Readings in the Theory of Economic Development

Dilip Mookherjee
Boston University

Debraj Ray
New York University

November 1999

1 Introduction

The purpose of this reader is to provide an introduction to new ways of thinking about the problem of economic development. The emphasis throughout is on economic *theory*, and that too a selective sort of theory, but one we feel will define and shape the conceptual landscape of development research for some years to come.

While the readings speak for themselves, we felt the need to knit them together and thereby expound a consistent thread of research themes, in the hope that this will benefit both the interested scholar seeking an entry into development economics, as well as the teacher planning a course outline in the subject.

The volume is divided into two parts. The first part is devoted to broader conceptual themes in development economics, and economy-wide perspectives on the nature of the development "problem". The second part deals with a variety of market imperfections pervasive in the developing world, and informal institutions that arise to fill the resulting gaps. The readings selected represent a small sample from a burgeoning literature. They have been selected primarily for their theoretical completeness, accessibility and clarity. While many of these papers address policy implications, they have not been chosen primarly for their usefulness in policy design. This is based on our conviction that policy analysis must be preceded by an adequate understanding of the way in which an economy functions. Moreover, a satisfactory treatment of policy issues would have to give equal attention to instititutional problems faced by the state or activist agencies with respect to information and incentives, a topic which deserves a separate volume in its own right.

2 The Development Problem: A Macro Perspective

There is one central, simple question in the study of economic development: why are some countries developed, and others less so? In other words, what accounts for the phenomenal disparities in living standards around the world? In a related vein, will developing countries eventually catch up? If not, why not? If yes, how long will it take? Does policy have an important role in the process, or can development be left to the market?

Beginning with the work of Robert Solow [1956], the dominant paradigm for several decades in the neoclassical tradition has been one of inter-country (or inter-personal) *convergence*. This fundamental model has led to an explosion of literature — largely empirical — on the subject. The basic idea is described as follows. By the law of diminishing returns to inputs, poor capital-scarce countries should exhibit higher rates of return to capital. Consequently, *assuming* that savings and fertility rates are the same across countries, per capita incomes in poor countries should grow faster, and eventually living standards in all countries must converge. Reversing the argument: if disparities

between rich and poor countries persist over time, it must be because other things, such as savings and population growth rates, are not equal.

This hypothesis, it should be noted, does not postulate the refutable — and empirically refuted? — view that per capita incomes *do* converge. It allows for persistent differences, but then the cause of such differences are laid at the door of "fundamental" or "exogenous" variation across countries. In these respects the neoclassical approach is similar to the more classical view of Arthur Lewis [1954]. Both stress factors promoting convergence, as well as the role of key parameters in affecting the speed of development. These parameters range from variables that no one would question are endogenous in some deeper analysis, such as fertility and savings rates, through variables that some genuinely believe are exogenous to economic discourse, such as corruption, traditions, and incompetent government, and finally to less disputed (but, alas, a small number of) fundamental variables, such as geography or climate. In short, by truncating our willingness to go further, this sort of analysis has two important implications:

(1) It appears to suggest that societies are somehow basically different, for instance with regard to underlying attitudes, preferences or culture, and these account for their differences in economic development. For instance, some may believe that Confucianism or the Protestant ethic breeds economic success. Or it may be argued that the feudal *zamindari* system in West Bengal was inimical to the development of entrepreneurship. Other more naive approaches may blandly assert that overpopulation and endemic corruption are at the heart of the development problem. These views may or may not be correct in some sense, but the way in which they are phrased suggests a comfortable distinction between our notions of what is exogenous and what is endogenous, a distinction that the papers in this volume are less willing to make.

(2) This approach also generates a particular set of attitudes towards economic *policy*. By stressing the role of factors such as savings, population growth or levels of corruption that might actually be symptoms rather than causes of underdevelopment, they direct policy concerns in superficial and often wrong directions, missing deeper sources of the problem. If these factors are truly endogenous variables — the *result* of underdevelopment rather than its cause — they are unlikely to be prone to manipulation by simple-minded policy tinkering. And even if the policies are effective, such approaches can lead to misjudgment on the required duration of necessary interventions. For instance, if it is truly Hindu fatalism that keeps Indian savings rates low, a policy of encouraging savings (say, through tax breaks) will certainly have an effect on growth rates. But the tax breaks would have to be offered indefinitely to preserve their effect. In contrast, an interactive approach to the savings-growth problem may suggest permanent effects of one-time interventions, an issue we shall return to below in more detail.

The papers included in the first part of this reader take a different approach. Briefly, one might interpret this literature as implicitly arguing — or at least taking as its working

basis — that *there are no fundamental differences across people of different societies.* One can draw a line arbitrarily down through the middle of some country and endow the two parts with different initial conditions, captured perhaps by the initial state of their economic systems. The convergence hypothesis would predict that disparities in living standards in the two parts would shrink over time and eventually disappear. Alternative theories in contrast predict that the disparities would persist, absent suitable shocks or policy interventions: the very same fundamental society can be subsequently locked into very different, self-reinforcing modes of behavior. Initial historical legacies — by pinning down expectations or some other, more concrete variable, such as the distribution of assets — can precipitate one or more of these self-reinforcing outcomes. Yet, happily, the bifurcation process can be reversed: the gaps in standard of living between the two parts of the formerly same country might eventually vanish if the disparities in the economic systems are removed, perhaps by means of a one-time policy. A variety of such theories emphasizing the role of history are included in this section.

It will be useful to distinguish between two forms of *self reinforcing equilibria.* The first is simply based on a widespread coordination failure. Put another way, there are multiple equilibria (associated with differing levels of development) that are driven by alternative degrees of optimism or pessimism. The equilibria, in turn "justify" these beliefs. Beliefs that a "bad" outcome will occur do come true, in the sense that such beliefs precipitate the bad outcome. Such beliefs may exhibit enormous inertia; hence we use the term *inertial self-reinforcement* to describe this outcome.

The second form of self-reinforcement arises from the possibility that historically given initial conditions can influence current outcomes to the extent that certain historical legacies are reinforced. Here concrete historical experiences account for differences in development, rather than differences in beliefs of current agents. Given a particular historical experience, the subsequent evolution of the economy may be uniquely defined; hence this approach is not necessarily based on the pheneomenon of multiple equilibria. We use the term *historical self-reinforcement* to decribe this phenomenon.

The papers in the first part of this reader fit conveniently under these two categories.

2.1 Inertial Self-Reinforcement

Writing in 1943, Paul Rosenstein-Rodan argued that economic development could be thought of as a massive *coordination failure*, in which several investments do not occur simply because other complementary investments are not made, and these latter investments are not forthcoming simply because the former are missing. Thus one might conceive of two equilibria *under the very same fundamental conditions*, one in which active investment is taking place, with each industry's efforts motivated and justified by the expansion of other industries, and another equilibrium involving persistent stagnation,

in which the inactivity of one industry seeps into another. This serves as a potential explanation of why similar economies may behave very differently, depending on the nature of beliefs held by agents in different sectors concerning the actions of each other.

It should be obvious that for this sort of situation to arise, there must be interactive effects or *externalities* across industries. Broadly speaking, these externalities can take two forms. First, two industries could be *linked*, in the sense that the expansion of one may provoke a greater demand for the product of the other (a demand link), or facilititate the production of the second industry (a supply link). These links receive particular emphasis in the work of Albert Hirschman [1958], in an old debate on "balanced" versus "unbalanced" growth, and in the related concept of leading sectors. Unfortunately, there has been little serious formalization of this all-important theme.

The second form that externalities might take is more indirect. Industries generate income, and income generates demand for other industries. Because no individual firm internalizes these effects, a coordination failure, reinforced by pessimistic expectations, may generate a low level of economic activity. As Henry Ford wrote in his autobiography, higher wages paid by firms will actually be good for those firms in the long run, because the income returns in the form of greater demand. [But that does not mean, of course, that a *particular* firm will unilaterally raise its wages to please Henry Ford and the other firms in the economy.]

The argument, then, is that an enhanced level of economic activity generates greater national income, and the generation of national income may create additional demand to justify that activity. As Scitovsky [1954] later clarified, these externalities are 'pecuniary' rather than 'technological' that are inconsistent with the traditional Arrow-Debreu paradigm of a complete set of perfectly competitive markets. A full set of forward, contingent markets would enable these interdependencies to be mediated through the price mechanism, eliminating the possibility of multiple Pareto-ordered equilibria. So these pecuniary externalities are particularly pervasive in early stages of development, when well developed financial markets are yet to emerge. This phenemomenon is formalized in the reading by Acemoglu and Zilibotti [4] discussed further below.

These broad pecuniary externalities need not work through demand alone. Suppose that the expansion of some sectors contributes to the generation of a skilled, reliable, educated workforce. Then the supply of a labor pool of high quality will stimulate the development of other industries. Alternatively, they may stimulate the development of infrastructure such as transport, power and communication which foster the development of other industries previously stymied by the absence of such infracstructure. These are general externalities that work by facilitating production, not by raising the demand for products. Since they involve the development of a range of nontraded services, they cannot be circumvented via international trade.

