will have to be based either on interest rate multiplicities rooted in the lack of capital market integration, or political economy channels, rather than market-based occupational choice mechanisms.

6.4.10 Proofs. *Proof of Proposition 6.7.* Define $\bar{\kappa}(\lambda) \equiv u(\bar{w}(\lambda)) - u(\bar{w}(\lambda) - x)$; this is the utility cost of acquiring education when the parent is high. Define a similar utility cost for the low parent: $\underline{\kappa}(\lambda) \equiv u(\underline{w}(\lambda)) - u(\underline{w}(\lambda) - x)$. Let $b(\lambda) \equiv u(\bar{w} - x) - u(\underline{w})$ be the one-period gain to being high (assuming that the high parent also invests in her child and the low parent does not), and define $B(\lambda) \equiv (1 - \delta)^{-1} b(\lambda)$.

Finally, for any sequence $\{\lambda_s\}$ and for any date t, define

$$B_t \equiv \sum_{s=t}^{\infty} \delta^{s-t} b(\lambda_s).$$

This is the lifetime gain between a currently high and a currently low dynasty (starting from any date t), *assuming* that dynasties never switch their skill status.

The reason why B_t acquires salience is given by the following simple observation, which states that at every date, the *equilibrium* lifetime utility of the high (and low) must be equal to the utility they would have received *were their descendants never to switch status*. [To be sure, along the equilibrium path, switching of status will generally occur nevertheless.]

Lemma 6.1. *If $\{\bar{w}_t, \underline{w}_t, \lambda_t\}_{t=0}^{\infty}$ is a competitive equilibrium, then for each date t,*

$$\bar{V}_t = \sum_{s=t}^{\infty} \delta^{s-t} u(\bar{w}_s - x) \tag{6.23}$$

and

$$\underline{V}_t = \sum_{s=t}^{\infty} \delta^{s-t} u(\underline{w}_s), \tag{6.24}$$

so that in particular,

$$\bar{V}_t - \underline{V}_t = B_t \text{ for all } t. \tag{6.25}$$

Proof. It suffices to show that for each $t \geq 0$,

$$\bar{V}_t = u(\bar{w}_t - x) + \delta \bar{V}_{t+1}$$

and

$$\underline{V}_t = u(\underline{w}_t) + \delta \underline{V}_{t+1}.$$

To prove this, apply Observations 6.1 and 6.2. By Observation 6.1 and our restriction on λ_0, $\lambda_t \in (0, \tilde{\lambda})$ for all $t \geq 0$. Now using Observation 6.2, we may conclude that at all dates, some of the high people stay high, while some of the low people stay low. This is enough to establish the result. ∎

Thus along any competitive equilibrium, no dynasty will strictly prefer to switch skills, though it may well be the case that it strictly prefers to stay where it is. Lemma 6.1 yields, in turn

Lemma 6.2. *If* $\{\bar{w}_t, \underline{w}_t, \lambda_t\}_{t=0}^{\infty}$ *is a competitive equilibrium, then for every* t, $(\bar{w}_t, \underline{w}_t)$ *are the wages associated with* λ_t, *and*

$$\bar{\kappa}(\lambda_t) \le \delta B_{t+1} \le \underline{\kappa}(\lambda_t), \tag{6.26}$$

with

$$\lambda_{t+1} > \lambda_t \text{ only if } \delta B_{t+1} = \underline{\kappa}(\lambda_t) \tag{6.27}$$

and

$$\lambda_{t+1} < \lambda_t \text{ only if } \delta B_{t+1} = \bar{\kappa}(\lambda_t) \tag{6.28}$$

Proof. Let $\{\bar{w}_t, \underline{w}_t, \lambda_t\}$ be a competitive equilibrium. Then by definition, $(\bar{w}_t, \underline{w}_t)$ must be the wages associated with λ_t for every t. Using (6.23) and utility maximization, we see that

$$\bar{V}_t = \sum_{s=t}^{\infty} \delta^{s-t} u(\bar{w}_s - x) \ge u(\bar{w}_t) + \delta \underline{V}_{t+1},$$

so that (using (6.23) again)

$$u(\bar{w}_t) - u(\bar{w}_t - x) \le \delta[\bar{V}_{t+1} - \underline{V}_{t+1}]$$

with equality holding whenever a switch from "high" to "low" does occur along the equilibrium path. Invoking (6.25) of Lemma 6.1, we get half of (6.26) as well as (6.28). The same argument applied to a currently low dynasty gets us the other half of (6.26) and (6.27). ∎

The next step is central:

Lemma 6.3. *If* $\{\bar{w}_t, \underline{w}_t, \lambda_t\}_{t=0}^{\infty}$ *is a competitive equilibrium, then for every* t,

$$\max\{B(\lambda_t), \frac{1}{\delta}\bar{\kappa}(\lambda_t)\} \ge B_{t+1} \ge \min\{B(\lambda_t), \frac{1}{\delta}\underline{\kappa}(\lambda_t)\}. \tag{6.29}$$

Proof. It suffices to prove the result for $t = 0$. I first show that

$$\max\{B(\lambda_0), \frac{1}{\delta}\bar{\kappa}(\lambda_0)\} \ge B_1 \tag{6.30}$$

Suppose this assertion is false. Then I claim that there exists a *first* date $T \ge 0$ such that $\lambda_0 = \cdots = \lambda_T$ and

$$B_{T+1} \ge \frac{1}{\delta}\underline{\kappa}(\lambda_T). \tag{6.31}$$

Of course, if $B_1 \ge (1/\delta)\underline{\kappa}(\lambda_0)$, the claim is automatically true; otherwise $B_1 < (1/\delta)\underline{\kappa}(\lambda_0)$. We also have (by virtue of the presumption that (6.30) does not hold) that $B_1 > (1/\delta)\bar{\kappa}(\lambda_0)$, so that

$$\frac{1}{\delta}\underline{\kappa}(\lambda_0) < B_1 < \frac{1}{\delta}\bar{\kappa}(\lambda_0)$$

which — by Lemma 6.2 — implies that $\lambda_0 = \lambda_1$. Moreover, $B_1 = b(\lambda_1) + \delta B_2 = b(\lambda_0) + \delta B_2$, while $B(\lambda_0) = b(\lambda_0) + \delta B(\lambda_0)$, so that

$$B_2 - B(\lambda_0) = \frac{1}{\delta}[B_1 - B(\lambda_0)],$$

and simple manipulation of this equality shows that

$$B_2 = B_1 + \frac{1-\delta}{\delta}\epsilon, \tag{6.32}$$

where $\epsilon \equiv B_1 - B(\lambda_0) > 0$ (again, by the failure of (6.30)).

Following the same reasoning leading up to (6.32), as long as $B_{t+1} < (1/\delta)\underline{\kappa}(\lambda_t)$ — and as long as this is also true of all dates before t — we have $\lambda_0 = \cdots = \lambda_t = \lambda_{t+1}$, and

$$B_{t+1} = B_t + \frac{1-\delta}{\delta}\epsilon, \tag{6.33}$$

It follows that there must be a first date T such that $\lambda_0 = \cdots = \lambda_T$, and (6.31) holds, as claimed.

On the other hand, (6.31) cannot hold with strict inequality, as this would surely violate (6.26) of Lemma 6.2, so it must be that

$$B_{T+1} = \frac{1}{\delta}\underline{\kappa}(\lambda_T). \tag{6.34}$$

In turn, this means that $\lambda_{T+1} \geq \lambda_T$ (see (6.28)). Moreover, $\delta B_{T+2} = B_{T+1} - b(\lambda_{T+1})$, so that

$$\begin{aligned} \delta B_{T+2} - \underline{\kappa}(\lambda_{T+1}) &= B_{T+1} - b(\lambda_{T+1}) - \underline{\kappa}(\lambda_{T+1}) \\ &\geq B_{T+1} - b(\lambda_T) - \underline{\kappa}(\lambda_T) \\ &= B_{T+1} - b(\lambda_T) - \delta B_{T+1} \\ &= B_{T+1} - B_T > 0, \end{aligned} \tag{6.35}$$

which contradicts (6.26) at date $T+1$. This establishes (6.30).

Now we prove that

$$B_1 \geq \min\{B(\lambda_0), \frac{1}{\delta}\underline{\kappa}(\lambda_0)\} \tag{6.36}$$

The argument runs closely parallel to the previous one. Suppose (6.36) is false. Then I claim that there exists a *first* date $T \geq 0$ such that $\lambda_0 = \cdots = \lambda_T$ and

$$B_{T+1} \leq \frac{1}{\delta}\bar{\kappa}(\lambda_T). \tag{6.37}$$

The steps are very similar to those used to establish (6.31), and are omitted.

Continuing the parallel argument, (6.37) cannot hold with strict inequality, so we have

$$B_{T+1} = \frac{1}{\delta}\bar{\kappa}(\lambda_T). \tag{6.38}$$

In turn, this means that $\lambda_{T+1} \leq \lambda_T$. These two observations establish, however, that

$$\delta B_{T+2} - \bar{\kappa}(\lambda_{T+1}) < 0,$$

(following steps parallel to those establishing (6.35)), which contradicts (6.26) and completes the proof of the lemma. ■

The following observation is a simple consequence of Lemma 6.3:

Lemma 6.4. *Along any equilibrium, if* $B(\lambda_t) \le (1/\delta)\bar{\kappa}(\lambda_t)$*, then* $B_{t+1} = (1/\delta)\bar{\kappa}(\lambda_t)$*. Similarly, if* $B(\lambda_t) \ge (1/\delta)\underline{\kappa}(\lambda_t)$*, then* $B_{t+1} = (1/\delta)\underline{\kappa}(\lambda_t)$*.*

Proof. We prove the first part; the second part uses a completely analogous argument. If $B(\lambda_t) \le (1/\delta)\bar{\kappa}(\lambda_t)$, then $\max\{B(\lambda_t), (1/\delta)\bar{\kappa}(\lambda_t)\} = (1/\delta)\bar{\kappa}(\lambda_t)$, so that by (6.29), $(1/\delta)\bar{\kappa}(\lambda_t) \ge B_{t+1}$. On the other hand, Lemma 6.2 tells us that $(1/\delta)\bar{\kappa}(\lambda_t) \le B_{t+1}$, and the proof is complete. ■

With these steps in hand, we may complete the proof of the theorem. There are three possibilities to consider (each a restriction on the initial value λ_0):

I. $(1/\delta)\bar{\kappa}(\lambda_0) < B(\lambda_0) < (1/\delta)\underline{\kappa}(\lambda_0)$. Then by (6.29), $B(\lambda_0) = B_1$. In particular,

$$(1/\delta)\bar{\kappa}(\lambda_0) < B_1 < (1/\delta)\underline{\kappa}(\lambda_0)$$

so that by Lemma 6.2, $\lambda_0 = \lambda_1$. Continuing the argument recursively, we see that $\lambda_t = \lambda_0$ for all t.

II. $B(\lambda_0) \le (1/\delta)\bar{\kappa}(\lambda_0)$. Then by Lemma 6.4, $B_1 = (1/\delta)\bar{\kappa}(\lambda_0)$ and so by Lemma 6.2, $\lambda_1 \le \lambda_0$. Suppose, in fact, that strict inequality holds. Then $(1/\delta)\bar{\kappa}(\lambda_1) < (1/\delta)\bar{\kappa}(\lambda_0)$ and $(1/\delta)\underline{\kappa}(\lambda_1) > (1/\delta)\underline{\kappa}(\lambda_0)$, so that

$$(1/\delta)\underline{\kappa}(\lambda_1) > B_1 > (1/\delta)\bar{\kappa}(\lambda_1). \tag{6.39}$$

Now I claim that in fact,

$$(1/\delta)\underline{\kappa}(\lambda_1) > B(\lambda_1) > (1/\delta)\bar{\kappa}(\lambda_1). \tag{6.40}$$

Suppose not. First suppose that $B(\lambda_1) \le (1/\delta)\bar{\kappa}(\lambda_1)$. Then by Lemma 6.4, $B_2 \le (1/\delta)\bar{\kappa}(\lambda_1)$, so that $B_1 = (1-\delta)B(\lambda_1) + \delta B_2 \le (1/\delta)\bar{\kappa}(\lambda_1)$, which contradicts (6.39). In exactly the same way, one can rule out the possibility that $B(\lambda_1) \ge (1/\delta)\underline{\kappa}(\lambda_1)$, so (6.40) is established.

Now we are in Case I, and λ must remain constant thereafter. So in Case II, we move to a steady state *in at most one step*.

III. $B(\lambda_0) \ge (1/\delta)\underline{\kappa}(\lambda_0)$. Then by Lemma 6.4, $B_1 = (1/\delta)\underline{\kappa}(\lambda_0)$ and so by Lemma 6.2, $\lambda_1 \ge \lambda_0$.

To bring out the contrast with Case II, I claim that $B(\lambda_1) \ge (1/\delta)\underline{\kappa}(\lambda_1)$, with strict inequality if the corresponding inequality at date 0 also holds strictly.

Suppose on the contrary that $B(\lambda_1) < (1/\delta)\underline{\kappa}(\lambda_1)$. Then

$$B_1 = (1/\delta)\underline{\kappa}(\lambda_0) \ge (1/\delta)\underline{\kappa}(\lambda_1) > B(\lambda_1), \tag{6.41}$$

while it is also true that

$$B_1 = (1/\delta)\underline{\kappa}(\lambda_0) \ge (1/\delta)\underline{\kappa}(\lambda_1) \ge B_2. \tag{6.42}$$

But (6.41) and (6.42) together contradict the fact that $B_1 = (1-\delta)B(\lambda_1) + \delta B_2$.

If strict inequality holds at date 0 — $B(\lambda_0) > (1/\delta)\underline{\kappa}(\lambda_0)$ — then we can arrive at a contradiction simply under the weaker condition $B(\lambda_1) \leq (1/\delta)\underline{\kappa}(\lambda_1)$. For then $\lambda_1 \neq \lambda_0$ and therefore (because $\lambda_1 \geq \lambda_0$) it must be that $\lambda_1 > \lambda_0$. Therefore the weak inequality in (6.41) holds strictly, and we obtain the same contradiction.

So the claim is established and we can apply Case III repeatedly to argue that $\lambda_{t+1} \geq \lambda_t$ for all t in this case.

Moreover, if $B(\lambda_0) > (1/\delta)\underline{\kappa}(\lambda_0)$, then $B(\lambda_t) > (1/\delta)\underline{\kappa}(\lambda_t)$ for all t subsequently, so convergence *cannot ever occur in finite time*, in contrast to the "one-step" property of Case II.

Finally, observe that in the case $B(\lambda_0) > (1/\delta)\underline{\kappa}(\lambda_0)$, in which $\lambda_{t+1} > \lambda_t$ for all t, there is no t and no $\lambda \in [\lambda_t, \lambda_{t+1}]$ which is a steady state. For suppose there were; then in particular, $\underline{\kappa}(\lambda) \geq B(\lambda)$, so that

$$\underline{\kappa}(\lambda_t) \geq \underline{\kappa}(\lambda) \geq B(\lambda) \geq B(\lambda_{t+1}) > B_{t+1}, \tag{6.43}$$

where the very last inequality follows from the fact that $\lambda_s < \lambda_{s+1}$ for all s. But (7.42) contradicts the equilibrium condition (6.26). This proves that in Case III, convergence occurs to the *smallest* steady state to the right of λ_0. ■

Proof of Observation 6.3. Let W_t denote the wealth of a typical member of generation t. We claim that if $W_{t+1} < W_t$, $W_{s+1} \leq W_s$ for all $s > t$. Suppose not; then there exist dates τ and s with $s > \tau$ such that (a)$W_{\tau+1} < W_\tau$, (b) $W_{s+1} > W_s$, and (c) $W_m = W_{\tau+1} = W_s$ for all intermediate dates $\tau+1 \leq m \leq s$ (this last requirement is, of course, vacuous in case $s = \tau+1$). But then a *strictly* higher wealth (W_τ compared to W_s) generates a strictly lower descendant wealth ($W_{\tau+1}$ compared to W_{s+1}), which contradicts a familiar "single-crossing" argument based on the strict concavity of the utility function. Therefore dynastic wealth converges in this case.

Next suppose that $W_{t+1} > W_t$. Then a reversal of the same logic implies that family wealth is nondecreasing across generations. ■

Lemma 6.5. *In the Becker-Tomes benchmark, $\tilde{W} - W$ has exactly the same sign (including equality) as*

$$(1+r) - \frac{U'(\frac{rW+w}{1+r})}{V'(W)}. \tag{6.44}$$

Proof. Simply recall the first-order conditions (8.42) and the concavity of U and V. ■

Proof of Observation 6.4. We first show that under (H),

$$U'(\frac{rW+w}{1+r})/V'(W) \text{ is increasing in } W, \tag{6.45}$$

for all $w > 0$ and $r > -1$. It is sufficient to show that

$$V'(W)\frac{r}{1+r}U''(\frac{rW+w}{1+r}) - U'(\frac{rW+w}{1+r})V''(W) > 0$$

which is equivalent to

$$-\frac{U''(\frac{rW+w}{1+r})}{U'(\frac{rW+w}{1+r})}\left(\frac{r}{1+r}\right) < -\frac{V''(W)}{V'(W)}.$$

Given (H), this reduces to the condition that

$$\frac{r}{1+r} < \frac{\alpha+\beta[\frac{rW+w}{1+r}]}{\alpha+\beta W},$$

which is always true *unless* $w = 0$ *and* $\alpha = 0$. This immediately verifies [LP], barring these exceptional cases.

If $w = 0$ and $\alpha = 0$, then $U'(\frac{rW+w}{1+r})/V'(W)$ is unchanging in W, so that [LP] is verified for all but one value of r. ■

To prepare for the proofs of the remaining propositions, we record several lemmas, and we presume (often implicitly) that [R], [E] and [LP] apply where needed.

Lemma 6.6. *In the Becker-Tomes benchmark, for every $w > 0$ and $r > -1$:*

(a) *If $w \le W < \Omega(w,r)$, then $W < \tilde{W}(W;w,r) \le \Omega(w,r)$. If $W > \Omega(w,r)$, then $\Omega(w,r) \le \tilde{W}(W;w,r) < W$.*

(b) $\Omega(w,r)$ *is nondecreasing in w.*

(c) $\Omega(w,r)$ *is continuous at every (w,r) at which it is finite.*

Proof. Part (a) follows simply from the fact that $\tilde{W}(W;w,r)$ is nondecreasing in W. If, for instance, $\tilde{W}(W;w,r) \le W < \Omega(w,r)$, or if $\tilde{W}(W;w,r) > \Omega(w,r) > W$, limit wealth starting from W can never equal $\Omega(w,r)$. This proves the first assertion in part (a); the second assertion follows in similar fashion.

Part (b). We claim first that $\tilde{W}(W;w,r)$ is nondecreasing in w. To show this, it suffices to assume that $\tilde{W} > w$, but then it must be the case that

$$U'\left(W - \frac{\tilde{W}-w}{1+r}\right) \le (1+r)V'(\tilde{W}). \tag{6.46}$$

Now suppose that w goes up to w', but $\tilde{W}$ falls to $\tilde{W}'$. The new situation must therefore have strictly positive consumption for the parent, so that the new first-order condition has the opposite weak inequality:

$$U'\left(W - \frac{\tilde{W}'-w'}{1+r}\right) \ge (1+r)V'(\tilde{W}'). \tag{6.47}$$

But (6.46) and (6.47) together contradict the assumption that U and V are strictly concave.

With this claim in place, the result follows immediately from [LP].

Part (c). Suppose that $\Omega(w,r)$ is finite. By part (a), $\tilde{W}(W;w,r)$ (viewed as a function of W) must "strictly intersect the 45^0 line" at the value $\Omega(w,r)$. The continuity of $\tilde{W}(W;w,r)$ now assures us that $\Omega(w,r)$ must be locally continuous. ■

Lemma 6.7. *For any steady state wage function:*

(a) *If h is inhabited, $x(h) = x(h')$ implies $w(h) \geq w(h')$.*

(b) *If h is inhabited, then $x(h) > x(h')$ implies $w(h) - w(h') \geq (1+r)[x(h) - x(h')]$.*

Proof. If (a) is false and $x(h) = x(h')$ while $w(h) < w(h')$ then any parent selecting occupation h for her child would do better to select occupation h' instead. The same is true if (b) were false: the parent could switch to occupation h' for the child, combined with a higher financial bequest so as to leave the child's wealth unaffected, while increasing her own consumption. ■

As in the main text, it will often be convenient to write the wage as a function of training cost: $w(x)$. Part (a) of Lemma 6.7 informs us that we can certainly do this right away for inhabited training costs.

Lemma 6.8. (a) *The set of all inhabited training costs — call it T — is a subset of* $[0, X]$ *of full measure, and* $\mathbf{w}$ *is continuous on T.*

(b) *There is a unique continuous extension of* $\mathbf{w}$ *on T to all of* $[0, X]$*, and it forms an equivalent representation.*

Proof. Part (a). Let T be the set of all inhabited training costs. Since a steady state must have positive output by definition, it follows from [R] and [E] that T must be of full measure. Moreover, $\mathbf{w}$ (viewed as a function of x) must be continuous on T. For if not, we can select training costs x and x' in T that are arbitrarily close, but such that their wage difference is bounded away from zero. In that case no parent would select the (almost identical) training cost with a lower wage.

Part (b). Because $\mathbf{w}$ is nondecreasing on T and T is of full measure, there is a unique continuous extension of $w(x)$ to all of $[0, X]$. Use this extension to define $\hat{w}(h)$ for all occupations, inhabited or not. That is, $\hat{w}(h) = w(h)$ if h is inhabited, and equals the continuous extension otherwise. We first claim that $\hat{w}(h) \geq w(h)$ for all h that are uninhabited. For if not, we have a contradiction in a manner similar to part (a): we can find an inhabited occupation h' arbitrarily close to h but with wages bounded below that of $w(h)$, which means that all occupiers of h' would prefer h, a contradiction.[24]

By this claim, if we replace $\mathbf{w}$ by $\hat{\mathbf{w}}$, no firm will wish to change its desired input mix (unused inputs have not become any cheaper). To complete the proof of equivalence, observe that no family occupying h' finds it strictly profitable to switch to an uninhabited occupation h once its wage has been replaced by $\hat{w}(h)$. For if this were true, then by the definition of continuous extension we can find a third *inhabited* occupation h'' with wage and training cost arbitrarily close to that of h, such that the family must therefore also find it profitable to switch from h' to h''. But this is a contradiction, since that option is already available in the going steady state. ■

[24] This argument, as well as its counterpart for (a), can be made entirely precise by showing that the preference for h over h' can be made uniform over all families, irrespective of their wealth.

In what follows we focus exclusively on this equivalent representation and call it $\mathbf{w}$ instead of $\hat{\mathbf{w}}$. Define $w = w(0)$.

Lemma 6.9. *Every family attains a wealth of at least* $\Omega(w, r)$.

Proof. Let W be the steady state wealth of some family. Suppose that it incurs training cost x and leaves bequest b; then by stationarity, $W = (1+r)b + w(x)$. Now, *given* the choice of x, b must maximize

$$U(W - x - b) + V(w(x) + (1+r)b) \tag{6.48}$$

subject to $b \geq 0$. Defining $B \equiv b + x$, this can equivalently be written as: B is chosen to maximize

$$U(W - B) + V(w' + (1+r)B) \tag{6.49}$$

subject to $B \geq w(x)$, where $w' \equiv w(x) - (1+r)x$. It is obvious that the solution to this problem involves a total bequest B at least as large as the value that would obtain if the constraint $B \geq w(x)$ were replaced by $B \geq 0$, i.e., if we were in a Becker-Tomes benchmark world with w' and r. Remembering that W is also next period's wealth in the problem (6.49), we conclude that

$$W \geq \tilde{W}(W; w', r). \tag{6.50}$$

We now claim that

$$W \geq \Omega(w', r). \tag{6.51}$$

If this is false, then $W < \Omega(w', r)$. Combining this with (6.50), we may conclude that $\tilde{W}(W; w', r) \leq W < \Omega(w', r)$, but this contradicts part (a) of Lemma 6.6.

Now recall the definition of w' and invoke part (b) of Lemma 6.7 to conclude that $w' \geq w$. Use this fact and part (b) of Lemma 6.6 to observe that $\Omega(w', r) \geq \Omega(w, r)$. Combine this last inequality with (6.51) to obtain the desired result. ∎

Lemma 6.10. *Let W be any lower bound on stationary wealth across all families.*

(a) *If for any occupation h, we have $w + (1+r)x(h) \leq W$, then $w(h) = w + (1+r)x(h)$.*

(b) *The stationary wealth of any family selecting an occupation h with $w + (1+r)x(h) \leq W$ must be $\Omega(w, r)$.*

Proof. Part (a). Suppose, on the contrary, that $w + (1+r)x(h) \leq W$, but $w(h) \neq w + (1+r)x(h)$. By Lemma 6.7, part (b), and the fact that $\mathbf{w}$ is a continuous equivalent representation, it must be that

$$w(h) > w + (1+r)x(h).$$

so that for some $x' < x(h)$ but close to it,

$$w(x') > w + (1+r)x'. \tag{6.52}$$

Now, we know that there is a sequence of inhabited occupations $\{h^k\}$ such that $x(h^k) \downarrow 0$. By assumption, W is a lower bound on stationary wealth across all families, and $w + (1+r)x' <$

$w + (1 + r)x(h) \leq W$. Therefore (for k large enough) such families are almost exclusively leaving financial bequests of magnitude at least x' to their children at rate of return r. They would be better off, instead, selecting an occupation with cost x' for their children (which yields a return of strictly more than r, by (6.52), and supplementing the remainder with financial bequests, a contradiction.

For each family selecting occupation h with $w + (1 + r)x(h) \leq W$, we have $w(h) = w + (1 + r)x(h)$. Therefore the realized rate of return to *all* the choices of such a family, financial and educational, is exactly r. Define $\hat{x}$ by

$$w + (1 + r)\hat{x} = W. \tag{6.53}$$

Once again, using Lemma 6.7, part (b), and the fact that $\mathbf{w}$ is a continuous equivalent representation, we also know that wages for training costs beyond $\hat{x}$ yield *no less* a return than r for all educational investments beyond $\hat{x}$. Yet these families choose (by part (a)) not to utilize such regions of educational investment. They must therefore be behaving in exactly the same way as in a Becker-Tomes benchmark world with parameters (w, r). We must conclude that their stationary wealth equals $\Omega(w, r)$. ■

Lemma 6.11. *In an equal steady state, the common wealth of all families must be $\Omega(w, r)$, and $\Omega(w, r) \geq w + X(1 + r)$.*

Proof. Let W denote the (common) wealth of each family in an equal steady state. We claim that

$$W \geq w + (1 + r)X. \tag{6.54}$$

For if not, we know from Lemma 6.9, part (b) (and the fact that $\mathbf{w}$ is a continuous equivalent representation) that there are inhabited occupations with training costs x arbitrarily close to X, and that

$$w(x) \geq w + (1 + r)x.$$

Consequently, if (6.54) were to be false, we would find inhabited occupations with $w(h) > W$. Because the wealth of families in such occupations is at least $w(h)$, this is a contradiction to equality.

It remains to prove that $W = \Omega(w, r)$. Because (6.54) is true, and because W is (trivially) a lower bound on stationary wealth, Lemma 6.10, part (b) applies to all families, and $W = \Omega(w, r)$. ■

Lemma 6.12. *A steady state is equal if and only if the continuous equivalent representation of the wage function is linear:*

$$w(h) = w + (1 + r)x(h) \tag{6.55}$$

for all occupations h.

Proof. For necessity, combine Lemma 6.10, part (a), and Lemma 6.11. For sufficiency, note that if (6.55) holds, then we are in a Becker-Tomes benchmark world and all limit wealths must be the same. ■

The remaining steps concern unequal steady states. Define θ by

$$w + (1+r)\theta = \Omega(w, r). \tag{6.56}$$

Lemma 6.13. *In an unequal steady state, it must be that $\theta < X$.*

Proof. By Lemma 6.9, $\Omega(w, r)$ is a lower bound on stationary wealth for all families. If, contrary to our assertion, $\theta \geq X$, then part (a) of Lemma 6.10 applies for *every* training cost, when W is replaced by $\Omega(w, r)$. Therefore $\mathbf{w}$ satisfies (6.55), and the sufficiency direction of Lemma 6.12 implies that the steady state must be equal, a contradiction. ∎

We are interested in the shape of $\mathbf{w}$ in the region $U \equiv [\theta, X]$.

Lemma 6.14. *Let I be some subinterval of U such that no financial bequests are made by any family that occupies some occupation with training cost in U. Then $\mathbf{w}$ satisfies* (6.19) *on I.*

Proof. Fix any $x \in I$, with $x < \sup I$. For $\epsilon > 0$ but small enough, $x + \epsilon \in I$ as well. Assume provisionally that both x and $x + \epsilon$ are inhabited. Then family wealth at x (resp. $x + \epsilon$) is merely $w(x)$ (resp. $w(x + \epsilon)$). Using the two optimality conditions, one for families with wealth $w(x)$ and the other for families with wealth $w(x + \epsilon)$, we see that

$$U(w(x)-x)-U(w(x)-(x+\epsilon)) \geq V(w(x+\epsilon))-V(w(x)) \geq U(w(x+\epsilon)-x)-U(w(x+\epsilon)-(x+\epsilon)). \tag{6.57}$$

Now, using the fact that $\mathbf{w}$ is a continuous equivalent representation, and invoking [R] and [E], we can see that (6.57) actually applies to *all* x and $x + \epsilon$ in I, not just those that are inhabited.[25]

Dividing these terms throughout by ϵ, applying the concavity of the utility function to the two side terms, and the mean value theorem to the central term, we see that

$$U'(w(x) - (x + \epsilon)) \geq V'(\gamma(\epsilon))\frac{w(x + \epsilon) - w(x)}{\epsilon} \geq U'(w(x + \epsilon) - x),$$

where $\gamma(\epsilon)$ lies between $w(x)$ and $w(x + \epsilon)$. Now send ϵ to zero and use the continuous differentiability of U and V to conclude that the right-hand derivative of w with respect to x — call it $w^+(x)$ — exists, and

$$w^+(x) = U'(w(x) - x)/V'(w(x)).$$

By exactly the same argument applied to x (greater than $\inf I$) and $x - \epsilon$, we may conclude the same of the left-hand derivative, which verifies (6.19). ∎

Lemma 6.15. *Any two-phase wage function has $w'(x) > 1 + r$ for almost all $x > \theta$.*

Proof. The continuous differentiability of U and V, and the continuity of $\mathbf{w}$ together imply that $\mathbf{w}$ is continuously differentiable in its second phase, where it follows (6.19). Note moreover that $w'(\theta) = 1 + r$. Therefore, if the assertion is false, there is an interval $[x_1, x_2]$, with $x_1 \geq \theta$, on which $w'(x) \leq 1 + r$, while $w'(x_1) = 1 + r$ and $w(x_1) - (1 + r)x_1 \geq 0$.[26] Applying

[25] Given [R] and [E], we can approach both x and $x + \epsilon$ by a sequence of inhabited training cost pairs in I, and for each such pair the inequality (6.57) holds.

[26] One way to assure the existence of an interval with all these properties is to take x_1 to be the minimum of the values among those greater than θ for which $w'(x) \leq 1 + r$.

(6.19) at x_1 and using $w'(x_1) = 1 + r$, we see that

$$U'(w(x_1) - x_1)) = (1 + r)V'(w(x_1)). \tag{6.58}$$

Define $\hat{w} = w(x_1) - (1 + r)x_1$, which is nonnegative by construction. Then (6.58) reduces to

$$U'\left(\frac{rw(x_1) + \hat{w}}{1 + r}\right) = (1 + r)V'(w(x_1)), \tag{6.59}$$

By the same logic as (6.58), except that at x_2 we have $w'(x_2) \leq 1 + r$, we see that

$$U'(w(x_2) - x_2)) \leq (1 + r)V'(w(x_2)). \tag{6.60}$$

Now observe that $w(x_1) + (1 + r)(x_2 - x_1) \geq w(x_2)$ (because $w'(x) \leq 1 + r$ on $[x_1, x_2]$), or equivalently, using the definition of $\hat{w}$, $x_2 \geq \frac{w(x_2)-\hat{w}}{1+r}$. Consequently,

$$w(x_2) - x_2 \leq \frac{rw(x_2) + \hat{w}}{1 + r},$$

and using this in (6.60) along with the concavity of U, we must conclude that

$$U'\left(\frac{rw(x_2) + \hat{w}}{1 + r}\right) \leq (1 + r)V'(w(x_2)). \tag{6.61}$$

Suppose that $w(x_1) > 0$. Then (6.59) means that $w(x_1)$ is a positive limit wealth in the Becker-Tomes benchmark with $(\hat{w}, r)$, while Lemma 6.5 tells us that $\tilde{W}(w(x_2); \hat{w}, r) \geq w(x_2)$. Because $w(x_2) > w(x_1)$, this means that $w(x_1)$ *cannot* be a limit wealth starting from initial wealth $w(x_2)$. This contradicts [LP].

On the other hand, if $w(x_1) = 0$, then $w = 0$ and $\theta = 0$ as well, which means that $\Omega(0, r) = 0$: the limit of Becker-Tomes wealth from all *positive* initial wealths is zero when $w = 0$.[27] But this fact is contradicted by Lemma 6.5 applied to (6.61), because $\hat{w}$ is also 0 in this case. ■

If a wage function satisfies $w(x) - w(x') = (1 + r)(x - x')$ for all x and x' in some interval, say that it is *r-linear* over that interval. We know, for instance, that any two-phase wage function with a nondegenerate first phase indeed r-linear over $[0, \theta(w)]$.

Lemma 6.16. *Suppose that a family in steady state, inhabiting training cost x at some date, also makes a financial bequest at that date. That is, it possesses (and bequeaths) total wealth W, where $W > w(x)$. Then $\mathbf{w}$ is r-linear over all $x' \geq x$ with $w(x) + (1 + r)(x' - x) \leq W$:*

$$w(x') = w(x) + (1 + r)(x' - x). \tag{6.62}$$

Proof. The proof is very similar to that of Lemma 6.10. Pick any x' with $w(x)+(1+r)(x'-x) \leq W$. Our family is making a financial bequest of at least $x' - x$. If (6.62) were to fail, then by Lemma 6.7, part (b), and the fact that $\mathbf{w}$ is a continuous equivalent representation, we must have

$$w(x') > w(x) + (1 + r)(x' - x),$$

which means that our family would certainly be better off choosing x' instead of x and supplementing the remainder (if any) with financial bequests, a contradiction. ■

[27] Remember: this is a different statement from the one that asserts that 0 is a Becker-Tomes limit wealth when $w = 0$, which is always trivially true.

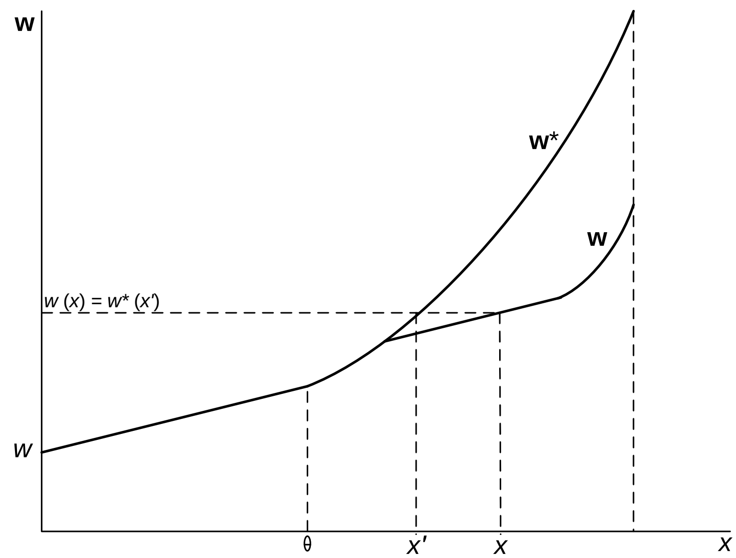


FIGURE 6.7. Proof of Lemma 6.18

The next lemma summarizes what we know so far about an unequal steady state.

LEMMA 6.17. *The continuous equivalent-representation of any unequal steady state wage function is r-linear up to θ, followed by combinations of intervals over which either the differential equation* (6.19) *is obeyed, or r-linearity holds.*

PROOF. Combine Lemmas 6.14 and 6.16. ■

However, we now establish a stronger property:

LEMMA 6.18. *In an unequal steady state, the continuous, equivalent-representation wage function must be two-phase.*

PROOF. Let **w** be a (continuous) steady state wage function, starting from w. Denote by $\mathbf{w}^*$ the two-phase wage function starting from the same point. We know already that the two functions coincide at least up to θ. Suppose, contrary to the assertion, that $w^*(x) \neq w(x)$ for some $x \in (\theta, X]$. Then there is some *first* r-linear segment "after" θ at which **w** departs from $\mathbf{w}^*$.

By Lemma 6.15, **w** must lie below $\mathbf{w}^*$ in this segment. Use Figure 6.7 as a guide in what follows.

Pick some inhabited x in the interior of the r-linear segment; then pick $x' \in (\theta, x)$ such that

$$w^*(x') = w(x) > 0. \tag{6.63}$$

Pick any family that inhabits x at any date, and has stationary wealth W. The first-order conditions for utility maximization tell us that

$$U'(\frac{rW + w}{1 + r}) = (1 + r)V'(W),$$

so that $W > 0$ is a limit wealth in the Becker-Tomes benchmark with baseline wage $w' = w(x) - (1+r)x$ and rate of return r. By [LP] applied to this benchmark, a family with starting wealth $W' = w(x) > 0$ in this benchmark world must converge to the very same limit wealth. Because $W' = w(x) \leq W$, convergence to W must require that $\tilde{W}(W'; w', r) \geq W'$. Invoking Lemma 6.5 and writing $w(x)$ in place of W', we must conclude that

$$(1+r)V'(w(x)) \geq U'(\frac{rw(x) + w'}{1+r}) = U'(w(x) - x).$$

Now recall the definition of x' from (6.63). Replacing $w(x)$ by the same value $w^*(x')$, replacing x by the smaller value x', and using the strict concavity of U, we see that

$$(1+r)V'(w^*(x')) > U'(w^*(x') - x'),$$

but this contradicts the fact that $\mathbf{w}^*$ satisfies (6.19) at x'. ■

Proof of Proposition 6.9. Lemma 6.8 establishes that there is a continuous equivalent representation to the wage function in every steady state. Lemma 6.11 shows that in any equal steady state, $\Omega(w, r) \geq w + (1+r)x$, where w is the lowest wage in that steady state wage function. Lemma 6.55 shows that the wage function must be r-linear for equal steady states. Lemma 6.18 shows that unequal steady state wage functions must have the two-phase property: it is r-linear up to θ, which is defined in (6.56), and follows the differential equation (6.19) thereafter. Lemma 6.15 shows that the second phase must exhibit a rate of return that is almost everywhere higher than r. ■

Proof of Observation 6.5. The differential equation (6.19) in the exponential utility case reduces to

$$w'(x) = \frac{1}{\delta\alpha}\exp(\alpha x) \tag{6.64}$$

from which the stated result follows. Applying (6.19) to the constant elasticity case, we see that for all $x \geq \theta$,

$$w'(x) = \frac{1}{\delta}\left[\frac{w(x)}{w(x) - x}\right]^{\sigma}. \tag{6.65}$$

Differentiation of this equality shows us that

$$w''(x) = \sigma\left[\frac{w(x)}{w(x) - x}\right]^{\sigma - 1}\frac{w(x) - xw'(x)}{[w(x) - x]^2},$$

so that $w''(x)$ is continuous and has precisely the same sign as $w(x)/x - w'(x)$. Notice that

$$\frac{w(x)}{x} > w'(x)$$

at $x = \theta$. So $w'(x)$ increases just to the right of θ, while — using (6.65) — $w(x)/x$ monotonically falls. But it must be the case throughout that $w(x)/x$ continues to exceed $w'(x)$, otherwise the very changes described in this paragraph cannot occur to begin with. Therefore $w'(x)$ rises throughout, establishing strict convexity to the right of θ.

