Theories of International Migration: A Review and Appraisal

DOUGLAS S. MASSEY
JOAQUÍN ARANGO
GRAEME HUGO
ALI KOUAOUCCI
ADELA PELLEGRINO
J. EDWARD TAYLOR

Over the past 30 years, immigration has emerged as a major force throughout the world. In traditional immigrant-receiving societies such as Australia, Canada, and the United States, the volume of immigration has grown and its composition has shifted decisively away from Europe, the historically dominant source, toward Asia, Africa, and Latin America. In Europe, meanwhile, countries that for centuries had been sending out migrants were suddenly transformed into immigrant-receiving societies. After 1945, virtually all countries in Western Europe began to attract significant numbers of workers from abroad. Although the migrants were initially drawn mainly from southern Europe, by the late 1960s they mostly came from developing countries in Africa, Asia, the Caribbean, and the Middle East.

By the 1980s even countries in southern Europe—Italy, Spain, and Portugal—which only a decade before had been sending migrants to wealthier countries in the north, began to import workers from Africa, Asia, and the Middle East. At the same time, Japan—with its low and still declining birth rate, its aging population, and its high standard of living—found itself turning increasingly to migrants from poorer countries in Asia and even South America to satisfy its labor needs.

Most of the world’s developed countries have become diverse, multiethnic societies, and those that have not reached this state are moving decisively in that direction. The emergence of international migration as a basic structural feature of nearly all industrialized countries testifies to the strength and coherence of the
underlying forces. Yet the theoretical base for understanding these forces remains weak. The recent boom in immigration has therefore taken citizens, officials, and demographers by surprise, and when it comes to international migration, popular thinking remains mired in nineteenth-century concepts, models, and assumptions.

At present, there is no single, coherent theory of international migration, only a fragmented set of theories that have developed largely in isolation from one another, sometimes but not always segmented by disciplinary boundaries. Current patterns and trends in immigration, however, suggest that a full understanding of contemporary migratory processes will not be achieved by relying on the tools of one discipline alone, or by focusing on a single level of analysis. Rather, their complex, multifaceted nature requires a sophisticated theory that incorporates a variety of perspectives, levels, and assumptions.

The purpose of this article is to explicate and integrate the leading contemporary theories of international migration. We begin by examining models that describe the initiation of international movement and then consider theories that account for why transnational population flows persist across space and time. Rather than favoring one theory over another a priori, we seek to understand each model on its own terms in order to illuminate key assumptions and hypotheses. Only after each theory has been considered separately do we compare and contrast the different conceptual frameworks to reveal areas of logical inconsistency and substantive disagreement. In undertaking this exercise, we seek to provide a sound basis for evaluating the models empirically, and to lay the groundwork for constructing an accurate and comprehensive theory of international migration for the twenty-first century.

The initiation of international migration

A variety of theoretical models has been proposed to explain why international migration begins, and although each ultimately seeks to explain the same thing, they employ radically different concepts, assumptions, and frames of reference. Neoclassical economics focuses on differentials in wages and employment conditions between countries, and on migration costs; it generally conceives of movement as an individual decision for income maximization. The “new economics of migration,” in contrast, considers conditions in a variety of markets, not just labor markets. It views migration as a household decision taken to minimize risks to family income or to overcome capital constraints on family production activities. Dual labor market theory and world systems theory generally ignore such micro-level decision processes, focusing instead on forces operating at much higher levels of aggregation. The former links immigration to the structural requirements of modern industrial economies, while the latter sees immigration as a natural consequence of economic globalization and market penetration across national boundaries.
Given the fact that theories conceptualize causal processes at such different levels of analysis—the individual, the household, the national, and the international—they cannot be assumed, a priori, to be inherently incompatible. It is quite possible, for example, that individuals act to maximize income while families minimize risk, and that the context within which both decisions are made is shaped by structural forces operating at the national and international levels. Nonetheless, the various models reflect different research objectives, focuses, interests, and ways of decomposing an enormously complex subject into analytically manageable parts; and a firm basis for judging their consistency requires that the inner logic, propositions, assumptions, and hypotheses of each theory be clearly specified and well-understood.