Our first reading presents a simple, coherent formalization of this broader externality in a general equilibrium setting (Murphy, Shleifer and Vishny [1]). In this paper, the

authors go through a succession of models that attempt to capture "indirect externalities" across firms and industries. There are two aspects of this paper that governed its choice for inclusion in this volume, both of which merit explicit attention. First, the paper is noteworthy for its systematic theoretical exploration of the Rosenstein-Rodan hypothesis, along the second route outlined above. The pecuniary externalities arise from the lack of perfect competition inherent in the economies of scale and indivisibilities associated with infrastructural service. While the models are highly stylized and simplistic, they are consciously so, and allow us to see the coordination failure in particularly stark form.

Second, the paper shows that not all forms of income generation lead to the possibility of a coordination failure. For instance, income generation in the form of additional *profits* means higher demand for the products of other firms. Moreover, this is an externality that the profit-creating firm cares nothing about. Nevertheless, this externality cannot precipitate a coordination failure. The reason is that the externality does not distort a firm's *actions* away from the socially desirable level — as long as marginal profits are positive, output will indeed be expanded, and that is all that's needed to avoid the coordination failure. But this is not true of other forms of income, such as additional wage income. Not only does this create an externality, it also distorts actions away from the social optimum — a firm will not take actions to maximize the wage income it generates.

The Murphy-Shleifer-Vishny formulation is echoed in several other papers of independent interest. Notable among them are Rodríguez-Clare [1996] and Ciccone and Matsuyama [1996] (see also the special issue of the *Journal of Development Economics* in which these last two papers appear).

Inertial self-reinforcement can appear even in dynamic models of economic growth, as our next reading illustrates. To be sure, the concept must now be restated in terms of *growth paths*. It is the multiplicity of such paths that signal the possibility of self-reinforcement driven by pessimistic expectations. The reading we chose to illustrate this, Romer [2], is a classic in its own right. The source of the externality in Romer's model is the assumption that all the social benefits of economic knowhow cannot, in general, be expropriated by the creator(s) of that knowhow. The circle is completed by observing that one possible impact of expanded knowledge is a greater incentive to invest in knowledge-creation. This leads to the possibility of multiple equilibrium. As Romer writes, "...the private response to an aggregate increase in the stock of knowledge will be to reinforce its effects rather than to dampen them. Since the rate of growth of the stock of knowledge is increasing in the level, this kind of disturbance causes the stock of knowledge to be larger at all future dates. Thus small current or anticipated future differences can potentially have large, permanent, aggregate effects."

Simple exercises based on the Romer paper allow us to take the idea of self-reinforcement beyond one that's just based on expectational inertia. For instance, while it is true that there may be multiple equilibria from some initial stocks in the Romer model, it is also

true that that the *set* of possible equilibria also vary from stock to stock, and that this variation may persist into the indefinite future. For instance, it is possible to check (under some conditions) that "high enough" initial stocks of knowledge *must* be associated with subsequent growth (though a variety of such paths may coexist, as already discussed). But the same is not true of "low enough" initial stocks — there may be some paths emanating from these that grow, and others that stagnate. For even lower stocks, it is possible that growth paths do not exist at all. There is an interplay here between history and expectations, a theme that we now turn to.

2.2 What creates the Inertia of Expectations?

A fundamental issue associated with inertial self-reinforcement concerns the use of the word "inertia". This has to do with the fact that at any given moment of time, a *particular* equilibrium is in force, and has possibly been in force in that society in the medium- or long-run past. What causes the past to stick? How is a particular equilibrium pinned down by the force of historical inertia and what will it take to unpin it?

Unfortunately, the multiple equilibrium or coordination-game paradigm is not of much use in this regard beyond the demonstration that multiplicities may exist. In some sense, it avoids altogether any answer to the question: why is one society less developed than another, and what can be done about it? For this would require a theory of where the pessimistic beliefs originally came from, or how they could be manipulated by policy interventions. The paradigm is also at a loss for *explaining* historical inertia: repeat a multiple equilibrium story and numerous dynamic equilibria emerge, including those in which the society jumps between the bad and good equilibria in all sorts of deftly coordinated ways. We lack good economic theory that actually identifies the "stickiness" of equilibria. Rosenstein-Rodan and Hirschman were certainly concerned with this issue in a very central way though they did not make much serious progress on it.

A small literature — too small, in our opinion — exists on this topic (see, for instance, Krugman [1991], Matsuyama [1991], and Adserà and Ray, Reading [3]). There is also a corresponding smattering of literature among macroeconomists studying business-cycle models based on coordination failure (see, e.g., Chamley and Gale [1994] and Cooper [1999]). We have chosen the Adserà-Ray paper for its clear exposition of the basic problem. The paper embeds a coordination game into a real-time model of "intersectoral choice" (the choices corresponding to the actions of the static coordination game). Now agents may switch sectors (more than once, if they so desire), and their returns are added over time, by applying a discount factor. The objective of the paper is to give meaning to the notion of inertia, to the idea that historical predominance of a "sector" might impede the development of a Pareto-superior "sector". The main result is that if externalities manifest themselves with a lag (which may be arbitrarily small), and if there are no congestion costs in intersectoral migration, then initial conditions *do* pin

down equilibria — there is inertia. The paper suggests a research program in which the study of lagged externalities may be fruitful, as also the study of moving costs (a topic given more emphasis in the Krugman and Matsuyama papers).

2.3 Historical Self-Reinforcement: The Inequality-Development Interaction

Inertial self-reinforcement is a story of *multiple equilibrium*, and therefore directs our attention to the beliefs or expectations of the economic agents which shore up one or another of the equilibria. In particular, one might ask — and we do ask this above — how the formation of such expectations may be significantly conditioned by history.

This view of underdevelopment may be usefully complemented by a related, though distinct approach. This is the observation that historical legacies may actually select among *different sets* of equilibria (quite apart from the possible multiplicities in each set). In short, past history may echo persistently into the future, and not just via the determination of expectations.

We have already seen how this works in the Romer model. The same sort of historical pressure may be exerted by factors as diverse as legal structure, traditions, or group reputations (see, e.g., the review in Ray [1998] or Hoff and Stiglitz [1999]). But of all these, perhaps the darkest shadow is cast by initial inequalities in the distribution of asset ownership. With imperfect capital markets, the poor are limited in their access to credit necessary for production and investment. Hence increased inequality can exert negative effects on both levels and growth rates of per capita income. High initial inequalities may also create conditions for self-perpetuation, generating a lock-in effect with economic stagnation. The very same fundamental economy would perform differently were initial inequality to be altered.

The three readings that we have chosen emphasize various aspects of this connection between inequality and development. These are the papers by Galor and Zeira [4], Banerjee and Newman [5], and Lundqvist [6]. For related literature, see, for example, Aghion and Bolton [1997], Banerjee and Newman [1991], Bowles and Gintis [1994, 1995], Dasgupta and Ray [1986, 1987, reprinted as Reading [9]], Hoff [1994], Hoff and Lyon [1995], Legros and Newman [1996], Mookherjee [1997, reprinted as reading [13]), Piketty [1997], and Ray and Streufert [1993].

The Galor-Zeira model shows how the convergence prediction of the neoclassical growth model can be overturned by dropping the assumptions of a convex technology and perfect capital markets. With setup costs in the acquisition of certain occupations or skills, and borrowing constraints for poor agents, the initial distribution of wealth will influence the aggregate skill composition of the economy and total output, resulting in self-reinforcement. Poor families will not find it worthwhile to invest in the education of their children, locking their descendants into a poverty trap. High initial inequalities

thus tend to perpetuate themselves. Moreover, countries with a historially higher poverty rate will have a persistently lower per capita income.

Capital market imperfections alone would not generate this result. The question of the interconnection between credit market imperfections and the dynamics of inequality was earlier explored by Loury [1981] in a pioneering paper. He assumed that the technology was convex, and showed that the convergence occurs nonetheless, despite the complete absence of capital markets. The nonconvex technology in the Galor-Zeira model is therefore essential to their result concerning the persistence of inequality and underdevelopment. If, for instance, skills could be upgraded gradually and continuously across successive generations, the poverty traps would disappear in the long run, even if capital markets were entirely missing.

The simple demonstration of possible lock-ins to underdevelopment and poverty in the simple version of the Galor-Zeira model can be subjected to further criticism. Even in the presence of indivisibilities in investment, substantial stochastic perturbations might restore ergodicity, by simply permitting different wealth levels to communicate (though possibly with very small probability). For instance, in the presence of random elements reflecting luck, a poor family may tip over the required threshold and join the ranks of the prosperous, just as wealthy families may encounter a string of failures and temporarily drift into poverty.

A rebuttal to this criticism would argue that under the conditions of the Galor-Zeira model, those in poverty would remain locked there for a long period of time; the problem would be manifested instead by a low degree of wealth mobility. In part this is a signal that ergodicity (and convergence, more generally) is itself a problematic concept, a topic that we return to below. But in part, it points to a second inadequacy of these simple models, which is that they are not interactive across agents. The economy is just several copies of isolated agents (or families) running in parallel. Then inequality has no aggregate effects that are not simply trivial sums of individual effects. The model misses the interdependence in the evolution of fortunes of different families in a given society, which may strengthen the tendency towards lock-in.