However, w' cannot go to ∞, as another perusal of (6.65) will readily reveal. Indeed, w' converges to a finite limit, which is computed by setting both $w'(x)$ and $w(x)/x$ equal to the same value in (6.65). ■

Proof of Proposition 6.10. The necessity of [P] is obvious, given the characterization in Proposition 6.9, so we establish sufficiency.

Index each two-phase wage function $\mathbf{w}$ by its starting wage w, and define $c^*(w) = c(r, \mathbf{w})$. Condition P assures us that $c^*(w_1) \leq 1$ for some w_1. We claim that $c^*(w_2) > 1$ for some w_2. Suppose not; then $c^*(w) \leq 1$ for all w. Send $w \uparrow \infty$, then to maintain $c^*(w) \leq 1$ it must be that the associated cost-minimizing $\boldsymbol{\lambda}$ — call it $\boldsymbol{\lambda}(w)$ — converges weakly to 0. Fix any $k > 0$. Then for w large enough, [E] implies that

$$f(k, \boldsymbol{\lambda}(w))/k < r.$$

For all such w, concavity of f in k tells us that the associated cost-minimizing capital input $k(w)$ must be bounded. But now the continuity of f (together with [E]) atells us that output goes to zero as $w \to \infty$, which contradicts unit cost minimization. This proves the claim.

Because c^* is continuous,[28] there exists w^* between w_1 and w_2 such that $c^*(w^*) = 1$.

We prove that the two-phase wage function $\mathbf{w}$ emanating from w^* satisfies all the conditions for a steady state wage function. To this end, we specify a steady state wealth and bequest distribution, and occupational choice.

First, let $\boldsymbol{\lambda}^*$ be the input mix associated with the supporting wage function $\mathbf{w}$. Arrange the population over occupations according to $\boldsymbol{\lambda}^*$. Let $\theta = \theta(w^*)$.

If a family i is assigned to occupation h with $x(h) \leq \theta$, set that family's wealth equal to $\Omega(w^*, r)$, its educational bequest equal to $x(h)$, and its financial bequest equal to $[\Omega(w^*, r) - x(h)]/(1+r)$.

Otherwise, if occupational assignment h has $x(h) > \theta$, set that family's wealth equal to $w^*(x(h))$, its educational bequest equal to $x(h)$, and its financial bequest equal to 0.

To complete the proof, we must show that each family chooses an optimal bequest. First pick a family located at occupation h with $x = x(h) \geq \theta$. Because $\mathbf{w}^*$ has a slope of at least $1+r$ in x, this family has no need to make financial bequests. Let $M(x, x') \equiv U(w(x) - x') + V(w(x'))$ be this family's expected payoff from leaving an educational bequest x', and let $N(x) \equiv M(x, x)$. Then N is differentiable and it is easy to see that

$$N'(x) \geq U'(w(x) - x)w'(x) \text{ for all } x, \text{ with equality if } x \geq \theta. \tag{6.66}$$

[28]This is a consequence of the maximum theorem and the assumption that production is continuous in the weak topology over occupational distributions on $\mathcal{H}$.

For any $x' \geq x \geq \theta$, then, using the equality in (6.66)

$$M(x,x) = M(x',x') - \int_x^{x'} U'(w(z)-z)w'(z)dz \tag{6.67}$$

$$\geq M(x',x') - \int_x^{x'} U'(w(z)-x')w'(z)dz \tag{6.68}$$

$$= M(x',x') + U(w(x)-x') - U(w(x')-x') \tag{6.69}$$

$$= M(x,x'). \tag{6.70}$$

Similarly, for $x' \leq x$, using the inequality in (6.66),

$$M(x,x) = M(x',x') + \int_{x'}^{x} N'(z)dz \tag{6.71}$$

$$\geq M(x',x') + \int_{x'}^{x} U'(w(z)-z)w'(z)dz \tag{6.72}$$

$$\geq M(x',x') + \int_{x'}^{x} U'(w(z)-x')w'(z)dz \tag{6.73}$$

$$= M(x',x') + U(w(x)-x') - U(w(x')-x') \tag{6.74}$$

$$= M(x,x'). \tag{6.75}$$

Therefore $M(x,x) \geq M(x,x')$ for all x', so that the family at x behaves optimally by bequeathing x.

Now consider a family located at $x \in [0,\theta]$. We know that its total wealth equals $\Omega(w^*,r) = w(\theta) \leq w(x)$ for all $x \geq \theta$, so by a standard single-crossing argument, and the observations of the previous paragraph, that family will never bequeath more than θ. Therefore this family must behave just as in a Becker-Tomes benchmark world with prices (w^*,r), so its assigned bequest in optimal. ■

Proof of Proposition 6.11: Suppose, on the contrary, that there are two steady state wage functions (modulo equivalent representations). Denote these by $\mathbf{w}_1$ and $\mathbf{w}_2$, and observe from Proposition 6.9 that each of them must belong to the two-phase family. Let w_1 and w_2 be the two baseline wages, and let $\theta_i = \theta(w_i)$, for $i = 1,2$. Without loss of generality suppose that $\theta_2 \geq \theta_1$.

These two wage functions *must* cross, otherwise if the profit-maximization (support) condition is satisfied at one of them it will not be satisfied at the other. Beyond θ_2 both wage functions satisfy the same differential equation (6.19), which rules out a crossing in this region. The functions also cannot cross below θ_1 since both wage functions are r-linear in this region. So $\theta_1 < \theta_2$, and the functions cross at some $\bar{x} \in (\theta_1,\theta_2)$.

Now proceed as in the proof of Lemma 6.18. Pick some x in $(\bar{x},\theta_2)$, inhabited under $\mathbf{w}_2$; then pick $x' \in (\bar{x},x)$ such that $w_1(x') = w_2(x) > 0$.

By construction, $\Omega(w_2,r) > w_2(x)$ (x is on the r-linear segment of $\mathbf{w}_2$), and so $\tilde{W}(w_2(x);w_2,r) > w_2(x)$. By Lemma 6.5,

$$(1+r)V'(w_2(x)) > U'(w_2(x)-x).$$

Using $w_1(x') = w_2(x)$, $x' < x$ and the concavity of U, we must conclude that

$$(1+r)V'(w_1(x')) > U'(w_1(x') - x'),$$

but this contradicts the fact that $\mathbf{w}_1$ satisfies (6.19) at x'. ∎

Proof of Observation 6.6. Condition P tells us that for some two-phase wage function $\mathbf{w}$, $c(r, \mathbf{w}) \leq 1$. Define a new wage function $\hat{\mathbf{w}}$ that is r-linear from the same baseline wage as that for $\mathbf{w}$; then by Lemma 6.15, $\hat{w}(x) \leq w(x)$ for all x. It follows that $c(r, \hat{\mathbf{w}}) \leq 1$ as well. The existence of the required baseline wage a now follows from the same argument used in the proof of Proposition 6.10. ∎

Proof of Proposition 6.12. First assume that (6.21) fails; using the same technique as in the proof of Proposition 6.10, it is easy to see that the r-linear wage function starting at a is an equal steady state. Given Proposition 6.11, this completes the proof.

Indeed, by the characterization result of proposition 6.9, an equal steady state wage function *must* be the wage function that starts at a. If, therefore, (6.21) holds, that proposition assures us that an equal steady state cannot exist. ∎

6.4.11 Informal Observations on Equilibrium Dynamics.

[This section needs rewriting to harmonize notation and cross-references with the material discussed earlier. What follows has been extracted from Mookherjee and Ray (AER 2002).]

We now turn to the question of non-steady state dynamics. The basic ideas are simple, though the details are complicated and thus suppressed below. A distribution of wealth (past bequests plus current income) prevails at any date; this will map into a distribution of wealth for the next generation. Of course, the analysis will be different depending on whether condition (**??**) holds or fails, and in addition there are several different kinds of initial conditions to consider. We report on a single case, but one that holds particular interest. We assume that initial wealth is *perfectly equally distributed*, yet we suppose that a steady state is incompatible with perfect equality; i.e. that (**??**) holds.

Begin, then, with a single wealth level which we shall call W_0, commonly held by every member of generation 0. The key to understanding the dynamics is the following simple but powerful observation.

Observation 6.7. *Suppose that next period's wage function is given by $\mathbf{w}_1$, with lowest wage $\underline{w}_1$. Let $W_1 \equiv \Psi(W_0, \underline{w}_1)$. Then for every x such that $w_1(x) \leq W_1$, it must be the case that $w_1(x) = \underline{w}_1 + (1+r)x$.*

This follows from similar arguments made earlier: the rate of return on educational investment must equal r over the entire range spanned by financial bequests for any given occupation. Observation 6.7 has an interesting corollary.

Observation 6.8. *For any N, there exists a threshold such that starting from any equal initial wealth above this threshold, there is perfect equality for at least N generations.*

To see this, recall the wage function w^* constructed earlier . For any N, define a threshold — call it $\tilde{W}$ — such that if $W_0 \geq \tilde{W}$, then $\Psi^{(N)}(W_0, \underline{w}^*) \geq \bar{w}^*$, where $\Psi^{(N)}$ is the function Ψ

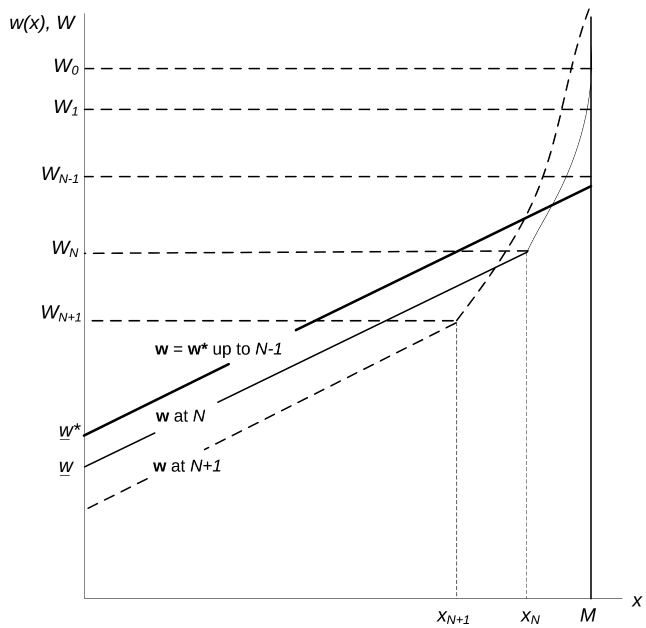


FIGURE 6.8. SYMMETRY-BREAKING ALONG EQUILIBRIUM PATHS.

iterated N times. In words, we choose an initial wealth high enough so that N iterations will not suffice to bring wealth down to $\bar{w}^*$, in the pure bequest model with baseline wage $\underline{w}^*$. If the wage function for the next N periods is given by $\mathbf{w}^*$, this is perfectly consistent with equilibrium behavior for those N periods. All occupations are equally valuable, so individuals are indifferent over these choices. Moreover, there will be full equality over this epoch.

Observation 6.7 implies that this is the *only* possible outcome for the first N periods. Any other candidate wage function would have to exhibit a return of r over occupations in which the wage fell short of wealth. This implies right away that no other equilibrium wage function can exist with the same starting wage $\underline{w}^*$. Using Observation 6.7 again and employing an argument similar to that used to establish uniqueness, we may also rule out any candidate wage function with a different starting wage than $\underline{w}^*$.

So perfect equality can prevail for a substantial number of periods. But — if (**??**) holds — it cannot prevail forever. Because $\Omega(\underline{w}^*) < \bar{w}^*$, there must eventually come a date T when the recursion $\psi^{(T)}(W_0, \underline{w}^*)$ dips below $\bar{w}^*$. At this stage, the symmetry of equilibrium is broken and the economy must depart from the wage function $\mathbf{w}^*$. If not, all occupations with wages that exceed $\psi^{(T)}(W_0, \underline{w}^*)$ would remain unoccupied. This cannot be: the wages of such occupations must rise, yielding rates of return that exceed their counterpart on financial bequests. For all other occupations, Observation 6.7 is applicable in full force, and the occupational rate of return remains anchored to the financial rate.

Figure 6.8 describes the resulting "distortion" in the wage function. The diagram displays N initial periods in which the wage function $\mathbf{w}^*$ prevails, and there is no inequality. At date N, wealth dips into the support of $\mathbf{w}^*$, and no occupation more expensive than x_N can be supported under the old rates of return. The economy must react by lowering the wages of all occupations below x_N, and by raising the rates of return to all other occupations. These higher rates encourage the acquisition of such professions, and in so doing must generate inequality. Of course, all generation-N individuals are still indifferent between the various choices. But this is the last generation for which all payoffs are equal. The higher rates of return on the more expensive occupations must inevitably result in inequality among the next generation.

Once such inequality sets in, it will not go away. Familiar single-crossing arguments guarantee that descendants of higher-wealth individuals must occupy the richer professions and receive higher payoffs. The subsequent dynamics are complicated by the fact that the wealth distribution is no longer degenerate. In the special case where there are just two occupations (involving different training costs), it can be shown that the economy converges to an unequal steady state, with inequality rising over time. Moreover, the rise in inequality is augmented in the presence of financial bequests: it turns out that individuals in the unskilled occupation at any date are less wealthy than their parents, and make smaller financial bequests to their children than they received.

In the continuum case, the wealth distribution at any date must retain a mass point at its lower bound, and it is this mass point which will be spread over all the "low" occupations, which continue to bear the financial rate of return r. In the relatively "high" occupations, the rate of return will continue its departure from r, attracting more individuals into this zone.

Figure 6.8 illustrates this with yet another iteration for date $N+1$. The new wage function is given by the dashed curve. It must be linear (with slope $1+r$) up to the new threshold x_{N+1}, and then rises even more steeply than before, intersecting both the previous wage functions from below. This rise induces some fresh symmetry-breaking, as new dynasties from the "cheap" occupations seek the higher rates of return. If this process converges (a subject of our current research), it must be to precisely the steady state we have described earlier. At this steady state, there is a mass point of individuals with identical wealth, and among such individuals there is a simple tradeoff between occupational choice and financial bequests, among which they are indifferent. But there will also be a positive measure of individuals arrayed over varying levels of wealth (and utility).

CHAPTER 7

Polarization and Conflict

> As the struggle proceeds, the whole society breaks up more and more into two hostile camps, two great, directly antagonistic classes: bourgeoisie and proletariat. The classes *polarize*, so that they become internally more homogeneous and more and more sharply distinguished from one another in wealth and power. (Morton Deutsch, 1971)

7.1 Introduction

Conflict is a fundamental part of life in many developing countries. This section of the course will try to set up models of conflict and also study its correlates.

Some basic facts: between 1945–1999, battle deaths in 25 interstate wars were approximately 3.33m. Compared to that, deaths in civil wars have totalled around 16.2m. Such wars have been experienced in 73 states, with a median duration of six years. In 1999, there were 25 ongoing conflicts. It is reasonable to say that matters have only become worse in the twenty first century.

These numbers are symptomatic of ongoing unrest, violence around the world. They obviously have big effects for GDP and growth: one writer places the narrow economic costs of conflict at around 8% of GDP.

By conflict we don't just mean violent conflict. Demonstrations, strikes, processions: all of these also cause a substantial amount of economic waste.

One sort of conflict may simply come from class antagonisms: the presence of rich and poor subclasses in the population. This is classic redistributive conflict where redistribution occurs, or is sought, from rich to poor. But it has also been argued that "ethnic divisions" — broadly defined — may be a significant determinant of conflict. For instance, writers such as Samuel Huntington (1996) have argued for a cultural perspective on conflict and war. After all, there is no particular reason why redistribution must occur across *economic* classes. It could happen between religious, ethnic, or spatially separated groups. Resources are resources, no matter where they come from.

The monumental threatise on ethnic conflict is by Donald Horowitz (1985), who summarizes thus:

> "The Marxian concept of class as an inherited and determinative affiliation finds no support in [the] data. Marx's conception applies with far less distortion to ethnic groups. Ethnic membership us generally given at birth, . . . and have considerable power to generate conflict . . .
>
> In much of Asia and Africa, it is only modest hyperbole to assert that the Marxian prophecy has had an ethnic fulfillment."

As I've said above, this doesn't mean that ethnicity is necessarily a "primordial" or intrinsic source of conflict. Ethnicity may simply be a relevant marker for carving out a larger share of a smaller pie. Very often, ethnic groups are not particularly ranked by income: Sinhalese/Tamils in Sri Lanka, Malays/Chinese in Malaysia, Hausa/Yoruba in Nigeria, Serbs/Muslims in Bosnia, Dalits/low-income-Muslims in Gujarat, Basques/others in Spain.

In other situations, marked socio-economic differences do exist: Hutu/Tutsi in Rwanda and Burundi, caste divisions in Bihar, Bengalis in pre-Bangladesh Pakistan, Muslims in the Philippines, Kurds in Iraq, Sikhs in the Punjab, or the Naxalites in Eastern and SouthEastern India.

Some interesting questions:

What sort of distributions make for conflict?

Distributions over *what* ? Ethnic groups or income classes? The answer would presumably depend on what the salient marker is for conflict.

How might we decide whether unranked ethnicity or ranked class is salient in conflict?

7.2 What Creates Conflict?

7.2.1 Variables. Much (though not all) of the empirical literature on conflict throws in the kitchen sink to see what washes up. On the left hand side goes in one of several alternative dependent variables: many measures of "conflict", ranging all the way from incidence of demonstrations, procesions, or strikes, through riots and on to civil war.

Even with a specific choice such as civil war (used here) we need defining criteria (see Singer-Small (1982), Licklider (1993), Doyle-Sambanis (2000), Fearon-Laitin (2003)). One important distinction to be made is between *onset* and *incidence*. Did the conflict start in that year of measurement, or was it ongoing? As we will see later this makes a big difference. Another important question is the minimum number of deaths before a conflict is classifiable as such: for civil wars, this is a matter of ongoing debate.

On the right hand-side goes in a slew of explanatory variables:

Economic: per-capita income, inequality of income or wealth, resource holdings . . .

Geographical: mountainous terrain, separation from capital city . . .

Political: “extent of democracy”, prior war . . .

And, of course

Ethnic: Measures of ethnolinguistic diversity drawn from the World Christian Encyclopedia, Encyclopedia Britannica, Atlas Narodov Mira, CIA FactBook. Measures of religious diversity drawn from L’Etat des Religions dans le Monde, World Christian Encyclopedia, The Statesman’s Yearbook.

This last information is typically put into an index of diversity. An index with particularly wide currency is the *ethnolinguistic fractionalization index*, or ELF (of course, there’s nothing to stop its use for religious diversity as well).

To define this, say there are M groups. n_j is the population share of group j. Then

$$E = \sum_{j=1}^{M} n_j(1 - n_j).$$

This has an obvious relationship to the Gini index of inequality if we take inter-group distance to be equal to one.

ELF is widely used in empirical work (see Taylor and Hudson (1972), Mauro(1995), Easterly and Levine (1997), Alesina *et al.* (2003), Vigdor (2002), Collier and Hoeffler (2002), Fearon and Laiton (2003), and many others).

7.2.2 Some Findings. Perhaps the most robust finding in many different empirical studies is that per-capita income is negatively related to conflict. A negative and significant association is to be observed in the cross-sectional studies carried out by Collier and Hoeffler (2002), Fearon and Laitin (2003), Montalvo and Reynal-Querol (2005) and several others. As Fearon and Laitin (2003) summarize,

> “Per capita income (measured as thousands of 1985 U.S. dollars and lagged one year) is strongly significant in both a statistical and a substantive sense . . . ”

It hardly needs mentioning that no causal inference can be drawn from such an association. Relatively speaking, the best study that addresses this question is Miguel, Satyanath and Sergenti (2004), which instruments for GDP growth shocks using rainfall. Such an instrunment would be unacceptably weak in many of the richer developing countries, so they restrict attention to a sample of 41 sub-Saharan African countries over the period 1981-1999. Their main results are best summarized in their own words:

> “Using the comprehensive new database of conflicts developed by the International Peace Research Institute of Oslo, Norway, and the University of Uppsala, Sweden, we find that GDP growth is significantly negatively related to the incidence of civil conflict in sub-Saharan Africa during the period 1981–99 across a range of regression specifications, including some with country fixed effects. The relationship between GDP growth and the incidence of civil wars is extremely strong: a five-percentage-point drop

> in annual economic growth increases the likelihood of a civil conflict (at least 25 deaths per year) in the following year by over 12 percentage points — which amounts to an increase of more than one-half in the likelihood of civil war. Other variables that have gained prominence in the recent literature — per capita GDP level, democracy, ethnic diversity, and oil exporter status — do not display a similarly robust relationship with the incidence of civil wars in sub-Saharan Africa. In the second main result, we find — perhaps surprisingly — that the impact of income shocks on civil conflict is not significantly different in richer, more democratic, more ethnically diverse, or more mountainous African countries or in countries with a range of different political institutional characteristics."

To be sure, the exclusion restriction behind the use of rainfall as an instrument could be violated if rainfall has some direct effect on conflict via some other pathway. For instance, floods could destroy the road network and make it harder to contain government troops — more conflict (though this is a bias that runs the other way from the instrumental direction, which may actually strengthen the results). But of course, one could cook up other connections, such as floods make it harder for government troops and rebels to engage each other, which reduces conflict (in terms of deaths). The authors attempt to take care of some of these effects in the paper.

What lies behind the connection behind poverty and conflict? What comes to mind most immediately is the fact that poverty reduces the opportunity cost of conflict labor, making it easier to use labor for some other activity. Of course, with poverty all around there may not be much to appropriate, so this suggests that it might be more interesting to examine the effects of (growth shocks)*inequality.

Some evidence that opportunity costs lie at the heart of the story comes from the paper by Dube and Vargas (2007), which studies the effect of coffee and oil prices on conflict in Colombia. Oil prices affect government revenues which are there to be grabbed, so one would imagine that higher prices for oil positively affect conflict. In contrast, if coffee-growing is an economic activity that competes with resources given over to conflict (such as labor!), then an increase in coffee prices would *reduce* conflict. Thus if the opportunity cost story is right, one would expect that oil and coffee have opposite effects on conflict, and indeed this is what Dube and Vargas find.

The opportunity cost argument is explicitly made by Collier and Hoeffler (1998), though it is fairly obvious and appears implicitly in many papers. In contrast, Fearon and laitin (2003) have argued that poverty is likely correlated with low government capabilities and infrastructure that makes it harder to crack down on insurgencies. In support of this line of reasoning they show that "mountainous terrain" is significant in "explaining" conflict, showing thereby that infrastructure and geography matter. This line of argument requires more investigation, however.

A second important finding is that the *inequality* of income appears to have at best an ambiguous effect on conflict. Some early papers on this subject are Nagel (1974) for Vietnam, Midlarski (1988) and Muller, Seligson and Fu (1989) (the last two are on land inequalities). Specifically, under several measures of inequality such as the Gini, conflict appears to be low *both* for low and for high values of inequality. This is a theme to which we will return in more detail below.

As a related matter, we've already noted that measures of ethnic dispersion are closely related to "inequality" — i.e., diversity — across ethnic groups. These, too, fail to matter. The same findings appear and reappear in Collier and Hoeffler (1998, 2002), Fearon and Laitin (2003), as well as Miguel et al (2004). Fearon and Laitin (2003) conclude thus:

> "The estimates for the effect of *ethnic* and *religious fractionalization* are substantively and statistically insignificant . . . The empirical pattern is thus inconsistent with . . . the common expectation that ethnic diversity is a major and direct cause of civil violence."

This isn't to say that ethnic or religious fractionalization cannot *indirectly* affect conflict. It might, *via* reduced GDP (Alesina et al. (2003)), reduced GDP growth (Easterly and Levine (1997)), or poor governance (Mauro (1995)). The claim really is that there is no *direct* effect.

An important theme that we pursue below is the development of a measure of "polarization", which we show is different from inequality. Is a Gini-like index like ELF a good measure? Listen to Horowitz again:

> "I have intimated at various points that a system with only two ethnic parties . . . is especially conflict prone . . . In dispersed systems, group loyalties are parochial, and ethnic conflict is localized; it 'could put one of a series of watertight compartments out of order, but it could not make the ship of state sink . . . ' The demands of one group can sometimes be granted without injuring the interests of others . . . "

On the other hand, continues Horowitz,

> "A centrally focused system [with few groupings] possesses fewer cleavages than a dispersed system, but those it possesses run through the whole society and are of greater magnitude. When conflict occurs, the center has little latitude to placate some groups without antagonizing others."

The conflictual power of broad cleavages is of course an older theme: read the quotation at the beginning of this chapter!

7.3 The Identity-Alienation Framework

Following Esteban and Ray (1991, 1994) and Duclos, Esteban and Ray (2004), I develop below the *identification-alienation* framework. The idea is simple: polarization is related to the alienation that individuals and groups feel from one another, *but such alienation is fuelled by notions of within-group identity*. In concentrating on such phenomena, I do not mean to suggest that instances in which a single isolated individual runs amok with a machine gun are rare, or that they are unimportant in the larger scheme of things. It is just that these are not the objects of our enquiry. We are interested in the correlates of organized, large-scale social unrest — strikes, demonstrations, processions, widespread violence, and revolt or rebellion. Such phenomena thrive on differences, to be sure. But they cannot exist without notions of group identity either.

This brief discussion immediately suggests that inequality, inasmuch as it concerns itself with interpersonal alienation, captures but one aspect of polarization. To be sure, there are some obvious changes that would be branded as both inequality- and polarization-enhancing. For instance, if two income groups are further separated by increasing economic distance, inequality and polarization would presumably both increase. However, *local* equalizations of income differences at two different ranges of the income distribution will most likely lead to two better-defined groups — each with a clearer sense of itself and the other. In this case, inequality will have come down but polarization may be on the rise.

Imagine, then, that society is divided into "groups" (economic, social, religious, spatial...)

Identity. There is "homogeneity" *within* each group.

Alienation. There is "heterogeneity" *across* groups.

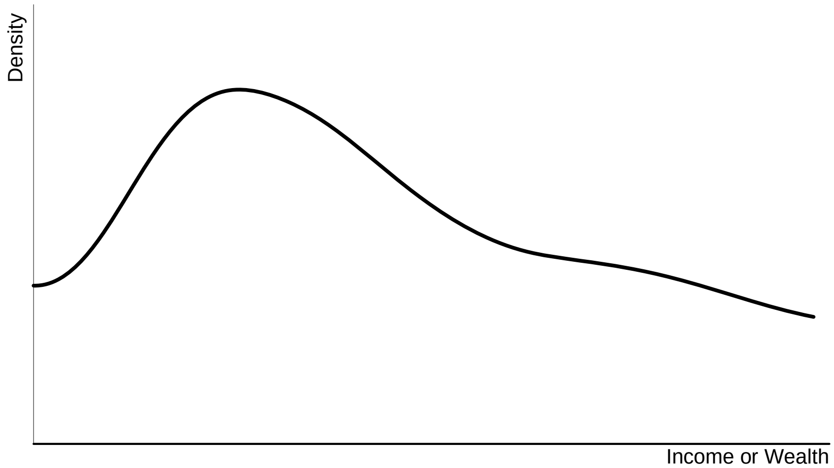
The IA framework *presumes* that such a situation is inherently conflictual. In the words of Esteban and Ray (1994),

> "We begin with the obvious question: why are we interested in polarization? It is our contention that the phenomenon of polarization is closely linked to the generation of tensions, to the possibilities of articulated rebellion and revolt, and to the existence of social unrest in general . . . At the same time, measured inequality in such a society may be low."

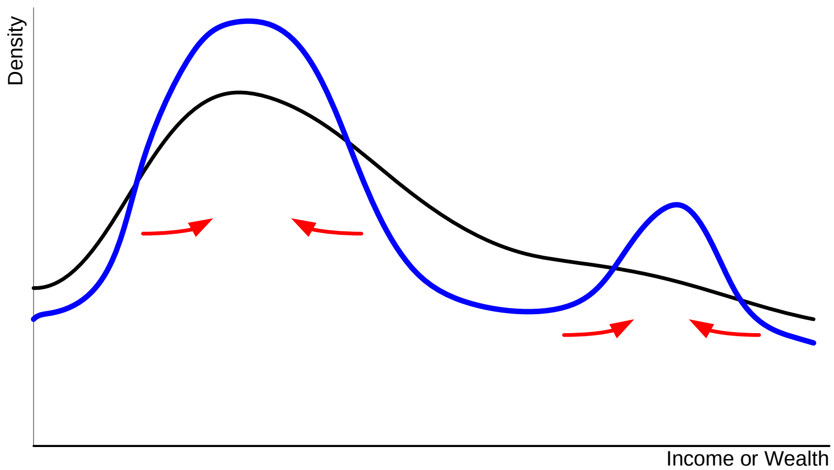
Does the standard theory of inequality measurement fit? Recall the

Pigou-Dalton Transfers Principle. A transfer of resources from a relatively poor to a relatively rich individual must raise income inequality.

This principle forms the building block for all measures of inequality. But now look at this example. Start here:

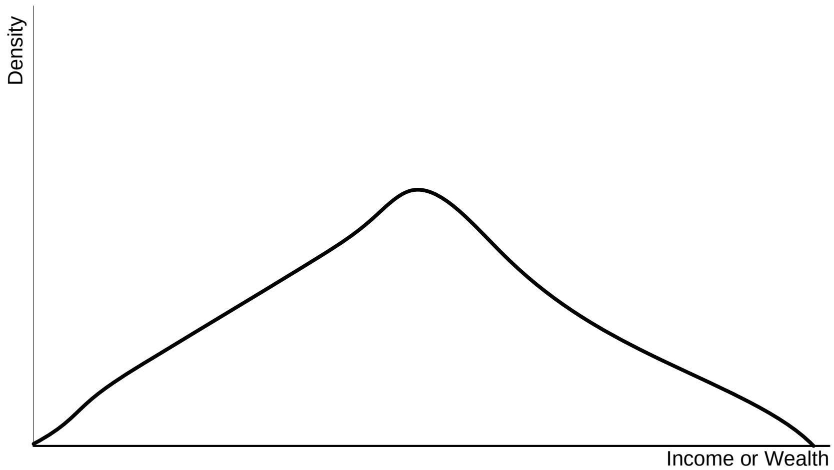


and go here:

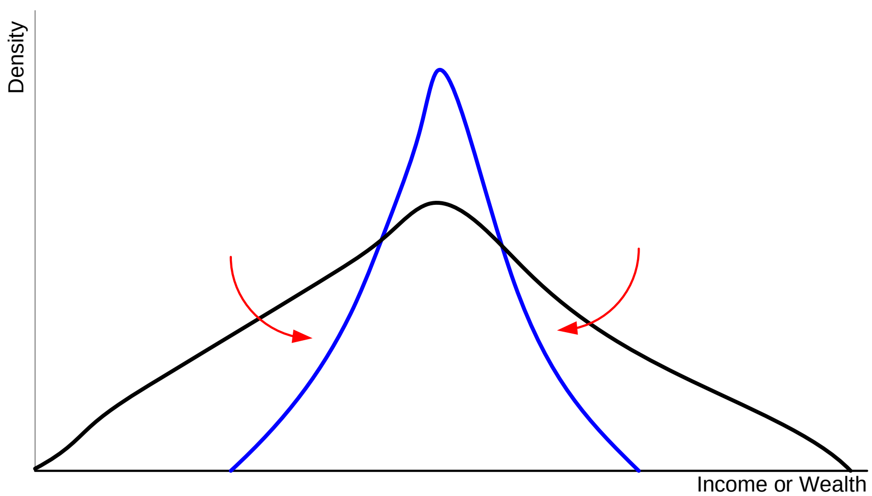


By all measures of inequality that are consistent with the Lorenz criterion (or equivalently, second-order stochastic domination), inequality has come *down*. Yet there is something disturbing about that: social tensions could be going *up* as two well-defined and distinct groups begin to form. There could be situations here in which polarization is going up.

Of course, this isn't to suggest that polarization — whatever it is — is *always* different from inequality. If there is a "global compression" of the distribution, we would expect both inequality and polarization to fall. See these diagrams. Start here:



and go here:



But these two examples also tell us something else. It tells us that polarization may not be a "local construct". In both the examples, we have a compression taking place, yet they have very different implications. A "local move" may have different effects depending on the overall distribution. In contrast, inequality — as captured by the Pigou-Dalton principle — *is* a local construct.

Observe, too, that the notion of "groups" may be quite general:

Economic: income- or wealth-based (class)

Social: religious, linguistic, geographical, political groupings.

But a natural notion of "distance" across economic groups makes income-based polarization an easier starting point.

7.4 A Measure of Polarization

The task of this section is to develop axiomatically a measure of polarization. Our "inputs" are various distributions of income or wealth on different populations (more precisely, *density functions* with varying populations).

Our "output" is a measure of polarization for each distribution.

As we've tried to motivate earlier, each individual feels two things:

Identification with people of "similar" income.

[Use as proxy the height of density $f(x)$ at income x.]

Alienation from people with "dissimilar" income.

[Income distance $|y - x|$ of y from x.]

We therefore describe the *effective antagonism* of x towards y as a function $T(i, a)$, where i is the identification that a person at income x feels, and a is the income distance between x and y. We suppose that $T(0, a) = T(i, 0) = 0$: both a sense of identification and alienation is needed to fuel an effective sense of antagonism.

We view *polarization* as the "sum" of all such antagonisms over the population:

$$P(f) = \int \int T\left(f(x), |x - y|\right) f(x) f(y) dx dy$$

This is not very useful as it stands. Way too much depends on the choice of the function T. But hopefully it is a good starting point.

The axioms we use are based on densities that are unions of one or more *basic densities*. These are symmetric, unimodal density functions f with compact support.

[By symmetry we mean that $f(m - x) = f(m + x)$ for all $x \in [0, 1]$, and by unimodality we mean that f is nondecreasing on $[0, m]$.]

A basic density (or indeed any density)

can be population-scaled: $g(y) = pf(y)$.

can be income-scaled: $g(y) = (1/\mu)f(x/\mu)$.

can undergo a slide to the right or left: $g(y) = f(y - x)$, and

can be squeezed.

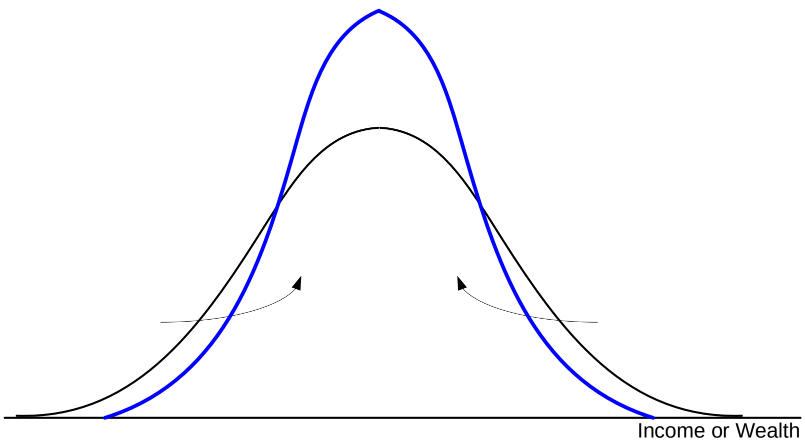
Take a closer look at this last one. Let f be any density and let λ lie in $(0, 1]$. A *λ-squeeze* of f is a transformation f^λ:

(7.1) $$f^\lambda(x) \equiv \frac{1}{\lambda} f\left(\frac{x - [1-\lambda]m}{\lambda}\right).$$

Scalings, slides and squeezes partition the space of all densities. Each element of the partition can be associated with a *root*, a basic density on $[0, 2]$ with mean 1.

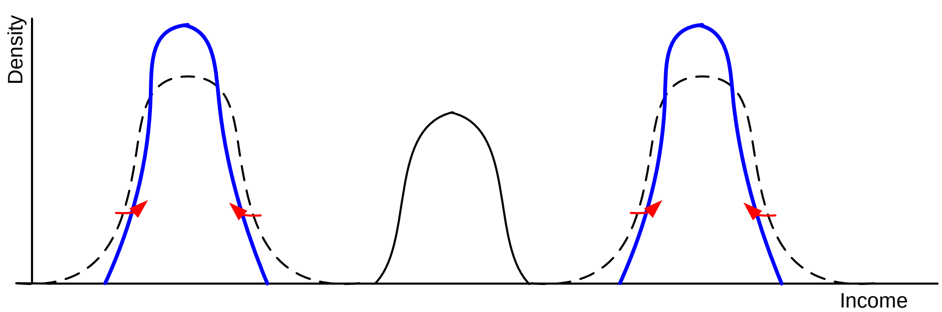
Axiom 1. If a distribution is just a single basic density, a "global compression" of that density cannot increase polarization.

That is: global compression (the move described in the diagram below) cannot raise polarization.



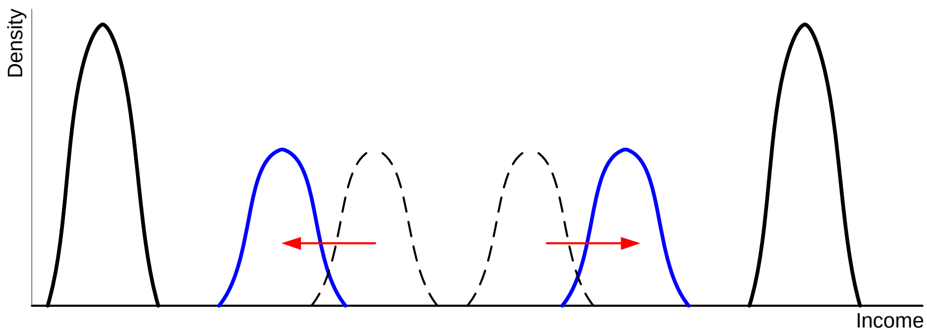
Axiom 2. If a *symmetric* distribution is composed of three disjoint scalings of the same basic density, then a compression of the *side* densities cannot reduce polarization.

This is illustrated in the diagram below.



Axiom 3. Consider a *symmetric* distribution composed of four basic densities drawn from the same root. Slide the two middle densities to the side as shown. Then polarization must go up.

This is illustrated in the diagram below:



Our last axiom states that polarization rankings must be invariant to population scalings.

Axiom 4. [Population Neutrality.] Polarization comparisons are unchanged if both populations are scaled up or down by the same percentage.

We can now state:

THEOREM 7.1. *A polarization measure satisfies Axioms 1–4 if and only if it is proportional to*

$$\int\int f(x)^{1+\alpha} f(y)|y-x|dydx, \tag{7.2}$$

where α lies between 0.25 and 1.

7.5 Proof of Theorem 7.1

Note: this section is technical (though illustrative), and should probably be omitted unless you are comfortable with long and detailed arguments.

First, we show that axioms 1–4 imply (7.2). The lemma below follows from Jensen's inequality; proof omitted.