More specifically, the "isolationist" model has an important implication. Historical self-reinforcement in such a model can only be the outcome of a lack of *individual* mobility (as might be the case with individual nonconvexities). If individual mobility is present (as it is, for instance in the convex case or even in the nonconvex case with sufficient stochastic variation), the final outcome is ergodic in a macroeconomic sense as well.

In contrast, the more complicated interactive models — such as those in the later part of the Galor-Zeira paper, and more fully explored in Readings [5] and [6] by Banerjee and Newman, and Lundqvist respectively — shows that there need be no contradiction between individual mobility and aggregate hysteresis. With interdependent evolution of the wealth of different families, there are numerous additional sources of historical lock-in.

The wide range of forms of interdependence can be illustrated in the following simple overlapping generations model, which also serves to explain the connection between the seemingly different models analysed in the three readings.[1] Let H be some list of *occupations*, over which a population of unit size is distributed at any date t. The date t is to be interpreted as the lifetime of the generation alive at t.

For each λ, λ', to be interpreted as occupational distributions (of successive generations), a *wage function* $w = \{w(h)\}_{h \in H}$ is defined on H. These define the incomes earned by different occupations.

A wage function w on H in turn helps determine a *cost function* $x = \{x(h)\}_{h \in H}$ is defined on H. This can be interpreted as the cost, payable in the current date, of acquiring skills necessary for occupation h for members of the next generation.

Thus given a sequence $\{\lambda_t\}_{t=0}^{\infty}$ of occupational distributions on H, we obtain a sequence $\{w_t, x_t\}_{t=0}^{\infty}$ of wage and cost functions defined on H, where each wage function w_t depends on the neighboring occupational distributions $(\lambda_t, \lambda_{t+1})$, and each cost function x_t is determined in turn by this wage function. We can then say that $\{w_t, x_t\}_{t=0}^{\infty}$ is *generated* by $\{\lambda_t\}_{t=0}^{\infty}$.

Individuals only foresee the wage-cost sequence (the actual generation of this sequence is of little import to them). They care about their own income, and those of their descendants. For an individual i (or current representative of family i) with $h_0(i)$ given, the problem is to

$$\max \sum_{t=0}^{\infty} \beta^t u(c_t) \tag{1}$$

subject to the constraints

$$y_t = w_t(h_t) \tag{2}$$

and

$$y_t = c_t + x_t(h_{t+1}) \tag{3}$$

for all t. Similar to Loury [1981], this formulation presumes that parents care about the utility (rather than just the consumption or income levels) of their descendants in a consistent fashion, so bequests or educational investments in children will be nonpaternalistic, thus removing one potential source of market imperfection. However, capital markets are missing: investments must be financed entirely from current income. The maximization problem above will result in a sequence of occupational choices made by successive generations, which we may denote by $\{h_t(i)\}_{t=0}^{\infty}$ for each family i.

Aggregate these occupational choices across families by defining, for each t, $\lambda_t(h) \equiv \mu\{i : h_t(i) = h\}$. [Of course, the distribution λ_0 is exogenously given.] This generates a sequence of occupational distributions: say that $\{\lambda_t\}_{t=0}^{\infty}$ is an *aggregate response* to $\{w_t, x_t\}_{t=0}^{\infty}$ (for given λ_0).

[1] What follows ins based on Mookherjee annd Ray [1999].

An *equilibrium* (given the historical distribution λ_0) is a sequence of succeeding occupational distributions, income and cost profiles $\{\lambda_t, w_t, x_t\}_{t=0}^{\infty}$ such that (a) $\{w_t, x_t\}_{t=0}^{\infty}$ is generated by $\{\lambda_t\}_{t=0}^{\infty}$, and (b) $\{\lambda_t\}_{t=0}^{\infty}$ is an aggregate response to $\{w_t, x_t\}_{t=0}^{\infty}$. In such an equilibrium all families have perfect foresight concerning the future evolution of the economy and the returns to different occupations; their optimal responses in turn justify their beliefs.

It is possible to embed several well-known models — as well as the readings included in this volume — within this framework. Consider the following examples:

[1] **Models of Noninteracting Agents.** H is the set of all capital stocks, $w(h)$ is *independent* of the occupational distribution, and equals some production function $f(h)$, while $x(h) = h$. This is the framework (with uncertainty added) studied in Loury [1981], under the assumption that f is a "standard" concave production function. Alternatively, one might interpret H as some discrete set of skills. This is the first model studied in the Galor-Zeira reading (they also use a simpler paternalistic "warm-glow" formulation of the bequest motive).

[2] **Entrepreneurship.** $H = \{1, 2\}$. 1 stands for worker; 2 stands for employer. $x(h)$, the cost function, is independent of the wage function: it is 0 if $h = 1$, and is S, a setup cost for entrepreneurship, if $h = 2$. To determine the wage function, suppose that there is a production function F defined on the amount of employed labor. Each entrepreneur chooses L to

$$\max F(L) - w(1)L,$$

where $w(1)$ is the wage rate for labor. In equilibrium, L is just the employment per capitalist, which is $\lambda(2)/\lambda(1)$. So $w(1)$ is given by

$$F'\left(\frac{\lambda(2)}{\lambda(1)}\right) = w(1),$$

while $w(2)$ is the resulting profit:

$$w(2) = F\left(\frac{\lambda(2)}{\lambda(1)}\right) - F'\left(\frac{\lambda(2)}{\lambda(1)}\right)\frac{\lambda(2)}{\lambda(1)}.$$

This is essentially the Banerjee and Newman [1993] model. Like Galor and Zeira, they employ a warm-glow model of bequests, and assume a fundamental indivisibility in the occupational structure (i.e., there are two discretely different occupations). The evolution of wealth and of occupational decisions is, however, fundamentally interdependent across different families. The resulting dynamics are complicated. Banerjee and Newman manage to describe the nature of this dynamic in a number of special cases, and show how distinct occupational structures and related production systems (such as the

factory system rather than independent cottage production) may evolve in the long run, depending on historical conditions.

Further developments of a related model with a divisible investment technology and random shocks were subsequently explored by Piketty [1997], who showed that the interactive nature of the wealth dynamic may still result in multiple long run steady states from different historical conditions. In this sense historical lock-in can persist even in the presence of wealth mobility at the level of individual families, and the presence of a convex technology.

[3] **Demand Effects.** H is a finite set of commodities. A person with occupation h can produce one unit of the specialized commodity h. Again, take $x(h)$ as independent of other variables.

Let $\mathbf{p} = \{p(h)\}_{h \in H}$ be a *price vector* on H. Given income y, a consumer generates a demand vector $c(\mathbf{p}, y)$ on H.

An equilibrium price vector will equate supply and demand. But the demand by occupants of occupation h is just $c(\mathbf{p}, p(h))\lambda(h)$, so that equilibrium prices must be given by the solution to the system

$$\sum_{h \in H} c(\mathbf{p}, p(h))\lambda(h) = \lambda.$$

By constant returns to scale, take $p(h) = w(h)$ for all h. A model of this kind is studied by Mani [1998].

[4] **Labor Skills.** This is the approach followed in reading [7] by Lundqvist. $H = \{1, 2\}$. 1 stands for unskilled worker; 2 stands for skilled worker. The production function $F(a_1, a_2)$ defines output produced by a_1 and a_2 units of unskilled and skilled labor respectively. This determines the wage pattern:

$$w(h) = F_h\left(a(1), a(2)\right)$$

for $h = 1, 2$. The function $x(h)$ defining the cost of training for different occupations in turn depends on the wage function: it is 0 if $h = 1$, and is $\alpha w(2)$ if $h = 2$. The idea is that to acquire skill a worker needs to be trained by α units of currently skilled workers, who need to be paid their opportunity cost of not working in the production sector and earning the wage $w(2)$. Skilled workers in the economy thus divide themselves between the production and training sectors, depending on the demand in the two sectors. Unskilled workers work only in the production sector. In equilibrium, the occupational distributions at successive dates will determine the allocation of skilled workers in the following manner. Let λ and λ' denote the occupational distributions for succeeding generations. Then notice that

$$a(1) = \lambda(1),$$

while

$$a(2) = \lambda(2) - \alpha\lambda'(2),$$

so that the wage function is ultimately related to the successive occupational distributions:

$$w(h) = F_h\left(\lambda(1), \lambda(2) - \alpha\lambda'(2)\right)$$

for $h = 1, 2$.

It is precisely the dependence of the wage and training cost functions on the occupational distribution that generates new insights. For instance, *even if there is perfect equality to start with, the subsequent evolution of inequality is inevitable.* To illustrate this, suppose all individuals in a particular generation have equal wealth. Is it possible for all of them to make the same *choices*? The answer is, in general, no. If all of them choose to leave their descendants unskilled, then the return to skilled labor will become enormously high, encouraging some fraction of the population to educate their children. Similarly, it is not possible for all parents to educate their children, if unskilled labor is also necessary in production. Thus identical agents are forced to take nonidentical actions, precisely because of the interdependence of decisions made by different families. This means, of course, that in the *next* generation some inequality must emerge. This is why all steady states in the Lundqvist model involve inequality.

This inequality, in turn, leads to a lack of efficiency. Individuals cannot simply compensate for their unequal positions by taking recourse to a credit market. In the models studied here, there isn't a credit market; or if there is one, it is imperfect. It is this imperfection that underlies the inefficiency of inequality. Individuals with low wealth may be unable to take advantage of profitable opportunities open to them, be these in the form of skill acquisition, certain occupational advantages, or remunerative investment opportunities.

Moreover, as we have already noted, inequality fundamentally affects the working of equilibrium prices — broadly defined — and in so doing it affects the dynamic fate of individuals in a way that cannot be disentangled by simple stochastic perturbations of individual outcomes. The Banerjee-Newman reading makes clear that a multiplicity of unequal distributions may all be self-reinforcing. One must be careful not to interpret these as "multiple equilibria" — *given* the historical circumstances each economy follows a unique path. And the seeming quibble is important: with multiple equilibria, economic policy must overcome expectational inertia. The policy implications here are quite different: the intervention will require, in general, a change in the distribution of asset ownership, and a one-time intervention can have effects that last into the indefinite future.

2.4 Self-Reinforcement as Slow Convergence.

There is a third view on convergence that we have referred to so far only in passing. It is a viewpoint that has not received as much attention in the literature as it deserves. Consider, again, the stochastic version of the neoclassical growth model. It predicts convergence, to be sure, but how "soon" is the long-run? Why do ergodic distributions receive so much attention, if they do not matter to the relevant future of current generations?

One answer to this question is based on considerations of tractability: we don't know how to usefully analyze "non-steady state" behavior. It should be noted that this is even more true of the new generation of models discussed above than of the neoclassical growth models, where some form of nonsteady state analysis, such as turnpike theory, has traditionally been carried out. But even in the earlier literature, the analysis raises a relatively limited set of questions, concerning eventual convergence to a steady state, and the speed of such convergence.

The Acemoglu-Zilibotti [8] reading has been chosen because it does not display this preoccupied focus on the steady state or on eventual convergence to it. The appropriate stochastic process governing economic evolution *is* ergodic in their model. They describe, instead, the arduous and difficult period that an economy can go through in the process of transition to this ergodic distribution. The example they use is the development of financial depth. An added attraction of their model is the endogenous explanation for incompleteness of the market structure of an economy, and of how this evolves in the process of development, formalizing the ideas of Scitovsky [1954] concerning the role of pecuniary externalities in the development process.

Consider a society in which there are several *potential* sectors that can be opened, though to open any one of them a scale requirement for minimum investment in that sector must be met. The Acemoglu-Zilibotti model is crafted so that the opening of an additional sector does nothing to expected returns (all sectors have the same mean return), but permits additional diversification of financial assets (because sectoral returns are only imperfectly correlated).

Now suppose that individual savings are divided among the (risky) securities floated on the basis of investment in the productive sectors, and some safe sector (such as storage). The supply of savings to the risky financial sector depends, then, on two factors: the income of the savers, which raises savings, and the allocation of savings between the safe and risky sectors.

Begin with the very short run, in which income is given. Then it is easy to see that savings will be low, so that the economy will only be able to sustain a relatively small number of open financial sectors. This limitation comes from the minimum scale requirements for each sector. Consequently, diversification is limited, so that a poor economy is capable of enormous variability in its income process.

Two factors might exacerbate this situation. First, *given* the number of open sectors, an individual investor will be interested in spreading his risky portfolio evenly over these sectors. But this evenness, in the aggregate, may prevent some sectors from opening up if different sectors have different scale requirements. No individual — indeed, no financial subcoalition — internalizes the externality created from opening a new sector. Second — and somewhat less interesting in the sense that we have already seen similar phenomena in the other readings — the depth of the risky financial sectors may be a self-reinforcing variable in the following sense. A larger number of open sectors creates greater diversification and pulls savings away from the safe asset. The greater supply of savings may then "justify" the larger number of open sectors. These pecuniary externalities can conceivably provide a rationale for interventionist policies.

What the Acemoglu-Zilibotti reading brings out with particular power, however, is the observation that poor societies may languish in their state of poverty for an inordinately large period of time before finally recieving a series of lucky draws that pulls them into the limiting distribution. This is because poor societies generate low levels of savings, and low levels of savings make for limited diversification. It is therefore possible — perhaps even likely — that poor societies will be often faced with calamitous outcomes in which very low incomes are generated, while these outcomes reinforce, in turn, the likelihood of a similar calamity being repeated in the next period. To be sure, sooner or later, there will be a string of lucky successes, which will create high incomes. The resulting high savings will then create a self-reinforcing move towards greater diversification, insulating the society from low income shocks in the future.

Notice that *ultimately*, all societies converge. But that convergence may be a long time coming, and is not half as interesting as the lingering, self-reinforcing phase that precedes the diversification-based jump to maturity. This is why we find this reading an excellent example of "self-reinforcement as slow convergence", and offer it to our readers in the hope of provoking greater interest in the subject.

3 Market Incompleteness and Informal Institutions: A Micro Perspective

We turn now to a more fine-grained analysis of market imperfections in developing countries, and the nature of informal institutions that arise to fill the resulting gaps. While many of the macroeconomic models described in the preceding section rely on some of these imperfections, particularly in the credit market, they do not adequately address where these imperfections stem from. The literature in this area has witnessed a remarkable explosion in the last two decades, drawing on advances in game theory and the economics of information. The chief contribution of this literature is to explain the source of distinctive institutional characteristics of the informal economy in developing coun-

tries extensively documented by empirical economists, sociologists and anthropologists. These include:

(a) fragmented labor and credit markets — e.g., large variations in wages and interest rates within a narrow geographic region, despite the presence of competition;

(b) persistent lack of market clearing, such as unemployment or credit rationing, despite absence of any regulations that prevent prices from adjusting flexibly;

(c) pervasiveness of long term contracts between borrowers and lenders, employers and employees, or farmers and traders;

(d) coexistence of diverse contractual forms: e.g., tenancy contracts some of which involve fixed rents and others sharecropping;

(e) unequal treatment of observationally similar workers or borrowers: e.g., dual labor markets in which some workers enter into long term contracts while others are employed to carry out similar tasks on a casual basis at substantially lower wages;

(f) interlinked transactions and exclusive dealing between specific agent pairs across multiple markets, such as bundling of credit with tenancy, employment or marketing contracts;

(g) importance of asset ownership in access to credit, tenancy or employment markets: e.g., limited access of the poor to credit owing to lack of collateral, to tenancy owing to higher risks of rent default, and to employment owing to malnutrition and absence of human capital;

(h) higher yields achieved by small farms *vis-a-vis* large farms, despite superior access of the latter to credit and technology;

(i) thinness of certain markets, such as the market for land sales, causing persistence of tenancy and unequal landownership distributions despite the superior productivity of small family farms;

(j) importance of informal cooperatives and kinship networks in determining access to essential productive inputs such as credit, insurance, technological information, water and common lands.

These phenomena are typically difficult to explain within the traditional neoclassical theory of a complete set of Arrow-Debreu markets. Neither are most of them consistent with textbook versions of monopoly or oligopoly. Accordingly many traditional scholars have inferred the irrelevance of neoclassical economics to the context of developing countries, and the need for alternative paradigms. The new economics of information

does, however, provide a cogent explanation of many of these institutional characteristics, within the context of an analytical framework grounded on the same methodology as the traditional approach. Apart from intellectual coherence, the grounding in explicit microfoundations allows assessment of counterfactuals; of the implications of these phenomena for efficiency, equality and growth; and of the effects of policy interventions. The models also serve as a useful basis for empirical testing and measurement.

3.1 Implications for Agrarian Organization

Eswaran and Kotwal [8] discuss the implications of imperfections in credit and labor markets for the nature of agrarian organization. The principal focus is on the emergence of different "classes" within a given agrarian economy, as conventionally defined by sociologists or Marxist scholars in terms of ownership of assets such as land (the "means of production"). The paper also highlights the importance of class structure for the performance of the economy (e.g., levels of farm productivity). Specifically, the model can explain some empirical regularities (such as the inverse farm size-productivity relationship), why the poor may be restricted in their productive choices, and how inequality in landholding affects class structure and agricultural productivity.

In order to focus on the principal implications, the paper assumes empirically plausible forms of credit and labor market imperfections. The credit market imperfection is represented by a constraint on the amount that can be borrowed by any given household to finance working capital needs, which depends on the amount of land owned by the household. The labor market imperfection is represented by the need to supervise hired workers. Households are assumed similar in all respects apart from their endowment of land: this assumption permits the effects of differences in landownership *per se* to be isolated. Competitive markets for fixed rent tenancy and of labor are also assumed, and farm cultivation is assumed to involve a (possibly small) fixed cost.

When there is an imperfection in only one market, credit or labor, all cultivating households select the same efficient farm size and land-labor ratio. The resulting allocation is efficient from the point of view of production. Poor households lease in land from richer households, and the market for hired labor is thin, as all farms are cultivated entirely by family labor. All households select the same labor effort. In this world, therefore, unequal landownership patterns have no productive implications, affecting only the distribution of income.