LEMMA 7.1. *Let g be a continuous real-valued function defined on R such that for all $x > 0$ and all δ with $0 < \delta < x$,*

$$g(x) \geq \frac{1}{2\delta}\int_{x-\delta}^{x+\delta} g(y)dy. \tag{7.3}$$

Then g must be a concave function.

In what follows, remember that our measure only considers income *differences* across people, so that we may slide any distribution to left or right as we please.

LEMMA 7.2. *The function T must be concave in a for every $i > 0$.*

Proof. Fix $x > 0$, some $i > 0$, and $\delta \in (0, x)$. Consider three basic densities as in Axiom 2 but specialize as shown in Figure 7.1; each is a transform of a uniform basic density. The bases

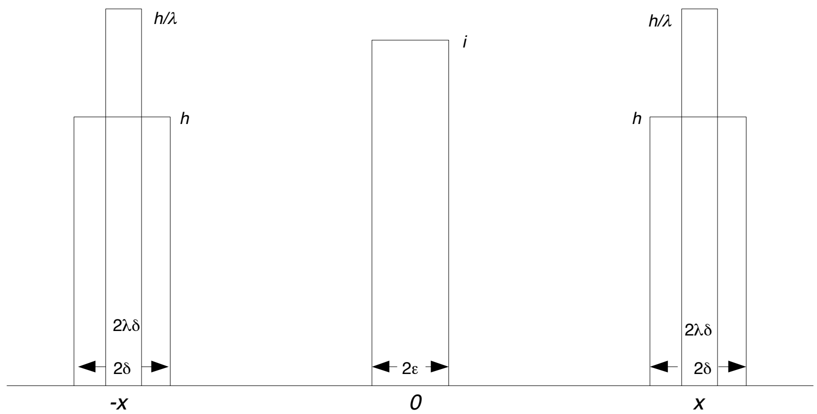


FIGURE 7.1

are centered at $-x$, 0 and x. The side densities are of width 2δ and height h, and the middle density is of width 2ϵ and height i. We shall vary ϵ and h but to make sure that Axiom 2 applies, we choose $\epsilon > 0$ such that $\delta + \epsilon < x$. A λ-squeeze of the side densities simply contracts their base width to $2\lambda\delta$, while the height is raised to h/λ. For each λ, decompose the measure (**??**) into five components. (a) The "internal polarization" P_m of the middle rectangle. This component doesn't vary with λ so there will be no need to explicitly calculate it. (b) The "internal polarization" P_s of each side rectangle. (c) Total effective antagonism, A_{ms} felt by inhabitants of the middle towards each side density. (d) Total effective antagonism A_{sm} felt by inhabitants of each side towards the middle. (e) Total effective antagonism A_{ss} felt by inhabitants of one side towards the other side. Each of these last four terms appear twice, so that (writing everything as a function of λ),

$$P(\lambda) = P_m + 2P_s(\lambda) + 2A_{ms}(\lambda) + 2A_{sm}(\lambda) + 2A_{ss}(\lambda), \tag{7.4}$$

Now we compute the terms on the right hand side of (7.4). First,

$$P_s(\lambda) = \frac{1}{\lambda^2} \int_{x-\lambda\delta}^{x+\lambda\delta} \int_{x-\lambda\delta}^{x+\lambda\delta} T(h/\lambda, |b' - b|) h^2 db' db,$$

where (here and in all subsequent cases) b will stand for the "origin" income (to which the identification is applied) and b' the "destination income" (towards which the antagonism is felt). Next,

$$A_{ms}(\lambda) = \frac{1}{\lambda} \int_{-\epsilon}^{\epsilon} \int_{x-\lambda\delta}^{x+\lambda\delta} T(i, b' - b) ih db' db.$$

Third,

$$A_{sm}(\lambda) = \frac{1}{\lambda} \int_{x-\lambda\delta}^{x+\lambda\delta} \int_{-\epsilon}^{\epsilon} T(h/\lambda.b - b') hi db' db,$$

And finally,

$$A_{ss}(\lambda) = \frac{1}{\lambda^2} \int_{-x-\lambda\delta}^{-x+\lambda\delta} \int_{x-\lambda\delta}^{x+\lambda\delta} T(h/\lambda, b' - b) h^2 db' db.$$

The axiom requires that $P(\lambda) \geq P(1)$. Equivalently, we require that $[P(\lambda) - P(1)]/2h \geq 0$ for all h, which implies in particular that

$$\liminf_{h \to 0} \frac{P(\lambda) - P(1)}{2h} \geq 0. \tag{7.5}$$

If we divide through by h in the individual components calculated above and then send h to 0, it is easy to see that the only term that remains is A_{ms}. Formally, (7.5) and the calculations above must jointly imply that

$$\frac{1}{\lambda} \int_{-\epsilon}^{\epsilon} \int_{x-\lambda\delta}^{x+\lambda\delta} T(i, b' - b) db' db \geq \int_{-\epsilon}^{\epsilon} \int_{x-\delta}^{x+\delta} T(i, b' - b) db' db, \tag{7.6}$$

and this must be true for all $\lambda \in (0, 1)$ as well as all $\epsilon \in (0, x - \delta)$. Therefore we may insist on the inequality in (7.6) holding as $\lambda \to 0$. Performing the necessary calculations, we may conclude that

$$\frac{1}{\epsilon} \int_{-\epsilon}^{\epsilon} T(i, x - b) db \geq \frac{1}{\epsilon} \int_{-\epsilon}^{\epsilon} \int_{x-\delta}^{x+\delta} T(i, b' - b) db' db \tag{7.7}$$

for every $\epsilon \in (0, x - \delta)$. Finally, take ϵ to zero in (7.7). This allows us to deduce that

$$T(i, x) \geq \int_{x-\delta}^{x+\delta} T(i, b') db'. \tag{7.8}$$

As (7.8) must hold for every $x > 0$ and every $\delta \in (0, x)$, we may invoke Lemma 7.1 to conclude that T is concave in x for every $i > 0$.

Q.E.D.

Lemma 7.3. *Let g be a concave, continuous function on R_+, with $g(0) = 0$. Suppose that for each a and a' with $a > a' > 0$, there exists $\bar{\Delta} > 0$ such that*

$$g(a + \Delta) - g(a) \geq g(a') - g(a' - \Delta) \tag{7.9}$$

for all $\Delta \in (0, \bar{\Delta})$. Then g must be linear.

The proof is straightforward and is omitted.

Lemma 7.4. *There is a continuous function $\phi(i)$ such that $T(i, a) = \phi(i)a$ for all i and a.*

Proof. Fix a and a' with $a > a' > 0$, and $i > 0$. Consider four basic densities as in Axiom 3 (see Figure **??**) but specialize as shown in Figure 7.2; each is a transform of a uniform basic density. The bases are centered at locations $-y$, $-x$, x and y, where $x \equiv (a - a')/2$ and $y \equiv (a + a')/2$. The "inner" densities are of width 2δ and height h, and the "outer" densities are of width 2ϵ and height i. We shall vary different parameters (particularly x) but to ensure disjoint support we assume throughout that $\epsilon < x$ and $\delta + \epsilon < y - x - \bar{\Delta}$ for some $\bar{\Delta} > 0$. Again, decompose the polarization measure (**??**) into several distinct components. (a) The "internal polarization" of each rectangle j; call it P_j, $j = 1, 2, 3, 4$. These components are unchanged as we change x so there will be no need to calculate them explicitly. (b)

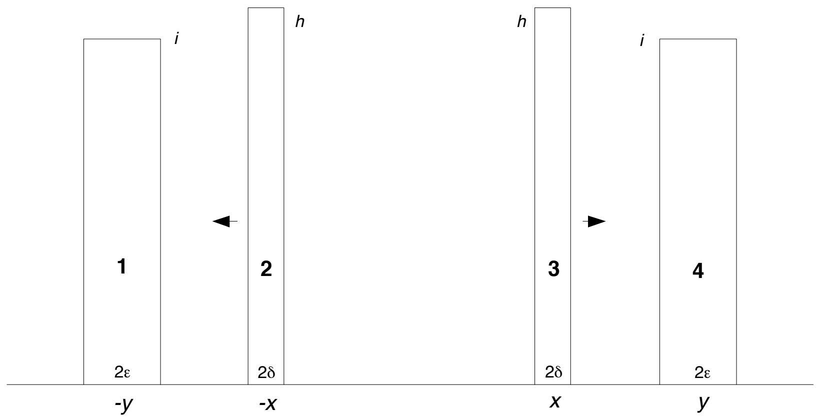


FIGURE 7.2

Total effective antagonism $A_{jk}(x)$ felt by inhabitants of rectangle j towards rectangle k (we emphasize dependence on the parameter x). Thus total polarization $P(x)$ is given by

$$\begin{aligned} P(x) &= \sum_{j=1}^{4} P_j + \sum_j \sum_{k \neq j} A_{jk}(x) \\ &= \sum_{j=1}^{4} P_j + 2A_{12}(x) + 2A_{13}(x) + 2A_{21}(x) + 2A_{31}(x) + 2A_{23}(x) + 2A_{14}, \end{aligned}$$

where the second equality simply exploits obvious symmetries and A_{14} is noted to be independent of x. Let's compute the terms in this formula that do change with x. We have

$$A_{12}(x) = \int_{-y-\epsilon}^{-y+\epsilon} \int_{-x-\delta}^{-x+\delta} T(i, b' - b) ih db' db,$$

$$A_{13}(x) = \int_{-y-\epsilon}^{-y+\epsilon} \int_{x-\delta}^{x+\delta} T(i, b' - b) ih db' db,$$

$$A_{21}(x) = \int_{-x-\delta}^{-x+\delta} \int_{-y-\epsilon}^{-y+\epsilon} T(h, b - b') ih db' db,$$

$$A_{31}(x) = \int_{x-\delta}^{x+\delta} \int_{-y-\epsilon}^{-y+\epsilon} T(h, b - b') ih db' db,$$

and

$$A_{23}(x) = \int_{-x-\delta}^{-x+\delta} \int_{x-\delta}^{x+\delta} T(h, b - b') h^2 db' db.$$

Now, the axiom requires that $P(x + \Delta) - P(x) \geq 0$. Equivalently, we require that $[P(x + \Delta) - P(1)]/2ih \geq 0$ for all h, which implies in particular that

$$\liminf_{h \to 0} \frac{P(x + \Delta) - P(x)}{2ih} \geq 0.$$

Using this information along with the computations for $P(x)$ and the various $A_{jk}(x)$'s, we see (after some substitution of variables and transposition of terms) that

$$\begin{aligned}
&\int_{-y-\epsilon}^{-y+\epsilon}\int_{x-\delta}^{x+\delta}[T(i,b'-b+\Delta)-T(i,b'-b)]\,db'db \\
\geq\ &\int_{-y-\epsilon}^{-y+\epsilon}\int_{-x-\delta}^{-x+\delta}[T(i,b'-b)-T(i,b'-b-\Delta)]\,db'db,
\end{aligned}$$

Dividing through by δ in this expression and then taking δ to zero, we may conclude that

$$\int_{-y-\epsilon}^{-y+\epsilon}[T(i,x-b+\Delta)-T(i,x-b)]\,db \geq \int_{-y-\epsilon}^{-y+\epsilon}[T(i,-x-b)-T(i,-x-b-\Delta)]\,db,$$

and dividing this inequality, in turn, by ϵ and taking ϵ to zero, we see that

$$T(i,a+\Delta)-T(i,a)\geq T(i,a')-T(i,a'-\Delta),$$

where we use the observations that $x+y=a$ and $y-x=a'$. Therefore the conditions of Lemma 7.3 are satisfied, and $T(i,.)$ must be linear for every $i>0$ since $T(0,a)=0$. That is, there is a function $\phi(i)$ such that $T(i,a)=\phi(i)a$ for every i and a. Given that T is continuous by assumption, the same must be true of ϕ.

Q.E.D.

Lemma 7.5. *$\phi(i)$ must be of the form Ki^{α}, for constants $(K,\alpha)\gg 0$.*

Proof. As a preliminary step, observe that

(7.10) $$\phi(i)>0 \text{ whenever } i>0.$$

otherwise Axiom 3 would fail for configurations constructed from rectangular basic densities of equal height i. We first prove that ϕ satisfies the fundamental Cauchy equation

(7.11) $$\phi(p)\phi(p')=\phi(pp')\phi(1)$$

for every $(p,p')\gg 0$. To this end, fix p and p' and define $r\equiv pp'$. In what follows, we assume that $p\geq r$.[1] Consider a configuration with two basic densities, both of width 2ϵ, the first centered at 0 and the second centered at 1. The heights are p and h (where $h>0$ but soon to be made arbitrarily small). A little computation shows that polarization in this case is given by

$$\begin{aligned}
P &= ph[\phi(p)+\phi(h)]\{\int_{-\epsilon}^{\epsilon}\int_{1-\epsilon}^{1+\epsilon}(b'-b)db'db\} \\
&\quad +[p^2\phi(p)+h^2\phi(h)]\{\int_{-\epsilon}^{\epsilon}\int_{-\epsilon}^{\epsilon}|b'-b|db'db\} \\
&= 4\epsilon^2 ph[\phi(p)+\phi(h)]+\frac{8\epsilon^3}{3}[p^2\phi(p)+h^2\phi(h)], \qquad (7.12)
\end{aligned}$$

where the first equality invokes Lemma 7.4. Now change the height of the first rectangle to r. Using (7.10) and $p\geq r$, it is easy to see that for each ϵ, there exists a (unique) height $h(\epsilon)$

[1] If $r\geq p$, simply permute p and r in the argument below.

for the second rectangle such that the polarizations of the two configurations are equated. Invoking (7.12), $h(\epsilon)$ is such that

$$ph[\phi(p) + \phi(h)] + \frac{2\epsilon}{3}[p^2\phi(p) + h^2\phi(h)]$$
$$= rh(\epsilon)[\phi(r) + \phi(h(\epsilon))] + \frac{2\epsilon}{3}[r^2\phi(r) + h(\epsilon)^2\phi(h(\epsilon))]. \tag{7.13}$$

By Axiom 4, it follows that for all $\lambda > 0$,

$$\lambda^2 ph[\phi(\lambda p) + \phi(\lambda h)] + \frac{2\epsilon}{3}[(\lambda p)^2\phi(\lambda p) + (\lambda h)^2\phi(\lambda h)]$$
$$= \lambda^2 rh(\epsilon)[\phi(\lambda r) + \phi(\lambda h(\epsilon))] + \frac{2\epsilon}{3}[(\lambda r)^2\phi(\lambda r) + [\lambda h(\epsilon)]^2\phi(\lambda h(\epsilon))]. \tag{7.14}$$

Notice that as $\epsilon \downarrow 0$, $h(\epsilon)$ lies in some bounded set. We may therefore extract a convergent subsequence with limit h' as $\epsilon \downarrow 0$. By the continuity of ϕ, we may pass to the limit in (7.13) and (7.14) to conclude that

$$ph[\phi(p) + \phi(h)] = rh'[\phi(r) + \phi(h')] \tag{7.15}$$

and

$$\lambda^2 ph[\phi(\lambda p) + \phi(\lambda h)] = \lambda^2 rh'[\phi(\lambda r) + \phi(\lambda h')]. \tag{7.16}$$

Combining (7.15) and (7.16), we see that

$$\frac{\phi(p) + \phi(h)}{\phi(\lambda p) + \phi(\lambda h)} = \frac{\phi(r) + \phi(h')}{\phi(\lambda r) + \phi(\lambda h')}. \tag{7.17}$$

Taking limits in (7.17) as $h \to 0$ and noting that $h' \to 0$ as a result (examine (7.15) to confirm this), we have for all $\lambda > 0$,

$$\frac{\phi(p)}{\phi(\lambda p)} = \frac{\phi(r)}{\phi(\lambda r)}. \tag{7.18}$$

Put $\lambda = 1/p$ and recall that $r = pp'$. Then (7.18) yields the required Cauchy equation (7.11). To complete the proof, recall that ϕ is continuous and that (7.10) holds. The class of solutions to (7.11) (that satisfy these additional qualifications) is completely described by $\phi(p) = Kp^\alpha$ for constants $(K, \alpha) \gg 0$ (see, e.g., Aczél [1966, p. 41, Theorem 3]).

Lemmas 7.4 and 7.5 together establish "necessity", though it still remains to establish the bounds on α. We shall do so along with our proof of "sufficiency", which we begin now.

LEMMA 7.6. *Let f be a basic density with mass p and mean μ on support $[a, b]$. Let $m \equiv \mu - a$ and let f^* denote the root of f. Then, if f^λ denotes some λ-squeeze of f,*

$$P(F^\lambda) = 4kp^{2+\alpha}(m\lambda)^{1-\alpha}\int_0^1 f^*(x)^{1+\alpha}\left\{\int_0^1 f^*(y)(1-y)dy + \int_x^1 f^*(y)(y-x)dy\right\}dx \tag{7.19}$$

for some constant $k > 0$.

Proof. Recall that a slide of f has no effect on the computations, so we may as well set $a = 0$ and $b = 2m$, where $m = \mu - a$ is now to be interpreted as the mean. Given (7.2),

$$P(F) = k\int\int f(x)^{1+\alpha} f(y)|y - x|dydx \tag{7.20}$$

for some $k > 0$. Using the fact that f is symmetric, we can write

$$\begin{aligned} P(F) &= 2k \int_0^m \int_0^{2m} f(x')^{1+\alpha} f(y')|x'-y'|dy'dx' \\ &= 2k \int_0^m f(x')^{1+\alpha} \left\{ \int_0^{x'} f(y')(x'-y')dy' + \int_{x'}^m f(y')(y'-x')dy' \right. \\ &+ \left. \int_m^{2m} f(y')(y'-x')dy' \} dx' \right\}. \end{aligned} \tag{7.21}$$

Examine the very last term in (7.21). Change variables by setting $z \equiv 2m - y'$, and use symmetry to deduce that

$$\int_m^{2m} f(y')(y'-x')dy' = \int_0^m f(z)(2m - x' - z)dz.$$

Substituting this in (7.21), and manipulating terms, we obtain

$$P(F) = 4k \int_0^m f(x')^{1+\alpha} \left\{ \int_0^m f(y')(m-y')dy' + \int_{x'}^m f(y')(y'-x')dy' \right\} dx'. \tag{7.22}$$

Now suppose that f^λ is a λ-squeeze of f. Note that (7.22) holds just as readily for f^λ as for f. Therefore, using the expression for f given in (7.1), we see that

$$\begin{aligned} P(F^\lambda) &= 4k\lambda^{-(2+\alpha)} \int_{(1-\lambda)m}^m f\left(\frac{x'-(1-\lambda)m}{\lambda}\right)^{1+\alpha} \left\{ \int_{(1-\lambda)m}^m f\left(\frac{y'-(1-\lambda)m}{\lambda}\right)(m-y')dy' \right. \\ &+ \left. \int_{x'}^m f\left(\frac{y'-(1-\lambda)m}{\lambda}\right)(y'-x')dy' \} dx' \right\}. \end{aligned}$$

Perform the change of variables $x'' = \frac{x'-(1-\lambda)m}{\lambda}$ and $y'' = \frac{y'-(1-\lambda)m}{\lambda}$. Then it is easy to see that

$$P(F^\lambda) = 4k\lambda^{1-\alpha} \int_0^m f(x'')^{1+\alpha} \left\{ \int_0^m f(y'')(m-y'')dy'' + \int_{x''}^m f(y'')(y''-x'')dy'' \right\} dx''.$$

To complete the proof, we must recover the root f^* from f. To this end, first population-scale f to h, where h has mass 1. That is, $f(z) = ph(z)$ for all z. Doing so, we see that

$$P(F^\lambda) = 4kp^{2+\alpha}\lambda^{1-\alpha} \int_0^m h(x'')^{1+\alpha} \left\{ \int_0^m h(y'')(m-y'')dy'' + \int_{x''}^m h(y'')(y''-x'')dy'' \right\} dx''.$$

Finally, make the change of variables $x = x''/m$ and $y = y''/m$. Noting that $f^*(z) = mh(mz)$, we get (7.19).

Q.E.D.

LEMMA 7.7. *Let f and g be two basic densities with disjoint support, with their means separated by distance d, and with population masses p and q respectively. Let f have mean μ on support $[a, b]$. Let $m \equiv \mu - a$ and let f^* denote the root of f. Then for any λ -squeeze f^λ of f,*

$$A(f^\lambda, g) = 2kdp^{1+\alpha}q(m\lambda)^{-\alpha} \int_0^1 f^*(x)^{1+\alpha}dx, \tag{7.23}$$

where $A(f^\lambda, g)$ denotes the total effective antagonism felt by members of f^λ towards members of g.

Proof. Without loss of generality, let f have support $[0, 2m]$ (with mean m) and g have support $[d, d+2m]$ (where $d \geq 2m$ for disjoint supports). Using (7.20),

$$\begin{aligned} A(f,g) &= k\int_0^{2m} f(x)^{1+\alpha}\left[\int_d^{d+2m} g(y)(y-x)dy\right]dx \\ &= k\int_0^{2m} f(x)^{1+\alpha}\left[\int_d^{d+m} g(y)(y-x)dy + \int_{d+m}^{d+2m} g(y)(y-x)dy\right]dx \\ &= k\int_0^{2m} f(x)^{1+\alpha}\left[\int_d^{d+m} g(y)2(m+d-x)dy\right]dx \\ &= kq\int_0^{2m} f(x)^{1+\alpha}(m+d-x)dx \\ &= 2dkq\int_0^{m} f(x)^{1+\alpha}dx, \end{aligned}$$

where the third equality exploits the symmetry of g,[2] the fourth equality uses the fact that $\int_d^{d+m} g(y) = q/2$, and the final equality uses the symmetry of f.[3] To be sure, this formula applies to any λ -squeeze of f, so that

$$\begin{aligned} A(f^\lambda, g) &= 2dkq\int_0^{m} f^\lambda(x')^{1+\alpha}dx' \\ &= 2dkq\lambda^{-(1+\alpha)}\int_{(1-\lambda)m}^{m} f\left(\frac{x'-(1-\lambda)m}{\lambda}\right)^{1+\alpha}dx', \end{aligned}$$

and making the change of variables $x'' = \frac{x'-(1-\lambda)m}{\lambda}$, we may conclude that

$$A(f^\lambda, g) = 2dkq\lambda^{-\alpha}\int_0^m f(x'')^{1+\alpha}dx''.$$

To complete the proof, we must recover the root f^* from f. As in the proof of Lemma 7.6, first population-scale f to h, where h has mass 1. That is, $f(z) = ph(z)$ for all z. Doing so, we see that

$$A(f^\lambda, g) = 2dkp^{1+\alpha}q\lambda^{-\alpha}\int_0^m h(x'')^{1+\alpha}dx''.$$

Finally, make the change of variables $x = x''/m$. Noting that $f^*(z) = mh(mz)$, we get (7.23). ∎

Lemma 7.8. *Define, for any root f and $\alpha > 0$,*

$$\psi(f,\alpha) \equiv \frac{\int_0^1 f(x)^{1+\alpha}dx}{\int_0^1 f(x)^{1+\alpha}\left\{\int_0^1 f(y)(1-y)dy + \int_x^1 f(y)(y-x)dy\right\}dx}. \tag{7.24}$$

Then — for any $\alpha > 0$ — $\psi(f,\alpha)$ attains its minimum value when f is the uniform root, and this minimum value equals 3.

[2]That is, for each $y \in [d, d+m]$, $g(y) = g(d+2m-(y-d)) = g(2d+2m-y)$. Moreover, $[y-x]+[(2d+2m-y)-x] = 2(d+m-x)$.

[3]That is, for each $x \in [0, m]$, $f(x) = f(2m-x)$. Moreover, $[m+d-x]+[m+d-(2m-x)] = 2d$.

Proof. It will be useful to work with the inverse function

$$\zeta(f,\alpha) \equiv \psi(f,\alpha)^{-1} = \frac{\int_0^1 f(x)^{1+\alpha}\left\{\int_0^1 f(y)(1-y)dy + \int_x^1 f(y)(y-x)dy\right\}dx}{\int_0^1 f(x)^{1+\alpha}dx}.$$

Note that $\zeta(f,\alpha)$ may be viewed as a weighted average of

$$L(x) \equiv \int_0^1 f(y)(1-y)dy + \int_x^1 f(y)(y-x)dy \tag{7.25}$$

as this expression varies over $x \in [0,1]$, where the "weight" on a particular x is just

$$\frac{f(x)^{1+\alpha}}{\int_0^1 f(z)^{1+\alpha}dz}$$

which integrates over x to 1. Now observe that $L(x)$ is *decreasing* in x. Moreover, by the unimodality of a root, the weights must be nondecreasing in x. It follows that

$$\zeta(f,\alpha) \leq \int_0^1 L(x)dx. \tag{7.26}$$

Now

$$\begin{aligned} L(x) &= \int_0^1 f(y)(1-y)dy + \int_x^1 f(y)(y-x)dy \\ &= \int_0^1 f(y)(1-x)dy + \int_0^x f(y)(x-y)dy \\ &= \frac{1-x}{2} + \int_0^x f(y)(x-y)dy. \qquad (7.27) \end{aligned}$$

Because $f(x)$ is nondecreasing and integrates to 1/2 on [0,1], it must be the case that $\int_0^x f(y)(x-y)dy \leq \int_0^x (x-y)/2dy$ for all $x \leq 1$. Using this information in (7.27) and combining it with (7.26),

$$\begin{aligned} \zeta(f,\alpha) &\leq \int_0^1 \left[\frac{1-x}{2} + \int_0^x \frac{x-y}{2}dy\right]dx \\ &= \int_0^1 \left[\int_0^1 \left[\frac{1-y}{2}\right]dy + \int_x^1 \left[\frac{y-x}{2}\right]dy\right]dx \\ &= \zeta(u,\alpha), \qquad (7.28) \end{aligned}$$

where u stands for the uniform root taking constant value 1/2 on [0,2]. Simple integration reveals that $\zeta(u,\alpha) = 1/3$.

Q.E.D.

LEMMA 7.9. *Given that $P(f)$ is of the form (7.20), Axiom 1 is satisfied if and only if $\alpha \leq 1$.*

Proof. Simply inspect (7.19).

Q.E.D.

Lemma 7.10. *Given that $P(f)$ is of the form (7.20), Axiom 2 is satisfied if and only if $\alpha \geq 0.25$.*

Proof. Consider a configuration as given in Axiom 2: a symmetric distribution made out of three basic densities. By symmetry, the side densities must share the same root; call this f^*. Let p denote their (common) population mass and m their (common) difference from their means to their lower support. Likewise, denote the root of the middle density by g^*, by q its population mass, and by n the difference between mean and lower support. As in the proof of Lemma 7.2, we may decompose the polarization measure (7.20) into several components. First, there are the "internal polarizations" of the middle density (P_m) and of the two side densities (P_s). Next, there are various subtotals of effective antagonism felt by members of one of the basic densities towards another basic density. Let A_{ms} denote this when the "origin" density is the middle and the "destination" density one of the sides. Likewise, A_{sm} is obtained by permuting origin and destination densities. Finally, denote by A_{ss} the total effective antagonism felt by inhabitants of one side towards the other side. Observe that each of these last four terms appear twice, so that (writing everything as a function of λ), overall polarization is given by

$$P(\lambda) = P_m + 2P_s(\lambda) + 2A_{ms}(\lambda) + 2A_{sm}(\lambda) + 2A_{ss}(\lambda). \tag{7.29}$$

Compute these terms. For brevity, define for any root h,

$$\psi_1(h,\alpha) \equiv \int_0^1 h(x)^{1+\alpha}\left\{\int_0^1 h(y)(1-y)dy + \int_x^1 h(y)(y-x)dy\right\}dx$$

and

$$\psi_2(h,\alpha) \equiv \int_0^1 h(x)^{1+\alpha}dx.$$

Now, using Lemmas 7.6 and 7.7, we see that

$$P_s(\lambda) = 4kp^{2+\alpha}(m\lambda)^{1-\alpha}\psi_1(f^*,\alpha),$$

while

$$A_{ms}(\lambda) = 2kdq^{1+\alpha}pn^{-\alpha}\psi_2(g^*,\alpha).$$

Moreover,

$$A_{sm}(\lambda) = 2kdp^{1+\alpha}q(m\lambda)^{-\alpha}\psi_2(f^*,\alpha),$$

and

$$A_{ss}(\lambda) = 4kdp^{2+\alpha}(m\lambda)^{-\alpha}\psi_2(f^*,\alpha),$$

(where it should be remembered that the distance between the means of the two side densities is $2d$). Observe from these calculations that $A_{ms}(\lambda)$ is entirely insensitive to λ. Consequently, feeding all the computed terms into (7.29), we may conclude that

$$P(\lambda) = C\left[2\lambda^{1-\alpha} + \frac{d}{m}\psi(f^*,\alpha)\lambda^{-\alpha}\{\frac{q}{p}+2\}\right] + D,$$

where C and D are positive constants independent of λ, and

$$\psi(f^*,\alpha) = \frac{\psi_2(f^*,\alpha)}{\psi_1(f^*,\alpha)}$$

by construction; see (7.24) in the statement of Lemma 7.8. It follows from this expression that for Axiom 2 to hold, it is necessary and sufficient that for *every* three-density configuration of the sort described in that axiom,

$$2\lambda^{1-\alpha} + \frac{d}{m}\psi(f^*,\alpha)\lambda^{-\alpha}\left[\frac{q}{p}+2\right] \tag{7.30}$$

must be nonincreasing in λ over $(0,1]$. An examination of the expression in (7.30) quickly shows that a situation in which q is arbitrarily close to zero (relative to p) is a necessary and sufficient test case. By the same logic, one should make d/m as small as possible. The disjoint-support hypothesis of Axiom 2 tells us that this lowest value is 1. So it will be necessary and sufficient to show that for every root f^*,

$$\lambda^{1-\alpha} + \psi(f^*,\alpha)\lambda^{-\alpha} \tag{7.31}$$

is nonincreasing in λ over $(0,1]$. For any f^*, it is easy enough to compute the necessary and sufficient bounds on α. Simple differentiation reveals that

$$(1-\alpha)\lambda^{-\alpha} - \alpha\psi(f^*,\alpha)\lambda^{-(1+\alpha)}$$

must be nonnegative for every $\lambda \in (0,1]$; the necessary and sufficient condition for this is

$$\alpha \geq \frac{1}{1+\psi(f^*,\alpha)}. \tag{7.32}$$

Therefore, to find the necessary and sufficient bound on α (uniform over all roots), we need to minimize $\psi(f^*,\alpha)$ by choice of f^*, subject to the condition that f^* be a root. By Lemma 7.8 , this minimum value is 3. Using this information in (7.32), we are done. ■

LEMMA 7.11. *Given that $P(f)$ is of the form (7.20), Axiom 3 is satisfied.*

Proof. Consider a symmetric distribution composed of four basic densities, as in the statement of Axiom 3. Number the densities 1, 2, 3 and 4, in the same order displayed in Figure 7.2. Let x denote the amount of the slide (experienced by the inner densities) in the axiom. For each such x, let $d_{jk}(x)$ denote the (absolute) difference between the means of basic densities j and k. As we have done several times before, we may decompose the polarization of this configuration into several components. First, there is the "internal polarization" of each rectangle j; call it P_j, $j = 1,2,3,4$. [These will stay unchanged with x.] Next, there is the total effective antagonism felt by inhabitants of each basic density towards another; call this $A_{jk}(x)$, where j is the "origin" density and k is the "destination" density. Thus total polarization $P(x)$, again written explicitly as a function of x, is given by

$$P(x) = \sum_{j=1}^{4} P_j + \sum_j \sum_{k\neq j} A_{jk}(x)$$

so that, using symmetry,

$$P(x) - P(0) = 2\{[A_{12}(x) + A_{13}(x)] - [A_{12}(0) + A_{13}(0)]\} + [A_{23}(x) - A_{23}(0)] \tag{7.33}$$

Now Lemma 7.7 tells us that for all i and j,

$$A_{ij}(x) = k_{ij}d_{ij}(x),$$

where k_{ij} is a positive constant which is independent of distances across the two basic densities, and in particular is independent of x. Using this information in (7.33), it is trivial to see that

$$P(x) - P(0) = A_{23}(x) - A_{23}(0) = k_{ij}x > 0,$$

so that Axiom 3 is satisfied. ▮

Given (7.20), Axiom 4 is trivial to verify. Therefore Lemmas 7.9, 7.10 and 7.11 complete the proof of the theorem.

Q.E.D.

7.6 Polarization and Inequality

Our measurement of polarization is given by

$$\text{Pol} = \int\int f(x)^{1+\alpha} f(y)|y-x|dydx,$$

where α lies between 0.25 and 1.

Compare with the Gini coefficient / fractionalization index:

$$\text{Gini} = \int\int f(x)f(y)|y-x|dydx,$$

It's the value of α that makes all the difference. The lower bound on α comes from Axiom 2, which directly runs against inequality. Otherwise the other axioms are broadly in agreement in inequality.

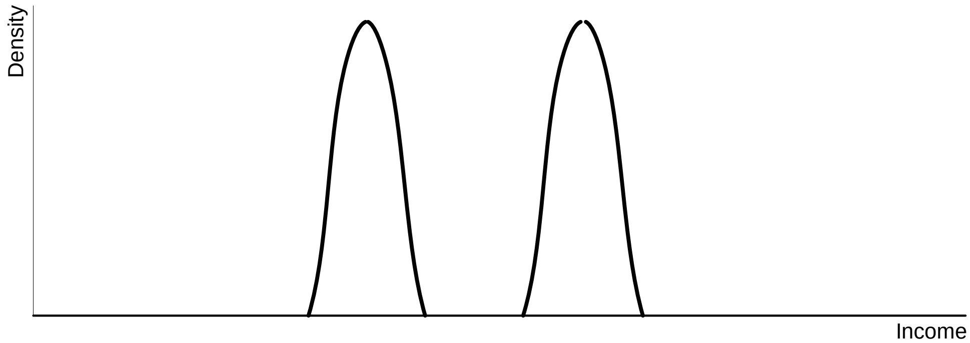
Note by the way that there also must be an upper bound on α. For if this were not the case, identification effects will swamp alienation effects and global squeezes in the distribution will increase polarization, which makes no sense. This is the job of Axiom 1.

7.7 Three New Properties

1. Bimodality. Polarization maximal for bimodal distributions, but defined of course over all distributions.

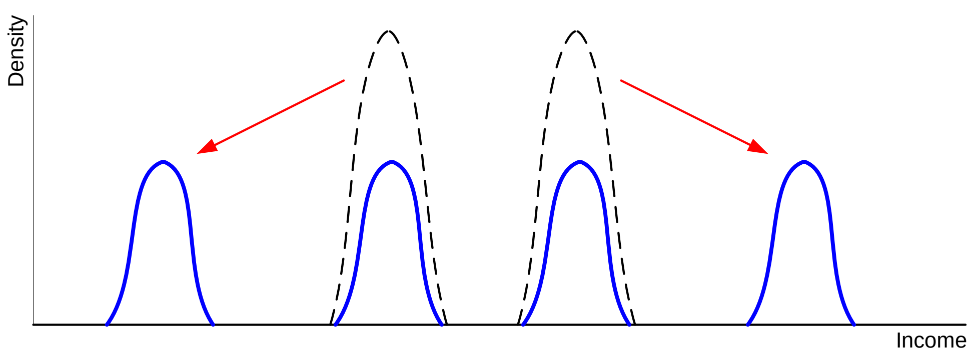
2. Globality. The local merging of two "groups" *raises* polarization if there is a third "group" of significant size, but *lowers* conflict in the absence of such a group.

3. Nonlinearity. Same direction of population or income movements may cause polarization to go down or up, depending on context.



3. Nonlinearity. Same direction of population or income movements may cause polarization to go down or up, depending on context.

3. Nonlinearity. Same direction of population or income movement may cause polarization to go down or up, depending on context.

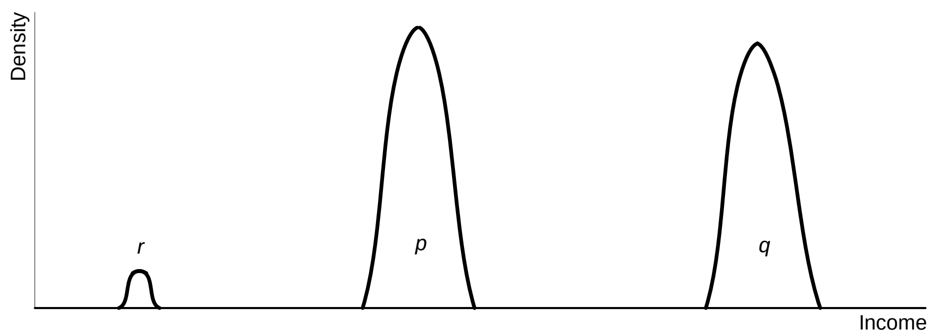


7.8 Other Remarks on the Measure

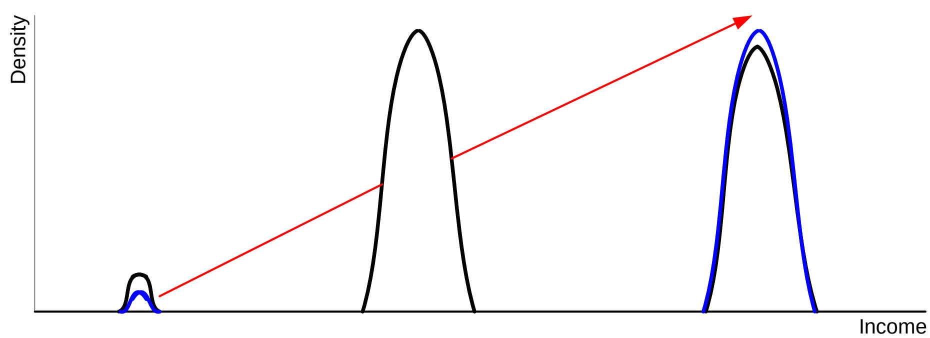
7.8.1 More on α. As we've discussed, there is a family of possible values of α, subject to the lower and upper bounds of 0.25 and 1. We've already explained where these bounds come from. The bounds can be narrowed further using additional axioms, though it is unclear how compelling the additional steps are. Here is an example of the axiomatic approach:

Axiom 5. If $p > q$ but $p - q$ is small and so is r, a small shift of mass from r to q cannot reduce polarization.

Diagrammatically, go from here:



to here:



Theorem 7.2. *Under the additional Axiom 5, it must be that $\alpha = 1$, so the unique polarization measure that satisfies the five axioms is proportional to*

$$\int \int f(x)^2 f(y)|y - x| dy dx.$$

In a later section, we argue that there are strong *behavioral* reasons to focus on $\alpha = 2$.

7.8.2 Scaling.

Note that our main theorem states that polarization must be proportional to a particular family of measures. There is therefore a scale factor here which can be employed as we please. While this is not entirely precise, we can exploit this degree of freedom to make the measure scale-free. Normalize the measure by $\mu^{\alpha-1}$, where μ is mean income.

This procedure is equivalent to one in which all incomes are normalized by their mean. Esteban and Ray (1994) begin, in contrast, by using the log of incomes and imposing axioms on this variable.

7.8.3 Importance of the IA Structure..

Both the axioms and the IA structure needed to pin down P.

It can be checked that several other candidates satisfy Axioms 1–4. These details are to be included in future editions of the notes.

7.8.4 Partial Ordering.

α varies between bounds, but *can* vary. So theorem gives us a partial ordering. Ordinal description of this ordering is an open question.

7.8.5 Identification Windows..

"Identification" here is based on the point density. More generally, individuals may possess a "window of identification". Individuals within this window would be considered "similar" — possibly with weights decreasing with the distance — and would contribute to a sense of group identity.