In the more realistic scenario with imperfections in both credit and labor markets, poor households are unable to borrow sufficiently to meet their rental payments, thus forcing inefficiently small scales of cultivation. Those with little or no land derive their livelihood from hiring out their labor services, and constitute a class of *workers*. Those owning more land can viably enter cultivation, but small landowners face a credit crunch in meeting their working capital needs. To alleviate these problems, they supplement

cash inflows by hiring out their labor to other farms, and avoid hiring in any non-family labor from the market. This corresponds to a second class of *worker-cultivators*. Within this class, the richer households cultivate larger farms and hire out less labor. After a further landownership threshold is reached, they cease hiring out any labor at all, devoting themselves entirely to working on their own farms. This generates the next class of *owner-cultivators*. With further increases in landownership in this class, the credit constraint is progressively eased. Labor shortage is now the overriding constraint. Families supplement family labor with hired labor, resulting in the emergence of a fourth class of *small capitalists*. Further increases in wealth will result in the household hiring in more workers and diverting its own labor away from cultivation to supervision. Eventually a wealth level is reached when the household cultivates the farm entirely on the basis of hired workers, and devotes all its time to supervision, creating the fifth class of *large capitalists*.

Different classes are characterized by differing farm sizes, technology and productivity. Poorer households cultivate smaller farms and cultivate their lands more intensively using their own labor, resulting in high levels of productivity per acre. As households become wealthier, the stringency of credit constraints is eased, while labor shortages become endemic. The result is a progressive substitution of land for labor and adoption of less intensive methods of production, resulting in lower farm yields. A more equal distribution of land will, *ceteris paribus*, increase the average productivity in the economy, and alter the class structure: there will be fewer landless workers, more family farms, and fewer capitalist farms. Eswaran and Kotwal carry out a number of illustrative numerical simulations of the model to illustrate the effects of land redistribution and credit interventions, incorporating attendant general equilibrium effects, in order to confirm these predictions. In sum, the model explains the observed differences between agrarian structure of the Latin American *latifundias* based on large inequality in landholdings and large capitalist farms, with the more equal ownership patterns in Asian agriculture dominated by small family-farms.

3.2 Labor Market Imperfections

The readings in this section focus on imperfections in the labor market, such as involuntary umemployment and dual labor markets.

3.2.1 Malnutrition and Efficiency Wages

Dasgupta and Ray [9] consider the phenomenon of nutrition-based *efficiency wages*, and its resulting implications for the labor market, a topic which goes back to earlier work by Leibenstein [1957], Prasad [1970], Mirrlees [1976], Stiglitz [1976] and Bliss and Stern [1978].

The phenomenon of involuntary unemployment poses a challenge for conventional economic theory. If wages are flexible in the downward direction, any excess supply ought to be eliminated by corresponding wage cuts. For instance, unemployed workers could undercut the going wage by offering to do the same work for less pay, an offer that should be accepted by profit-maximizing employers. What prevents such arbitrage? The efficiency wage theory provides one answer to this conundrum: if the productive efficiency of the worker depends on the wage, a wage cut will be accompanied by a drop in the worker's efficiency, thus rendering the arbitrage worthless to the employer.

Dasgupta and Ray embed this story into a general equilibrium setting, permitting analysis of the effects of land endowment patterns on unemployment and productivity. Assuming the existence of a mass of landless individuals, they show that the supply curve of productive 'tasks' is subject to a discontinuity at the threshold piece rate at which it is just feasible for the landless to enter the labor market. Below this rate, the poor are excluded because the incomes they would earn at any given level of productive effort would be insufficient to finance the nutritional intake required to sustain that level of effort. The nature and position of the supply curve depends on the landholding distribution, while the position of the demand curve depends on the total endowment of land, the product price and the nature of the technology.

If the equilibrium involves a piece rate at or below the threshold, some or all of the landless cannot participate in the labor market, though there are others very similar who do get employed, and are strictly better off. Hence if the demand for labor is low — owing either to a low endowment of land, or absence of technological innovations —- involuntary unemployment and malnutrition results, despite the operation of perfectly competitive markets. *Laissez faire* cannot then be relied on to "solve" problems of unemployment and poverty.

Indeed, competitive markets tend to amplify the effects of unequal asset ownership: at a common piece rate a wealthier worker will be able to perform more tasks and thus earn a higher wage income. Alternatively, employers may only employ workers above a certain wealth threshold, while poorer individuals are excluded from the market. This suggests that the market mechanism can be the source of increasing disparity between *ex ante* identical workers. For instance, of two landless workers starting their working careers, one may be employed and the other unemployed while young, causing the former to accumulate greater wealth and thus be favored by employers in middle age as well, while the latter drifts into destitution.

What are the policy implications? This question is addressed by Dasgupta and Ray in the sequel to the first paper, where they show that the competitive equilibrium is Pareto efficient. So there is no scope for external interventions to improve the welfare of the poor and malnourished, without making the non-poor better off. However, if the government were to redistribute land from those that are rich enough to not want to work, to the landless or to the landed working force, the effect would be to increase total

output. There are three effects of this land redistribution: a larger range of workers can now participate in the labor market; those that do participate are more productive owing to improved nutrition, and some of the "landed gentry" who were previously voluntarily unemployed may now join the labor force. The total effect is to shift the labor supply function rightward. Hence output must increase, whereas unemployment may or may not increase. Dasgupta and Ray, however, do show that there exist economies where complete equalization of landownership must eliminate involutary unemployment and malnutrition. In such cases, inequality of landownership can be said to be the real cause of these problems. Differing asset distributions can therefore serve to explain differences in agricultural output and productivity across time or across regions.

3.2.2 Incentive-based Efficiency Wages

The next reading by Eswaran and Kotwal [10] analyzes an alternative source of efficiency wages, stemming from the problem of eliciting trustworthy behavior from employees. Certain tasks in agriculture require the application of effort which is difficult to monitor. This is especially true of ploughing, planting seeds, weeding, and many household needs. Certain other tasks that are seasonal, such as harvesting, are less subject to worker moral hazard, since the product of the worker's effort (e.g., fruit plucked or crop harvested) is easily monitored. Piece rates may suffice for harvesting labor, but not for labor hired for the first category of tasks. The performance of the worker on these tasks can be ascertained by the employer only much later, perhaps at the end of the year or in future years, whereas wages have to be paid upfront. Moreover workers' performance may not be verifiable by third party contract enforcers. For either of these reasons, wages for the first category of tasks will be independent of performance levels; accordingly trust plays a significant role.

The employer will therefore seek to employ family members or other kin for these tasks. If hired hands are employed for these tasks, they have to be induced to behave in a trustworthy fashion. This is made possible by an implicit long term contract, which is renewed in future years only if the employer verified the employees' performance to have been satisfactory. To give the employee a stake in the continuation of the employment relationship, long term workers have to be treated better than short term workers hired for harvesting tasks. This implies in turn that the market for long term contracts will be characterized by involuntary unemployment: all workers will queue up for long term contracts but employers will typically be willing to employ a fraction of the entire labor force in long term contracts, the remaining workers being forced into the residual short term sector. The unemployment will not be eliminated despite wage flexibility, since wage cuts will reduce the stake of long term workers in the subsequent continuation of the relationship, inducing them to abuse their employers' trust.

This explanation for long term contracts is similar to earlier theories advanced by

Simon [1951], Klein and Leffler [1979], Shapiro [1983] and Shapiro and Stiglitz [1984]. What is of particular interest here is the explanation of coexistence of long term and short term workers, and how the composition of the work force shifts in response to demand and technology changes. The number of workers on long term contracts is determined by the demand for such workers. A principal result of the model is that under reasonable assumptions on parameter values, an increase in the supply of labor (relative to derived demand) lowers the relative wage of short-term workers, thus inducing employers to substitute long term workers by short term workers. Long term workers will therefore form a minority in overpopulated countries. On the other hand, an expansion in the scale of output (resulting from growth in market demand or technological innovations) will increase the demand for labor and the relative pay of short-term workers, inducing employers to switch to permanent workers. Hence the process of development will be accompanied by an expansion in the proportion of workers on long-term contracts. Eswaran and Kotwal cite historical accounts of such processes in Chile, Germany and Russia in the 19th century. As development continues, short-term workers will form a minority, though dualism can never entirely vanish in order to preserve effort incentives.[2]

3.3 Credit and Land

3.3.1 The Credit Market

The Eswaran and Kotwal reading **[8]** clarifies the central role of credit market imperfections in determining the nature of agrarian organization. Yet it does not explain why such imperfections arise in the first place. Empirical studies of credit markets in developing countries frequently describe numerous institutional characteristics such as the existence of credit rationing; wide variations in access to credit and in interest rates within narrow geographical regions and between formal and informal sectors; the importance of collateral and interlinkage with transactions in tenancy, employment or marketing contracts; the prevalence of long term exclusive credit relationships of borrowers with single lenders; the significance of screening and enforcement costs in the costs incurred by lenders; and the importance of ethnic and kinship networks in determining credit access.[3]

Many of these features can be explained by problems of asymmetric information and contract enforcement. The informational problems are typically either of two kinds: *adverse selection* — borrowers are better informed about their default risks — and *moral hazard*, wherein borrowers have to be induced to take adequate precautions to avoid defaults. Owing to these problems, access to credit may have to be limited. For instance, the credit rationing theory of Stiglitz and Weiss [1981] is based on adverse selection:

[2] Mukherjee and Ray [1995] temper this view by analyzing another class of long-term labor contracts that actually decline with development. These are contracts which are long-term, not because of the nature of the tasks that are entailed, but as a way of offering security against seasonal fluctuations.