At the same time, individuals would feel alienated only from those outside the window. Thus, broadening one's window of identification has two effects.

Can capture these two effects somewhat in our seemingly narrower model.

Suppose that each individual at x "perceives" an individual with income y to be at the point $(1-t)x + ty$. Thus the parameter t is inversely proportional to "breadth of identification". The "perceived density" of y from the vantage point of an individual located at x is then

$$\frac{1}{t} f\left(\frac{y-(1-t)x}{t}\right)$$

It is easy to see that the polarization measure resulting from this extended notion of identification is proportional to our measure by the factor $t^{1-\alpha}$.

It is also possible to directly base identification on the average density over a non-degenerate window of width w. When we take w to zero, not only is P attained in the limit, but P is a first-order approximation to P^w, in the sense that $\partial P^w/\partial w|_{w\to 0} = 0$.

Nevertheless, the question of identification windows deserves more attention.

7.8.6 Comparing Distributions.

Our characterization shows that a comparison across distributions should depend on average alienation, average identification and on their joint co-movement. We formalize this very quickly.

Let $a(y) = \int |y - x| dF(x)$, which is the alienation felt by y, and define average alienation $\bar{a}$ by

$$\bar{a} = \int a(y) dF(y) = \int \int |y - x| dF(x) dF(y).$$

Similarly, $f(y)^\alpha$ is the identification felt by y, so define the *average* α-identification $\bar{\iota}$ by

$$\bar{\iota}_\alpha \equiv \int f(y)^\alpha dF(y) = \int f(y)^{1+\alpha} dy.$$

Finally, let ρ be the *normalized covariance* between identification and alienation:

$$\rho \equiv \mathrm{cov}_{\iota_\alpha, a} / \bar{\iota}_\alpha \bar{a}.$$

Then it is easy to verify that

$$P_\alpha(f) = \bar{a}\bar{\iota}_\alpha \left[1 + \rho\right].$$

7.9 Social Polarization

Our polarization measure is easily applicable to ethnolinguistic or religious groupings. Suppose that there are M "social groups", based on region, kin, ethnicity, religion... Let n_j be the number of individuals in group j, with overall population normalized to one. Let F_j describe the distribution of income in group j (with f_j the accompanying density), unnormalized by group population. One may now entertain a variety of "social polarization measures".

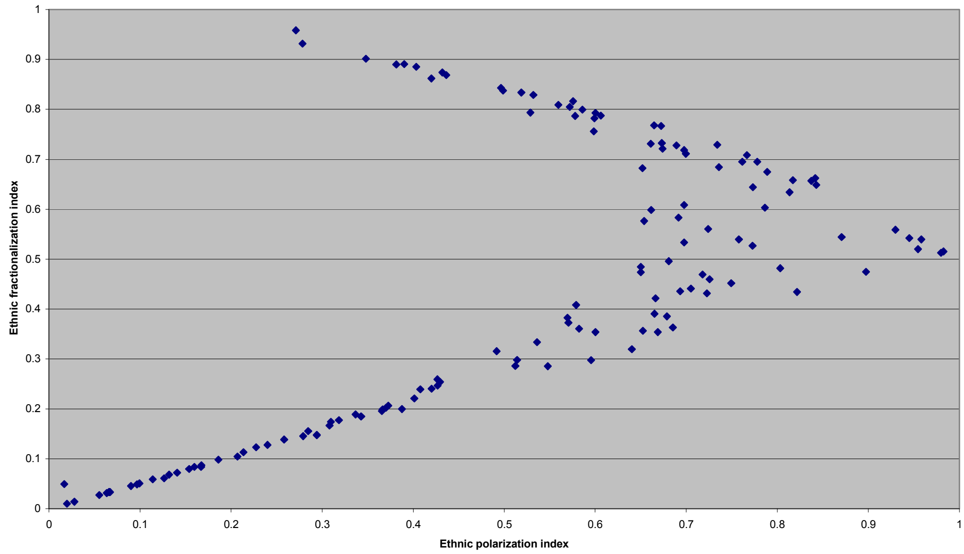
7.9.1 Pure Social Polarization. Consider, first, the case of "pure social polarization", in which income plays no role. Assume that each person is "fully" identified with every other member of his group. Likewise, the alienation function takes on values that are specific to group pairs and have no reference to income. For each pair of groups j and k denote this value by Δ_{jk}. Then a natural transplant of (7.2) yields the measure

$$P_s(\mathbf{F}) = \sum_{j=1}^{M} \sum_{k=1}^{M} n_j^{1+\alpha} n_k \Delta_{jk}. \tag{7.34}$$

Even this sort of specification may be too general in some interesting instances in which individuals are interested only in the dichotomous perception Us/They. In particular, in these instances, individuals are not interested in differentiating between the different opposing groups. Perhaps the simplest instance of this is a pure contest (Esteban and Ray [1999]), which yields the variant[4]

$$\tilde{P}_s(\mathbf{F}) = \sum_{j=1}^{M} n_j^{1+\alpha}(1 - n_j). \tag{7.35}$$

[4]See Reynal-Querol [2002] for a similar analysis. D'Ambrosio and Wolff [2001] also consider a measure of this type but with income distances across groups explicitly considered.



If we specialize to the case of $\alpha = 1$, then a special index of "pure" social polarization is given by

$$\text{Pol} = \sum_{j=1}^{M} \sum_{k=1}^{M} n_j^2 n_k = \sum_{j=1}^{M} n_j^2(1 - n_j).$$

This is the measure used by Montalvo and Reynal-Querol, *American Economic Review* (2005), a paper that we discuss in more detail below. It is instructive to recall the ELF, which is given by

$$\text{ELF} = \sum_{j=1}^{M} n_j(1 - n_j),$$

and compare the two.

For instance, if all groups are of equal size, then polarization peaks when the number of groups equals 2, and steadily declines thereafter. Fractionalization rises throughout.

As Montalvo and Reynal-Querol show, it matters empirically too . . .

Guatemala and Sierra Leone are examples of countries in which ethnic polarization is high but ethnic fractionalization is low. Nigeria and Bosnia are examples of countries in which religious polarization is high but religious fractionalization is low.

Once the two extremes — pure income polarization and pure social polarization — are identified, we may easily consider several hybrids. As examples, consider the case in which notions of identification are mediated not just by group membership but by income similarities as well, while the antagonism equation remains untouched. Then we get what one might call *social polarization with income-mediated identification*:

$$(7.36) \qquad P_s(\mathbf{F}) = \sum_{j=1}^{M}(1-n_j)\int_x f_j(x)^{\alpha} dF_j(x).$$

One could expand (or contract) the importance of income further, while still staying away from the extremes. For instance, suppose that — in addition to the income-mediation of group identity — alienation is also income-mediated (for alienation, two individuals must belong to different groups *and* have different incomes). Now groups have only a demarcating role — they are necessary (but not sufficient) for identity, and they are necessary (but not sufficient) for alienation. The resulting measure would look like this:

$$(7.37) \qquad P^{*}(\mathbf{F}) = \sum_{j=1}^{M}\sum_{k\neq j}\int_x\int_y f_j(x)^{\alpha}|x-y| dF_j(x) dF_k(y).$$

Note that we do not intend to suggest that other special cases or hybrids are not possible, or that they are less important. The discussion here is only to show that social and economic considerations can be profitably combined in the measurement of polarization. Indeed, it is conceivable that such measures will perform better than the more commonly used fragmentation measures in the analysis of social conflict. But a full exploration of this last theme must await future research (though see the chapter on ethnic conflict below).

7.10 Empirics Revisited: Polarization, not Fragmentation?

We now discuss the recent results of Montalvo and Reynal-Querol (2005). They employ the same basic specification as Fearon-Laitin (2003) and others, but this time with polarization instead of fractionalization indices.

They study 138 countries over 1960–1995. The dependent variable is incidence of a civil war over a five year period. They use what is known as the PRIO25 criterion for civil war, at least 25 yearly deaths. (Refine this in later versions.) Their *explanatory variables* include per-capita income, population size, terrain (proxy for ease of insurgency), primary exports (proxy for payoff in event of victory), democracy indicators, and of course indices of ethnic or religious polarization

First run a logit of civil war on ethnic fractionalization. Table 1 reports on the results. Observe how fractionalization matters in the first column but loses significance completely as variables such as per-capita income are included. This is the standard result that we had described earlier.

Now for the logit using ethnic polarization. Table 2 reports. Ethnic polarization is not just significant through all the variants; the effect is pretty big too. For instance, if polarization is raised from 0.51 (the average) to 0.95 (Nigeria) the predicted probability of conflict doubles. [An increase by one standard deviation (0.24) raises conflict probability by 50%.]

Now try the same logit with religious variables instead. Table 3 shows that just like ethnic fractionalization, religious fractionalization starts out significant but matters quickly degenerate when additional controls are thrown in. This finding is to be contrasted with

	[1]	[2]	[3]	[4]
EthFrac	0.81	0.22	-0.18	0.49
	(2.04)	(0.53)	(0.16)	(0.97)
LogPcGdp	-0.62	-0.76	-0.79	-0.93
	(5.07)	(5.90)	(5.96)	(5.40)
Constant	2.47	-0.42	-0.18	1.57
	(2.47)	(0.38)	(0.16)	(0.94)
LogPop		0.46	0.46	0.35
		(6.75)	(6.03)	(3.69)
PrimExp			0.25	0.50
			(0.26)	(0.48)
Mountains				0.00
				(1.67)
NonContiguous				-0.20
				(0.61)
Democracy				0.49
				(1.87)
Pseu R^2	0.07	0.15	0.15	0.14
Obs	860	860	840	741

TABLE 1. Ethnic Fractionalization and Conflict

what happens when a religious polarization variable is employed (Table 4). Religious polarization stays significant through the various specifications, just as ethnic polarization did.

These observations are robust to several different specifications. Ethnic polarization is significant when entered into same regression with ethnic fractionalization; the latter is not. The same is true if a measure of ethnic dominance (Collier (2001) and Collier and Hoeffler (2002)) is used instead. Both these observations are still true if "ethnic" is replaced by "religious".

The analysis is also robust to the use of different datasets. The World Christian Encyclopedia is used here to construct ethnic polarization indices. Alternatively, the Encyclopedia Britannica or the Atlas Nadorov Mira could be used. My main worry here is that I am not sure how the authors have constructed their ethnic groupings. Clearly they ahev aggregated some of the highly disaggregated information. But what if they've done so in some way that's endogenous?

The results are also robust to "joint indices" of ethnic and religious polarization.[Measure along each dimension, pick the max.] Finally, it appears to be robust to alternative definitions of civil war.

	[1]	[2]	[3]	[4]
EthPol	1.56	1.95	1.98	1.82
	(3.31)	(3.76)	(3.71)	(3.23)
LogPcGdp	-0.71	-0.77	-0.78	-0.93
	(6.16)	(6.53)	(6.57)	(5.50)
Constant	2.65	-1.56	-1.43	-0.93
	(3.01)	(1.47)	(1.27)	(0.16)
LogPop		0.49	0.48	0.38
		(7.15)	(6.46)	(4.33)
PrimExp			-0.09	0.17
			(0.09)	(0.16)
Mountains				0.00
				(1.13)
NonContiguous				-0.00
				(0.00)
Democracy				0.41
				(1.58)
Pseu R^2	0.09	0.17	0.17	0.16
Obs	860	860	840	741

Table 2. Ethnic Polarization and Conflict

7.11 Next Steps

Let's return to the old question: is ethnicity primordial or instrumental as a determinant of conflict?

An economist's instincts suggests that these things are essentiall;y instrumental but I suppose one never knows . . .

Recall that the findings on per-capita income certainly support an instrumentalist position on civil war. That suggests a test for ethnic instrumentality: see if economic differences across groups predicts conflict. But there is an important and fundamental reason why such a test may be problematic.

Two kinds of economic conflict: "vertical" versus "horizontal"

The "vertical war" certainly exists but is harder to spot and infrequently delineated by ethnicity. Caste is a good counterexample. With ethnicity and religion, the conflict is often horizontal: attacks on competing businesses, reduction of labor supply, reallocation of specific public goods. Listen to Horowitz again:

> "In study after study, it has been assumed that ethnic relations are necessary relations between superiors and subordinates . . . In fact, many ethnic groups *are* enmeshed in a system of subordination. But the relations of

	[1]	[2]	[3]	[4]
RelFrac	1.41	0.53	0.35	0.92
	(2.31)	(0.76)	(0.49)	(1.17)
LogPcGdp	-0.61	-0.84	-0.87	-1.03
	(4.91)	(5.75)	(5.85)	(5.27)
Constant	1.53	-1.24	-1.15	0.45
	(1.42)	(0.97)	(0.86)	(0.25)
LogPop		0.50	0.51	0.41
		(6.41)	(5.88)	(4.09)
PrimExp			0.63	1.15
			(0.61)	(1.04)
Mountains				0.01
				(2.17)
NonContiguous				0.10
				(0.31)
Democracy				0.36
				(1.29)
Pseu R^2	0.10	0.16	0.16	0.16
Obs	853	853	833	734

TABLE 3. Religious Fractionalization and Conflict

many other ethnic groups — on a global scale, most ethnic groups — are not accurately defined as superior-subordinate relations . . .

Unlike ranked groups, which form part of a single society, unranked groups constitute incipient whole societies. It is not so much the politics of subordination that concerns them, but rather the politics of inclusion and exclusion."

How then to augment our polarization measure for wealth differences across and within groups? The answer may depend on the observer's feel for the sort of conflict that is relevant.

If conflict is "vertical", income differences *across* groups are conducive to conflict, and so is income homogeneity *within* groups

On the other hand, if conflict is "horizontal", income or occupational *similarities* across groups may drive conflict and so might income *inequality* within groups (the buying of "conflict labor", as with Dalits in the Gujarat carnage). I will come back to these matters later.

The interaction of economics and ethnicity creates new conceptual challenges for the measurement of polarization.

	[1]	[2]	[3]	[4]
RelPol	1.09	0.71	0.65	1.06
	(2.93)	(1.71)	(1.50)	(2.20)
LogPcGdp	-0.57	-0.76	-0.78	-0.98
	(4.46)	(5.22)	(5.26)	(5.08)
Constant	1.17	-1.93	-1.85	0.17
	(1.10)	(1.52)	(1.40)	(0.10)
LogPop		0.49	0.50	0.39
		(6.36)	(5.75)	(3.94)
PrimExp			0.41	0.93
			(0.39)	(0.84)
Mountains				0.01
				(2.12)
NonContiguous				0.16
				(0.47)
Democracy				0.35
				(1.26)
Pseu R^2	0.10	0.17	0.17	0.17
Obs	853	853	833	734

TABLE 4. Religious Polarization and Conflict

7.12 Summary So Far

1. Several authors, notably Samuel Huntington, have argued for cultural explanations of economic development (or lack thereof).

2. Extending this line of thinking, authors such as Donald Horowitz have suggested links between conflict and ethnic differences.

3. But statistical studies that employ a well-known measure of ethnic and religious fragmentation show no links with conflict. [Though there are links with economic growth.]

4. In this chapter, I argue for the use of a measure very different from fragmentation — a *polarization index*.

5. The measure has a philosophical foundation — the *identity-alienation* framework — which may turn out to be useful in other applications.

6. I then discuss an empirical study which uses this polarization measure to exhibit a robust and positive relationship between (ethnic or religious) polarization and the incidence of conflict.

7. This does *not* mean that we buy Huntington on the primordial nature of cultural differences. An economic war may still be waged behind the ethnic veil.

7.13 A Behavioral Approach to Polarization

No measurement theory can do justice to the claim that there is indeed a link between polarization and conflict.

Two potential links:

(1) *Empirical*: fit the measures to the data and see they are related to measured indicators of social conflict.
(2) *Theoretical*: write down a "natural" theory of conflict in which the level of equilibrium conflict is related to distributional polarization.

Here, we concentrate on the latter (with an eye on the former, of course).

7.14 Benchmark Model of Conflict

A unit measure of individuals, situated in G groups. n_i: number of individuals in group i, so that $\sum_{i=1}^{G} n_i = 1$.

u_{ij}: utility derived by a member of group i if issue j is chosen by society. $u_{ii} > u_{ij}$ for all i, j with $i \neq j$.

r_i: resources expended per-capita by group i, so total is $n_i r_i$.

[Will take a closer look at this later.]

Let $R \equiv \sum_{i=1}^{G} n_i r_i$. R is our measure of societal *conflict*.

Per-capita cost of supplying effort r is $c(r)$: continuous, increasing, smooth, strictly convex, with $c'(0) = 0$.

7.15 Equilibrium

p_j: the probability that issue j will be chosen. A member of group i who expends r_i gets expected payoff given by

(7.38)
$$\sum_{j=1}^{G} p_j u_{ij} - c(r_i)$$

To complete the description of the basic model, we assume that

(7.39)
$$p_j = s_j \equiv \frac{n_j r_j}{\sum_{k=1}^{G} n_k r_k} = \frac{n_j r_j}{R}$$

for all $j = 1, \ldots, G$, if $R > 0$ (any probability vector otherwise).

$(r_1^*, \ldots, r_G^*)$ is an *equilibrium* if for every i the maximization of the expression in (7.38), subject to (7.39), is well-defined, and r_i solves this problem.

Equilibrium conflict is $R^* = \sum_{i=1}^{G} r_i^* n_i$.

The equilibrium *resource shares* are $s_i^* = r_i^* n_i / R^*$.

Neglect internal free-rider problem.

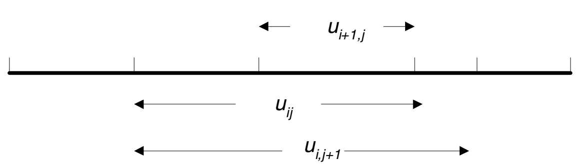
Also assume that groups cannot commit to support issues that are not their favorite.

7.16 Special Cases

(1) *Pure Contests*: $u_{ij} = 0$ for all i, j, with $i \neq j$, and $u_{ii} = 1$ for all i.

(2) *The Line*: There is an ordering $1, 2, \ldots, G$ of the groups such that for all i and j with $i < j$, $u_{ij} \leq u_{i,j+1}$ (if $j < G$) and $u_{ij} \geq u_{i+1,j}$.

ace*0.2in



7.17 Background Results

1. *Best Response Condition.* For each group i, let $v_{ij} \equiv u_{ii} - u_{ij}$ for all j.

If $r_j > 0$ for some $j \neq i$, then $r_i > 0$, and is given by

$$\sum_{j=1}^{G} s_i s_j v_{ij} = c'(r_i) r_i$$

2. *Existence and Uniqueness.* An equilibrium exists. If, in addition, $c'''(r) \geq 0$ for all r, then the equilibrium is unique.

Remark. $c'''(r) \geq 0$ is "necessary" for uniqueness.

7.18 Special Case: Quadratic Costs

Let $c(r) = \frac{1}{2}r^2$ for $r \geq 0$.

First order conditions from lemma:

$$\sum_{j=1}^{G} s_i s_j v_{ij} = c'(r_i) r_i$$

Rewrite for the quadratic case:

$$\sum_{j=1}^{G} s_j v_{ij} = \frac{c'(r_i)R}{n_i} = \frac{r_i R}{n_i} = \frac{s_i R^2}{n_i^2}.$$

Cross-multiply:

$$\sum_{j=1}^{G} s_j n_i^2 v_{ij} = s_i R^2.$$

$$\sum_{j=1}^{G} s_j n_i^2 v_{ij} = s_i R^2.$$

Form the matrix

$$\begin{pmatrix} n_1^2 v_{11} & n_1^2 v_{12} & \dots & n_1^2 v_{1G} \\ \vdots & \vdots & \vdots & \vdots \\ n_i^2 v_{i1} & n_i^2 v_{i2} & \dots & n_i^2 v_{iG} \\ \vdots & \vdots & \vdots & \vdots \\ n_G^2 v_{G1} & n_G^2 v_{G2} & \dots & n_G^2 v_{GG} \end{pmatrix}$$

Then R (squared) is the unique real eigenvalue of this matrix.

[The associated share vector is the unique positive eigenvector (on the simplex) corresponding to this eigenvalue.]

Remark. Observe the squaring of the population coefficients. This is related to polarization in a way that we will explain later.

7.19 Connections with Polarization

Several connections between behavioral model and the axiomatic approach:

1. Bimodality. Both conflict and measured polarization are maximal for symmetric bimodal distributions.

2. Globality. The local clustering of two groups raises conflict if there is a third group of significant size, but lowers conflict in the absence of a third group. Same true of polarization.

3. Nonmonotonicity. Start with a uniform distribution of population across G groups, where $G \geq 4$. Transfer population mass from one of the groups to the others, until a uniform distribution over $G - 1$ groups is obtained. Then conflict — and polarization — are higher at the "end" of this process, but may go down in the "intermediate" stages.

7.20 IsoElastic Costs: $c(r) = (1/\beta)r^\beta$

Use isoelasticity to rewrite first-order conditions:

$$s_i \sum_j s_j v_{ij} = r_i^\beta$$

Manipulating, we obtain

$$\left(\frac{n_i}{s_i}\right)^\beta s_i^2 \sum_j s_j v_{ij} = s_i R^\beta.$$

and adding over all groups i,

$$\boxed{R^\beta = \sum_i \sum_j \left(\frac{n_i}{s_i}\right)^\beta s_i^2 s_j v_{ij}}$$

Compare this result:

$$\boxed{R^\beta = \sum_i \sum_j \left(\frac{n_i}{s_i}\right)^\beta s_i^2 s_j v_{ij}}$$

with axiomatically derived polarization measure for the discrete case:

$$\boxed{P = \sum_i \sum_j n_i^{1+\alpha} n_j v_{ij}}$$

Pretty close for the case $\alpha = 1$.

Main difference: former is *not* a closed-form solution, because there are endogenous variables in it. This is the additional richness imparted by a behavioral model.

7.21 Four Research Questions

A. Group Formation and Interaction.

In many situations, groups are effectively given — men and women, Hindus and Muslims, importers and exporters ...

In others, there is a wider dispersion of preferences and (sharp) groups arise as a result of explicit membership decisions — political parties, trade unions, environmental coalitions.

Can be modeled as a two-stage game. For theory, see Bloch [1996, 1997], Ray and Vohra [1997, 1999, 2001], Konishi and Ray [2003], Yi [1997] ...

[1] In "stage 1", individuals, possibly endowed with widely varying characteristics, form groups. (They take "similar" actions, form coalitions or clubs, write binding agreements...)

[2] In "stage 2", groups "interact" in the way described earlier.

For some recent attempts for "conflict games" see Bloch and Sanchez (2003), Esteban and Sákovicz (2003), Tan and Wang (2000).

B. Group Salience.

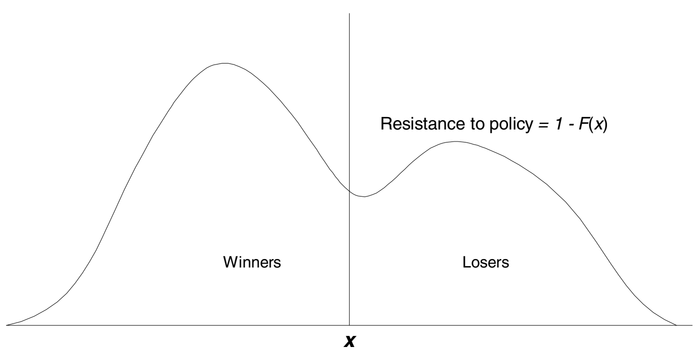
Return to the case of given groups. Many intersecting dimensions: what determines which group classification is *salient*?

To some extent, the answer must depend on the set of available *policies*.

One might say that a society is polarized if *the average resistance over a set of policies is high*.

Can apply this notion quite easily once we fix a space of policies and a measure over that space.

E.g., look at a very simple policy x, for which the winners are to one side of the income level x and the losers are on the other.



If we equate polarization to average resistance over a distribution G of policies, we get: $P = \int \min\{F(x), 1 - F(x)\} dG(x)$.

Might argue that G is uniform or even equal to F. Still begs the question: what determines the space of policies?

C. Effort and Money.

Behavioral model looks pleasantly general, but in fact the explicit introduction of income (or wealth) poses new challenges.

How are resources expended in the lobbying process?

[In a "perfect" democracy where all questions are solved by referenda, this problem becomes irrelevant.]

The poor use effort; the rich use money.

The political system determines a *political exchange rate* between effort and money.

Individual characteristics (abilities, access, wealth) determine an *economic exchange rate* between effort and money.

The cutoff for different types of participation (effort vs. money) will depend on the political system.

The structural features of such a model will impose useful restrictions on empirical analysis. Typically, instances of open conflict will be observable while the use of money to influence policies will not be.

Suggests the use of modified polarization measures in which the symmetry between different groups is broken if there are wealth differences.

D. Lobbies as Signals.

Based on Esteban and Ray (2006).

[1] Governments play a role in the allocation of resources.

[2] Governments lack information — just as private agents do — regarding which sectors are worth pushing in the interests of economic efficiency.

[3] Agents (sectoral interests, industrial confederations, R&D coalitions ...) lobby the government for preferential treatment.

[4] A government — even if it honestly seeks to maximize economic efficiency — may be confounded by the possibility that *both* high wealth and true economic desirability create loud lobbies.

Connects inequality and lobbying to resource allocation.

7.22 Summary

I have discussed two approaches to study of polarization and conflict.

The axiomatic approach delivers a new set of measures which may be useful in empirical work.

The behavioral approach complements the axiomatic approach by explicitly laying down a model of conflict.

While the axiomatic approach is possibly of greater empirical relevance, the behavioral approach cannot be dispensed with as a conceptual check on the axioms.

In particular, issues of group formation, group salience, the use of alternative forms of resources in a conflictual process, and role of lobbying as a signaling device can be usefully analyzed under the behavioral approach.

7.23 Other Issues in the Theory of Conflict

7.23.1 Public Versus Private Goods. So far we have assumed that conflict takes place over the allocation of public goods. One might also be interested in situations in which the good to be allocated is private or partly private. For simplicity, suppose that there are only two groups, with population sizes n_1 and n_2 summing to one. We also suppose that the cost function is isoelastic, $c(r) = (1/\alpha)r^\alpha$ for $\alpha \geq 2$ (thus satisfying our third-derivative condition).

There is a budget of G which is up for grabs. Assume that a fraction λ of this is public, while the remaining fraction is private and divided equally among the winning group. Group i then chooses r_i to maximize

$$\frac{n_i r_i}{R}\left[\lambda G + (1-\lambda)\frac{G}{n_i}\right] - c(r).$$

Define $P(n) = \left[\lambda G + (1-\lambda)\frac{G}{n_i}\right]$; then the FOC for this problem is

$$P(n_i)\left[\frac{n_i}{R} - \frac{n_i^2 r_i}{R^2}\right] = r_i^{\alpha-1},$$

or

$$P(n_i)n_i n_j = R^2\frac{r_i^{\alpha-1}}{r_j}. \tag{7.40}$$

Similarly,

$$P(n_j)n_i n_j = R^2\frac{r_j^{\alpha-1}}{r_i}. \tag{7.41}$$

Raising both sides of (7.41) to the power $\alpha - 1$ and multiplying with (7.40), we see that

$$P(n_i)n_i n_j[P(n_j)n_i n_j]^{\alpha-1} = R^{2\alpha} r_j^{\alpha(\alpha-2)}. \tag{7.42}$$

If $\alpha = 2$, we see right away that

$$R^4 = P(n_i)n_i^2 P(n_j)n_j^2, \tag{7.43}$$

and it is very easy to now see that this expression is maximized at the symmetric bipolar distribution in which $n_i = n_j = 1/2$. Thus in the quadratic case the degree to which the prize is public or private matters not at all in determining which distribution is the most conflictual.

When $\alpha \neq 2$ an extended version of (7.43) holds, which we now derive. Use equation (7.42) and the assumption that $\alpha \neq 2$ to write

$$r_j = \left[P(n_i)P(n_j)^{\alpha-1}\right]^{1/\alpha(\alpha-2)} (n_i n_j)^{1/(\alpha-2)} R^{-2/(\alpha-2)}. \tag{7.44}$$

Of course, a parallel equation holds for r_i:

$$r_i = \left[P(n_i)^{\alpha-1}P(n_j)\right]^{1/\alpha(\alpha-2)} (n_i n_j)^{1/(\alpha-2)} R^{-2/(\alpha-2)}. \tag{7.45}$$

Multiplying both sides of (7.44) by n_j and both sides of (7.45) by n_i, and adding the two equations, we see that

$$R = R^{-2/(\alpha-2)} \left[P(n_i)P(n_j)\right]^{(\alpha-1)/\alpha(\alpha-2)} (n_i n_j)^{1/(\alpha-2)} \left\{P(n_i)^{-1/\alpha} n_j + P(n_j)^{-1/\alpha} n_i\right\},$$

and after manipulating (quite) a bit, we conclude that

$$\begin{aligned} R^\alpha &= \left[P(n_i)P(n_j)\right]^{(\alpha-1)/\alpha} (n_i n_j) \left\{P(n_i)^{-1/\alpha} n_j + P(n_j)^{-1/\alpha} n_i\right\}^{\alpha-2} \\ &= \left[P(n_i)n_i P(n_j)n_j\right]^{1/\alpha} (n_i n_j)^{(\alpha-1)/\alpha} \left\{P(n_j)^{1/\alpha} n_j + P(n_i)^{1/\alpha} n_i\right\}^{\alpha-2} \end{aligned} \tag{7.46}$$

and this is the appropriate extension of (7.43), which was derived for $\alpha = 2$ (in fact, just put $\alpha = 2$ in (7.46) to recover (7.43)).

It can be shown that for $\alpha > 2$, this function is maximized at $n_i = n_j = 1/2$.[5] Details relegated to Digression below.

Digression. Do you really want to read this? Well, carry on if you must. We are going to show that the last line in (7.46) is maxed at $n_i = n_j = 1/2$. There are three terms in this last line, each raised to different powers. It is very easy to verify that the first and second of these, $P(n_i)n_i P(n_j)n_j$ and $n_i n_j$, are each maximized when $n_i = n_j = 1/2$. The third term (which is raised to the power $\alpha - 2$) takes a bit more work. It will suffice to show that the expression

$$P(n)^{1/\alpha} n = \left[\lambda G n^\alpha + (1-\lambda) G n^{\alpha-1}\right]^{1/\alpha}$$

is concave in n. It is actually trivial to verify this either for $\lambda = 0$ or $\lambda = 1$ but the intermediate steps require some calculus. Differentiate the expression above with respect to n to obtain the first derivative proportional to

$$\begin{aligned} \Delta(n) &= \frac{1}{\alpha}\left[\lambda n^\alpha + (1-\lambda)n^{\alpha-1}\right]^{(1-\alpha)/\alpha}\left[\lambda\alpha n^{\alpha-1} + (1-\lambda)(\alpha-1)n^{\alpha-2}\right] \\ &= \lambda\left[n^{-\alpha}\left(\lambda n^\alpha + (1-\lambda)n^{\alpha-1}\right)\right]^{(1-\alpha)/\alpha} + \frac{(1-\lambda)(\alpha-1)}{\alpha}\left[n^{\frac{(\alpha_2)\alpha}{1-\alpha}}\left(\lambda n^\alpha + (1-\lambda)n^{\alpha-1}\right)\right]^{(1-\alpha)/\alpha} \\ &= \lambda\left[\lambda + (1-\lambda)n^{-1}\right]^{(1-\alpha)/\alpha} + \frac{(1-\lambda)(\alpha-1)}{\alpha}\left[\lambda n^{\alpha/(\alpha-1)} + (1-\lambda)n^{1/(\alpha-1)}\right]^{(1-\alpha)/\alpha} \\ &= \left[\lambda + \frac{(1-\lambda)(\alpha-1)}{\alpha}n^{-1}\right]\left[\lambda + (1-\lambda)n^{-1}\right]^{(1-\alpha)/\alpha}. \end{aligned}$$

---

[5] I don't believe this result holds when $\alpha < 2$.

To complete the argument we must show that the very last line above is negatively related to n. To do this it suffices to show that the derivative of

$$\left[\lambda + \frac{(1-\lambda)(\alpha-1)}{\alpha}x\right][\lambda + (1-\lambda)x]^{(1-\alpha)/\alpha}$$

with respect to x is nonnegative. That derivative is given by

$$\begin{aligned} & [\lambda + (1-\lambda)x]^{(1-\alpha)/\alpha}\frac{(1-\lambda)(\alpha-1)}{\alpha} + \left[\lambda + \frac{(1-\lambda)(\alpha-1)}{\alpha}x\right]\frac{1-\alpha}{\alpha}[\lambda + (1-\lambda)x]^{(1-2\alpha)/\alpha}(1-\lambda) \\ = \; & \frac{(1-\lambda)(\alpha-1)}{\alpha}[\lambda + (1-\lambda)x]^{(1-\alpha)/\alpha}\left\{1 - \frac{\lambda + \frac{(1-\lambda)(\alpha-1)}{\alpha}x}{\lambda + (1-\lambda)x}\right\} \\ \geq \; & 0. \end{aligned}$$

We've therefore shown that the bipolarization result is quite robust when it comes to explaining the intensity of conflict, *conditional on it happening in the first place*. Whether the good to be fought over is public or private really does not matter too much.

7.23.2 Conflict Initiation. The italicized phrase in the last paragraph is important, however. Highly polarized situations may not exhibit any conflict to begin with, because everyone knows that the conflict is going to be very costly if it happens. Therefore, a little bit more attention has to be paid to theories that attempt to relate conflict to polarization.

Let us explore these matters a bit further as they have interesting implications for the identity of the initiating party. Suppose again that there are two groups, and that group i receives a peacetime payoff in which it obtains a share s_i of the good. Then it is easy to see that its peacetime payoff is simply given by $s_i P(n_i)$, where we define $P(n) = \left[\lambda G + (1-\lambda)\frac{G}{n}\right]$ as in the previous section. It follows that group i will want to engage in conflict provided that

$$P(n_i)\frac{n_i r_i}{R} - c(r_i) > s_i P(n_i), \tag{7.47}$$

where r_i and R are to be interpreted as equilibrium values in the ensuing conflict.

Once again adopt the isoelastic specification and recall the first-order condition (7.40) with respect to r_i. But write it in a form that we used for our conflict and distribution study:

$$P(n_i)p_i p_j = r_i^{\alpha}, \tag{7.48}$$

where p_i (and likewise p_j) is just the win probability or conflict share $n_i r_i / R$. We may use this to obtain a quasi-closed-form for the net payoff from conflict:

$$P(n)p_i - c(r_i) = P(n_i)\left[p_i - \frac{1}{\alpha}p_i(1-p_i)\right]. \tag{7.49}$$

Now we proceed to an analysis of conflict initiation. Consider the equivalent formulation of the first-order condition for group i (which essentially reproduces (7.40)):

$$P(n_i)n_i n_j = R^2 \frac{r_i^{\alpha-1}}{r_j}. \tag{7.50}$$

If the good in question is purely public, then we may simply normalize $P(n) = 1$, so that

$$n_i n_j = R^2 \frac{r_i^{\alpha-1}}{r_j}.$$

Of course, exactly the same condition applies to r_j, and it therefore follows that $r_i = r_j = R$, and so $p_i = n_i$. Using this information in (7.49) above and invoking (7.47), we obtain the following condition for conflict initiation:

$$n_i - \frac{1}{\alpha} n_i(1 - n_i) > s_i.$$

It is easy to see that the left-hand side of this expression is strictly increasing in n_i,[6], and that it must therefore cross the right-hand side from below (provided s_i lies strictly between 0 and 1). We conclude then that *large* groups initiate conflict when the battle is over public goods.

Place s_i at 1/2 for symmetry. Observe that the condition *never* holds when $n_i = 1/2$. This is the most polarized case but it does not exhibit conflict. On the other hand, *once* conflict breaks out it is monotonically related to polarization, as discussed in the previous section. So equilibrium conflict jumps up at the threshold and then declines again monotonically to zero as $n_i \to 1$.

An interesting point to note is that $n_i > \sqrt{s}$ is always sufficient for conflict (simply make the LHS as small as possible in α by putting $\alpha = 1$). This is a more general point that is independent of the cost function.[7] In any case the important point is that large groups initiate in the public goods case.

This observation is, however, reversed when the good at stake is private. To see this, recall the first-order condition for group i, (7.50), and normalize $P(n) = 1/n$ to see that

$$n_j = R^2 \frac{r_i^{\alpha-1}}{r_j}.$$

Dividing by the corresponding first-order condition for group j, we may conclude that

$$\frac{r_i}{r_j} = \left(\frac{n_j}{n_i}\right)^{1/\alpha},$$

which captures part of the Olson intuition that small groups lobby more per-capita.[8] Therefore the probability that i wins the conflict is given by

$$p_i \equiv \frac{n_i r_i}{n_i r_i + n_j r_j} = \frac{n_i^{(\alpha-1)/\alpha}}{n_i^{(\alpha-1)/\alpha} + n_j^{(\alpha-1)/\alpha}}, \tag{7.51}$$

[6]The reason is that the derivative of $n(1-n)$ never exceeds 1.

[7]To prove this, note that the more general first-order condition is just $n_i n_j = R^2 c'(r_i)/r_j$ for both i and j, so it is still true that $r_i = r_j = R$. It follows that $c'(R)R = n_i n_j$. Now recall the conflict initiation condition (7.47), and note that by convexity $c'(R)R \le c(R)$. Therefore a sufficient condition for (7.47) to hold is that $n_i - n_i n_j > s_i$, which proves the claim.

[8]This is, however, far from a full examination of the Olson paradox as we also have to allow for free-riding within the group. For a model that does this, see Esteban and Ray (APSR 2001).

so in particular large groups still have a higher win probability.

Finally, invoke (7.49) and (7.47) to write down the condition for conflict initiation in the private goods case, which is

$$\left[p_i - \frac{1}{\alpha}p_i(1-p_i)\right] > s_i,$$

where p_i is now given by (7.51).

In contrast to $s_i = 1/2$, the symmetric case for peacetime is now $s_i = n_i$. Let us use this benchmark to understand conflict initiation. Rewriting the condition above with $s_i = n_i$, we obtain

$$\frac{1}{\alpha}p_i^2 + \frac{\alpha-1}{\alpha}p_i > n_i. \tag{7.52}$$

Combine this inequality and (7.51) to now establish the following observation. There is a unique population threshold *below* which a group will initiate conflict.

To prove this claim, recall (7.51) and take a closer look at p (dropping subscripts):

$$p(n) = \frac{n^k}{n^k + (1-n)^k},$$

where $k \equiv (\alpha - 1)/\alpha$. The function p has an interesting "reverse-logistic" shape. It starts above the 45^0 line and at the point $n = 1/2$ crosses it and dips below. The derivatives at the two ends are infinite. To check these claims, note that

$$\frac{n^k}{n^k + (1-n)^k} \geq n$$

if and only if $n \leq 1/2$ (simply cross-multiply and verify this), and that

$$p'(n) = \frac{kn^{k-1}(1-n)^{k-1}}{[n^k + (1-n)^k]^2},$$

which is infinite both at $n = 0$ and $n = 1$. Now recall (7.52) and write it as

$$(1-k)p(n)^2 + kp(n) > n.$$

By the arguments just made on derivatives, the LHS starts out higher than the RHS and ends up lower than the RHS. This means that conflict is preferable for small minorities and not so for large majorities, in contrast already to the results for public goods.

Indeed, we can strengthen that last argument to show that

$$(1-k)p(n)^2 + kp(n) < n$$

for *any* $n \geq 1/2$, that conflict becomes strictly bad for any weak (nonunanimous) majority. Suppose this is false for some $1 > n \geq 1/2$. By the properties of p already established, we know that $n \geq 1/2$ implies $n \geq p(n)$, so that

$$(1-k)n^2 + kn \geq n,$$

but this can never happen when $n < 1$, a contradiction. So conflict can never be preferable for a weak majority of the population.

It remains finally to show that there is a unique intersection (crossing from above to below) in the interior. Let $n^* \in (0, 1)$ be an interior solution; then

$$(1 - k)p(n^*)^2 + kp(n^*) = n^*.$$

Differentiate to show that the crossing is "from above".

CHAPTER 8

Inequality and Incentives

In this chapter we study collective action problems (such as team production or the voluntary provision of public goods) and identify three channels through which inequality affects incentives and therefore overall efficiency.

The general structure of a team production problem is as follows. There is a group of n agents, each of whom contributes resources (money, effort) r_i. The joint output is given by

$$Y = F(\mathbf{r}) \tag{8.1}$$

where $\mathbf{r} = (r_1, \ldots, r_n)$ is a (nonnegative) vector.

Depending on the application, Y could be a public good or a private good.

I now introduce the three channels through which inequality might function:

A. *Wealth.* Suppose that the team good is public, and that person i has wealth endowment w_i. The person i's payoff may be written as

$$Y + u(y_i - r_i).$$

The question then is: how does inequality in wealth affect incentives?

B. *Access.* Suppose that all agents are identical in endowment, and the good is private. But it is unequally distributed. Then different individuals have different shares of Y.

In this example, our proxy for inequality is a vector of *shares* $\ell \equiv (\lambda_1, \lambda_2, \lambda_n)$, which sum to unity. For instance, this may be a problem of joint maintenance of an irrigation network, or of cooperative farming, in which case the shares would proxy the amount of land holdings of the agent.

How might inequality in access affect incentives?

C. *Inequality Tolerance.* Attitudes to inequality might affect incentives in situations where commitment is not possible. To illustrate this problem suppose that there are two regions A and B which receive funding from an equality-minded central government. Suppose that region i invests r_i and produces $y_i = f(r_i)$ (think of y_i as regional tax revenues and r_i as resources put into the generation of local revenues).

Suppose that the center must choose transfers t_i to each of the regions, satisfying some budget constraint $t_A + t_B = T$. Suppose that it has welfare function

$$W(y_A + t_A) + W(y_B + t_B).$$

If the center can commit to the transfers then in a symmetric optimum it would set $t_i = T/2$ and then each region i would choose r_i to

$$\max t_i + f(r_i) - r_i;$$

this would implement the first-best without a problem. However, if the center does not commit to the transfers then the states move first and choose y_i. This creates a dilemma if the center *then* picks t_A and t_B to

$$\max W(y_A + t_A) + W(y_B + t_B),$$

with y_A and y_B given. Ignoring corner solutions, we see that this would entail

$$y_A + t_A = y_B + t_B.$$

This creates a disincentive problem for the regions. Each region i will anticipate this compensatory behavior by the center and so choose t_i to maximize

$$y_i + r_i(y_A, y_B).$$

It is easy to see that this leads to the first-order condition

$$\frac{1}{2} f'(r_i) = 1,$$

which implies substantial underinvestment relative to the first best.

Clearly, much of this depends on how the center reacts to the choices of y_A and y_B, which in turn depends on the degree of egalitarianism felt by the center, which is embodied in W. It also depends on whether the center plans to "compensate" the regions for their investment of r. This leads to the general class of questions:

When a principal cannot commit to a reward function, how does her tolerance for inter-agent inequality affect the incentives of participating agents?

8.1 Inequality in Endowments

I build on a classic paper on the voluntary provision of public goods by Bergstrom, Blume and Varian (1981). There are n agents, each of whom contributes resources r_i to a public good. Output is given by

$$G = f(r), \tag{8.2}$$

where $r = \sum_i r_i$, and f is smooth, increasing and concave. (One can take a more general specification but this will do to illustrate our main points.)

The individual utility function is

$$u(c_i) + G,$$

where $c_i = w_i - r_i$ for all i, and u is smooth, increasing and strictly concave.

An equilibrium is a vector $\mathbf{r}^*$ such that every individual i choose r_i^* optimally, given that resources contributed by the others is r_j^* for $j \neq i$.

Proposition 8.1. *There exists a unique equilibrium.*

Proof. Suppose, on the contrary, that two distinct vectors $\mathbf{r}$ and $\mathbf{r}'$ are both equilibria. Without loss of generality $r' \geq r$. Pick i such that $r_i' > r_i$. Because $r_i' > 0$, we must have

$$u'(w_i - r_i') = f'(r'), \tag{8.3}$$

while

$$u'(w_i - r_i) \geq f'(r). \tag{8.4}$$

Because $r_i' > r_i$, we have that $u'(w_i - r_i') > u'(w_i - r_i)$. Using this in (8.3) and (8.4), we must conclude that

$$f'(r') = u'(w_i - r_i') > u'(w_i - r_i) \geq f'(r),$$

so that by the concavity of f, $r' < r$. This is a contradiction. ■

Our next proposition describes an interesting fact about the Bergstrom-Blume-Varian model, which is that — to the extent that contributions are interior — it is a distribution-neutral theory.

Proposition 8.2. *Suppose that under a wealth distribution* $\mathbf{w}$ *every individual makes strictly positive contributions. Consider another wealth distribution* $\mathbf{w}'$, *with the same aggregate wealth, such that no individual's wealth is reduced by more than her contribution under* $\mathbf{w}$. *Then* $r = r'$: *overall contributions are unaffected.*

Proof. Define $\epsilon_i \equiv w_i' - w_i$ for every i, and then define $r_i' \equiv r_i + \epsilon_i$. Because $\sum_i \epsilon_i = 0$, we have $r' = r$. Also, $r_i' \geq 0$ for all i because $r_i \geq -\epsilon_i$, by assumption. Now recall that

$$u'(w_i - r_i) = f'(r)$$

for all i, because $r_i > 0$, so we must conclude that

$$u'(w_i' - r_i') = f'(r')$$

as well. This means that $\mathbf{r}'$ is an equilibrium under the new wealth distribution, and by Proposition 8.1, it is the only one. ■

Proposition 8.3. *In any equilibrium, everyone who makes positive contributions has their consumptions and utilities fully equalized, regardless of their wealth.*

Proof. For r_i and r_j both positive, we must have

$$u'(w_i - r_i) = u'(w_j - r_j) = f'(r),$$

so that $c_i = c_j$ and $u_i = u_j$. ■

This is a strange result, and it cannot be broken even with very general utility functions such as $u(c, G)$. It does break down if f is asymmetric across agents or if the utility function is heterogeneous across agents.

This motivates a careful study of corner solutions in the voluntary contributions model. These solutions are not at all exceptional: it is quite reasonable to suppose that some poor agents do not make contributions, while the richer agents do. So let $\mathbf{w}$ be some wealth distribution. Let I be the set of all agents who make positive contributions: $I = \{i | r_i > 0\}$. Observe that small redistributions of wealth within I have no effect on aggregate outcomes, just as in the neutrality proposition studied earlier.

What about a redistribution from $j \notin I$ to some $i \in I$?

Proposition 8.4. *Greater inequality in wealth either leaves aggregate contributions unchanged, or pushes them towards higher levels.*

Proof. To prove this, it suffices to consider a small transfer ϵ from $j \notin I$ to $i \in I$, such that the set I remains unchanged. (This follows from Proposition 8.2 on distribution-neutrality.) I claim that in the new equilibrium, $r' > r$. Suppose not; then $r' \leq r$. Now

$$u'(w_k - r'_k) = f'(r')$$

for $k \in I$, $k \neq i$. Because $f'(r') \geq f'(r)$, we must conclude that $r'_k \geq r_k$ for all such k. For i, we have that

$$u'(w_i - r'_i) = u(w_i + \epsilon - r'_i) = f'(r')$$

so by the same argument, $r'_i - \epsilon \geq r_i$. Putting all this together, we must conclude that $r' = \sum_k r'_k \geq \sum_k r_k + \epsilon > r$, a contradiction. ■

Therefore in the Bergstrom-Blume-Varian model, we must conclude that greater wealth inequality must lead to greater provision of the public good. Roughly speaking, wealth inequality raises the contribution costs of noncontributors and lowers the cost to contributors. The former has no effect: noncontributors are noncontributors and it does not matter whether their costs go up or not. The latter increases contributions; hence the result.

It should be pointed out that the effect on total (utilitarian) surplus is uncertain. After all, utilities are concave and disequalizing them through regressive transfers of wealth has inherently negative consequences. However, sometimes one can argue that overall welfare goes up. A disequalization of wealth has three effects: (i) it increases rich utilities (positive); (b) it lowers poor utilities (negative) and (c) it increases public goods supply (positive). To compute net effects think of a two-person example in which 1 is a contributor and 2 is not. If we are very close to the contribution threshold for 2, we have that

$$w_1 - r_1 \simeq w_2$$

by the equalization proposition above. Now a disequalization of wealth will have no first-order effect on the sum of utilities, while it will have a first-order positive effect on public good provision.

In the next section, we will see an important sense in which Proposition 8.4 may not be robust.

8.2 Inequality in Access

In this section, we assume that the output is a private good, though with a little bit of clever renormalization most of the analysis goes through with mixed public-private goods.

The objective is to analyze *access inequalities*, so we will have the produced output divided in some given way among the contributing agents. The question is: how does the inequality in this division affect efficiency?

Our focus will be on an aspect of production that we did not consider in the last section: we will allow for the inputs of different agents to be complements in producing the joint output. The maintenance of an irrigation network is a good generic example: the maintenance of different stretches of an irrigation channel is complementary, and the idea of access inequalities also makes sense (think of irrigation being important in proportion to the amount of land being farmed).

Other than complementarities, we keep everything very simple. Individual utility functions are identical, and taken to be linear both in consumption and in contributions (which, in keeping with the maintenance example) we will now call effort e. Thus each agent i seeks to maximize $c_i - e_i$, and chooses e_i to

$$\max_{e_i} \lambda_i F(e_i, \mathbf{e}_{-i}) - e_i \tag{8.5}$$

where the notation $\mathbf{e}_{-i}$ stands for the vector $\mathbf{e}$ with the ith component removed.

An *equilibrium* is an effort vector $\mathbf{e}^*$ with the property that for every i, e_i^* solves (8.5), given $\mathbf{e}^*_{-i}$.

An effort vector $\hat{\mathbf{e}}$ is *efficient* if it maximizes the expression

$$\hat{S} \equiv F(\mathbf{e}) - \sum_{i=1}^{n} e_i$$

over all possible effort vectors. Thus $\hat{S}$ is the maximal surplus that can be generated in the economy. Assume that appropriate end-point conditions hold so that the maximization problem above is well-defined.

Define the surplus S^* associated with any equilibrium $\mathbf{e}^*$ by the expression $F(\mathbf{e}^*) - \sum_{i=1}^{n} e_i^*$. Obviously, $S^* \leq \hat{S}$ in any equilibrium. Take $\hat{S} - S^*$ to be our measure of the inefficiency of an equilibrium.

This definition naturally induces a class of inefficiency measures for a given level of "access inequality", as proxied by ℓ. We adopt the following definition: the inefficiency $I(\ell)$ of access inequality ℓ is given by

$$I(\ell) \equiv \min\{\hat{S} - S^* | \quad S^* \text{ is the surplus associated with some equilibrium } \mathbf{e}^* \text{ under } \ell\}. \tag{8.6}$$

The question that we seek to explore is: which levels of access inequality minimize inefficiency? Writers such as Mancur Olson have stressed that unequal sharing rules are good for efficiency, because they minimize the free-rider problem. We may summarize this intuition in the following

Proposition 8.5. *Suppose that F is an increasing, concave function of the sum of efforts: $F(\mathbf{e}) = f(\sum_{i=1}^n e_i)$ for some increasing differentiable concave f satisfying the Inada endpoint conditions. Let $\ell = (\lambda_1, \ldots, \lambda_n)$ and $\ell' = (\lambda'_1, \ldots, \lambda'_n)$ be two different access inequalities, with the property that $\max_i \lambda_i > \max_i \lambda'_i$. Then $I(\ell) < I(\ell')$. Moreover, $I(\ell) \to 0$ as ℓ converges to any of the unit vectors.*

Proof. Let $M(\ell) \equiv \max_i \lambda_i$. Then in any equilibrium $\mathbf{e}^*$ under ℓ, e_i^* is positive only if $\lambda_i = M(\ell)$, and consequently, if $E^* \equiv \sum_{i=1}^n e_i^*$, E^* must maximize $M(\ell)f(E) - E$ with respect to E. Thus $I(\ell)$ is simply $\hat{S} - M(\ell)f(E^*) + E^*$. It is easy to check that $I(\ell)$ is monotonically decreasing in $M(\ell)$ and indeed, that $I(\ell) \to 0$ along any sequence such that $M(\ell) \to 1$. ■

Thus under some conditions, inequality of sharing is conducive to efficiency. These conditions require that output be a function of the *sum* of efforts. This assumption may be particularly cogent in the case of lobbying, where the effectiveness of lobbying may be a function of the sum of monetary contributions.

However, in many production activities, the efforts of different individuals may be complements in production. The following equally simple proposition considers the extreme Leontief case, to provide a stark contrast to Proposition 8.5.

Proposition 8.6. *Suppose that F is an increasing concave function of the scale of activity, where scale is determined by equi-proportional contribution of efforts: $F(\mathbf{e}) = f(\min_i e_i)$ for some increasing differentiable concave f satisfying the Inada endpoint conditions. Let $\ell = (\lambda_1, \ldots, \lambda_n)$ and $\ell' = (\lambda'_1, \ldots, \lambda'_n)$ be two different access inequalities, with the property that $\min_i \lambda_i > \min_i \lambda'_i$. Then $I(\ell) < I(\ell')$.*

Proof. Note that in this case we have a continuum of equilibria for each possible level of access inequality ℓ. These are characterized as follows.

Define $m(\ell) \equiv \min_i \lambda_i$. Then for each ℓ, $\mathbf{e}^*$ is an equilibrium vector of efforts if and only if $e_1^* = e_2^* = \ldots = e_n^* = e$ (say), and $m(\ell)f'(e) \geq 1$. Thus

$$I(\ell) = \hat{S} - m(\ell)f(e) + ne.$$

It is easy to check that $I(\ell)$ is a decreasing function of $m(\ell)$. ■

Thus in this case, we have exactly the opposite result: *equality* is conducive to efficiency, though unlike the case of proposition 8.5, we never obtain full efficiency (it is easy to check that $I(\ell)$ is bounded away from zero). We may therefore conjecture that the least inefficient access inequality varies in some systematic way with the degree of complementarity in production.

8.2.1 More on Access Inequalities. The analysis above suggests that the degree of substitution will play a role in determining whether or not inequality is good for efficiency. To capture this idea, parameterize production so it runs the gamut from perfect to zero substitutability of effort is to use a CES structure for the vector of efforts. Thus we may take F to have the form

$$F(\mathbf{e}) = \left[\sum_{i=1}^{n} e_i^{1-\sigma}\right]^{\frac{\alpha}{1-\sigma}}, \tag{8.7}$$

where $\sigma \geq 0$ measures the degree of effort substitution, and $\alpha \in (0,1)$ is a scale parameter.

Direct computation reveals that the maximal surplus $\hat{S}$ in this model is given by

$$\hat{S} = \alpha^{\frac{\alpha}{1-\alpha}}(1-\alpha)n^{\frac{\alpha\sigma}{(1-\alpha)(1-\sigma)}}. \tag{8.8}$$

To compute the Nash equilibrium, note first that for $\sigma > 0$, best responses for each i are strictly positive provided that all other efforts are strictly positive. So let us write down interior first-order conditions to the maximization problem (8.5), for each i, which we may rephrase here as follows:

$$\max_{e_i} \lambda_i \left[\sum_{j=1}^{n} e_j^{1-\sigma}\right]^{\frac{\alpha}{1-\sigma}} - e_i \tag{8.9}$$

Simple computation reveals the first-order conditions to be

$$e_i^{\sigma} = \alpha\lambda_i \left[\sum_{j=1}^{n} e_j^{1-\sigma}\right]^{\frac{\alpha+\sigma-1}{1-\sigma}}, \tag{8.10}$$

and summing over all i in (8.10) and simplifying, we see that

$$\sum_{j=1}^{n} e_j^{1-\sigma} = \alpha^{\frac{1-\sigma}{1-\alpha}} \left[\sum_{j=1}^{n} \lambda_j^{\frac{1-\sigma}{\sigma}}\right]^{\frac{\sigma}{1-\alpha}} \tag{8.11}$$

We may substitute (8.11) in (8.10) to conclude that

$$e_i = \lambda^{\frac{1}{\sigma}} \alpha^{\frac{1}{1-\alpha}} \left[\sum_{j=1}^{n} \lambda_j^{\frac{1-\sigma}{\sigma}}\right]^{\frac{\alpha+\sigma-1}{(1-\alpha)(1-\sigma)}}. \tag{8.12}$$

Equipped with these equations, we may now obtain a closed form for the surplus $S^*(\ell)$ in any interior Nash equilibrium under ℓ:

$$S^*(\ell) = \alpha^{\frac{\alpha}{1-\alpha}} \left[\sum_{j=1}^{n} \lambda_j^{\frac{1-\sigma}{\sigma}}\right]^{\frac{\alpha+\sigma-1}{(1-\alpha)(1-\sigma)}} \left[\sum_{j=1}^{n} \lambda_j^{\frac{1-\sigma}{\sigma}} - \alpha \sum_{j=1}^{n} \lambda_j^{\frac{1}{\sigma}}\right]. \tag{8.13}$$

The problem that we are interested in thus reduces to the seemingly innocuous maximization exercise: maximize $S^*(\ell)$, as given by (8.13), with respect to ℓ.

It is possible to use the equations (8.8) and (8.13) to run some simple consistency checks. Propositions 8.5 and 8.6 may be obtained by the taking of appropriate limits in the above maximization exercise. The more general issue is to calculate the values of ℓ to intermediate problems.

We do not have a complete solution to this problem. But here are some observations.

Proposition 8.7. *If $\sigma \geq 1/2$, then perfect equality maximizes constrained surplus.*

Proof. Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_n)$ be a vector of nonnegative shares summing to unity. We already have a closed form for the surplus $S^*(\boldsymbol{\lambda})$ in any interior Nash equilibrium, which I write here slightly differently:

$$S^*(\boldsymbol{\lambda}) = \alpha^{\frac{\alpha}{1-\alpha}} \frac{\sum_{j=1}^n \lambda_j^\theta (1 - \alpha\lambda_j)}{\left(\sum_{j=1}^n \lambda_j^\theta\right)^\gamma}, \tag{8.14}$$

where $\theta \equiv \frac{1-\sigma}{\sigma}$, and $\gamma \equiv \frac{1-\alpha-\sigma}{(1-\alpha)(1-\sigma)}$. Notice that γ is only well-defined if $\sigma \neq 1$, which we shall assume in what follows. [The logarithmic case $\sigma = 1$ will be discussed at the end of the proof.]

We will use this expression to establish the proposition. To this end, let us compute the derivative of $S^*(\boldsymbol{\lambda})$ with respect to any one of the shares, say λ_i. This derivative is given by

$$\frac{\partial \alpha^{\frac{-\alpha}{1-\alpha}} S^*(\boldsymbol{\lambda})}{\partial \lambda_i} = \frac{\Delta(\lambda_i)}{\left(\sum_{j=1}^n \lambda_j^\theta\right)^\gamma},$$

where (after some computation) it can be seen that

$$\Delta(\lambda_i) = \theta\lambda_i^{\theta-1}\left\{\frac{\sum_{j=1}^n \lambda_j^\theta [1 - \gamma(1 - \alpha\lambda_j)]}{\sum_{j=1}^n \lambda_j^\theta}\right\} - \alpha(\theta + 1)\lambda_i^\theta. \tag{8.15}$$

Note: we will be done if we can show that whenever $\lambda_i < \lambda_k$, then $\Delta(\lambda_i) > \Delta(\lambda_k)$. For this would mean that a small transfer from the higher share to the lower share would raise total surplus. Because the terms within curly brackets in (8.15) are unchanged in this comparison, it will suffice to show that the function

$$\Delta(x) = \theta x^{\theta-1}\left\{\frac{\sum_{j=1}^n \lambda_j^\theta [1 - \gamma(1 - \alpha\lambda_j)]}{\sum_{j=1}^n \lambda_j^\theta}\right\} - \alpha(\theta + 1)x^\theta \tag{8.16}$$

is strictly decreasing in x as x varies over $[0, 1]$. To this end, differentiate $\Delta(x)$ to see that

$$\Delta'(x) = \theta(\theta - 1)x^{\theta-2}\left\{\frac{\sum_{j=1}^n \lambda_j^\theta [1 - \gamma(1 - \alpha\lambda_j)]}{\sum_{j=1}^n \lambda_j^\theta}\right\} - \alpha\theta(\theta + 1)x^{\theta-1}. \tag{8.17}$$

It will suffice to show that this expression is strictly negative for $x \in (0, 1)$.

We distinguish between two cases.

Case 1. $\sigma \in (1/2, 1)$. In this case it is easy to see that $\theta \in (0, 1)$ and that $\gamma < 1$. Using this information in (8.17), it follows right away that $\Delta'(x) < 0$.

Case 2. $\sigma > 1$. In this case it is easy to see that $\theta < 0$ and that $\gamma > 1$.[1] Using part of this information (the fact that $\theta < 0$), we see from (8.17) that $\Delta'(x) < 0$ if

$$(\theta - 1)\{\frac{\sum_{j=1}^{n} \lambda_j^{\theta}[1 - \gamma(1 - \alpha\lambda_j)]}{\sum_{j=1}^{n} \lambda_j^{\theta}}\} > \alpha(\theta + 1)x,$$

or equivalently, if we can establish the inequality

$$\{\frac{\sum_{j=1}^{n} \lambda_j^{\theta}[1 - \gamma(1 - \alpha\lambda_j)]}{\sum_{j=1}^{n} \lambda_j^{\theta}}\} < \frac{\theta + 1}{\theta - 1}\alpha x = \frac{1}{1 - 2\sigma}\alpha x. \tag{8.18}$$

However, notice that

$$\frac{\sum_{j=1}^{n} \lambda_j^{\theta}[1 - \gamma(1 - \alpha\lambda_j)]}{\sum_{j=1}^{n} \lambda_j^{\theta}} \leq 1 - \gamma(1 - \alpha) = -\frac{\alpha}{\sigma - 1},$$

where use is made of the fact that $\lambda_j \leq 1$ for each j. With (8.18), this means that it will suffice to establish the inequality

$$-\frac{\alpha}{\sigma - 1} < \frac{1}{1 - 2\sigma}\alpha x,$$

or equivalently,

$$\frac{\alpha}{\sigma - 1} > \frac{1}{2\sigma - 1}\alpha x,$$

But this inequality follows from direct inspection, and we are done.

Notice that the logarithmic case $\sigma = 1$ remains uncovered. I believe this should be established by separate computation. ∎

Proposition 8.8. *If $\sigma < 1/2$, there exist population sizes for which equal division cannot be optimal.*

Proof. As a construction for the proof, set up the equal minority problem with m people receiving equal shares $1/m$: call this vector $\boldsymbol{\lambda}_m$. Then it is easy to see that

$$S^*(\boldsymbol{\lambda}_m) \simeq m^b(m - \alpha), \tag{8.19}$$

where $b \equiv \frac{\alpha\sigma+\sigma-1}{(1-\alpha)(1-\sigma)}$.

Define m^* by the smallest integer such that

$$m^* \geq \frac{1 - \alpha\sigma - \sigma}{1 - 2\sigma} \tag{8.20}$$

I will now show that $S^*(\boldsymbol{\lambda}_m) > S^*(\boldsymbol{\lambda}_n)$ for any $n > m^*$, which proves that perfect equality is impossible when n exceeds m^*. To this end, we pretend that m is a continuous variable in (8.19). It will suffice to show that the derivative of $S(m)$ with respect to m is negative for all $m > m^*$. Differentiating (8.19) with respect to m, we need to show that

$$(b + 1)m - \alpha b < 0$$

[1] To check the observation for γ, be a little careful. First note that when $\sigma > 1$, γ can be written as $\frac{\sigma+\alpha-1}{(\sigma-1)(1-\alpha)}$, and now examine this.

for all $m > m^*$. Using the definition of b and m^*, this is a matter of simple algebra. ■

The simulations (see below) show a strange area where two players receive different shares. It looks subtle. The following proposition takes us part of the way towards resolving these strange issues.

Proposition 8.9. *There can be no more than two distinct positive values of the share in any constrained optimum. Moreover, there can be no more than one person endowed with the lower of the two positive values of the share.*

Proof. To prove this, recall the expression for $\Delta(\lambda_i)$ in (8.15), and note the following:

Lemma 8.1. *Let $\boldsymbol{\lambda}$ be a share vector. If $\Delta(\lambda_i)$ and $\Delta(\lambda_k)$ are different for two positive values λ_i and λ_k in the share vector, then that share vector cannot be a constrained optimum.*

The proof of this lemma is trivial. Simply transfer shares from the one with lower Δ-value to the one with the higher Δ-value. Since the shares in question are both positive, there is no constraint to doing this in any "direction" we please. ■

Given Lemma 8.1, our question boils down to this: for how many distinct and positive values of x can $\Delta(x)$ have the same value? Recall that $\Delta(x)$ is defined in (8.16);[2] it may be written as

$$\Delta(x) = \theta x^{\theta-1} - Cx^{\theta}, \tag{8.21}$$

where

$$C \equiv \frac{\alpha(\theta+1)\sum_{j=1}^{n}\lambda_j^{\theta}}{\sum_{j=1}^{n}\lambda_j^{\theta}[1-\gamma(1-\alpha\lambda_j)]} > 0.$$

Now we prove the following observation: *the function $\Delta(x)$, for $x \geq 0$ is "single-peaked" (though not necessarily concave): first rising, then falling.*

To show this be a little careful, because the function is not necessarily strictly concave (try $\theta > 2$). But it is easy to do (details omitted).

Now we are done with the first part of the proposition. Such a function can exhibit the same value for at most two distinct points in the domain.

To complete the proof, let a denote the smaller of the two positive values of the share. We will show that if two (or more) persons are given a, we can improve the surplus by transferring some share from one of them to the other. To this end, think of the share of these two individuals as variables x and y. "Initially", $x = y = a$. Holding all other shares constant, we may think of the aggregate surplus simply as a function $S(x, y)$. For some $\epsilon > 0$, we know from the mean value theorem for multivariate functions (see, e.g., Hoffman (1975, Section 8.4, Theorem 6)) that

$$S(a+\epsilon, a-\epsilon) - S(a, a) = \epsilon[S_1(\hat{x}, \hat{y}) - S_2(\hat{x}, \hat{y})] \tag{8.22}$$

[2]It is true that the shares also appear within the curly brackets. But this does not matter as we are simply comparing for different shares "over the cross-section" and the share vector does not change.

where superscripts denote the appropriate partial derivatives, and $\hat{x}$ can be chosen to be *strictly* larger than $\hat{y}$.[3] Now we know that

$$S_1(\hat{x}, \hat{y}) = \theta \hat{x}^{\theta-1} - C(\hat{x}, \hat{y})\hat{x}^{\theta}, \tag{8.23}$$

and

$$S_2(\hat{x}, \hat{y}) = \theta \hat{y}^{\theta-1} - C(\hat{x}, \hat{y})\hat{y}^{\theta}, \tag{8.24}$$

where $C(\hat{x}, \hat{y})$ is exactly the same C as before (see definition just following (8.21)), but with two of the arguments made explicit to remind the reader that $\hat{x}$ and $\hat{y}$ enter there as well. Recall now that

$$\theta z^{\theta-1} - C(a, a)z^{\theta}$$

is strictly increasing in z around $z = a$. It follows that the same is true of the slightly perturbed function

$$\theta z^{\theta-1} - C(\hat{x}, \hat{y})z^{\theta}$$

around $z = a$. Using this information in (8.23) and (8.24), it is therefore easy to see that for $\epsilon > 0$ but small enough,

$$S_1(\hat{x}, \hat{y}) > S_2(\hat{x}, \hat{y}).$$

Applying this inequality to (8.22), we may conclude that $S(a + \epsilon, a - \epsilon) > S(a, a)$: surplus is increased by a small transfer between two persons with the same lower share, and the proof is complete. ■

Tentative Notes on Simulations:

Unfortunately, we could not obtain analytically the parameter values for which the unequal pair dominates equal minority with $m = 1, 2$. We therefore turned towards simulations, the results of which we now examine. The following observations were obtained through the simulations:

(i) The area in the parameter space (α, σ) such that the inefficiency-minimizing distribution of shares corresponds to the equal minority case is large. In other words, for most values of the parameters, sharing equally among a limited (smaller than n) number of agents is the least inefficient distribution of shares. When $\sigma \geq \frac{1}{2}$, perfect equality is the (second-best) optimal distribution of shares.

(ii) There is a non-trivial area in the parameter space (α, σ) such that inefficiency is minimized by distributing shares unequally, and between only two agents. However, the two shares tend to be more equal as the degree of complementarity between efforts increases.

(iii) For $\sigma < 0.2179$, the optimal distribution of shares corresponds to the one of perfect inequality, where one agent concentrates all the shares, for all values of α. Thus, if substitutability between efforts is high enough, perfect inequality *a la* Olson appears to minimize inefficiency. This is a subtle observation that requires greater analytical investigation.

[3] The reason for this is that it is possible to take $\hat{x} \in (a, a + \epsilon)$ and $\hat{y} \in (a - \epsilon, a)$. While the statement of the theorem in Hoffman (1975) does not make this clear, the proof — see last two lines — does.

8.3 Egalitarianism and Incentives

Note. What follows has been taken from Ray and Ueda (1996) and needs to be shortened for the purpose of these lecture notes.

A social planner's concern for egalitarianism might lead to a dilution of incentives, and therefore a loss in efficiency. In an important class of situations, the efficiency loss arises because the planner cannot credibly commit to a future course of action, such as the decision *not* to tax an individual or group making efficiency-enhancing investments. So these investments are not made, or more generally, undersupplied.

Consider, then, the following class of situations. A group of agents is collectively engaged in a joint production activity, where the output from production is to be distributed among the members of the group. The agents (represented by a social planner, perhaps) are interested in maximizing the value of a Bergson-Samuelson social welfare function defined on their own utilities. However, while this welfare function represents their *social values*, individual actions are taken on an entirely selfish basis. Suppose that to achieve the desired outcome, each agent must take an observable action, followed by some *collective* action — the "social planner's move". Suppose, moreover, that a collective action (contingent on individual decisions) *cannot* be credibly committed in advance.[4]

Specifically, define a *soft mechanism* to be one that specifies a second best division of the output (relative to the social welfare function) conditional on *every* possible input vector. We wish to compare the resulting equilibria of the "soft game" so induced, with the first best under the very same welfare function.

8.3.1 Model.

8.3.1.1 Technology and Individual Preferences Consider a group of n individuals ($n \geq 2$) producing a single output. Output is produced by the joint efforts of these individuals according to the production function

$$Y = F(\mathbf{e}) \tag{8.25}$$

where $\mathbf{e} = (e_1, \ldots, e_n)$ is a (nonnegative) vector of *efforts*. We assume

(A.1) *F is continuous and concave with $F(\mathbf{0}) = 0$ and differentiable whenever $F > 0$, with* $\frac{\partial F(\mathbf{e})}{\partial e_i} \equiv F_i(\mathbf{e}) > 0$.

Each individual has preferences over pairs of consumption (c) and leisure (l). We assume

[4] A central example is the problem of allocating funds from a Central Government to different State Governments. Typically, each state carries out a number of different expenditure programs, which are financed from central funds and state revenues. The Center would like to channel proportionately greater funds to the poorer states. At the same time, the Center would like to induce each state to carry out activities that will raise *per capita* income in that state. But the former goal places limits on the punishments that a Center can credibly impose on a state for not taking actions to further certain desired goals.

(A.2) *For each i, preferences are representable by an increasing utility function $u^i : R^2_+ \to R$ which is C^2 and strictly quasiconcave, with $(u^i_c(c,l), u^i_j(c,l)) \gg 0$ whenever $u(c,l) > u(0,0)$. Moreover, leisure is a normal good.*[5]

Each individual has a labour endowment L_i. Thus, for each i,

(8.26) $$L_i = l_i + e_i$$

Let $\mathcal{E}$ be the product of intervals $[0, L_i]$ over all i. Note that whenever $e \in \mathcal{E}$, we shall have l_i given by (2), for $i = 1, \ldots, n$. Keeping this in mind, define an *outcome* as a pair $(\mathbf{c}, \mathbf{e}) = ((c_1, \ldots, c_n); (e_1, \ldots, e_n))$, such that $c \geq 0$, $e \in \mathcal{E}$, and

(8.27) $$c_1 + c_2 + \ldots + c_n = F(\mathbf{e})$$

8.3.1.2 Social Preferences The social planner is presumed to possess Bergson-Samuelson preferences satisfying standard restrictions:

(A.3) (i) *Society's preferences are representable by a C^2 welfare indicator $W : \mathbb{R}^{2n}_+ \to \mathbb{R}$, defined on $2n$-tuples of consumption-leisure vectors. This welfare function is strictly quasiconcave. Moreover, for each $\mathbf{l} \in \mathcal{E}$ and each i, c_i is not an inferior good under the function $W(., \mathbf{l})$.*

(ii) *There exists a function V (which will be C^2 by part (i) and (A.2)) such that $W(c,l) = V(u^1(c_1,l_1), \ldots, u^n(c_n,l_n))$. Moreover, $\frac{\partial V(u)}{\partial u^i} \equiv V_i(u) > 0$ for all i.*

8.3.1.3 First Best The planner would like to maximize social welfare. This is achieved by an outcome $(\mathbf{c}, \mathbf{e})$ that maximizes $W(\mathbf{c}, \mathbf{L} - \mathbf{e})$, where $\mathbf{L} \equiv (L_1, L_2, \ldots, L_n)$ is the vector of labor endowments. From Assumptions 1 and 3 it follows that there is a *unique* outcome $(\mathbf{c}^*, \mathbf{e}^*)$ that solves this problem. Call this outcome the *first best*.

Some minor restrictions on the first-best are summarized in the following assumption.[6]

(FB) *There is at least one individual i such that $e^*_i > 0$ and $l^*_i \equiv L_i - e^*_i > 0$. Moreover, the first best output is larger than the optimal output that would be chosen by any one individual acting on his own: $F(\mathbf{e}^*) > F(0, \ldots, 0, \hat{e}_j, 0, \ldots, 0)$, where $\hat{e}_j$ solves*
$\max_{0 \leq e_j \leq L_j} u^j(F(0, \ldots, 0, e_j, 0, \ldots, 0), L_j - e_j)$. *Finally, under the first-best, each individual obtains at least as much utility as total inaction (which yields $u^i(0, L_i)$).*

8.3.1.4 The Soft Mechanism The idea of a no-commitment mechanism, or *soft mechanism*, is based on the postulate that the planner cannot avoid maximizing welfare *ex post*, even though this may be detrimental to the maximization of welfare *ex ante*. Formally, for each *given* vector $\mathbf{e} \in \mathcal{E}$, consider the problem:

(8.28) $$\max_{\mathbf{c}} W(\mathbf{c}, \mathbf{L} - \mathbf{e})$$

subject to the constraint that $(\mathbf{c}, \mathbf{e})$ must be an outcome.

[5] The assumption that leisure is a normal good implies the restriction $u^i_c u^i_{cl} - u^i_l u^i_{cc} > 0$.

[6] All the restrictions are not used in all the propositions, but the conditions are so minor that we do not feel it useful to separate this list into separate components.

By (A.1) and (A.3), there is a unique vector $c(\mathbf{e})$ that solves (4). We will refer to $(\mathbf{c}(\mathbf{e}), \mathbf{e})$ as an *ex post outcome*. This captures, in an extreme way, the inability to impose arbitrary punishments on deviants. The lack of commitment creates limitations on the incentive mechanisms that can be used to stimulate production.

The collection of ex post outcomes $(\mathbf{c}(\mathbf{e}), \mathbf{e})$, for $\mathbf{e} \in \mathcal{E}$, induces a *soft game* in the obvious way: Player i chooses $e_i \in [0, L_i]$. If the vector $\mathbf{e}$ is chosen, i consumes $c_i(\mathbf{e})$, thus generating the payoff $u^i(c_i(\mathbf{e}), L_i - e_i)$.

8.3.1.5 An Additional Restriction We now have sufficient terminology to introduce an additional joint assumption on technology and preferences, which will be used to derive one of the main results. To motivate this assumption, first consider a standard property of the first best outcome, which can be easily verified: If $e_i^* < L_i$,

$$u_c^i(c_i^*, l_i^*)F_i(e^*) \le u_l^i(c_i^*, l_i^*) \tag{8.29}$$

(indeed, with equality holding if $e_i^* > 0$).

The intuition is simple. If this inequality did not hold, i's effort could be raised a little with all the additional output being credited to him. His utility would be higher, with every other utility remaining constant. Social welfare goes up, a contradiction to the fact that we have a first best outcome to start with.

The additional assumption that we wish to make is related closely to (5). Specifically, we suppose:

(A.4) *For each ex post outcome* $(c(\mathbf{e}), \mathbf{e})$ *such that* $F(\mathbf{e}) \ge F(\mathbf{e}^*)$, *there is i with* $e_i > 0$, *and*

$$u_c^i(c_i, l_i)F_i(\mathbf{e}) \le u_l^i(c_i, l_i) \tag{8.30}$$

where $c_i(\mathbf{e}) \equiv c_i > 0$.

This assumption looks plausible, because as output moves above the first best, we would expect that marginal products do not increase, while the marginal rate of substitution between consumption and leisure tilts in favor of leisure. Thus if (5) already holds at the first best, we expect this relationship to be maintained (for at least one individual) for outcomes with higher output.

However, it is only fair to point out (A.4) is *not* automatically implied by (A.1) - (A.3). However, experimentation with different functional forms suggests that it is implied by a large subclass of welfare functions, individual preferences and production technologies.

For instance, suppose (in addition to (A.1)–(A.3)) that individual utilities are separable as the sum of concave functions of consumption and leisure, that total output is some concave, smooth function of the *sum* of individual efforts, and that the social welfare function has a separable and concave representation in utilities.

Consider some ex post outcome $(\mathbf{c}(\mathbf{e}), \mathbf{e})$, distinct from the first best, but with the property that at least as much output is being produced as in the first best: $F(\mathbf{e}) \ge F(\mathbf{e}^*)$. There must be *some* individual i with $e_i > e_i^*$. In the case under consideration, it will be the case that at least *one* such individual gets $c_i(\mathbf{e}) \ge c_i^*$ (remember that at least as much output is being produced). Furthermore, since output is no lower, marginal product cannnot have increased relative to the first-best. Putting all this together with (5) and using the convexity properties of preferences and technology, it should be the case that

$$u^i_c(c_i(\mathbf{e}), L_i - e_i)F_i(\mathbf{e}) \le u^i_l(c_i(\mathbf{e}), L_i - e_i)$$

for this individual, which is (6).

Consider, therefore, the following formalization of the separable case:

(A.4*) (i) u^i *is additively separable in* (c, l), (ii) $F(\mathbf{e})$ *is of the form* $f(e_1 + \ldots + e_n)$, for *some* $f : \mathbb{R}_+ \to \mathbb{R}_+$, and (iii) V is separable in individual utilities.