[3] For an overview, see, for example, Hoff and Stiglitz [1993].

higher interest rates cause low risk borrowers to drop out of the applicant pool. Competitive equilibrium can then be consistent with credit rationing: only a fraction of loan applicants actually receive loans. Lenders do not react to the excess demand for loans by raising the interest rate because of the adverse selection effect.

This theory has been subjected to numerous criticisms (see, for instance, Bester [1985] and Riley [1987]). It can be argued (as Bester does) that the rationing stems from restrictions on the instruments available to lenders to screen borrowers of differing risks. One such instrument is collateral: with interest rates conditioned on collateral posted by borrowers, high risk borrowers would self-select into high-interest-rate-low-collateral loans, while low risk borrowers would opt for low-interest-rate-high-collateral loans. This separation would free up interest rates to clear the market and eliminate credit rationing. The adverse selection theory has also been frequently criticized as implausible in the context of traditional close-knit agrarian societies where lenders are well informed about the characteristics of potential borrowers.

Several recent studies have relied instead on the moral hazard of credit markets. As far as informal credit markets are concerned, one might distinguish between two aspects of the moral hazard problem. First — because of limited liability on the part of the borrower — loan repayment may be conditional on the realization of particular outcomes, such as the success of the project or (more generally) the solvency of the borrower. Moreover, the probability of occurrence of such outcomes may be influenced by borrower effort. This creates the well-known problem of debt overhang (see, for example, Aghion and Bolton [1997], Piketty [1997], and Mookherjee [13]).

Second, the contract itself may not be honored: the borrower may default irrespective of his ability to repay. This is the problem of *enforcement* rather than *information*. This sort of model appears (in different ways) in Eaton and Gersowitz [1981], Banerjee and Newman [1993] and Ghosh and Ray [1996, 1999]. As the research of Udry [1994] for rural credit in Northern Nigeria demonstrated, credit imperfections can arise entirely from this source, rather than any form of imperfect information.

Because there is no one paper that adequately summarizes the moral hazard view, we include a paper especially written for this reader (Ghosh, Mookherjee and Ray [11]). This paper studies the two aspects of the moral hazard problem described above. As in the adverse selection model, this approach gives rise to quantity restrictions of various kinds, and moreover, such restrictions may occur both at the level of the individual and at an aggregative, economy-wide level. The former may take the form of (individualistic) *credit rationing*, in the sense an individual is offered terms under which he would have liked to have taken a bigger loan than the one he is offered. Or it may involve *loan pushing*, in the sense that an individual wants to take a smaller loan under the terms offered to him. The paper explores both these cases.

The latter — economy-wide — mode of credit rationing is quite different. As reading 11 reveals, such rationing stems mainly from macroeconomic considerations that keep

the credit market functional in environments where information flows are limited. This sort of aggregative rationing serves to lower the outside option available to borrowers contemplating wilful default, thus maintaining incentives for an ongoing credit relationship.

3.3.2 Interlinked Credit Contracts

Another institution commonly observed in many developing countries is the interlinking of credit transactions with tenancy, employment or marketing contracts. Poor farmers (resp. workers) are frequently observed to borrow exclusively from their landlords (resp. employers). Conversely, tenancy contracts are bundled with credit: the landlord leases land, provides credit to cover working capital and consumption needs in the lean season, contributes to certain farm inputs, and receives a share of the resulting farm output in lieu of rent, besides repayment of loans. In some cases the landlord may also act as marketing agent for the farmer's share of the crop. With the farmers dependent on the same landlord for so many diverse transactions, many scholars have viewed these patterns of interlinkage as symptomatic of *semi-feudalism* or implicit bondage.[4]

The Braverman and Stiglitz reading [12] serves as a canonical example of the interlinked-contracts literature. They study sharecropping tenancy and demonstrate why such contracts may be linked to credit relationships. In the Braverman-Stiglitz model (as in many others), sharecropping is viewed as the outcome of two countervailing forces: the need to attentuate the risks faced by a risk-averse tenant, and the need to provide incentives to the same tenant. As extreme examples: fixed rent contracts would impose too much risk on the tenant, while fixed wage contracts would provide him with insufficient incentive to work hard. Accordingly, an optimal tenancy contract must strike an appropriate balance between these two considerations. Sharecropping may be viewed as one such contract.

Once this view is adopted, it is hard to escape the observation that a tenant's effort incentive will *also* depend on the loans that fall due at the time of harvest. The direction this pushes effort is *a priori* unclear. If there is limited liability, a higher debt burden will reduce effort because of the debt overhang. On the other hand, if debts are repaid in all states of nature, the extra burden increases the *marginal* (utility) gain from higher income. This will encourage the farmer to work harder, with consequent benefits that accrue partly to the landlord.

Interlinkage arises as a way of internalizing this externality between the credit contract and the tenancy contract. For instance, in the second scenario described above, the landlord can provide the tenant with subsidized credit, encouraging the latter to borrow more, and subsequently work harder to make the larger repayments. In fact, Braverman

[4]For instance, Bhaduri [1973] suggested that such forms of interlinkage may result in technological stagnation: landlords may prevent the adoption of new innovations by the farmer for fear of reducing the latter's credit needs, and hence the landlord's income from usury.

and Stiglitz show that interlinkage is a property of *all* Pareto efficient contracts; it allows an outward shift of the entire utility possibility frontier. In particular, it has no necessary connection with the exercise of monopoly power by the landlord: one would expect it to observe it just as much in contexts where markets are competitive and tenants have considerable bargaining power. Likewise, in the first scenario highlighted above, landlords would wish to restrict loans to their tenants, or at least impose credit exclusivity as a precondition for a tenancy contract.

3.3.3 Land Markets

A related issue associated with sharecropping tenancy is its alleged inefficiency, an observation that goes back to classical economists such as Adam Smith and Alfred Marshall. Since the tenant receives only a fraction of the output produced, the sharecropping arrangement operates in a manner similar to that of a distortionary tax which dulls his incentive to work hard. In contrast fixed rent tenancy or owner cultivation awards the entire residual claimancy to the farmer, encouraging effort. This observation has received substantial empirical support in careful econometric analyses (see e.g., Shaban [1987] or Laffont and Matoussi [1995]). It has been at the basis of recommendations for land reform, since the redistribution of the land from landlord to the tenant would convert sharecropping to owner cultivation, with resulting benefits in both higher productivity and reduced inequality.

Yet, as the reading by Mookherjee [13] discusses, this case for land redistribution leaves a number of questions unanswered. First, if sharecropping arose in the first place as a mechanism for tenants to receive partial insurance against crop risk, the same need for insurance would be faced by the farmers once they received possession of the land. Presumably, credit relationships would then replace tenancy contracts as the suppliers of such insurance. But the possibility of default, limited liability and the consequent debt overhang would dampen effort (just as sharecropping did). It is not *a priori* clear why this effort incentive problem would be any less serious than under sharecropping.

Moreover, there is the task of explaining why landlords do not *sell* the land to their tenants if the subsequent transition to owner-cultivation results in superior productivity. Land markets would then automatically "solve" the inefficiency inherent in sharecropping by causing the institution to disappear. What prevents these markets from performing this role, and why can state intervention lead to superior outcomes? This sort of question provokes a related exercise: to identify the precise welfare effects of (nonmarket) land redistribution. If farmers have higher effort incentives, this may be the expense of higher risks borne by them. Are they necessarily better off?

The answer to these questions lies in the nature of informational rents inherent in moral hazard with limited liability. As explained in the simple example above, these rents arise from the need to provide adequate effort incentives. A transfer of ownership

of land to tenants will augment their bargaining power. This causes a larger weight to be accorded to the interests of the farmer *vis-a-vis* their contracting partners, which increases informational rents and effort incentives. This explains the superior productivity of owner-cultivation, which leads to a welfare improvement owing to the more successful internalization of the externality between landlord and farmer of the latter's informational rents. Despite this potential welfare improvement, the paper shows that landlords would never voluntarily sell their land to their tenants, as the latter would lack the credit necessary to finance the purchase.

This sort of argument is symptomatic of a wider problem concerning the applicability of the Coase Theorem, which asserts the irrelevance of the allocation of ownership rights for efficient incentives. The Coase Theorem does not apply because of the presence of credit constraints that prevent the use of side transfers to internalize the relevant externalities. For the same reason, land markets would not serve to "solve" the problem; coercive state intervention is necessary to empower farmers at the expense of landlords, which the latter would tend to politically resist. Coercive land redistribution programs cannot generate a Pareto improvement. But they can lead to an increase total surplus or efficiency in the Kaldor-Hicks sense: the gainers could potentially compensate the losers and still remain better off.

3.4 Cooperatives and Informal Institutions

We have seen that the formal sector may leave large gaps in the efficiency of resource allocation. It is natural to expect that the local informal sector will step in to fill in these gaps. This is particularly true when the failure of the formal sector stem from its paucity of information about the characteristics and activities of local agents, and its inability to enforce contractual agreements. In contrast, civil society and the rich web of social networks that such society comprises is frequently better informed and more able to enforce obligations through social sanctions. Local villagers have access to a large amount of information regarding each other from informal sources. This information base may be "soft" and "unverifiable" and sometimes unreliable — being based often on local gossip rather than on hard facts. But it is typically far richer than can be represented by the statistics and formal documentation that external agents in the formal sector will have recourse to. Local villagers will also interact on a frequent basis in a variety of social and economic dimensions, allowing deviant behavior to be punished via social sanctions of diverse forms and degree.

Individual behavior can be influenced by community pressure in traditional societies in extreme ways — going as far as inducing deviants to the point of suicide and self-immolation. The pressures that can be brought upon citizens by a distant state often pales in comparison. Thus repayments on even small informal loans and leases will be scrupulously observed by the very same people that routinely default on larger payments

due to the government or commercial banks. It is no suprise then that the informal economy will be better placed to discipline opportunistic behavior, especially in credit and insurance transactions, where problems of trust play such an important role.

3.4.1 Mutual Insurance

A leading example of these general considerations is the provision of insurance. Anthropologists and sociologists have frequently described and extolled the ability of traditional communities to provide mutual credit and insurance. Coate and Ravallion [14] cite relevant evidence in the context of mutual insurance, where villagers help each other in times of sickness or crop failures. Yet, as they note, it is important to avoid an overromanticized view of the "moral economy": the anthropological evidence and several case studies (see, for instance, Attwood and Baviskar [1993]) describe numerous instances when this form of mutual insurance tends to break down. Understanding the relative strengths and weaknesses of informal insurance is important not only for its own sake, but also as a guide to circumstances where external intervention by the formal sector may be warranted.

Coate and Ravallion develop a theory of self-enforcing mutual insurance, where a pair of *ex ante* identical agents face a stationary but random income process. Insurance from the formal sector is assumed absent, as also the possibility of self-insurance via savings (e.g., because of costs of storage). They focus on mutual insurance arrangements, which prescribes transfers from the individual obtaining a higher income realization at any given date to the other individual who is poorer at that date. The chief simplifying assumption is that these transfers are conditioned only on the *current* income realizations, rather than the past history of their incomes or transfers. Extensions of the model to incorporate credit-like elements via history-dependence have since been developed in the work of Fafchamps [1996], Kletzer and Wright [1995], Ligon, Thomas and Worrall [1998] and others.

In the absence of any enforcement problem, complete sharing is the optimal (symmetric) insurance arrangement, in which the individual with the better income draw gives away half the difference between the two income realizations, resulting in perfect *ex post* equality in consumption. However, when the income differences are large, equal sharing may not be viable as the richer individual may balk at the thought of making large donations, even if the deviation may result in the breakdown of the sharing norm in the future. In other words, equal sharing may not be a self-enforcing norm.

Drawing on some results in the theory of repeated games, Coate and Ravallion are able to provide a simple and elegant description of all self-enforcing sharing norms, and therefater describe the characteristics of the "best" self-enforcing norm. The results provide insight into the strengths and failures of informal insurance. For instance, it can provide reasonable insurance with respect to moderate idiosyncratic income shocks, but

limited insurance against more extreme shocks. The extent of insurance also tends to fall when times are collectively bad for both parties, or when they face risks that are less idiosyncratic. Hence the scope for mutual insurance is limited in times of drought or famine, and in regions more prone to common weather or market shocks.

Numerical simulations also reveal that the insurance arrangement may be unstable with respect to small perturbations in the parameters of the economy. For instance, a substantive insurance arrangement may entirely disappear with small changes in discount rates, owing to a "bootstrapping" character of such schemes. As the scope for such transfers are narrowed by small parametric changes that cause the enforcement constraints to bind more sharply, the mutual benefit to participants in future continuation of the scheme is lessened. This further reduces the level of enforceable transfers, leading to a cumulative unraveling of the entire mechanism.

3.4.2 Mutual Credit

Credit represents a different context where informal cooperatives play an important role, in the form of *rotating savings and credit associations* (ROSCAs). Such cooperatives are pervasive throughout the developing world; indeed, many modern commerical banks in the developed world derive their historical origins from them. This is the topic of the next reading by Besley, Coate and Loury [15]. In a ROSCA, a group of individuals get together at different dates and jointly contribute money to a common pool; at any date the entire pool is allocated to one of the members. The winner at any date may be picked randomly, or may be decided on the basis of bidding. Besley, Coate and Loury develop a theory of the role and nature of ROSCAs based on the existence of indivisibilities in desired purchases that make it useful to pool savings. For instance, suppose that a bicycle costs 100, and each member can only save 10 every month. In isolation each of them would have to wait ten months to buy the bicycle from their own savings. If ten of them form a ROSCA and contribute 10 every month to the pool, one of them can purchase the first month, another the second month, and so on. This enables earlier purchases for all but one member, resulting in a Pareto improvement. To be sure, this is not necessarily an *optimal* ROSCA, which might prescribe lower individual savings rates as a result of the greater pooling now made possible.

Less obvious is the question of how the ROSCA should be constituted. Besley-Coate-Loury compare the performance of random and bidding ROSCAs, and establish a general result: random ROSCAs always outperform bidding ROSCAs if the individuals are all alike *ex ante*, but the opposite will be true if they are sufficiently heterogeneous (e.g., with respect to time preference). The simpler version of their theory does not incorporate enforcement constraints, arising from possible opportunistic behavior by early winners who stop contributing thereafter. Clearly social sanctions play an important role in reining in such forms of opportunism. They show how enforcement constraints may also

explain why the size of the group may be limited, and why random ROSCAs rather than bidding ROSCAs are so commonly observed.

3.4.3 Microcredit and Joint Liability Loans

Microcredit programs represent yet another context where the advantages of peer monitoring and social sanctions within close-knit traditional societies play an important role. The success of group loans made by the Grameen Bank in Bangladesh to the rural poor have become well-known throughout the world. Ghatak and Guinnane [16] survey the experience with these programs as well as the theories explaining their success. As they argue, these practices have their origins almost a century ago in the practices of German credit cooperatives, and have been used in more recent times in numerous developing countries. Joint liability loans have typically succeeded in allocating credit in the form of small loans to poor borrowers lacking the assets to post collateral required typically by commercial and government banks. They rely instead on a form of social collateral: different members of the group are jointly liable for each other's loans.

Different authors have stressed a number of possible advantages of joint liability. Different members of a given social group typically know more about each other's creditworthiness than does the bank. By encouraging groups to self-form and making members jointly liable for each other's loans, safe borrowers are provided incentives to form groups with others like themselves, thus weeding out high risk borrowers. Joint liability can also relax the implications of limits to individual liability, lowering default risk faced by lenders. Moreover, to the extent that a default imposes costs on others in the same group that the borrower has social relationships with, it results in the borrower internalizing the social costs of default, alleviating moral hazard problems. These effects are illustrated with a sequence of simple examples in the theoretical section of Ghatak and Guinnane's paper.

They subsequently discuss problems experienced by attempts to transplant joint liability schemes to different countries and social contexts. In some countries, the problem is that the group sizes are too large for the scheme to work effectively. In urban settings and in more developed countries, the social bonds between participants are not particularly strong, with individualistic behavior more prominent in local culture. In yet other settings (such as 19th century Ireland) where community involvement played an important part, the schemes failed because members were loathe to impose social sanctions on defaulters. The schemes also require the threat of sanctions applied to the group to be credible and actually applied by lenders to all group members following the default of a single member. In addition, they require the prospect of being cut off from the program to exercise a strong punitive role, which would not be the case if they have access to alternate credit sources.

3.4.4 Social Learning

Learning from others about the effectiveness and appropriate use of new technologies represents another avenue by which social structure affects the process of development. Foster and Rosenzweig [17] present a model of learning about new cultivation practices from past experience. Farmers learn about appropriate fertilizer application levels in connection with new high yielding varieties (HYV) of seeds, both from their own past experience, and the experience of their neighbors. The learning process is modelled as the outcome of an optimal Bayesian updating of priors over the correct fertilizer level, in the light of past experience. It is assumed that the farmer learns the correct fertilizer dosage appropriate to any particular parcel of land at any given date from observing *ex post* the experienced yield on that parcel. This is a noisy signal of the optimal fertilizer level for any given parcel at any future date. Accordingly, one's own past experience with plantings of HYV seeds, as well as those of one's neighbors, allow the farmer to obtain a more precise estimate of the optimal fertilizer application over time. The model allows a precise expression of these experience effects: in general they are positive and diminishing over time, and there is a constant ratio between the extent of learning from one's own experience and neighbors' experience over time.