Our motivating argument above shows that under (A.1)—(A.3), (A.4*) implies (A.4).

8.3.2 Egalitarianism and Incentives.

8.3.2.1 Egalitarianism Yields Underproduction...

Proposition 8.10. *Under (A.1) - (A.3) and (FB), the first best cannot be achieved as some equilibrium of the soft game. Moreover, if (A.4) holds, every equilibrium of the soft game must involve underproduction.*

Proposition 8.10 verifies an intuitive conjecture: when a team cannot precommit to adequately punish deviants, the first best cannot be achieved. Specifically, in this model there is underproduction relative to the first best.

It is perhaps worth mentioning that while Proposition **??** appears intuitive (especially because it is related to other inefficiency notions, such as the Marshallian inefficiency of sharecropping, or the holdup problem), it is far from being obviously true in the present context. For instance, if individual utility has a separable linear representation in consumption, the proposition is, in general, false: there are equilibria that attain the first best.[7] Moreover, without (A.4), whether there is underproduction or not is an open question. Most significantly, the assumptions of Proposition 8.10 do not cover the case of Rawlsian social preferences, and as we shall see in the next section, our proposition fails in this case.

It might help, therefore, to sketch the proof of Proposition 8.10 for a special case. Assume that individual preferences have a strictly concave representation that is separable in consumption and leisure, and that given this representation, the Bergson-Samuelson welfare function has a representation that is strictly concave and separable in individual utilities. Furthermore, assume that total output is a concave, increasing function f of the *sum* of individual efforts. This is the case covered by (A.4*).

Now consider an ex post outcome $(\mathbf{c}, \mathbf{e})$ with *overproduction* relative to the first best. *Someone* must be putting in more effort relative to the first best. Call that someone i. Evaluated at the first best consumption vector, i must now have a higher marginal weight in the social welfare function than he had before. In the ex post outcome under consideration, therefore, i

[7] Such utility functions are ruled out by the assumption that leisure is a normal good (see (A.2)).

must enjoy more consumption, because total output is not lower. Putting all this information together with the convexity of preferences and technology, it is easy to see that

$$u^i_c f'(e_1 + \ldots + e_n) \leq u^i_l \tag{8.31}$$

which just means that (A.4) is automatically satisfied in this special case.

Now, suppose that i reduces his effort level a tiny bit. Perform the thought experiment of cutting i's consumption by *exactly* the resulting fall in output. By the inequality (8.31), this has a *nonnegative* first-order effect on i's utility. If the outcome is to be a Nash equilibrium, therefore, i's true consumption decline must be *at least* of this order. But now note that this decline must have a first-order effect on i's marginal utility of consumption, which rises. Moreover, by the supposition that i is no better off, the marginal welfare contribution of his utility is no lower. It is now easy to check that in this new situation, the first-order conditions for ex post welfare maximization are destroyed. For because i is being made to take at least the entire output loss (as we have argued above), there must be some other individual whose consumption is no lower. For this other person, exactly the reverse changes must occur in the marginal conditions. So, if the first order conditions were holding earlier (as they must have), they cannot be holding now! This contradicts ex post welfare maximization, and proves that the original outcome could not have been an equilibrium.

8.3.2.2 ...But Not With a Rawlsian Planner Our main theme relates the degree of egalitarianism to the degree of efficiency failure. It will be convenient to begin by exploring the most extreme form of egalitarianism: Rawlsian social preferences. In a later section, we will carry out the more complicated exercise of varying the welfare function over different degrees of egalitarianism.

Observe that the Rawlsian case is *not* covered by Proposition 8.10. That result rests on a postulate that is seemingly so innocuous that we have not emphasized it in the discussion (though, of course, it is formally stated in (A.3)). It is that social welfare is *strictly* increasing in *every* utility level, for each vector of utilities. Rawlsian social preferences do not satisfy this condition. Indeed, this observation has striking consequences, as we shall see.

We first define Rawlsian preferences. To do so we need a benchmark comparison. This is summarized in the following assumption:

(R) *Society is indifferent between the complete inaction of any two individuals. Thus if the Rawlsian welfare is written as*

$$V(\mathbf{u}) = \min_i u^i \tag{8.32}$$

where utilities have already been normalized so that $u^i(0, L_i) = u^j(0, L_j)$ for all i and j.

The first best Rawlsian optimum involves the maximization of the expression in (8.32), subject to the constraints that $u^i = u^i(c_i, l_i)$ for all i and that $(\mathbf{c}, \mathbf{e})$ is an outcome. Just as before, there is a unique first best outcome; call it (c^*, e^*).

Ex post outcomes are defined exactly as they were earlier. It is easy to see that for each $e \in \mathcal{E}$, there is a unique consumption vector $c(\mathbf{e})$ that solves the ex post maximization problem.

The collection of ex post outcomes induces a soft game just as before, and we may study its equilibria.

Proposition 8.11. *Under* (A.1), (A.2), (FB) *and* (R), *the Rawlsian first best is an equilibrium of the Rawlsian soft game. Furthermore, any equilibrium which gives at least one agent strictly more utility than the utility from inaction must be the Rawlsian first best outcome.*

(1) One might wonder whether the strikingly positive result of Proposition 8.11 is due to the fact that the Rawlsian egalitarian optimum is so devoid of efficiency properties that we do not have any incentive problem at all to maintain it. But this is not the case. For instance, consider any special case of this model that is *symmetric* across all agents. It is easy to see that the first best outcome is invariant across all symmetric quasi-concave welfare functions, including the Rawlsian one.

(2) There is only one qualification in Proposition 8.11. To prove that an equilibrium must be the Rawlsian first-best, we assume that at least one individual receives strictly more utility than he receives from inaction. There may exist an equilibrium involving total inaction. This will happen if the technology has the property that $F(\mathbf{e}) = 0$ whenever $n-1$ components of $\mathbf{e}$ equal zero. On the other hand, if the technology is such that output depends on the sum of the efforts (and if right-hand marginal utilities are defined and strictly positive everywhere) then our qualification can be dispensed with.[8]

Thus, in contrast to standard intuition, extreme egalitarianism might actually has pleasing incentive properties. Egalitarianism applies not only to the choice of the social optimum, but in the treatment of deviants from the optimum. A greater concern for egalitarianism goes hand in hand with the ability to credibly mete out stronger punishments.

Apart from the technicalities, the proof of Proposition 8.11 is very simple and general. Under convexity of the feasible set and preferences, the Rawlsian criterion has the following property. All individual utilities move in the *same* direction from one ex post outcome to another. Consequently, the Rawlsian first best is always an equilibrium of the Rawlsian soft game. For if someone could improve his utility by a deviation, he would improve the utility of everyone else in the process. This would contradict the fact that the earlier outcome was first best.[9]

The second part of the result — that *every* equilibrium must be first best — is model-specific in two respects. The convexity and the differentiability features of the model must both be exploited. These are used to guarantee that if an allocation is *not* first best, then there is some small, *unilateral* change in someone's effort level that creates an ex post outcome with

[8] Our requirement that the production function be differentiable at positive output is used in the proof. Concave, increasing production functions which produce no output when $n-1$ effort components equal zero are not (right hand) differentiable at 0, even though all partial (right hand) derivatives may exist at 0.

[9] It should be pointed out that this joint movement of utilities in the same direction is also a feature of the case in which utilities have a separable linear representation in consumption. To see this, let $u^i(c,l) = a_i c + b_i(l)$, where $a_i > 0$ and b_i is an increasing, smooth, concave function. Assume that $V(\mathbf{u}) = v_1(u_1) + \cdots + v_n(u_n)$, where each v_i is smooth, increasing, and strictly concave. Now look at the ex post outcomes of this model. If all consumptions are strictly positive, then the necessary and sufficient conditions characterizing ex post optimality are that for all i, j, $v_i'(u^i(c_i,l_i))a_i = v_j'(u^j(c_j,l_j))a_i$. The co-movement of utilities should now be apparent. So in this case as well, the first best can be supported as an equilibrium of the soft game.

a higher Rawlsian value. Again, using the Rawlsian criterion and the equal-utility property yielded by convexity, the individual who makes the change must participate in its benefits, thereby destroying the equilibrium possibilities of the given outcome.

8.3.2.3 Changing Egalitarianism The result of the previous section, and the discussion following it, suggest an even stronger observation: that as the extent of egalitarianism increases, the degree of underproduction should monotonically decline. Proposition 8.11 would be the limiting case of such an observation. The purpose of this section is to demonstrate such a possibility.

We consider a *symmetric* version of our model, described formally as follows. First, any permutation of $\mathbf{e}$ produces the same output level as $F(\mathbf{e})$. Second, every individual's preferences is represented by the same function u. Finally, the social welfare function is symmetric:

$$W(\mathbf{c}, \mathbf{l}) = V\left(u(c_1, l_1), ..., u(c_n, l_n)\right) \tag{8.33}$$

where V is symmetric. We will fix the individual representation u, and analyze the effect of changing egalitarianism by altering the form of V in a manner made precise below.

The following assumption (in addition to those already maintained) will be made on the fixed cardinal representation, u, of individual preferences.

(A.5) *There exists a cardinal representation such that u is strictly concave in c, and V is quasiconcave in $\mathbf{u}$.*[10]

We begin by discussing how to compare the "degree of egalitarianism" among different V's, or more generally, among different social welfare orderings on vectors of individual utility. Our definition, while not formally requiring symmetric social welfare orderings, is best viewed in this background. Let S and $S^{'}$ be two social welfare orderings for utility vectors $\mathbf{u}$. We will say that $S^{'}$ is *at least as egalitarian as* S if for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, and for all i and j, the condition

$$u_k = v_k \text{ for } k \neq i, j, \ \mathbf{v}S\mathbf{u}, \text{ and } |v_i - v_j| < |u_i - u_j| \tag{8.34}$$

implies $\mathbf{v}S'\mathbf{u}$.

That is, $S^{'}$ is more egalitarian than S if it prefers every redistribution plan to narrow the difference between two individuals' utility that R prefers as well . This can be viewed as an extension of the Pigou-Dalton principle for utility distributions (see, for example, Moulin [13]), applied to a comparison of any two social welfare orderings.

Some remarks on this last observation will clarify our definition. Observe that under our definition, a social welfare ordering S' is at least as egalitarian as the *utilitarian* welfare ordering if the condition

$$u_k = v_k \text{ for } k \neq i, j, \ u_i + u_j \leq v_i + v_j, \text{ and } |v_i - v_j| < |u_i - u_j|$$

[10] The fix on a cardinal representation that is strictly concave is not necessary for the results (as long as (A.3) is maintained), but it greatly simplifies the writing of the proofs.

implies $\mathbf{v}S'\mathbf{u}$. Note that the condition above is just a specialization of (8.34), when S is the utilitarian ordering. Specializing even further so that $u_i + u_j = v_i + v_j$ in the above condition, we obtain the Pigou-Dalton principle for utility distributions. In other words, a social welfare ordering which is at least as egalitarian as the utilitarian ordering satisfies the Pigou-Dalton principle. Our suggested relation is therefore an extension of this idea.

The argument above suggests that as we progressively move "towards egalitarianism" in this ordering, we obtain social welfare functions that are "more willing" to trade off total utility for interpersonal equality.

This definition, and the subsequent discussion leads to two general observations. First, in the class of social welfare functions considered in this paper, every such function is at least as egalitarian as the utilitarian function.[11] Second, the Rawlsian function is at least as egalitarian as any of these social welfare functions.[12] In short, the utilitarian and the Rawlsian social welfare orderings are two extremes of the social welfare functions we are considering, in the spectrum of the degree of egalitarianism defined here.

In our model, social welfare orderings are given by the functions satisfying (A.2) and (A.3), which includes differentiability. Let V and V' be two such *symmetric* functions. Then we can show that V' is at least as egalitarian as V if and only if

(8.35) $$\text{For all } \mathbf{u} \text{ and for all } i, j \in \{1, ..., n\}, \quad u_i \geq u_j \Leftrightarrow \frac{V_i'}{V_j'} \leq \frac{V_i}{V_j}$$

We omit a proof of this result (it is available on request).[13] But the interpretation should be quite natural. When one individual j has a lower utility than another individual i, the "more egalitarian" social welfare function assigns a higher (relative) marginal welfare to i. The social marginal rate of substitution between i and j is tilted in favor of j by the more egalitarian welfare function.

To obtain a clearer idea of the partial ordering proposed here, consider two classes of social welfare functions. The first is the Atkinson family, which is represented by the form

(8.36) $$[\sum_{i=1}^{n} u_i^{\rho}]^{1/\rho}$$

[11]This is because increasing, strictly quasiconcave welfare functions will indeed rank $\mathbf{v}$ over $\mathbf{u}$ whenever the specialization of condition (8.34) discussed in the text happens to hold.

[12]This is easily seen as follows. Suppose that for some social welfare ordering S, and utility vectors $\mathbf{u}$ and $\mathbf{v}$, condition (8.34) holds but nevertheless, $\mathbf{u}$ is preferred by the Rawlsian ordering to $\mathbf{v}$. This means, in particular, that $\min\{u_i, u_j\} > \min\{v_i, v_j\}$. But because $|u_i - u_j| > |v_i - v_j|$, this implies that (u_i, u_j) *vector dominates* (v_i, v_j). But now we have a contradiction to the assumption that $\mathbf{v}S\mathbf{u}$.

[13]It is worth noting that the asserted equivalence is an analogue of a result in Moulin [13, Lemma 2.2], which characterizes a differentiable social welfare function (called a "collective utility function") that satisfies the Pigou-Dalton principle for utility distributions.

where $\rho \in (-\infty, 1]$. It is easy to check that higher degrees of egalitarianism correspond to lower values of ρ. In particular, the case $\rho = 1$ yields the utilitarian, and $\rho = -\infty$, the Rawlsian social welfare ordering.

The second class is the "constant absolute risk-aversion" family, which is given by the form

$$-\sum_{i=1}^{n} \exp(-\eta u_i) \tag{8.37}$$

where $\eta \in (0, \infty)$. A higher value of η is associated with a more egalitarian function, and the two extreme values of η (0 and ∞) again give the utilitarian and the Rawlsian social orderings respectively.

Now, let us return to the model. An outcome is *symmetric* if all individuals work equally hard, and consume equally. In the symmetric model, symmetric effort by individuals induces a symmetric outcome, in which $\mathbf{c}(e, ..., e) = (c, ..., c)$ and $c = \frac{1}{n}F(e, ..., e)$. The first best (c^*, e^*) is also a symmetric allocation, and the allocation is independent of the social welfare function.

We will restrict our discussion to the class of symmetric equilibria of the soft game.

In what follows, it will be convenient to consider a modified soft game where all but the first individual choose the same level of effort. Then the ex-post outcome assigns the same level of consumption for those providing this effort. Denote the generated levels of consumption by

$$(c_1(e_1, e; V), c(e_1, e; V)) \tag{8.38}$$

where V is the social welfare function determining consumption, and e and c are the common values for all but the first individual. The first individual's best response to e is the set of solutions to the problem of maximizing his indirect utility function with respect to $e_1 \in [0, L]$. Denote this set by $B(e; V)$. $B(e; V)$ is a compact-valued and upper hemi-continuous correspondence of e . We will need to further assume:

(A.6) *$u(c_1(e_1, e; V), L - e_1)$ is single-peaked with respect to e_1.*

Then, $B(e; V)$ is convex-valued and the existence of a symmetric equilibrium for each V is guaranteed.[14]

By the *degree of underproduction* in any outcome, we refer to the difference between the first-best production level and the level of aggregate production under that outcome. The following proposition establishes that while the degree of underproduction is positive, the degree of underproduction must *fall* with rising egalitarianism.

Proposition 8.12. *Assume a symmetric model satisfying (A.1)—(A.3), (A.5), (A.6) and (FB). Then every symmetric equilibrium of the soft game involves underproduction. Now consider two welfare*

[14]It should be noted that our maintained assumptions *do not* automatically yield (A.6). However, all parametric forms that we have tried satisfy this assumption.

functions V and V', where V' at least as egalitarian as V. Then, for each symmetric equilibrium under V, there exists a symmetric equilibrium under V' such that the degree of underproduction is not more for the latter.

This verifies that the Rawlsian case discussed in Proposition 8.11 is not an exception or some quirky failure of continuity. For symmetric equilibria of the symmetric model, increased egalitarianism never increases the degree of underproduction.[15]

8.3.3 Extensions. We briefly consider some extensions, as well as possible objections to the setting in which these results have been derived.

1. Sinking past differences: In computing ex post optima, why do teams take account of the sunk efforts already incurred by its members? It may be argued that bygones are bygones, and that the social welfare function should ignore this. This point of view may be identified with the assertion that individual utilities "should" be separable in consumption and leisure, and that the social welfare function "should" be utilitarian. At least, that is the only way to justify the assertion if one sticks to the Bergson-Samuelson setting. Such particular functional forms are already accommodated as a special case of the paper, and if one insists on such an interpretation, an entire class of welfare functions (including the Rawlsian function) is simply removed from consideration.

On the other hand, one might take the sinking of past differences as a primitive, simply asserting that ex post output is shared equally (in a symmetric model) *irrespective* of the welfare function. We find it difficult to see what might justify such an assumption. Consider any resource allocation problem. As long as resources are not allocated to everybody at exactly the same point of time, this point of view leads to absurd allocations.

If one were to admit such a structure for the sake of argument, inefficiencies will *always* arise, of course. But even then, it can be shown that as long as *some* weight is given to the past, Proposition 8.12 will continue to hold. In particular, the Rawlsian welfare function will exhibit the lowest degree of underproduction.

2. Emotions and credible punishments: One might proceed in the exactly the opposite direction to (1). Deviants that do not adhere to the desired outcome might enjoy an entirely different (and reduced) weight in the ex post welfare function. Emotions such as anger or social disapproval might induce such changes, and in so doing, lend credibility to punishments (for similar ideas, see, e.g., Frank [6]). It is to be expected that the degree of inefficiency will be lowered in the presence of these emotions. But even so, as long as the first best is thereby not automatically achieved, more egalitarian welfare functions will possess better efficiency properties.

3. The observability of effort: The results of this paper rest critically on the assumption that efforts are observable. We suspect that the results would extend to noisy observability,

---

[15] It is possible to find even stronger versions of this proposition in special cases. For instance, if the best-response correspondence is single-valued, and social welfare functions are in the Atkinson class, we can show that increased egalitarianism *strictly* lowers the maximal degree of underproduction. The qualification "maximal" is due to the fact that equilibrium may not be unique.

though we have not checked this. Of course, if efforts are not observable *at all*, the results cease to have any relevance.

4. Repeated Relationships: In a dynamic situation, teams would recognize that a departure from the soft mechanism will serve them well in the longer run. It may be of interest to study such repeated relationships. But as a prelude to that study, it is surely important to analyze the "one-shot" relationship without precommitment, which is exactly what we do here.

8.3.4 Discussion. While the connections between egalitarianism and inefficiency has long been a subject of debate, there have been surprisingly few attempts to model the exact nature of the tradeoff. To be sure, the sources of various tradeoffs are manifold in nature. This paper investigates one potential source, and shows that the commonly held intuition is not valid, at least in this case.

Specifically, this paper studies the idea that egalitarianism fails to uphold proper incentives because credible punishments are thereby destroyed. This statement really has two parts to it: one is the familiar "dilemma of the Samaritan" induced by the inability to precommit. Organizations that cannot precommit, *yet derive their sense of goal-fulfillment or welfare from the welfare of its members*, are particularly prone to these potential inefficiencies. Indeed, organizational structures where the source of utility for the "principal" is directly opposed to that for the "agent" will certainly do better (in terms of efficiency) compared to the structures considered here.[16] This is the intuition upheld by Proposition 8.10. An inability to precommit the reward function results in inefficiency. While the *particular* result proved is, to our knowledge, new, there is nothing particularly surprising or novel about the underlying theme.

But this is only one half of the idea. The second part goes further. It states that *egalitarian* organizations faced with the inability to precommit are doubly cursed: they are inefficient on the additional count that they cripple the incentive system (already weakened by the inability to precommit) even further. In this paper, we argue that this assertion is wrong. Increased egalitarianism *restores* incentives that are damaged by the lack of commitment (Proposition 8.12). Indeed, in the extreme case of Rawlsian egalitarianism, the precommitment and no-precommitment yield exactly the same first-best outcomes.

The results in this section are provocative on two counts. First, they might inspire greater interest in a challenging and crucially important area of research: the connections between the ethic of equality and the yardstick of aggregate performance (such as GNP growth). The second aspect is one that we have not emphasized in this paper, but of great interest, we believe. This is the connection with the theory of implementation. By far the dominant

[16] Indeed, in the usual principal-agent setup the problem is trivial. If efforts are observable, the organization is best off using a forcing contract. The contract demands that each individual must supply (at least) his first best effort level. If this effort level is indeed supplied by *all* individuals, all individuals receive the first best output division. If, however, these effort levels are not forthcoming, the contract promises dire consequences for the individuals who have deviated. These threats indeed implement the first best (at least as one equilibrium), and there is no incentive problem worth serious analysis. Authors such as Holmstrom (Holmstrom [8]) have argued that the existence of a residual claimant (capitalist, manager) creates a credible threat to carry out these punishments, such as retention of produced output, in the event of breach.

approach in this literature presumes that the planner can implement outcomes without regard to ex post credibility. But there are many situations where it is natural to constrain mechanisms off the equilibrium path by the provision that they should not be suboptimal relative to the actions that have been taken and observed, and the social welfare function of the planner. This may have some bearing on narrowing the class of allowable mechanisms in implementation contexts.[17]

8.3.5 Proofs. *Note.* Equation numbering needs to be changed.

Lemma 8.2. *Consider some ex post outcome $(c(\mathbf{e}), \mathbf{e})$ such that $F(\mathbf{e}) \geq F(\mathbf{e}^*)$, with the additional property that for some individual i,*

$$u_c^i(c_i, l_i) F_i(\mathbf{e}) \leq u_l^i(c_i, l_i) \tag{8.39}$$

Then such an outcome cannot be an equilibrium of the soft game.

Proof. Suppose that $(\mathbf{c}(\mathbf{e}), \mathbf{e}) \equiv (\mathbf{c}, \mathbf{e})$ is an ex post outcome with $F(\mathbf{e}) > 0$. Denote by $W_i(\mathbf{c}, \mathbf{l})$ the partial derivative of W with respect to c_i. Then, under (A.1)–(A.3) and using the ex post maximization problem (4), $(\mathbf{c}, \mathbf{e})$ must satisfy the following property: if $c_k > 0$, then for all $j = 1, \ldots, n$,

$$W_k(\mathbf{c}, \mathbf{l}) \geq W_j(\mathbf{c}, \mathbf{l}) \tag{8.40}$$

with equality whenever $c_j > 0$.

We need a slightly stronger implication than (16). Let i be such that $e_i > 0$, and suppose that there is some k with $c_k = 0$. Suppose, further, that there exists $\epsilon > 0$ such that for all $e_i' \in (e_i - \epsilon, e_i)$, the ex post consumption vector $c(e_i', \mathbf{e}_{-i}) \equiv \mathbf{c}'$ has $c_k' > 0$. Then, indeed, we can say that (16) holds for this k *even if* $c_k = 0$. We exclude the verification of this simple observation.

Now, suppose that, contrary to the lemma, this outcome is an equilibrium of the soft game.

Let M be the set of all indices j such that either $c_j > 0$, or with the property that there is $\epsilon > 0$ such that for all $e_i' \in (e_i - \epsilon, e_i)$, the ex post consumption vector $c(e_i', \mathbf{e}_{-i}) \equiv c'$ has $c_j' > 0$. We are going to consider the effect (on i's utility) of a small reduction in e_i by differential methods. By the maximum theorem (and the uniqueness of ex post consumption (given effort)), $\mathbf{c}(\mathbf{e})$ is a continuous function. Therefore the set M is all that counts for the analysis, and for all $k, j \in M$, (16) holds with equality.

Note first that because $(\mathbf{c}, \mathbf{e})$ is an equilibrium and $e_i > 0$, we have $c_i > 0$. Consequently, $i \in M$. Without loss of generality, number the indices in M as $1, \ldots, m$, and let i be rechristened with the index 1. We now claim that $m \geq 2$. Suppose not. Then, because $(\mathbf{c}, \mathbf{e})$ is an equilibrium, we must have $e_j = 0$ for all $j \neq 1$. Therefore, because $F(\mathbf{e}) \geq F(\mathbf{e}^*)$ and because we have assumed that the first best output is larger than the output of any individual acting completely on his own, we must have

$$u_c^1(c_1, l_1) F_1(\mathbf{e}) < u_l^1(c_1, l_1) \tag{8.41}$$

[17] Maskin and Moore [11] study a related problem where the ability of the planner is limited by the possibility that agents may renegotiate the outcome.

Because $M = \{1\}$, it follows from (17) and the definition of M that a small reduction in e_1 will raise 1's welfare, because all other consumptions will continue at zero. This contradicts our supposition that $(\mathbf{c}, \mathbf{e})$ is an equilibrium, and shows that $m \geq 2$.

For all $i \in M$, we have the first order condition of the ex post maximization problem: for some $\lambda < 0$,

$$\begin{aligned} W_i(\mathbf{c}, \mathbf{l}) + \lambda &= 0 \\ c_1 + \ldots + c_m &= F(\mathbf{e}) \end{aligned} \tag{8.42}$$

We are interested in differentiating this system with respect to a parametric change in e_1, and studying $\frac{dc_1}{de_1}$.[18] Given our remarks above, the differentiation argument will reflect the true story for small changes in e_1, provided that we think of these changes as *reductions*.

Differentiating (18) and defining $W_{ij} \equiv \frac{\partial^2 W}{\partial c_i \partial c_j}$, we obtain the system

$$\begin{pmatrix} 0 & 1 & 1 & \ldots & 1 \\ 1 & W_{11} & W_{12} & \ldots & W_{1m} \\ 1 & W_{21} & W_{22} & \ldots & W_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_{m1} & W_{m2} & \ldots & W_{mm} \end{pmatrix} \begin{pmatrix} d\lambda \\ dc_1 \\ dc_2 \\ \vdots \\ dc_m \end{pmatrix} = \begin{pmatrix} F_1(\mathbf{e})de_1 \\ \frac{\partial W_1(c,\mathbf{l})}{\partial l_i} de_1 \\ \frac{\partial W_2(c,\mathbf{l})}{\partial l_i} de_1 \\ \vdots \\ \frac{\partial W_m(c,\mathbf{l})}{\partial l_i} de_1 \end{pmatrix} \tag{8.43}$$

Let us calculate $\frac{\partial W_i(\mathbf{c},\mathbf{l})}{\partial l_i}$. For $i = 1$,

$$\frac{\partial W_1(\mathbf{c}, \mathbf{l})}{\partial l_1} = \frac{\partial [V_1(\mathbf{u}) u_c^1(c_1, l_1)]}{\partial l_1} = V_{11}(\mathbf{u}) u_c^1 u_l^1 + V_1(\mathbf{u}) u_{cl}^1 \tag{8.44}$$

For $i \neq 1$, we have

$$\frac{\partial W_i(\mathbf{c}, \mathbf{l})}{\partial l_1} = \frac{\partial [V_i(\mathbf{u}) u_c^i(c_i, l_i)]}{\partial l_1} = V_{i1}(\mathbf{u}) u_c^i u_l^1 \tag{8.45}$$

Next, let us calculate $W_{i1}(\mathbf{c}, \mathbf{l})$. For $i = 1$,

$$W_{11}(\mathbf{c}, \mathbf{l}) = \frac{\partial [V_1(\mathbf{u}) u_c^1(c_1, l_1)]}{\partial c_1} = V_{11}(\mathbf{u}) [u_c^1]^2 + V_1(\mathbf{u}) u_{cc}^1 \tag{8.46}$$

while for $i \neq 1$, we see that

$$W_{i1}(\mathbf{c}, \mathbf{l}) = \frac{\partial [V_i(\mathbf{u}) u_c^i(c_i, l_i)]}{\partial c_1} = V_{i1}(\mathbf{u}) u_c^1 u_c^i \tag{8.47}$$

[18](8.42) defines $(\mathbf{c}, \lambda)$ implicitly as a function of the "parameter" $\mathbf{e}$. Because output is positive, it follows from (A.1)–(A.3) that this function is C^1. Moreover, because the bordered Hessian in (8.43) has non-zero determinant (see footnote 6), it follows from the implicit function theorem that $\mathbf{c}$ is a differentiable function of $\mathbf{e}$, and in particular of e_1.

Now, recalling that leisure is a normal good in individual preferences, we have $u_c^1 u_{cl}^1 - u_l^1 u_{cc}^1 > 0$ (recall (A.2) and footnote 7). Using this and manipulating (20), we have

$$\begin{aligned} \frac{\partial W_1(\mathbf{c},\mathbf{l})}{\partial l_1} &= \frac{u_l^1}{u_c^1}(V_{11}(\mathbf{u})[u_c^1]^2 + V_1(\mathbf{u})\frac{u_c^1 u_{cl}^1}{u_l^1}) \\ &> \frac{u_l^1}{u_c^1}(V_{11}(\mathbf{u})[u_c^1]^2 + V_1(\mathbf{u})u_{cc}^1) \\ &= \frac{u_l^1}{u_c^1} W_{11}(\mathbf{c},\mathbf{l}) \end{aligned} \tag{8.48}$$

while from (21) and (23) it is easy to see that for all $i \neq 1$,

$$\frac{\partial W_i(\mathbf{c},\mathbf{l})}{\partial l_1} = \frac{u_l^1}{u_c^1} W_{i1}(\mathbf{c},\mathbf{l}) \tag{8.49}$$

Now, let us return to (19) and write down the solution for $\frac{dc_1}{de_1}$. Let U be the determinant of the cofactor of the first 1 (from the left) in the first row of the matrix in (19), and U_i be the determinant of the cofactor of $W_{i1}, i = 1, \ldots, m$. Include in these the signs generated by cofactor expansion. Let U^* be the determinant of the matrix in (19). Then it can be seen that

$$\frac{dc_1}{de_1} = \frac{1}{U^*}[F_1(\mathbf{e})U + \frac{\partial W_1(\mathbf{c},\mathbf{l})}{\partial l_1}U_1 + \frac{\partial W_2(\mathbf{c},\mathbf{l})}{\partial l_1}U_2 + \ldots + \frac{\partial W_m(\mathbf{c},\mathbf{l})}{\partial l_1}U_m]$$

Recalling from the strict quasiconcavity of $W(.)$ and (18) that U_1 and U^* have different signs, and using (24) and (25), we have

$$\frac{dc_1}{de_1} < \frac{1}{U^*}[F_1(\mathbf{e})U + \frac{u_l^1}{u_c^1}(W_{11}U_1 + W_{21}U_2 + \ldots + W_{m1}U_m)] \tag{8.50}$$

Next, because c_1 is a noninferior good under $W(., l)$, we have $U/U^* \geq 0$. Consequently, combining (8.39) and (26),

$$\begin{aligned} \frac{dc_1}{de_1} &< \frac{u_l^1}{u_c^1}\frac{1}{U^*}[U + W_{11}U_1 + W_{21}U_2 + \ldots + W_{m1}U_m] \\ &= \frac{u_l^1}{u_c^1} \end{aligned} \tag{8.51}$$

Therefore, using (27),

$$\frac{du^1(c_1, l_1)}{de_1} = u_c^1 \frac{dc_1}{de_1} - u_l^1 < 0 \tag{8.52}$$

so that (28) proves that a small reduction in e_1 will raise the utility of 1 under the new ex post outcome. This contradicts our supposition that $(\mathbf{c}, \mathbf{e})$ is an equilibrium, and completes the proof of the lemma. ■

Proof of Proposition 1. Observe that the first best allocation is indeed an ex post outcome, with positive output, satisfying (8.39) in Lemma 1. Given the assumption that at the first best, there is i with $(e_i^*, l_i^*) \gg 0$, (8.39) follows as a necessary condition describing the first best. By Lemma 1, this outcome cannot be an equilibrium of the soft game.

If (A.4) holds, then consider any ex post outcome with $F(\mathbf{e}) \geq F(\mathbf{e}^*)$. It follows right away from (A.4) that (8.39) is satisfied. By Lemma 1, this outcome cannot be an equilibrium of the soft game. ■

Proof of Proposition 2. For any outcome $(\mathbf{c}, \mathbf{e})$, define its *Rawlsian value* as $R(\mathbf{c}, \mathbf{e}) \equiv \min_i u^i(c_i, l_i)$. Now note that under the first best outcome, we must have, for *every* i,

$$R(\mathbf{c}^*, \mathbf{e}^*) = u^i(c_i^*, l_i^*) \tag{8.53}$$

The reason is that for each i, either $c_i^* > 0$ or $l_i^* > 0$ (or both), because $u^i(c_i^*, l_i^*) \geq u^i(0, L_i)$. Therefore, if (29) were not true, we could always improve the Rawlsian value of the outcome by small changes.

Given this claim, we first prove that the first best outcome is indeed an equilibrium. Suppose not. Then for some i, there is $e_i \in [0, L_i]$ such that

$$u^i(c_i, l_i) > u^i(c_i^*, l_i^*) \tag{8.54}$$

where c_i is the ith component of $\mathbf{c} \equiv \mathbf{c}(\mathbf{e})$, where $\mathbf{e} \equiv (e_i, \mathbf{e}_{-i}^*)$.

Because $u^i(c_i^*, l_i^*) \geq u^i(0, L_i)$ and because (30) holds, it must be the case that $c_i > 0$. It follows that

$$R(\mathbf{c}, \mathbf{e}) = u^i(c_i, l_i) \tag{8.55}$$

For if not, the Rawlsian value of $(\mathbf{c}, \mathbf{e})$ can be improved by only changing the consumption allocation, a contradiction to the fact that $(\mathbf{c}, \mathbf{e})$ is an ex post outcome (use the fact that $c_i > 0$).

Combining (29), (30) and (31), we have

$$R(\mathbf{c}, \mathbf{e}) > R(\mathbf{c}^*, \mathbf{e}^*)$$

but this contradicts the fact that (c^*, e^*) is the Rawlsian first best.

Now we prove the second part of the proposition. First, we show that if $(\mathbf{c}, \mathbf{e})$ is an equilibrium, then for all $i = 1, \ldots, n$,

$$R(\mathbf{c}, \mathbf{e}) = u^i(c_i, l_i) \tag{8.56}$$

Suppose this is not true. Then for some pair i, j, $u^i(c_i, l_i) > u^j(c_j, l_j)$. But then it must be the case (by an earlier argument) that $c_i = 0$. Consequently, $u^j(c_j, l_j) < u^i(0, l_i) \leq u^i(0, L_i) = u^j(0, L_j)$, where the last equality follows from normalization. But then, $(\mathbf{c}, \mathbf{e})$ cannot be an equilibrium, for j can guarantee himself at least $u^j(0, L_j)$ by deviating.

With this established, we return to the main proof. Suppose that the proposition is false. Then there exists an equilibrium $(\mathbf{c}, \mathbf{e}) \neq (\mathbf{c}^*, \mathbf{e}^*)$, such that at least one player gets strictly more utility than inaction. By virtue of (32) and assumption (R), it follows that *every* player gets strictly more than inaction. This observation will be used below.

Because the first best is unique, $R(\mathbf{c}^*, \mathbf{e}^*) > R(\mathbf{c}, \mathbf{e})$. Define $\mathbf{e}(t) \equiv t\mathbf{e} + (1-t)\mathbf{e}^*$ for $t \in [0, 1]$, and $\mathbf{c}(t) \equiv \mathbf{c}(\mathbf{e}(t))$. Then it is easy to check that $R(\mathbf{c}(t), \mathbf{e}(t)) > R(\mathbf{c}, \mathbf{e})$ for all $t \in (0, 1)$ (use (A.1), (A.2), (29) and (32)). Using (32), it follows that for all $t \in (0, 1)$ and all i,

$$u^i(c_i(t), l_i(t)) > u^i(c_i, l_i) \tag{8.57}$$

Now we claim that the following is true:

There exists j such that *either* (i) $e_j = L_j$ and $u_c^j(c_j, l_j)F_j(\mathbf{e}) < u_l^j(c_j, l_j)$, or (ii) $e_j \in (0, L_j)$ and $u_c^j(c_j, l_j)F_j(\mathbf{e}) \neq u_l^j(c_j, l_j)$, or (iii) $e_j = 0$ and $u_c^j(c_j, l_j)F_j(\mathbf{e}) > u_l^j(c_j, l_j)$.

To prove the claim, first recall that each player gets strictly more than inaction at $(\mathbf{c}, \mathbf{e})$, so that by (A.2), $(u_c^i(c_i, l_i), u_l^i(c_i, l_i)) \gg 0$ for all i. Using (A.2) again, together with (33), we see that for all i,

$$(c_i(t) - c_i) + (l_i(t) - l_i)\frac{u_l^i}{u_c^i} > 0$$

or equivalently, for all i,

$$(c_i(t) - c_i) - (e_i(t) - e_i)\frac{u_l^i}{u_c^i} > 0 \tag{8.58}$$

Now, if the claim is false, then, noting that $e_i(t) \in [0, L_i]$ for all i, we can use (34) to deduce that

$$(c_i(t) - c_i) - (e_i(t) - e_i)F_i(\mathbf{e}) > 0$$

Summing this inequality over all i, we have

$$F(\mathbf{e}(t)) - F(\mathbf{e}) > (e_1(t) - e_1)F_1(\mathbf{e}) + \ldots + (e_n(t) - e_n)F_n(\mathbf{e}) \tag{8.59}$$

Now observe that because $(\mathbf{c}, \mathbf{e})$ provides more utility than inaction, $F(\mathbf{e}) > 0$. But then (35) contradicts (A.1), which assumes that F is differentiable whenever $F(\mathbf{e}) > 0$, and that F is concave. This establishes the claim.

Pick j as given by the claim. If part (i) of the claim is true, notice that because $(\mathbf{c}, \mathbf{e})$ is an equilibrium, $c_j > 0$. Then, a small reduction in e_j makes j *strictly* better off even if j pays for the entire reduction in output from his own consumption. This improvement in j's utility persists even if j must pay a small additional amount to each of the other agents. This proves that the Rawlsian value of $(\mathbf{c}, \mathbf{e})$ can be improved by a *unilateral* change made by j, by going to an ex post outcome $(\mathbf{c}', \mathbf{e}')$ (where $\mathbf{e}'$ differs from $\mathbf{e}$ only in the jth component). That is, $R(\mathbf{c}', \mathbf{e}') > R(\mathbf{c}, \mathbf{e})$. But note that, as a result, $u^j(c'_j, l'_j) \geq R(\mathbf{c}', \mathbf{e}') > R(\mathbf{c}, \mathbf{e}) = u^j(c_j, l_j)$ (where the last equality uses (32)). This contradicts the supposition that $(\mathbf{c}, \mathbf{e})$ is an equilibrium.

Finally, if parts (ii) or (iii) of the claim are true, use the same argument as above, if $u_c^j(c_j, l_j)F_j(\mathbf{e}) < u_l^j(c_j, l_j)$, or its reverse (increase e_j), if $u_c^j(c_j, l_j)F_j(\mathbf{e}) > u_l^j(c_j, l_j)$, to arrive at a contradiction.