In this setting, farmers are presumed to strategically select the number of parcels in which they will plant an HYV seed at any given date, as a function of the experience they have accumulated so far, and the corresponding planting strategies of their neighbors. The resulting dynamic equilibrium is difficult to characterize in closed form. But the model does indicate the complex interdependence among farmers that results from the presence of learning spillovers. Increased levels of adoption by one's neighbors allows a farmer to learn more about the HYV cultivation process, raising profitability of the new technology; this induces greater adoption. In this sense the social learning process encourages the diffusion of the technology. On the other hand the higher level of experience can cause the returns to additional future experience to diminish; this free-riding effect implies that the increased adoption by neighbors will reduce own-adoption incentives. The overall effect of the learning spillovers on the average speed of adoption will then depend on the trade-off between the profitability and free-riding effects, which cannot be determined at the level of theory.

Foster and Rosenzweig thereafter use data from HYV adoption decisions of farmers in north India in the late 1960s during the Green Revolution to empirically identify the relevant learning effects, as well as to test some of the restrictions imposed by the theory (e.g., the constancy of the own-learning effect relative to the social-learning effect over time). These restrictions are upheld in the data, providing support to the theory. Both forms of learning turn out to have significantly affected profitability of the HYV seeds. They also have roughly similar positive effects on adoption decisions, though the adoption effect of social learning is not statistically significant. Hence the social learning

process had a positive but weak effect on the adoption process on average. Simulations of adoption dynamics predicted by the model yield the familiar S-shaped diffusion curve, the exact position of which for any given farmer depends on the nature of interaction with neighbors. In particular, Foster and Rosenzweig find evidence for a significant free-riding effect among farmers with respect to adoption decisions. This suggests a possible role of enhanced coordination among farmers in the village, as well as public subsidies that allow the externalities associated with learning spillovers to be internalized. In common with the previous reading by Ghatak and Guinnane, this paper exemplifies how a proper understanding of the functioning of the informal economy can enable the design of enlightened interventionist policies that fruitfully complement the relative strengths of formal and informal sectors of the economy.

References

Aghion, P. and P. Bolton (1997). "A Theory of Trickle-Down Growth and Development," *Review of Economic Studies* **64**, 151–172.

Attwood, D. and B. Baviskar (1993). *Raising Cane: The Political Economy of Sugar in Western India*. New Delhi: Oxford University Press.

Banerjee, A. and A. Newman (1991). "Risk-Bearing and the Theory of Income Distribution," *Review of Economic Studies* **58**, 211–235.

Bester, H. (1985), "Screening vs. Rationing in Credit Markets with Imperfect Information," *American Economic Review* **75**, 850–855.

Bhaduri, A. (1973), "A Study in Agrarian Backwardness Under SemiFeudalism," *Economic Journal* **83**, 120–137.

Bliss, C. and N. Stern (1978). "Productivity, Wages and Nutrition, Part I: The Theory," *Journal of Development Economics* **5**, 331–362.

Bowles, S. and H. Gintis (1994). "Credit Market Imperfections and the Incidence of Worker Owned Firms," *Metroeconomica* **45**, 209–223.

Bowles, S. and H. Gintis (1995), "The Distribution of Wealth and the Assignment of Control Rights," mimeo., Department of Economics, University of Massachusetts, Amherst.

Chamley, C. and D. Gale (1994). "Information Revelation and Strategic Delay in a Model of Investment," *Econometrica* **62**, 1065–1085.

Ciccone, A. and K. Matsuyama (1996). "Start-up Costs and Pecuniary Externalities as Barriers to Economic Development," *Journal of Development Economics* **49**, 33–59.

Cooper, R. (1999). *Coordination Games: Complementarities and Macroeconomics*. Cambridge University Press.

Eaton, J. and M. Gersovitz (1981). "Debt with Potential Repudiation: Theoretical and Empirical Analysis," *Review of Economic Studies* **48**, 289–309.

Fafchamps, M. (1996). "Risk Sharing, Quasi-Credit, and the Enforcement of Informal Contracts." Mimeograph, Department of Economics, Stanford University.

Ghosh, P. and D. Ray (1996). "Cooperation in Community Interaction without Information Flows," *Review of Economic Studies* **63**, 491–519.

Ghosh, P. and D. Ray (1999). "Information and Repeated Interaction: Application to Informal Credit Markets." Mimeograph, Department of Economics, Texas A&M University.

Hirschman, A. O. (1958). *The Strategy of Economic Development*. New Haven, CT: Yale University Press.

Hoff, K. (1994). "The Second Theorem of the Second Best," *Journal of Public Economics* **45**, 223–242.

Hoff, K. and A. Lyon (1995), "Non-Leaky Buckets: Optimal Redistributive Taxation and Agency Costs," *Journal of Public Economics* **58**, 365–390.

Hoff, K. and J. Stiglitz (1993). "Imperfect Information in Rural Credit Markets: Puzzles and Policy Perspectives, " in K. Hoff, A. Braverman, and J. Stiglitz (eds.), *The Economics of Rural Organization: Theory, Practice and Policy*. London: Oxford University Press (for the World Bank).

Hoff, K. and J. Stiglitz (1999). "Modern Economic Theory and Development," in G. Meier and J. Stiglitz (eds), *Pioneers in Development*, Oxford University Press, forthcoming.

Klein, B. and K. Leffler (1981), "The Role of Market Forces in Assuring Contractual Performance," *Journal of Political Economy* **89**, 615–641.

Kletzer, K. and B. Wright (1995). "Sovereign Debt as Intertemporal Barter." Mimeograph, University of California, Santa Cruz.

Krugman, P. (1991). "History versus Expectations," *Quarterly Journal of Economics* **106**, 651–667.

Laffont, J-J. and M. Matoussi (1995). "Moral Hazard, Financial Constraints and Sharecropping in El Ouja," *Review of Economic Studies* **62**, 381–400.

Legros, P. and A. Newman (1996). "Wealth Effects, Distribution and the Theory of Organization," *Journal of Economic Theory* **70**, 312–341.

Leibenstein, H. (1957). *Economic Backwardness and Economic Growth*. New York: Wiley.

Lewis, W. A. (1954). "Economic Development with Unlimited Supplies of Labor," *The Manchester School of Economic and Social Studies* **22**, 139–191.

Ligon, E., Thomas, J. and T. Worrall (1998). "Informal Insurance Arrangements in Village Economies," mimeograph.

Loury, G. C. (1981). "Intergenerational Transfers and the Distribution of Earnings," *Econometrica* **49**, 843–867.

Mani, A. (1998). "Income Distribution and the Demand Constraint." Mimeograph, Department of Economics, Vanderbilt University.

Matsuyama, K. (1991). "Increasing Returns, Industrialization, and Indeterminacy of Equilibrium," *Quarterly Journal of Economics* **106**, 617–50.

Mirrlees, J. (1976). "A Pure Theory of Underdeveloped Economies," in L. Reynolds (ed.), *Agriculture in Development Theory*. New Haven, CT: Yale University Press.

Mookherjee, D. and D. Ray (1999). "Persistent Inequality and Endogenous Investment Thresholds," mimeograph.

Mukherjee, A. and D. Ray (1995). "Labor Tying," *Journal of Development Economics* **47**, 207–239.

Piketty, T. (1997), "The Dynamics of the Wealth Distribution and the Interest Rate with Credit Rationing," *Review of Economic Studies* **64**, 173–189.

Prasad, P. (1970). *Growth with Full Employment*. Bombay: Allied Publishers.

Ray, D. (1998). *Development Economics*. Princeton: Princeton University Press.

Ray, D. and P. Streufert (1993). "Dynamic Equilibria with Unemployment due to Undernourishment," *Economic Theory* **3**, 61–85.

Riley, J. (1987), "Credit Rationing: A Further Remark," *American Economic Review* **77**, 224–227.

Rodríguez-Clare, A. (1996). "The Division of Labor and Economic Development," *Journal of Development Economics* **49**, 3–32.

Scitovsky, T. (1954). "Two Concepts of External Economies," *Journal of Political Economy* **62**, 143–151.

Shaban, R. A. (1987). "Testing between Competing Models of Sharecropping," *Journal of Political Economy* **95**, 893–920.

Shapiro, C. (1983), "Premiums for High Quality Products as Returns on Reputation," *Quarterly Journal of Economics* **98**, 659–680.

Shapiro, C. and J. Stiglitz (1984). Equilibrium Unemployment as a Worker Discipline Device, *American Economic Review* **74**, 433–444.

Simon, H. (1951), "A Formal Theory of the Employment Relationship," *Econometrica* **19**, 293–305.

Solow, R. (1956). "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics* **70**, 65–94.

Stiglitz, J. (1976). The Efficiency Wage Hypothesis, Surplus Labour and the Distribution of Income in L.D.C.'s, *Oxford Economic Papers* **28**, 185–207.

Stiglitz, J. and A. Weiss (1981). "Credit Rationing in Markets with Imperfect Information," *American Economic Review* **71**, 393–410.

Udry, C. (1994). "Risk and Insurance in a Rural Credit Market: An Empirical Investigation in Northern Nigeria," *Review of Economic Studies* **61**, 495–526.