This completes the proof of the proposition. ■

Proof of Proposition 3. By our assumptions, the first best is symmetric and has

$$0 < e^* < L.$$

It follows from the first-order conditions characterizing the first best that

$$u_c(c^*, L-e^*)F_i(e^*,\ldots,e^*) = u_l(c^*, L-e^*),$$

and moreover, for every $e > e^*$ and $c \equiv \frac{1}{n}F(e,\ldots,e)$,

$$u_c(c, L-e)F_i(e,\ldots,e) \le u_l(c, L-e).$$

Moreover, these hold for all i. It follows from Lemma 1 that no symmetric $e \ge e^*$ can be an equilibrium of the soft game.

To prove the remainder of the proposition, fix a common utility level $\bar{u}$ for all but the first individual, and let u_1 denote the utility level of individual 1. Then by applying the characterization of increased egalitarianism for smooth welfare functions, we see that

$$\bar{u} \ge u_1 \Leftrightarrow \frac{V_2'}{V_1'} \le \frac{V_2}{V_1} \tag{8.60}$$

where a subscript 1 denotes the marginal social welfare contributed by the first individual's utility, and the subscript 2 is the corresponding contribution by the others. Define a "utility possibility frontier" (given efforts) in the modified soft game by $\Psi(\bar{u}, e_1, e)$:

$$\Psi(\bar{u}, e_1, e) \equiv u(F(e_1, e, ..., e) - (n-1)c, L - e_1) \tag{8.61}$$

where c is chosen such that $\bar{u} = u(c, L-e)$. Ψ is downward sloping and concave with respect to $\bar{u}$, by (A.5).

The social planner chooses consumption to maximize V on Ψ. Let us introduce some notation. We will denote by $U_1(e_1, V)$ the utility of agent 1 when the action taken by him is e_1, the welfare function is V, the action of the remaining agents is fixed at e (which is not explicitly carried in the notation), and the planner chooses ex-post consumption optimally. Likewise, we will denote by $U(e_1, V)$ the utility of each of the other agents under exactly the same state of affairs.[19] We make two observations. First,

$$c_1(e, e; V) = c(e, e; V) = \frac{1}{n}F(e, ..., e) \tag{8.62}$$

for any V. Second, it is possible to show, using the concavity of Ψ, the strict quasiconcavity of V and V', and (36),[20] that

$$U_1(e_1, V) \ge U(e_1, V) \Leftrightarrow U_1(e_1, V) \ge U_1(e_1, V') \ge U(e_1, V') \ge U(e_1, V), \tag{8.63}$$

when V' is at least as egalitarian as V. Moreover, precisely the opposite chain of inequalities hold on the RHS of (39) if the opposite inequality holds on the LHS of (39).

Let V' be at least as egalitarian than V. We claim that

[I] If $e \in B(e; V)$, then $e \le \max B(e; V')$.

[19] Thus $U_1(e_1, V) = u(c_1(e_1, e, V), L - e_1)$ and $U(e_1, V) = u(c(e_1, e, V), L - e)$.

[20] In fact, this is the only point at which (36) is used in the proof.

Suppose not. Then $\max B(e; V') < e$. So there exists $x \in [0, e)$ such that

$$U_1(x, V') > U_1(e, V'). \tag{8.64}$$

Because $e \in B(e, V)$, it must be the case that

$$U_1(x, V) \leq U_1(e, V) \tag{8.65}$$

Noting from (38) that $U_1(e, V) = U_1(e, V')$, we may combine (40) and (41) to obtain

$$U_1(x, V) < U_1(x, V') \tag{8.66}$$

Combine (42) with (39) and the claim immediately following (39). We may deduce that

$$U_1(x, V') \leq U(x, V') \tag{8.67}$$

Noting from (38) that $U_1(e, V') = U(e, V')$, and combining this observation with (40) and (43),

$$U(x, V') \geq U_1(x, V') > U_1(e, V') = U(e, V'). \tag{8.68}$$

In words, (40) tells us that under V', player 1 is better off choosing x rather than e. At the same time, (44) tells us that the rest of the players are *also* better off when player 1 chooses x instead of e. Thus a vector-inferior collection of efforts $(x, e, \ldots, e)$ instead of $(e, \ldots, e)$ (which in turn is lower than the first best $(e^*, \ldots, e^*)$) leads to a Pareto-improvement. Under our assumptions, this cannot be.

To see this more formally, note that if $x < e < e^*$, then

$$U_1(e, V') = U_1(e, V) > \Psi(\bar{u}, x, e) \tag{8.69}$$

as long as $\Psi(\bar{u}, x, e) \leq \bar{u}$. By putting $\bar{u} = U(x, V')$, $\Psi(\bar{u}, x, e) = U_1(x, V')$, and using (44), we contradict (45). This completes the proof of Claim [I].

Denote the maximal effort level among all symmetric equilibria under V' by $\bar{e}$. We have already shown that $\bar{e} < e^*$. We claim that

[II] $\max B(e; V') < e$ for all $e \in (\bar{e}, L]$.

To see this, first observe that $\max B(L, V') \leq L$ simply by definition. Now suppose that the claim is false. Then for some $e \in (\bar{e}, L]$, $\max B(e; V') \geq e$. Moreover, $B(., V')$ is convex-valued and upperhemicontinuous. But this establishes (using a simple argument analogous to the Intermediate Value Theorem) the existence of some $e' > \bar{e}$ such that $e' \in B(e'; V')$. Moreover, a fixed point of $B(, ; V')$ corresponds to a symmetric equilibrium. But this contradicts the definition of $\bar{e}$ as the *largest* symmetric equilibrium effort level under V'.

To complete the proof of the proposition, suppose that there exists $e \in (\bar{e}, L]$ which belongs to $B(e; V)$. Then $\max B(e; V') \geq e > \max B(e; V')$, where the first inequality follows from Claim [I], and the second inequality follows from Claim [II]. This is a contradiction, and the proof is complete. ■

CHAPTER 9

Notes on Inequality and Growth

Begin with the conceptual links between inequality and growth. I will say a little bit on each.

9.1 Inequality, Capital markets and Growth

We've already seen some of this from the earlier lectures. If capital markets are imperfect, inefficiencies will arise from the fact that some individuals will be unable or unwilling to take the socially optimal action — the imperfections generate a high or infinite utility cost of doing so. Whether or not the final outcome is truly inefficient in the sense of dynamic Pareto-efficiency remains to be seen (Mookherjee and Ray (2003) present some results to this effect) but it is certainly the case that output is lower than in the first-best.

Just what is needed to translate this sort of result into a lower *rate of growth* is unclear. Ideally, one would have to write down a model that allows for unbounded expansion — say, at some constant exponential rate — and then examine whether such an expansion rate is affected by the extent of capital market imperfections.

Such a model will also have to take account of the fixed cost barrier to setting up in a new occupation or business. This must also "move" with growth, otherwise all credit market imperfections must ultimately die away. One reason why such barriers may move in step with growth is that barriers are denominated in human capital (e.g., education for an occupation or professional staff for a business) which is getting more expensive with growth. for a related model, see Rigolini (2003).

9.2 Inequality, Public Allocation and Growth

With high inequality, high-income or high-wealth individuals may garner a disproportionate share of public assets. But doing so may be inefficient from the point of view of future growth. Here are three examples:

1. Bhagwati and Desai (1956), in their study of Indian planning, suggest that the richer business interests could easily buy up licences in a corrupt system, leading to the allocation

of valuable resources (such as import quotas) in their direction. New blood didn't get a chance. See also Banerjee (1997) on bureaucratic corruption.

2. Sokoloff and Engerman (2000) and Acemoglu, Johnson and Robinson (2001) have argued that an inefficient elite, once installed in power, might attempt to do all it can to keep the non-elite at bay. As Acemoglu (2005) argues, they might set fiscal policy for three reasons: (i) to generate revenue off the back of non-elite businesses, (ii) to make it more difficult for such non-elites to conduct business and this way lower the factor costs of production (e.g., wage rates) for their own use, and (iii) they might tax non-elites in order to reduce the resources available for political opposition. This is more a theory of how inequality in the access to political power (which may in turn stem from economic inequality in the past) might stifle overall growth.

3. Esteban and Ray (2005) argue that public allocations may be severaly distorted even in the absence of any corrupt motive. Imagine a world in which governments — just like private agents — are unsure which direction the economy will, or should take. for instance, they may not know whether textiles or software or call centers will be the wave of the future. If the government has a limited supply of "licences" (infrastructual support, subsisdies, tax breaks, etc.) that it can hand out, then it is rational for them to do so to sectors that lobby hard. But hard lobbiers can be the outcome of individuals being highly productive (in which case the government should support the lobby) or people just being rich (in which case there is nothing to be gained from support). Thus high inequality distorts the lobbying signal, and reduces the profitability of public allocation.

9.3 Inequality, Redistribution and Growth

The models of this section are diametrically opposed to the models of the previous section. In the previous argument, high inequality leads to greater amount of rent-seeking, exploitation, or regressive taxation of the non-elite, which in turn stifles growth through bad allocation. In contrast, the models of Alesina and Rodrik (1994) and Persson and Tabellini (1994) argue that high inequality sets up a clamor for redistributive taxation. That redistribution then ends up being bad for growth because distortionary taxes stifle investment and therefore growth.

These models are set up in a democratic context, and the median voter theorem is used in particular. The simplest way to see these models (here I follow Bénabou (2000)) is to imagine that income is distributed lognormally — $\ln y \sim N(m, \Delta^2)$. Then it is trivial to see that the log of the median income is precisely m, and also, if μ denotes mean income,

$$m = \ln \mu - \frac{\Delta^2}{2}. \tag{9.1}$$

Suppose that there are only two policies on the table: $t = 0$ (laissez faire) and $t = 1$ (with full lumpsum transfers to everyone; full redistribution). Everyone below the mean income will prefer full redistribution, so that the degree of support for full redistribution may be proxied by the value of the cfd evaluated at μ. Recalling (9.1) and using Ψ for the cdf of a standard normal, we see that the degree of support for redistribution is proportional to

$$p = \psi\left(\frac{\Delta^2/2}{\Delta}\right) = \psi\left(\frac{\Delta}{2}\right).$$

Of course this value is bigger than 1/2, but that isn't necessarily enough for full redistribution to be passed. After all, as Bénabou argues, a large fraction of the individuals at the lower end may not vote at all. For instance, if the π poorest agents don't vote, then majority requires that we need a support of $(1-\pi)/2$ counting from the bottom of those individuals that *do* vote. Adding back the π, this shows that the identity of the modified median voter is given by λ, where $\Psi(\lambda) = (1+\pi)/2$.

However, what appears (seemingly to be) robust is the point that $\partial P/\partial\Delta > 0$. That is, the greater is the amount of inequality in the system, the more widespread is the support for full redistribution.

What if one augments the model — as one needs to, to connect it with growth — by presuming that redistributions are costly? Continuing to follow Bénabou, the simplest depiction of that cost is one in which every income y is reduced equiproportionately from y to ye^{-B}, where $B > 0$, whenever redistribution takes place. Now the threshold income Y which will support full redistribution is no longer μ, but is given by

$$\ln Y = \ln\mu - B = m + \left(\frac{\Delta^2}{2} - B\right),$$

using (9.1). Consequently, the new degree of support is now given by

$$p = \psi\left(\frac{-B + \Delta^2/2}{\Delta}\right) = \psi\left(-\frac{B}{\Delta} + \frac{\Delta}{2}\right).$$

Of course, this reduces the level of support for redistribution (relative to the benchmark case) but what is more relevant is that the sensitivity of demand for redistribution to the underlying amount of inequality becomes even higher. To see this, observe that

$$\frac{\partial P}{\partial\Delta} = \psi'\left(-\frac{B}{\Delta} + \frac{\Delta}{2}\right)\left(\frac{B}{\Delta^2} + \frac{1}{2}\right)$$

and this exceeds the ealier sensitivity, which is just

$$\frac{1}{2}\psi'\left(\frac{\Delta}{2}\right)$$

on two counts, provided that $Y > m$.[1]

9.4 Inequality, Lack of Redistribution and Growth

Notice that the model of the previous section relies on very different predictions regarding the effect of redistribution on growth, relative to the arguments that went before. The Alesina-Rodrik view is that redistribution occurs via distortionary taxation and is therefore necessarily growth-reducing. But our earlier models argue that redistribution is actually a

[1]This will be the case anyway is at least a majority of the population is required to achieve redistribution.

good thing, either because it prevents inefficient public allocation, or leads to lower rent-seeking, or permits individuals to find their true economic niche even if capital markets are imperfect.

This forms the basis of Bénabou's argument, which is simply to entertain the opposite presumption that redistribution is *beneficial* to growth. Let's try out this idea on the very simple exercise presented above. Then instead of y being reduced by redistribution, let us suppose that it is raised to ye^B, where $B > 0$. Now the threshold income Y which will support full redistribution is given by

$$\ln Y = \ln \mu + B = m + \left(\frac{\Delta^2}{2} + B\right),$$

oncea again invoking (9.1). Therefore, the degree of support is given by

$$p = \psi\left(\frac{B + \Delta^2/2}{\Delta}\right) = \psi\left(\frac{B}{\Delta} + \frac{\Delta}{2}\right).$$

But now this affects the demand for redistribution, or rather, the way it reacts to inequality. To see this, note that

$$\frac{\partial P}{\partial \Delta} = \psi'\left(\frac{B}{\Delta} + \frac{\Delta}{2}\right)\left(-\frac{B}{\Delta^2} + \frac{1}{2}\right).$$

Now notice that for small Δ, the change in the support for redistribution (as inequality increases) is actually *negative*. The reason is not hard to see. By our assumed independence of B from Δ, we see that for very small degrees of inequality practically everyone benefits from redistribution. So there is near-universal support for it. As soon as Δ climbs, some of this support goes away as people prefer to retain their own incomes with laissez faire rather than succumb to full redistribution and undergo an increase in income which will not compensate them for the original loss. [For very high levels of inequality the relationship will again turn positive, because most people are now poor and the redistribution can only benefit them all.]

The interesting thing about this model is that it yields the same inequality-growth tradeoff, but for exactly the opposite set of reasons from the Alesina-Rodrik-Persson-Tabellini exercise! In particular, assuming inequality is not too high, an increase in it actually lowers the demand for redistribution, but lower redistribution is presumed to be bad for growth, so that the same reduced-form relationship between inequality and growth is obtained.

9.5 Inequality, Conflict and Growth

This is a connection emphasized by Alesina and Perotti (1994), Benhabib and Rustichini (1996), Svensson (1994) and others. According to this view, high inequality causes political instability, protests, violent demonstrations, coups, and rioting. These factors in turn make for low growth.

The literature on this issue is largely empirical (see, e.g., Perotti (1996)). The conceptual distinctions between inequality and polarization drawn by Esteban and Ray (1994, 1999) have yet to be usefully applied to instability, though see the recent work of Montalvo and Reynal-Querol (2005) on ethnic polarization and civil war.

9.6 Inequality, Status and Growth

Suppose that individuals accumulate in part because they derive utility from overtaking other individuals in the wealth distribution. Then very flat wealth distributions will imply low densities generally, so that the status effects of investment are likely to be relatively small at the margin. Conversely, if the distribution is very egalitarian, then small changes in investment will have large status effects, leading to a higher rate of growth.

This is only half the story, though, because the collective investment decisions made by everybody will determine the distribution of wealth. In a fully specified equilibrium, the distribution of wealth and the rate of growth would be jointly determined. the iintuitive argument thus suggests that there are multiple steady states, but across those states the rate of growth and the extent of inequality are negatively correlated.

Here is a very simple model that exhibits this effect. Suppose that individual utilities are given by

$$A\frac{(y-x)^{1-\sigma}}{1-\sigma}+\hat{F}(x(1+r)),$$

where y is initial wealth, x is investment (so $y-x$ is consumption), r is the rate of return to investment, and $\hat{F}$ is the cdf of wealth in the "next generation".

We study steady states, so suppose that everybody's wealth grows at some common rate g (though with different level coefficients). Define, then, a cdf F on the *normalized* variable $y(t)/(1+g)^t$. So we may rewrite individual utility above as

$$A\frac{(y-x)^{1-\sigma}}{1-\sigma}+F\left(\frac{x(1+r)}{1+g}\right),$$

where tomrrow's wealth is now in normalized form. For this to be a steady state, the above expression must be precisely maximized at $x(1+r)=(1+g)y$, or $x=(1+g)y/(1+r)$. Writing down the first-order conditions and making this substitution, we see that

$$Ay^{-\sigma}\left(\frac{r-g}{1+r}\right)^{-\sigma}=\frac{1+r}{1+g}F'(y),$$

and this solves out for F over various supports and for various values of g. It is easy to see that the negative correlation between inequality and growth is borne out in this model.

9.7 Positive Connections Between Inequality and Growth

All the models above suggest that there should be a negative reduced-form relationship between inequality and growth, though the channels of interaction are very different. Are there any reasons for believing that inequality might be growth-enhancing in some situations? There certainly are.

9.7.1 Setup Costs and Extreme Poverty. Imagine a very poor society in which everyone has low income, too low to incur the setup cost of a potentially profitable business. Then that business activity will be thoroughly undersupplied. Instead, if wealth is distributed unequally, some of the the individuals will be able to start the business activity, which may then have high-growth implications for society as a whole. To be sure, the additional inequality thus generated may be intolerable, but that is not the issue here. Inequality will be positively associated with growth.

9.7.2 Uneven Technical Progress, Inequality and Growth. The sources of technical progress are inherently uneven. Some sector (such as software design, biotechnology, or financial services) takes off, and there is a frenetic increase in demand for individuals with these skills. The economy as a whole registers growth, of course. But this growth is highly concentrated in a relatively small number of sectors. Then, at least over the short to medium run, inequality will be positively associated with growth. Whether this effect persists in the longer run will depend on the frequency of such technological "shocks" and the speed at which subsequent intersectoral adjustments occur.

9.7.3 The Tunnel Effect. The above argument is related somewhat to the "tunnel" effect described by Hirschman and Rothschild (1973). You are driving through a two-lane tunnel, where both lanes are in the same direction and, guess what, you get caught in a serious traffic jam. No car is moving in either lane as far as you can see. You are in the left lane and your spirits are not exactly high. After a while, however, the cars in the right lane begin to move. Do you feel better or worse? It depends on how long the right lane has been moving. At least initially, you might feel that the jam has cleared further ahead and that your turn to move will come soon.

It has been the experience of several developing economies that the level of inequality in the distribution of income increases over the initial phases of development. The responses to such a rise in inequality have been varied, both across economies as well as within the same economy at different points in time, and they have ranged from an enthusiastic acceptance of the growth process that accompanied the rise in inequality to violent protests against it in the form of social and political upheaval. Such differences in the tolerance for inequality may be explained with the help of the tunnel analogy. The individual's response to an uneven improvement will depend on his beliefs as to what it implies for his own prospects. If he believes such a rise in others' fortunes is indicative of brighter prospects for himself in the foreseeable future, then he may make complementary investments, further enhancing the growth correlations described in the previous section.

Of course, if such an improvement in the welfare of others were to persist for a sustained period of time, without any improvement in one's own welfare, initial acceptance of the improved condition of others would soon give way to anger and frustration, as in the tunnel example. Moreover, increased inequality may not be tolerated at all if the perceived link between the growing fortunes of others and the individual's own welfare is weak or nonexistent. The greater the extent of segregation in society to begin with, the higher the possibility of this outcome.

9.7.4 Inequality, Savings and Growth. A last positive connection between inequality and growth may arise from the effect of inequality on the ability to save, and subsequently on growth rates. the argument is based on a nonconvexity similar to the setup costs story of a previous subsection, but the nonconvexity this time arises from minimum consumption needs. This may cause the savings function to become convex, at least over some range, with the corollary that higher inequality (in this range of incomes) may be associated with greater savings rates. The higher savings rates would presumably translate into higher growth rates via standard argument.

9.7.5 Symmetric HoldUps, Inequality and Growth. the last argument that I present here, due to Banerjee and Duflo (2003), is different from all the others in that it argues for a nonmonotonic relationship between inequality and growth. This is analogous to the assertion of a Kuznets curve, though that curve is concerned with a nonmonotonic relationship between inequality and per-capita GDP.

Return to the Alesina-Rodrik view, discussed in an earlier section. This view may be thought of as a situation in which the poor essentially hold up the rich, demanding transfers in exchange for participating in the growth process where the benefits are flowing directly to the relatively rich (and so need to be transferred via redistribution). In the Alesina-Rodrik view, these transfers create distortionary incentives at the margin, thereby slowing down investment and growth.

Banerjee and Duflo attempt to argue that this sort of holdup may well be symmetric. Just as the poor may hold up the rich, the rich may hold up the poor. Excessive *equality* may create a situation in which the traditionally rich demand compensation for pro-poor growth, thus degrading that growth via analogous distortionary effects — in just the same way as redistributive taxation did. This sort of argument suggests an inverse association between growth and *changes* in inequality in *any* direction (the compensating transfers). Banerjee and Duflo argue that such changes are more likely when inequality is either very high or very low, and passing to the reduced form, this might suggest an inverse-U relationship between inequality and growth. The warning: don't impose linear relationhips in the empirics.

9.8 Empirical Results

The arguments above all rely on different "structural assertions": inequality clearly works through a variety of channels on its way to a final impact on growth. An initial exercise, then, is to simply address the reduced-form question: is inequality positively or negatively associated with growth? or perhaps more boldly, do *initial* inequalities retard or encourage *subsequent* growth?

What is a good proxy for "initial inequality"? We would like to get a handle on inequalities in *wealth* or assets at the beginning of the time period, but data on these are notoriously hard to come by. One proxy for wealth is inequality of *income* at that time, but we must recognize that this is an imperfect proxy. Wealth inequalities at some date are, in a sense, the sum total of all income inequalities up to that date, and there is no reason why the last of these inequalities should adequately mirror the history of all its predecessors.

Another proxy for wealth inequality is the inequality in some (relatively) easy-to-observe asset, such as land. Data on land inequality are easier to obtain, although they are plagued with problems of their own. Of these problems, the most serious is the distortion created in countries that are subject to a land reform measure through the imposition of land ceilings.

In such countries land belonging to a single individual or household may be held under a variety of names, thus creating the illusion of lower inequality than there actually is. Aside from this problem, land inequality can only be a good proxy for overall inequality in wealth if agriculture is either significantly important in the economy (for the beginning of the time period under consideration), or at the very least, has been of significant importance in the recent past. Fortunately for our purposes, this is a condition that is adequately satisfied by developing countries.

Alesina and Rodrik (1994) regress per capita income growth over the period 1960–85 on a variety of independent variables, such as initial per capita income and a measure of initial human capital. [these are well-known controls: the initial income variable is a proxy for possible convergence effects, while the human capital variable is used as a proxy for the endogenous growth effects studies by Uzawa, Lucas and others (see, e.g., Barro (1991)). Indeed, as far as these variables are considered, Alesina and Rodrik use the same data as Barro, which makes for cleaner comparison with existing work. In addition, they included data on initial inequality of income and initial inequality of land.[2]

Their results indicate a substantial negative relationship between initial inequality and subsequent growth. Particularly strong is the influence of the Gini coefficient that represents the initial inequality in *land* holdings. Their results suggested that an increase in the land Gini coefficient by 1 standard deviation (which is only an increase of 0.16 in this case) would decrease subsequent economic growth by as much as 0.8 percentage points per year. [The Gini coefficient on initial *income* is only significant at the 10% level.]

These results are unaltered once we allow for structural differences across democratic and nondemocratic political systems. What is more, the democracy dummy is insignificant both by itself and when interacted with the Gini coefficient on land. It does appear that political systems play little role in this relationship, so it is unclear whether a median-voter type argument is central here (more on that below).

The Alesina-Rodrik findings are confirmed with the use of a more comprehensive data set in Deininger and Squire (1996). Initial land inequality is more significant than initial income inequality and stays that way even under several variations on the basic regression exercise (such as the use of regional dummies). The insignificance of the political system also holds up under the Deininger-Squire investigation.

The study by Persson and Tabellini (1994) finds slightly different results with somewhat different specifications. They use the income share of the top quintile as their measure of inequality. They don't use land distributions. They carry out two kinds of regressions. One is a set of historical regressions for nine now-developed countries for which the data stretches well back.[3] The other is a cross-section study, where the independent variable is the share of the middle class (see also Perotti (1996)).

. . . to be completed

---

[2]The income inequality data are taken from Jain (1975) and Fields (1989). The land distribution data are drawn from Taylor and Hudson (1972).

[3]The countries are Austria, Denmark, Finland, Germany, the Netherlands, Norway, Sweden, the UK and the US.

CHAPTER 10

Notes on Credit Markets

10.1 Introduction

We now turn to a detailed study of credit markets. As explained in my text, most developing countries have a preponderance of informal credit markets (where the lenders are not formal financial institutions). We will take the following route.

[1] We begin with a simple study of interest rate variation. We show how — depending on the presence or absence of collateral — interest rates can be either high or low. Additional fuel for this sort of variability will also be found when we study interlinkage.

[2] We then turn to a study of quantity restrictions on credit. These may actually take two forms: credit rationing or loan pushing. The former involves the borrower wanting a larger loan than what he is given at the going rate of interest. The latter involves wanting a smaller loan. Finally, note that credit rationing can take the "micro form" that we have just described or can be of a more "macro variety": some people are entirely excluded from the credit market. We will discuss various sources of credit restrictions.

[3] We study the general equilibrium of credit markets. In the analysis so far, we take as given an outside option for borrowers. To be sure, this outside option will be endogenous in a larger context. We show that the presence of information regarding past dealings is crucial in determining these outside options, and indeed, that in some cases the credit market may entirely collapse.

[4] We study interlinkage and segmentation in credit markets, using largely rural examples. The idea is to interlink contracts to get around limited collateral, or limited incentives, by offering a package. We will see that there may be several advantages to doing so. First, some distortions can be avoided. Second, incentives (such as the threat of termination) can be made to do double-duty (for, say, credit repayment as well as tenancy). Finally, in the absence of monetizable collateral (such as unskilled labor or a small piece of land), interlockers may have a serious edge over non-interlinking moneylenders.

[5] Finally, we study group lending schemes such as those promoted by the Grameen Bank. We show that under some conditions, group lending can be used as an effective tool ton transfer information from borrowers to lenders, thereby making some new contracts feasible (and profitable) for both lender and borrower.

10.2 A Simple Example

Imagine that loans are forthcoming at an interest rate of 10%, and that there are alternative projects, each requiring a startup cost of 100,000 pesos. Suppose that the projects are arrayed in terms of their rate of return, and that there are two of them, with rates of return pegged at 15% and 20%. If there is no uncertainty about the projects, and all projects pay off fully in the next time period, this is tantamount to saying that the projects will return gross revenues of 115,000, and 120,000 pesos respectively.

Observe that in this case there is a perfect coincidence of interests between the bank and the borrower. The bank wants its 10% back, and would presumably also want the borrower to take up the optimal project. Given that the borrower wants to make as much money as he can, there is no reason for him not take up the project with a 20% return. Everyone is happy.

But now let us change matters around a little bit. Suppose that the return to the second project is *uncertain*. Thus keep the second project just the same as before, but suppose that the first project pays off 230,000 pesos with probability 1/2, and nothing with probability 1/2. The *expected* return is just the same as it was before.

Now let us think about the rankings of these projects from the viewpoints of borrower and lender. To do this, we shall assume *limited liability*: if a project fails, the borrower cannot return any money to the lender: she simply declares bankruptcy. The bank would like to fund the 20% project, just as before; indeed, more than it did before. This is because the 20% project pays off its interest (and principal) for sure, while this happens only with probability 1/2 in the case of the 15% project.

What about the borrower's expected return? Assuming that she is risk-neutral like the bank is,[1] it is $120,000 - 110,000 = 10,000$ for the safe project, and it is $(1/2)[230,000 - 110,000] + (1/2)0 = 60,000$ for the risky project. Her expected return is *much* higher under a *riskier* project with a *lower* rate of return! She will therefore try to divert the loan to this project, and this will make the bank very unhappy.

What went wrong with the market here? What is wrong is that the borrower has *limited liability*. In this example, she pays up if all goes well, but if the project fails, he does not repay anything (she does not get anything either but that is not the main point). In a sense, this creates an artificial tendency for a borrower to take on too much risk: she benefits from the project if it goes well, but is cushioned on the downside. The bank would like to prevent this risk from being taken. Often it cannot.

We will come back to this scenario in more detail later. It will form the basis for theories of credit rationing.

[1] A similar example can also be constructed even if the borrower is risk-averse.

Observe that if the borrower could somehow be made to repay the loan under every contingency, we would be back to a world that's equivalent to one of perfect certainty. The bank would not care what the borrower did with the money, and the borrower would choose the project with the highest expected rate of return. But who can repay in all (or most) contingencies? They are the relatively *rich* borrowers, who can dig into their pockets to repay even if the project goes badly. We see here, then, in particularly stark form, one important reason why banks discriminate against poor borrowers.

Thus institutional credit agencies often insist on collateral before advancing a loan. For a bank which is interested in making money, this is certainly a reasonable thing to do. For poor peasants, however, this usually makes formal credit an infeasible option. It is not that they lack collateral to put up. But the collateral is often of a very specific kind. A farmer may have a small quantity of land that he is willing to mortgage. But a bank may not find this acceptable collateral, simply because the costs of selling the land in the event of a default is too high for the bank. Likewise, a landless laborer may be seeking funds to cover a sudden illness in the family, and pledge his labor as collateral: he would *work* off the loan. But no bank will accept labor as collateral.

10.3 Interest Rate Variations

Classical explanation for a high rate of interest in informal credit markets rests on the lender's risk hypothesis (Bottomley [1963]). because there is risk, as in the previous section, and because lending costs need to be covered, the lender tacks on a premium over and above the opportunity cost of lending. This is a fairly obvious point but it should also be noted that the point is often wrong! The reason is that the interest rate premium will also systematically affect borrower behavior. There may be effects working through adverse selection, moral hazard, or the strategic repayment incentive. In the next section, we begin our study of these effects with the model of Stiglitz and Weiss (1981).

10.4 Quantity Restrictions

10.4.1 Quantity Restrictions Based on Adverse Selection. We base this on Stiglitz and Weiss [1981]. Assume away the strategic default problem. But there is involuntary default coupled with limited liability. To model this, assume that borrowers differ in some parameter θ measuring riskiness. Project returns R are given by a cdf $F(R,\theta)$, and it is assumed that higher values of θ generate the same means but higher risk. That is, if $\theta > \theta'$, then

$$\int_0^\infty R dF(R,\theta) = \int_0^\infty R dF(R,\theta'), \tag{10.1}$$

while

$$\int_0^y F(R,\theta) dR \geq \int_0^y F(R,\theta') dR \tag{10.2}$$

for all y with strict inequality holding for some y.

Assume that individuals know their own type, while the bank only knows the distribution of types and cannot recognize the individual riskiness of persons.

The project has a startup cost of B. There is limited liability: each individual has access to only some fixed collateral C. Now study the repayment decision. If the rate of interest charged is r, then an individual repays if and only if

$$R-(1+r)B \geq -C, \tag{10.3}$$

where we are assuming that the bank can seize both the total return R as well as the collateral. It follows that the borrower's return (when the realized return is R and the rate of interest is r) is given by

$$\pi(R,r) \equiv \max\{R-(1+r)B, -C\}, \tag{10.4}$$

which is convex in R, while the lender's return is

$$\rho(R,r) \equiv \min\{R+C, B(1+r)\}, \tag{10.5}$$

which is concave in R.

This permits us to display the *expected* payoff to a borrower of type θ; it is

$$\hat{\pi}(\theta,r) \equiv \int_0^\infty \pi(R,r)dF(R,\theta), \tag{10.6}$$

while the expected payoff to the lender (when facing a borrower of type θ) is

$$\hat{\rho}(\theta,r) \equiv \int_0^\infty \rho(R,r)dF(R,\theta). \tag{10.7}$$

Now here is the important point. To be sure, the expected payoff to a lender — $\hat{\rho}(\theta,r)$ — is surely decreasing in θ, which is immediately intuitive. But a bit more subtle than this is the observation that the expected payoff to a *borrower* — $\hat{\pi}(\theta,r)$ — is actually *increasing* in his own riskiness θ.

To put this another way, define, for each r the threshold $\hat{\theta}(r)$ given by

$$\hat{\pi}(\hat{\theta}(r),r) = 0. \tag{10.8}$$

Then the set of types applying for a loan at the rate of interest r is precisely $\{\theta : \theta \geq \hat{\theta}(r)\}$, and $\hat{\theta}(r)$ is increasing in r. *So increasing r lowers the average quality of loan applicants.* It follows that bank profits will — in general — be nonmonotonic with respect to r.

We may build on these concepts to define — relatively informally — a notion of competitive equilibrium. To do this, let G stand for the distribution of types. First, define the profit rate at r by

$$d(r) \equiv \frac{1}{B[1-G(\hat{\theta}(r))]}\left(\int_{\hat{\theta}(r)}^\infty \hat{\rho}(\theta,r)dG(\theta)\right), \tag{10.9}$$

where $\hat{\rho}(\theta,r)$ — which is expected profits when the borrower is of type θ — is defined in (10.7), and $\hat{\theta}(r)$ — the marginal borrower when the interest rate is equal to r — is defined by (10.8).

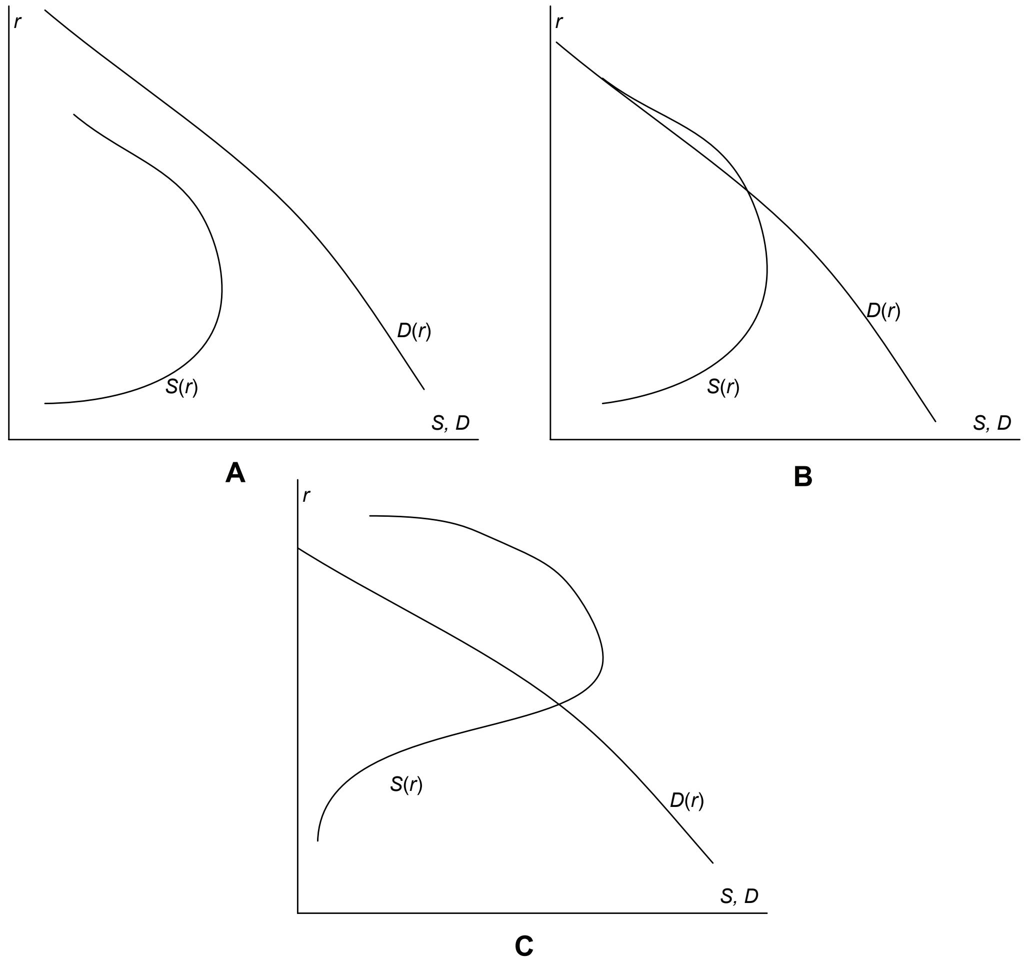


FIGURE 10.1. CREDIT RATIONING: THREE CASES

With free entry and exit of banks, the profit rate r will actually turn out to be the deposit rate that is paid to depositors (which is why we are using the d-notation).

Now for each deposit rate d, there is a supply of deposits $\hat{S}(d)$ from potential depositors. Let us assume that $\hat{S}(d)$ is an upward-sloping function. However, the supply of deposits *viewed as a function of* r will typically be nonmonotonic simply because d is generally a nonmonotonic function of r. Define $S(r) \equiv \hat{S}(d(r))$; then this is the effective supply of deposits as a function of the interest rate charged to borrowers (with the zero-profit condition already built in).

Finally, there is a demand curve for loans $D(r)$, which has the usual downward-sloping shape. Indeed, in our specific model, everyone who borrows borrows the same amount B, while the cutoff $\hat{\theta}(r)$ gets progressively higher with r. That means that the aggregate demand has the appropriate shape in r.

Define a competitive equilibrium to be a compact set of interest rates charged by banks such that no bank wants to deviate from those interest rates given optimal borrower behavior.

Proposition 10.1. Credit Rationing. *(a) Suppose that there is an interest rate r with $S(r) = D(r)$ such that $S(r') < S(r)$ for all $r' < r$. Then there is no equilibrium with credit rationing.*

(b) Suppose that for every interest rate r such that $D(r) = S(r)$, there exists $r' < r$ with $S(r') > S(r)$. Then there is credit rationing in equilibrium.

Figure 10.1 illustrates. In panel A, there is no crossing of the two curves and so condition (b) of the proposition holds trivially. In panel B, there is a crossing of the two curves but condition (b) is satisfied nevertheless. In panel C, condition (a) holds. Thus in cases A and B, there is credit rationing and in case C there is not.

It is easy to see how the proposition is proved. First assume case (a). Suppose, on the contrary, that there is an equilibrium with credit rationing. Let r^* be the lowest equilibrium rate; certainly credit is rationed there. Observe that all r' in the equilibrium set of interest rates — call it E — must be equally attractive to depositors, so that

$$d(r') = d(r^*) \equiv d^* \text{ for all } r' \in E$$

and that the *total* supply of deposits divided across *all* lending banks must be $S(r^*)$.

Fix r as given by condition (a).

First suppose that $r^* > r$. Then we have $S(r^*) < D(r^*) < D(r) = S(r)$, so we conclude that $S(r) > S(r^*)$ and consequently that $d(r) > d(r^*)$. But now this equilibrium can be broken by having a bank deviate to r from r^*. Because $d(r) > d(r^*)$, it can offer some $d' > d(r^*)$ and less than $d(r)$ and so make strictly positive profits. All the borrowers will come to him and all the depositors as well. This contradicts the assumption that we are at an equilibrium.

Therefore $r^* < r$. But then by (a), we know that $S(r) > S(r^*)$, so $d(r) > d^*$, the equilibrium deposit rate. Now consider a deviation by a bank to charging an interest rate r and a deposit rate $d \in (d^*, d(r))$. All the depositors will come to this bank, so the supply of funds is not a problem. To see if borrowers will be available, proceed as follows.

The *total* supply of loans to all banks is $S(r^*)$ (depositors get the same rate in all banks and simply divide themselves up). In particular: add up (or integrate) all depositor funds available in equilibrium at interest rates r or less: this is not any more than $S(r^*)$. On the other hand, the total demand for loans *just* by risk types above $\hat{\theta}(r)$ is $D(r) = S(r) > S(r^*)$. Thus demand (by these types alone) strictly exceeds total supply at interest rates r or less. Indeed, because banks that ration credit randomly reject applicants, we see that conditional on $\theta > \hat{\theta}(r)$, the distribution of risk types by borrowers is no different from G. These borrowers will come to our deviating bank, and the bank can make positive profit from them. We therefore have a contradiction, and the proof of part (a) is complete.

Now assume that the condition in (b) is true. If the conclusion is false, there is an equilibrium with interest rate r, such that $S(r) = D(r)$. We know that $S(r') > S(r)$ for some $r' < r$. Because $S(r') > S(r)$, we know that $d(r') > d(r)$. Therefore a bank can deviate by offering r' and some $d' > d(r)$ and less than $d(r')$ and still make profits. All the borrowers will come to him and all the depositors as well. This contradicts the assumption that we are at an equilibrium.

It still remains to prove that an equilibrium exists with credit rationing in the case where (b) holds. To this end, define $\bar{r}$ as the maximizer of $d(r)$ (there could be many but let's assume for ease of exposition that there is only one). Let r_1 denote the lowest Walrasian outcome (set it equal to infinity if there is no such outcome). If $\bar{r}$ is smaller than r_1, it is easy to check that all banks announcing $\bar{r}$ is an equilibrium, and that it involves credit rationing.

The case $\bar{r} = r_1$ is not consistent with case (b), so it remains to look at the case in which $\bar{r} > r_1$. By condition (b), there is some conditional maximizer $r^* < r_1$ of $d(r)$ (subject to the constraint that $r \le r_1$). Now define r_2 to be the smallest value of $r > r_1$ such that $d(r_2) = d(r^*)$ (such an r_2 must exist because r^* is a local but not a global maximizer). Because $S(r_2) = S(r^*) > S(r_1) = D(r_1) > D(r_2)$, there must be excess supply at r_2. Now figure out a division of banks among the two rates of interest so that all demand is soaked up. It is easy to check that this forms an equilibrium. Notice that in this equilibrium there *must* be credit rationing because there is a multiplicity of interest rates.

Some Observations.

(1) The absence of collateral, is, of course, critical to this sort of reasoning. But even if there is some collateral (not necessarily enough to cover the entire risk of default), the collateral can be used in a clever way to achieve some screening. Thus think now of a contract as a pair (r, C), where r is the rate of interest and C is the collateral that has to be put up in order to obtain loans at that rate of interest. Then the return to a borrower of type θ is

$$\begin{aligned}\hat{\pi}(\theta, r, C) &= \int_0^\infty \max\{R - (1+r)B, -C\} dF(R,\theta) \\ &= -CF((1+r)B - C, \theta) + \int_{(1+r)B-C}^\infty [R - (1+r)B] dF(R,\theta).\end{aligned}$$

Using this expression, look at the tradeoff between r and C that leaves a borrower of type θ on the same indifference curve. This is just

$$\frac{dr}{dC}|_{\hat{\pi}=\bar{\pi}} = -\frac{\partial\hat{\pi}/\partial C}{\partial\hat{\pi}/\partial r} = \frac{F((1+r)B - C, \theta)}{-B[1 - F((1+r)B - C, \theta)]}.$$

Look at the case in which $(1+r)B - C$ is a "low" return; e.g., assume that $(1+r)B-C$ is smaller than the average value of R and that all distributions are symmetric. In this case, $\frac{dr}{dC}|_{\hat{\pi}=\bar{\pi}}$ goes up as θ goes up. That is, low-risk types are less willing to accept a given interest rate increase for a given collateral reduction. This is because they fear less the loss of collateral, and expect to pay interest in more states of nature.

It follows that if banks are offering (r, C) pairs, separation will be possible. This is the criticism levied by Bester [AER 1985].

(2) Credit rationing is not inevitable in this model. For instance, if the supply of loanable funds inccreases, credit rationing disappears. Alternatively, if projects were divisible and there were constant returns to scale then we would get equal treatment: all applicants would get a loan, though not as much as they would like — back to micro credit rationing.

(3) Notice that mean and riskiness are uncorrelated in this model. If the correlation is positivem then the adverse selection effect would be weaker. Thus this form of credit rationing relies on specific distributional assumptions.

(4) This form of credit rationing may also not be empirically very plausible if there are several observationally distinguishable groups (along riskiness lines). Then at most one group would be rationed (Riley [AER 1987]).

10.4.2 Quantity Restrictions Based on Moral Hazard.

10.4.2.1 Moral Hazard in Project Choice Let's begin with a variant of the Stiglitz-Weiss model in which there is unobserved project choice (and therefore moral hazard) on the part of the borrower. Say that there are borrowers with various levels of (observable and seizable) collateral C. There is also a whole range of projects indexed by θ as before, but this time no particular θ is attached to any borrower; they all choose from this set of projects.

Assume that project returns are binary for every θ. Specifically, suppose that the return is θ with probability $p(\theta)$, and is 0 with probability $1 - p(\theta)$. The gross expected value of project θ is, then, $\theta p(\theta)$. We suppose without loss of generality that $p(\theta)$ falls with θ; if this did not happen over some range, those projects would be never be chosen anyway.

Assume, just as before, that each project requires the same loan size of B.

Now a borrower who puts down collateral C and faces a rate of interest r will choose θ to maximize

$$p(\theta)\,[\theta - B(1+r)] - [1 - p(\theta)]\,C,$$

where I assuming that the best θ exceeds $B(1+r)$, otherwise the borrower would not borrow at all.

This determines $\theta(C, r)$ as a function of C and r; assume that the maximized value is unique (it is easy to give suficient conditions for this to happen).

Observe that $\theta(C, r)$ is decreasing in C and increasing in r: more collateral induces safety; higher interest rates induce risk.

To prove this, define $Z = B(1 + r) - C$, and convince yourself that the borrower maximizes

$$p(\theta)[\theta - Z]$$

by choosing θ. Let $Z_1 > Z_2$ and θ_1 and θ_2 be the corresponding maxima. Then by the assumed uniqueness of the maximum,

$$p(\theta_1)[\theta_1 - Z_1] > p(\theta_2)[\theta_2 - Z_1],$$

while

$$p(\theta_2)[\theta_2 - Z_2] > p(\theta_1)[\theta_1 - Z_2].$$

Adding these two inequalities, we can conclude that

$$[p(\theta_1) - p(\theta_2)]\,(Z_1 - Z_2) < 0.$$

Because $p(\theta)$ is declining, we can conclude that θ is increasing in Z, which completes the proof.

Now look at a person with $C = 0$. Notice that as long as

$$\max_r p(\theta(0, r))(1 + r) < 1 + \bar{r}$$

where $\bar{r}$ is the opportunity rate of return to funds, such people will be completely rationed out, and all collaterals that are small enough will, by continuity, be completely rationed out as well. There is *no* rate of interest at which a loan will be forthcoming to them.

So moral hazard also generally leads to credit rationing. Under some conditions, it is also possible to predict what happens to the interest rate for different sizes of collateral. For instance, under competitive lending, it must be the case that

$$p(\theta(C, r))B(1 + r) + [1 - p(\theta(C, r))]C = B[1 + \bar{r}].$$

raise C; then by our previous result θ comes down so that $p(\theta)$ goes up. Obviously, $B(1+r) > C$ so in summary, the left hand side of the equality above must go up. To restore equality, the interest rate r must fall. Therefore under competition the interest rate on loans moves inversely with collateral.

With monopolistic lending, matters are not so clear, which makes sense. A monopolistic lender may well take advantage of a highly collateralized (and therefore safe) borrower to push up the interest rate. I leave it to you to examine whether r must rise with C over *some* stretch of collaterals.

10.4.2.2 Moral Hazard in Effort A *debt overhang* refers to a state of affairs in which an ongoing debt reduces the effort taken by the borrower to ensure a good output. This is the moral hazard problem. It is unclear whether one thinks of this as an instance of strategic or involuntary default — both elements are present.

In simplest terms, think of a consumption smoothing model. Let wealth be w_0 today, while today it has value w_1 with probability p and 0 otherwise. Assume that $w_1 > w_0$. If L is borrowed today and then R repaid tomorrow (only in the good state, of course), then expected utility is given by

$$u(w_0 + L) + \delta p u(w_1 - R), \tag{10.10}$$

where we have normalized $u(0)$ to 0.

The lender gets an expected value of

$$-L + \delta p R, \tag{10.11}$$

where this formulation presumes that the risk-free rate of interest is given by $\delta = 1/(1 + r)$.

Now imagine that p is exogenous and that we are trying to find a Pareto-optimal allocation. That is, maximize (10.11) while respecting some lower bound on the value in (10.10). It is easy to see that we must set

$$w_1 - R = w_0 + L, \tag{10.12}$$

so that there is perfect consumption smoothing (over those states that permit it). The exact values of R and L cannot be pinned down until we know the particular Pareto-optimal point that we are trying to characterize. For instance, if lenders operate in a competitive market with free entry, then lender's profits will be driven to zero, so that, by (10.11),

$$R = \frac{1}{\delta p} L, \tag{10.13}$$

which is just another way of saying that the rate of interest is driven down to the risk-adjusted value $\frac{1}{\delta p} - 1$.

In this case, equations (10.12) and (10.13) fully pin down the first best credit scheme.

Now suppose that the probability of success can be influenced by the deliberate application of effort. To keep things very simple, we shall suppose that there just two possible levels of effort, 0 and 1. The corresponding probabilities of success are $p(0)$ and $p(1)$ respectively, where $p(0) < p(1)$. The cost of setting effort equal to zero is 0, while the cost of setting effort equal to 1 is D.

Thus, ex post, imagine that there is an overhang of R and the borrower is deciding on effort input. He will choose the high level of effort if

$$[p(1) - p(0)]u(w_1 - R) \geq D,$$

or equivalently, if

$$R \leq \bar{R} \equiv w_1 - h\left(\frac{D}{p(1) - p(0)}\right), \tag{10.14}$$

where h is just the inverse function of u.

Now if $\bar{R}$ is higher than the first-best equilibrium, there is no problem. The first-best outcome still obtains.

On the other hand, there may be credit rationing especially if wealth is low or if D is high. In that case, one needs to compare the rationed contract with another possibility in which $e = 0$ and loans are advanced at the higher competitive rate $\frac{1}{\delta p(0)} - 1$. We would then expect that no credit rationing would occur both among the very poor and the very wealthy, while credit rationing would be observed for borrowers with intermediate levels of wealth.

We can easily extend these to cohesive notions of equilibrium (with free entry on the part of lenders). Notice that such equilibria would be second-best Pareto optimal: the government cannot do better if it must also repect the moral hazard constraint of the borrower.

However, this assumes that lenders can offer *exclusive* contracts. If exclusivity cannot be enforced, then these equilibria will break down.

To see this, notice that at the credit-rationed equilibrium, a borrower is under rationed *even relative to the higher interest rate* $\frac{1}{\delta p(0)} - 1$. The way to see this is to note that the loan repayment amount with perfect insurance under the higher interest rate must exceed its counterpart under the lower interest rate. Because this counterpart, in turn, exceeds the threshold $\bar{R}$, it follows that the credit-rationed borrower would love to take some additional loan (small enough) under the higher interest rate $\frac{1}{\delta p(0)} - 1$ (or even at a rate a little higher than that).

If a lender supplies this loan — which he will if exclusivity cannot be monitored — then this creates a negative externality on the "incumbent lender", and second-best optimality (at least relative to an ideal in which the government can enforce exclusivity) will break down.

10.4.3 Quantity Restrictions Based on the Default Problem. Consider the following model of working capital. Output simply depends on the amount of loan (L) taken via the function $F(L)$ satisfying standard properties. A borrower is engaged in a repeated relationship with a lender, but if he leaves the relationship he has an outside option given by a normalized value of v.

Consider only stationary contracts (L, R) for the moment (later, we explore the nonstationary case). Then what is the incentive constraint for the borrower? It is given by the restriction:

$$(1-\delta)F(L) + \delta v \leq F(L) - R,$$

or, rearranging:

$$\delta F(L) - R \geq \delta v. \tag{10.15}$$

Now suppose that given these outside options, the lender has all the bargaining power. Suppose, more over, that the opportunity interest rate of funds lent out is r. Then he maximizes his payoff

$$R - (1+r)L$$

by choice of (L, R), and subject to the incentive constraint (10.15).

It is easy to see, either by constructing the Lagrangean and differentiating, or drawing a diagram, that the solution involves the granting of a loan of size $\tilde{L}$, where this solves the equation

$$\delta F'(\tilde{L}) = 1 + r, \tag{10.16}$$

and then asking for a repayment $\tilde{R}$ satisfying

$$\delta F(\tilde{L}) - \tilde{R} = \delta v. \tag{10.17}$$

Now let us ask ourselves: is there credit rationing or loan pushing at this solution? Notice, to begin with, that $\tilde{L}$ is surely less than L^*, which solves $F'(L^*) = 1 + r$ and which would surely have been the optimal amount barring default considerations. So there is clearly a restruiction of the loan relative to this value. But this does not say anything *a priori* about whether there is credit rationing or loan pushing. To settle this we will have to look at the implict rate of interest on the loan, which is $\tilde{r} \equiv \frac{\tilde{R}}{\tilde{L}} - 1$.

The question is: would the borrower have preferred to borrow more or borrow less at this rate of interest $\tilde{r}$?

The answer to this question depends on the discount factor of the borrower. First note that a solution to the maximization problem exists if and only if

$$\max_{L \geq 0}[F(L) - \frac{1+r}{\delta}L] \geq v. \tag{10.18}$$

Suppose that we are *at* (or very close to) the existence borderline. In that case,

$$\tilde{R} \simeq (1+r)\tilde{L},$$

which is just another way of saying that

$$\tilde{r} \simeq r.$$

At the same time, $\tilde{L} < L^*$. So obviously for low discount factors we have *credit rationing* (relative to the implicit or explict rate of interest on the loan).

On the other hand, suppose that (10.18) holds for some threshold value of δ but that — in any case — δ is very close to one. Then the structure of the optimal contract — see (10.16) — tells us that

$$\tilde{L} \simeq L^*,$$

while

$$\tilde{R} > (1 + r)\tilde{L}.$$

[To understand this last inequality, recall from the optimality condition (10.17) that $\tilde{R} = \delta F(\tilde{L}) - \delta v > (1 + r)\tilde{L}$, where the strict inequality holds because (10.18) must hold strictly when δ is close enough to one.]

In other words, for δ sufficiently close to unity there must be *loan pushing*.

It turns out that loan pushing is a property of situations in which the creditor has some bargaining power and the borrower is sufficiently patient. For if we give the borrower all the bargaining power in this model, then loan pushing can never occur. To see this, consider the borrower's problem when he has all the bargaining power: lender profits must be zero, so the borrower maximizes

$$F(L) - (1 + r)L \tag{10.19}$$

subject to his own incentive constraint (10.15), which we may rewrite here as

$$\delta F(L) - (1 + r)L \geq \delta v. \tag{10.20}$$

Again, consider two cases. First, suppose that

$$\delta F(L^*) - (1 + r)L^* \geq \delta v. \tag{10.21}$$

In this case the borrower must choose the unrestricted maximum L^*. Then there are no quantity restrictions. Otherwise

$$\delta F(L^*) - (1 + r)L^* < \delta v. \tag{10.22}$$

But of course, it is still possible that (10.18) holds. On this assumption, the borrower will choose the loan size $L(\delta)$, which is the *maximum* loan size L such that (10.20) holds. In this case there will be credit rationing.

10.5 Information and Equilibrium in Credit Markets

The main point of this section is to argue, along with Ghosh and Ray [1996, 1998], that an informational breakdown in informal credit markets is likely in the process of development, and that this has implications for the structure of contracts, we well as for the very existence of such markets.

Consider v, the *outside option*. Where does it come from? Presumably, this is the expected (present-value) payoff conditional on the termination of an ongoing bilateral relationship.

The *reason* for this termination will typically matter. For instance, if a borrower has defaulted and this is the cause of termination, then if this matter becomes known to new lenders they may not be willing to advance a loan.[2]

Thus, v is some combination of punishment as well as the going equilibrium contract. There are obviously many ways to model this. Let us take a particularly simple version, where we assume that the lending market is competitive and borrowers have all the bargaining power. To review the usual stuff, recall that given the outside option v and the opportunity rate of interest r facing lenders, borrrowers maximize

$$F(L) - (1 + r)L \tag{10.23}$$

subject to the incentive constraint

$$F(L) - \frac{1 + r}{\delta}L \geq v, \tag{10.24}$$

where δ, it will be recalled, is the discount factor.

It is obvious, then, that in any solution L to the above problem, L must be at least as large as the value $\hat{L}$ that maximizes the LHS of the incentive constraint (10.24).

Let the optimum value of (10.23) be given by w (we emphasize that w depends on v).

Now let us try to endogenize v. Suppose that when a borrower defaults, he approaches a lender every period thereafter. The lender checks on the borrower's past and uncovers the default with probability p. In that case, the lender refuses the loan, and the borrower moves to the next period where exactly the same story repeats itself. If, on the other hand, the lender fails to uncover the default, the borrower enters into a new credit relationship with the lender. But in that case, the present value of the contract is w.

It follows that v — the expected value of the outside option — is given by

$$v = p\delta v + (1 - p)w = \frac{1 - p}{1 - \delta p}w. \tag{10.25}$$

Then we can write $v = (1 - \rho)w$, where

$$\rho \equiv \frac{p(1 - \delta)}{1 - \delta p} \tag{10.26}$$

can be viewed as the *scarring factor*. Notice that if p gets very close to one, so that a default is always recognized, then the scarring factor converges to one as well. On the other hand, for any p strictly between zero and one, the scarring factor goes to zero as δ goes to unity.

Proposition 10.2. *Define*

$$\rho^* \equiv \frac{(1 + r)(1 - \delta)\hat{L}/\delta}{F(\hat{L}) - (1 + r)\hat{L}}, \tag{10.27}$$

[2] *Why* new lenders may be unwilling to give a loan to past defaulters is actually a matter requiring some serious study. For instance, can one sustain this sort of outcome without any incomplete information at all on the part of the borrower? If there are a lot of borrowers and a lender can always find a fresh borrower, then this sort of assumption is relatively easy to maintain. Otherwise the equilibrium supporting this sort of penalty may be surprisingly convoluted — in the presence of complete information, at least.

where $\hat{L}$ is the maximizer of $F(L) - \frac{1+r}{\delta}L$.

Then a competitive equilibrium exists if and only if $\rho \geq \rho^*$.

Thus a certain minimum degree of scarring is needed before a competitive equilibrium can be guaranteed. This is intuitive. If scarring did not exist, individuals could always renege on loans and dive back into an anonymous pool.

Notice by the way that $\rho^* \in (0, 1)$.

Proof of Proposition. *Necessity.* Suppose, on the contrary, that $\rho < \rho^*$, but that a loan size L constitutes a competitive equilibrium. Then

$$\begin{aligned}
v &= (1-\rho)[F(L) - (1+r)L] \\
&= [F(L) - (1+r)L] - \rho[F(L) - (1+r)L] \\
&> [F(L) - (1+r)L] - \rho^*[F(L) - (1+r)L] \\
&= [F(L) - (1+r)L] - \frac{(1+r)(1-\delta)/\delta}{F(\hat{L})/\hat{L} - (1+r)}[F(L) - (1+r)L] \\
&\geq [F(L) - (1+r)L] - \frac{(1+r)(1-\delta)/\delta}{F(L)/L - (1+r)}[F(L) - (1+r)L] \\
&= [F(L) - (1+r)L] - \frac{(1+r)(1-\delta)L/\delta}{F(L) - (1+r)L}[F(L) - (1+r)L] \\
&= [F(L) - (1+r)L] - (1+r)(1-\delta)L/\delta \\
&= F(L) - \frac{1+r}{\delta}L,
\end{aligned}$$

where the first weak inequality in the string above uses the fact that $L \geq \hat{L}$. But now we have a contradiction to the incentive constraint (10.24).

Sufficiency. Conversely, assume that $\rho \geq \rho^*$. Define $\hat{w} \equiv F(\hat{L}) - (1+r)\hat{L}$.

For each $w \geq \hat{w}$ define $v \equiv (1-\rho)w$. First, consider $\hat{v} = (1-\rho)\hat{w}$. Notice that for $v = \hat{v}$, the maximization problem described by (10.23) and (10.24) has a feasible solution. For

$$\begin{aligned}
F(\hat{L}) - \frac{1+r}{\delta}\hat{L} &= F(\hat{L}) - (1+r)\hat{L} - \frac{(1+r)(1-\delta)}{\delta}\hat{L} \\
&= \hat{w}(1-\rho^*) \\
&\geq \hat{w}(1-\rho) = \hat{v}.
\end{aligned}$$

Now, for each $w \geq \hat{w}$, form the corresponding v (as we just did for $\hat{w}$) and keep noting the borrower's maximal w relative to that v. Call this mapping $h(w)$. Notice that there exists $\bar{w} \geq \hat{w}$ such that when we put $\bar{v} = (1-\rho)\bar{w}$,

$$F(\hat{L}) - \frac{1+r}{\delta}\hat{L} = \bar{v}.$$

Put another way, $\hat{L}$ is the *only* solution satisfying the incentive constraint when the outside option is $\bar{v}$. That means the borrower's maximum $w = h(\bar{w})$ at this point is simply $\hat{w}$. So we have found two points $\hat{w}$ and $\bar{w}$ such that $\hat{w} \le \bar{w}$ and

$$h(\hat{w}) \ge \hat{w} \text{ and } h(\bar{w}) = \hat{w} \le \bar{w}.$$

Noticing that h is continuous is between, we have proved existence. ■

Thus is the scarring rate is too low, equilibrium falls apart. What might restore equilibrium and still permit credit markets to function?

10.5.1 Credit Macro-Rationing. The first possibility is credit rationing at some macroeconomic level. To see how this fits, suppose that a past defaulter may be excluded from future loan dealings for two distinct reasons:

Targeted Exclusion. The defaulter's nasty behavior is unveiled bya new lender (which happens with probability p), and he is refused a loan. This is already present in the model above and we add nothing here.

Anonymous Exclusion. Whether or not a potential borrower has actually defaulted in the past, he may face difficulty in getting a loan. This is macro-rationing of credit, analogous to the equilibrium unempoloyment rate in Shapiro and Stiglitz [1984]. Let us denote the probability of such exclusion (in any period) by q. Notice that to build a coherent model in which $q > 0$, we really have to answer the question of why the market is not clearing — why lenders do not pounce upon available borrowers and ply them with loans. [Note, by the way, that the same issues come up when we attempt to explain $p > 0$ without taking recourse to any reputational factors.] One coherent model is given by the case in which lenders make zero expected profits, so that they are always indifferent between lending and not lending.

The main point is that anonymous exclusion may be an equilibrium-restoring device. To see this, let us calculate ρ, the effective scarring factor, when there is both targeted and anonymous exclusion. The corresponding equation is

$$\rho \equiv \frac{\pi(1-\delta)}{1-\delta\pi} \tag{10.28}$$

where π, now, is the overall probability of being excluded at any date. It is easy to see that

$$\pi = 1 - (1-p)(1-q). \tag{10.29}$$

Now notice that irrespective of the value of p, q can always adjust to guarantee that an equilibrium exists. [To be sure, the determination of q becomes an interesting question, but this is beyond the scope of the present exercise.]

10.5.2 Reputational Effects: Starting Small.

10.5.3 Nonstationary Contracts: Starting Small.

CHAPTER 11

Interlinked Contracts

We return to the question of borrower-lender matching. The data reveal that borrowers and lenders are ususally engaged in "compatible" occupations: the landlord lends to his tenant, the trader lends to farmers who supply to him, and so on. Moreover, one observes that the contacts are ofen *interlinked* with other contracts. A loan is often offered as part of a tenancy arrangement. A trader may advance loans to the farmers who supply to him. What explains these interlinked contracts?

[1] Interlinkage as Assurance. Suppose that the lender's principal occupation involves increasing returns to scale. Trading is the best example, because the trader must incur large fixed costs to transport the produce. Faced with uncertainty regarding the quantity of output he can transport to the market during harvest time, he might find it in his interest to "tie up" some farmers by giving them loans in exchange for the promise of output sales to him.

[2] Interlinkage as Enforcement. Interlinked relationships are sometimes useful for preventing *strategic* default as well. To see this, it will be useful to recall two stories with very similar features. First, think about the model of strategic default that we considered earlier. We noticed that to prevent default, the moneylender cannot drive down the borrower to his participation constraint. To avoid default, a certain surplus over the next best option had to be provided. The borrower will trades off the loss in this surplus in future dates with the one-time gain to be had from default.

Second, notice that a very similar story can be constructed for a tenancy contract (and we will do that in the last section of this course). A tenancy contract can be supplemented by intertemporal incentives which promise to renew the contract in the case of satisfactory performance, and end it otherwise. This firing threat can only be a threat, however, if the tenant earns by being fired than he does with his current (and long-term) landlord. Thus the permanent contract must be in the nature of a carrot, which can be used as a stick in the event of noncompliance with the contract.

The simple observation is that with an interlinked relationship, a single carrot can be used as two sticks, as long as deviations cannot be carried out *simultaneously* on both fronts. Combine the two scenarios. For instance, suppose that a landlord has a tenant to whom he offers a rental contract with threat of eviction in case the output is lower than some pre-understood minimum. Such a contract must, for the same reasons considered in the stories above, carry with it a certain surplus. Now it is easy to see that a loan to the tenant can be supported by an "interlinked threat": if the loan is not repaid, then the *tenancy* will be removed. The surplus in the tenancy thus serves a twin role. It assures the provision of appropriate effort in the tenancy contract, while at the same time it doubles as an incentive to repay loans. In this sense, the landlord is at a distinct advantage in advancing credit to his tenant, because he has at his disposal a pre-existing instrument of repayment. In contrast, a pure moneylender lending to the same tenant must offer additional incentives for repayment *through the credit contract itself*.

[3] Interlinkage and Nonmarketable Collateral. We've already talked about the possible nonmarketability of collateral such as labor or even land. This might create what *looks like* interlinked contracts, because it will be very natural that someone who has use for the borrower's collateral will be lending to him. But one should be careful to note that this — in itself — does not guarantee that the contract itself will be interlinked. But it does go some way towards explaining market segmentation.

[4]Interlinkages and Distortions. The example that follows is based on Gangopadhyay and Sengupta [OEP, 1987].

Suppose that a crop is grown with market price p. Farmers can produce the crop on their own and market it through a trader, but with working capital they can produce an even larger amount. One way to describe this is to assume that the famer has access to an indirect production function $F(L)$ defined on the amount borrowed, which is L. Think of $F(0)$ as the total output produced in the absence of working capital, assume that there are no other costs of production, and define $A \equiv pF(0)$. This is his *outside option*.

Now suppose that a trader has access to funds at some opportunity rate of interest i, and can lend it to the farmer. In principle he can offer a contract of the form (q, r), where q is the price at which he can pledge to buy output from the farmer and r is the rate of interest at which he offers the loan.

If the trader offers a simple contract in which only some rate of interest r is specified, then the farmer will choose L to maximize

$$F(L) - (1 + r)L,$$

yielding some return to the farmer — call it B. Note that $B > A$ if, for instance, $F'(0) > 1 + r$. It will also yield a return C to the trader, where

$$C = (r - i)L(r),$$

($L(r)$ being the optimal loan taken by the farmer at the rate of interest r).

To compare the surplus generated — which is $B+C$ — with the *maximal* total surplus that can be generated, introduce a fictional business in which the trader and the farmer are actually the same entity. Then this joint entity will choose L^* to maximize

$$F(L) - (1 + i)L.$$

Let S be the maximal value. This is the total surplus available in the system. It is easy to see that for any $r > i$, $S > B + C$. This is the distortion: the loan size $L(r)$ is distorted away from $L^* = L(i)$.

However, there is an *interlinked* contract which regains full efficiency, and the trader will have the incentive to offer it. Define (q^*, r^*) such that

$$q^*/p = (1 + r^*)/(1 + i), \tag{11.1}$$

and such that

$$\max_L q^* F(L) - (1 + r^*)L = A. \tag{11.2}$$

It is easy to see that there is a unique contract satisfying these two properties. It *must* involve $q^* < p$ and $r^* < i$. That is, the loan is advanced at a rate *below* the opportunity cost of funds for the trader, while the surplus from the loan is removed in the form of a lower output price.

This contract must maximize trader profits. By (11.1), the farmer will choose a loan size exactly equal to L^*, so this is the undistorted amount. By (11.2), the farmer is given *just* enough so that he will participate in the contract. So the trader will remove $S - A$, which is the highest that he can possibly get.

It is in this sense that interlinkage can remove distortions.

While this sort of argument has insight, it needs to be refined. For instance, there are other, noninterlinked contracts that are also efficient in this case. But these must be nonlinear contracts. For instance, consider the forcing contract under which the trader says:

Either borrow L^* at a high rate of interest R such that $pF(L^*) - (1 + R)L^* = A$, or don't borrow anything at all.

This will also implement the first best. In other words, the interlinkage result may not be robust to nonlinear contracts.

[5] The interlocker's Edge? Of course, notice that this sort of forcing contract only makes sense when there is no asymmetric information. The lender has to be able to calcuate the first-best loan. However, as Ray and Sengupta [1989] argue, there are several situations in which an interlocker has really no intrinsic advantage over a pure moneylender. To see this, look at the following cases.

Case 1. The pure moneylender cannot offer arbitrary nonlinear contracts. In this case, as we have seen in the discussion above, the interlocker has an advantage.

Case 2. The interlocker faces better terms in the market where he is active. In this case he can do better than the pure moneylender, and this pure gain may be transformed into much larger profits if there are epsilon costs of entry.

Consider the following example. Suppose that a laborer needs a loan to tide over the slack season. In the peak season he may be employed with probability p or unemployed with probability $1 - p$. The moneylender cannot do anything about the laborer's probability of employment but a large landowner who will hire labor in the peak season certainly can. This gives him the edge.

Case 3. Differential observability. Suppose a loan is given in the first period. In the second period, there is a standard principal-agent problem: the output is produced and the loan repaid. A fixed interest contract is not optimal then for the same reason that a fixed rent contract is not generally optimal. Lender can do better if he can observe output and condition repayments on these observations.

Case 4. Extend the trader-lender model so that the final output price is uncertain. If the lender is risk-neutral and the tenant is risk-averse, then the trader can provide perfect insurance by buying the crop at a fixed price. There is no need (for the trader) to observe the final price of the output. But if the moneylender cannot observe this final output, he cannot offer an equivalent contract.

Proposition 11.1. *If Cases 1–4 do not hold, then interlinking and pure moneylending yield the same returns.*

Proof. See Ray and Sengupta [1989].

CHAPTER 12

Credit Policy

12.1 Group Lending

Begin with an adverse selection model along the lines considered by Stiglitz and Weiss, but in even more simplified form (we base this on Ghatak [1999]). There are a continuum of borrowers each seeking a unit loan. The repayment on the loan (principal plus interest) is denoted by r. The opportunity cost of making the loan is denoted by i.

We consider only two possible outcomes: full repayment and no repayment. Borrowers are indexed by their probabilities of repayment, which we denote by p, and the amount that they repay conditional on success, which we denote by R. Without loss of generality we let p index borrower type and we assume that $pR(p)$ is a constant, R, for each type of borrower. Borrowers are distributed with support on the interval $[q, 1]$, where $q > 0$, and with density g on this support.

Each borrower has an outside option of u, and everybody is assumed to be risk-neutral.

If we assume that

$$R > i + u, \tag{12.1}$$

then it is first-best optimal to lend to all borrowers. But let us see how the equilibrium works out. Assume that there is a competitive market among lenders, so that we impose the zero-profit condition on them.

Suppose that the going rate of repayment is r. Then who will borrow? Only those types p such that $p \leq p(r)$, where $p(r)$ is given by the equation

$$R - p(r)r = u,$$

or

$$p(r) = \frac{R - u}{r}. \tag{12.2}$$

That is, the borrower pool will lie in the interval $(q, p(r)]$. It follows that the zero-profit equilibrium condition for the lender is

$$a(r)r - i = 0, \tag{12.3}$$

where

$$a(r) \equiv \frac{\int_q^{p(r)} sg(s)ds}{G(p(r))}. \tag{12.4}$$

where G is the cdf.

Proposition 12.1. *If (12.1) holds, then there is r such that $a(r)r - i = 0$. Moreover, there is a unique equilibrium (generically) given by the smallest of the solutions (in r) to this zero-profit equation.*

Proof. Begin by looking at $r = i$. In that case, we see that even the type $p = 1$ will benefit by taking a loan because of (12.1). It follows that $p(i) = 1$. But nevertheless, given that there is a spread of borrowers, $a(i) < 1$. Therefore

$$a(i)i - i < 0. \tag{12.5}$$

Now take r to the largest possible value which is consistent with the worst type borrower borrowing; call this value ρ. Then

$$\liminf_{r \to \rho} ra(r) = r \liminf_{r \to \rho} \frac{\int_q^{p(r)} sg(s)ds}{G(p(r))}.$$

Notice that both numerator and denominator of this expression go to zero as $r \to \rho$. So, using L'Hospital's Rule, we may conclude (taking derivatives in both numerator and denominator above) that

$$\liminf_{r \to \rho} ra(r) = \rho \liminf_{p \to q} \frac{pg(p)}{g(p)} = \rho q = R - u > i,$$

using (12.1) again. Consequently

$$\liminf_{r \to \rho} a(r)r - i > 0. \tag{12.6}$$

Noting that $a(r)r$ is continuous, we can conclude from (12.5) and (12.6) that there exists $r \in (i, \rho)$ such that $a(r)r - i = 0$. Notice that there is a compact set of such solutions, so that a minimum — call it r^* — is well-defined.

Notice that an equilibrium must belong to the set of zero-solutions. To see this, observe that generically there exists an interval of the form $(r^*, R^* + \epsilon)$ such that $a(r)r - i > 0$ in this interval.[1] Any bank can cut any other announcement to this value, get all the first pick at borrowers, and make positive profits. So the equilibrium must involve a common announcement of r^*. ■

Now we turn to the possibility of group lending. To begin with, simply look at lending to groups of two (without worrying about whether this can be an equilibrium or not). A contract is now a pair (r, c), such that r is the payment a borrower makes if he is successful

[1]Why is this generic? To see this, observe that the first zero of the function $a(r)r - i$ must be a cut from "below", because $a(i)i - i < 0$. generically, it must be a cut in which case the condition in the main text is satisfied. There is a nongeneric possibility that the first zero is a point of tangency between the $a(r)r - i$ line and the horizontal axis.

and c is an *extra* payment that he makes if he is successful but his partner is not. So the expected return to a type p when he is paired with a type p' is

$$\pi(p,p') \equiv pp'[R(p) - r] + p(1-p')[R(p) - r - c] = [R - rp] - p(1-p')c. \tag{12.7}$$

Of course, $\pi_2(p,p') > 0$: everybody wants a safer partner. But the important point is that $\pi_{12}(p,p') > 0$ as well: the safer person benefits more from a safer partner:

$$\begin{aligned} \pi_2(p,p') &= pc, \\ \pi_{12}(p,p') &= c. \end{aligned}$$

This means that in any group formation games, the matching of unqeual types can never be a core outcome, because

$$\pi(p,p) + \pi(p',p') > \pi(p,p') + \pi(p',p). \tag{12.8}$$

To see this, simply use (12.7) and grind away at the algebra — it's easy. Thus group lending gives rise to *positive assortative matching*.

Now we can study how joint liability contracts such as these may be used to overthrow the individual liability equilibrium. Assuming — using the above argument — that similar types will pair together, we may conclude that the return to a p-pair from a contract (r,c) is

$$\pi^*(p) \equiv [R - rp] - p(1-p)c. \tag{12.9}$$

Now consider a variation of (r,c) around the single liability contract $(r^*, 0)$ such that $r < r^*$, $c > 0$, and which leaves the marginal type (under the old equilibrium) — $p(r^*)$ unaffected. Writing $p^* = p(r^*)$, this means (using (12.9)) that

$$\Delta(p^*) \equiv -p^*dr - p^*(1-p^*)dc = 0.$$

From this expression, notice that if $p < p^*$, then (recalling that $dr < 0$ and $dc > 0$) $\Delta(p) < 0$ while if $p > p^*$ then $\Delta(p) > 0$ by the same token.

With this in mind, shift things a little bit so that $\Delta(p^*) > 0$ but $\Delta(p) < 0$ if $p \leq p^* - \epsilon$. Then all types p below this threshold would still go to the single liability contract. Thus a tiny change in the terms of the contract to a joint liability contract by one lender will completely change the mix of the applaicants that he gets, generating positive profits for him. It is in this way that an individual liability equilibrium is wiped out by the introduction of joint liability contracts.

12.1.1 An Informal Extension to Entrepreneurship. To illustrate this last point, let us redo and extend the two-skills example with private firms. What follows is just an informal discussion with Suppose that production requires a setup cost S that must be borne up front. Suppose that there are two wealth levels in the society, 0 and W (which exceed S). Assume, moreoever, that there are just two productivity levels, $\lambda = 0$ and $\lambda = 1$, and that the production function is given by $\lambda F(L)$, where L is the amount of labor hired. Finally, assume that the high-productivity type can cover the setup cost of production.

In this economy, production will be carried out by those individuals for whom wealth equals W *and* $\lambda = 1$. Let a be the fraction of such people. Note the following:

(i) People with low wealth who are productive will not be able to bear the setup cost, and this will be a loss to society. An increase in inequality that puts more people below the threshold needed to finance setup costs will result in greater economic loss at the aggregate level.

(ii) Assuming that the correlation between productivity and wealth is nonnegative, we see that the loss is greatest when the correlation is zero (this is when a assumes its highest value). The loss is minimal when productive types are only to be found among the wealthy (this may be true if wealth is an indicator of past productivity, and if productivity exhibits high serial correlation, both of which are strong assumptions).

(iii) The positive association between equality and greater efficiency depends critically on the *average* wealth level in the society. If the average wealth level is high, then an increase in equality puts more people above the threshold S needed to start up production. On the other hand, if the average wealth level is low, then an increase in equality may throw more individuals under the threshold, creating an even greater loss of efficiency. Thus poor societies may be caught in a double-bind: where the (intrinsic) need for redistribution is higher, the functional implications may be negative.

(iv) Denote the demand curve for labor by $L(v)$ (where the demand curve is obviously only defined for the high types and where v stands for the market wage). Then an increase in inequality (assuming that we are not in the low average wealth trap) will lower a and move the aggregate demand curve $aL(v)$ to the left. [So will a lowering of a caused by reduced correlation between productivity and wealth.] If the wage rate is endogenous, it will go down.

(v) This effect is heightened if those who cannot afford the setup costs to run a business must join the labor force — the supply curve of labor shifts to the right at the same time that the demand curve for labor shifts to the left. Effects (iv) and (v) are precisely the sort of interactive implications highlighted by authors such as Banerjee and Newman [1993]. Inequality affects macroeconomic variables, such as the level of employment or the wage rate.

(vi) Continuing this theme, we see that had we posited a slightly more complex model with a continuum of productivity types, then productivity types which are supposed to be inactive at the first best would enter production. This is because the equilibrium wage rate is low relative to the first best, so that certain productivity types which would have not covered the setup cost at the first best are now able to do so at the equilibrium wage rate. [These are wealthy people as well.]

(vii) Adopting a slight political-economy perspective, we can see why coalitions of producers might now want to oppose policies that provide better credit to the poor, even though these policies may not be overtly redistributive.