Neoclassical economics: Macro theory

Probably the oldest and best-known theory of international migration was developed originally to explain labor migration in the process of economic development (Lewis, 1954; Ranis and Fei, 1961; Harris and Todaro, 1970; Todaro, 1976). According to this theory and its extensions, international migration, like its local counterpart, is caused by geographic differences in the supply of and demand for labor. Countries with a large endowment of labor relative to capital have a low equilibrium market wage, while countries with a limited endowment of labor relative to capital are characterized by a high market wage, as depicted graphically by the familiar interaction of labor supply and demand curves. The resulting differential in wages causes workers from the low-wage country to move to the high-wage country. As a result of this movement, the supply of labor decreases and wages rise in the capital-poor country, while the supply of labor increases and wages fall in the capital-rich country, leading, at equilibrium, to an international wage differential that reflects only the costs of international movement, pecuniary and psychic.

Mirroring the flow of workers from labor-abundant to labor-scarce countries is a flow of investment capital from capital-rich to capital-poor countries. The relative scarcity of capital in poor countries yields a rate of return that is high by international standards, thereby attracting investment. The movement of capital also includes human capital, with highly skilled workers moving from capital-rich to capital-poor countries in order to reap high returns on their skills in a human capital–scarce environment, leading to a parallel movement of managers, technicians, and other skilled workers. The international flow of labor, therefore, must be kept conceptually distinct from the associated international flow of human capital. Even in the most aggregated macro-level models, the heterogeneity of immigrants along skill lines must be clearly recognized.

The simple and compelling explanation of international migration offered by neoclassical macroeconomics has strongly shaped public thinking and has provided the intellectual basis for much immigration policy. The perspective contains several implicit propositions and assumptions:
1. The international migration of workers is caused by differences in wage rates between countries.

2. The elimination of wage differentials will end the movement of labor, and migration will not occur in the absence of such differentials.

3. International flows of human capital—that is, highly skilled workers—respond to differences in the rate of return to human capital, which may be different from the overall wage rate, yielding a distinct pattern of migration that may be opposite that of unskilled workers.

4. Labor markets are the primary mechanisms by which international flows of labor are induced; other kinds of markets do not have important effects on international migration.

5. The way for governments to control migration flows is to regulate or influence labor markets in sending and/or receiving countries.

Neoclassical economics: Micro theory

Corresponding to the macroeconomic model is a microeconomic model of individual choice (Sjaastad, 1962; Todaro, 1969, 1976, 1989; Todaro and Maruszko, 1987). In this scheme, individual rational actors decide to migrate because a cost-benefit calculation leads them to expect a positive net return, usually monetary, from movement. International migration is conceptualized as a form of investment in human capital. People choose to move to where they can be most productive, given their skills; but before they can capture the higher wages associated with greater labor productivity they must undertake certain investments, which include the material costs of traveling, the costs of maintenance while moving and looking for work, the effort involved in learning a new language and culture, the difficulty experienced in adapting to a new labor market, and the psychological costs of cutting old ties and forging new ones.

Potential migrants estimate the costs and benefits of moving to alternative international locations and migrate to where the expected discounted net returns are greatest over some time horizon (Borjas, 1990). Net returns in each future period are estimated by taking the observed earnings corresponding to the individual’s skills in the destination country and multiplying these by the probability of obtaining a job there (and for illegal migrants the likelihood of being able to avoid deportation) to obtain “expected destination earnings.” These expected earnings are then subtracted from those expected in the community of origin (observed earnings there multiplied by the probability of employment) and the difference is summed over a time horizon from 0 to \(n\), discounted by a factor that reflects the greater utility of money earned in the present than in the future. From this integrated difference the estimated costs are subtracted to yield the expected net return to migration.

This decisionmaking process is summarized analytically by the following equation:
\[ ER(0) = \int_0^t [P_1(t)P_2(t)Y_d(t) - P_3(t)Y_o(t)]e^{-rt} dt - C(0) \] (1)

where \( ER(0) \) is the expected net return to migration calculated just before departure at time 0; \( t \) is time; \( P_1(t) \) is the probability of avoiding deportation from the area of destination (1.0 for legal migrants and <1.0 for undocumented migrants); \( P_2(t) \) is the probability of employment at the destination; \( Y_d(t) \) is earnings if employed at the place of destination; \( P_3(t) \) is the probability of employment in the community of origin; \( Y_o(t) \) is earnings if employed in the community of origin; \( r \) is the discount factor; and \( C(0) \) is the sum total of the costs of movement (including psychological costs).

If the quantity \( ER(0) \) is positive for some potential destination, the rational actor migrates; if it is negative the actor stays; and if it is zero, the actor is indifferent between moving and staying. In theory, a potential migrant goes to where the expected net returns to migration are greatest, leading to several important conclusions that differ slightly from the earlier macroeconomic formulations:

1. International movement stems from international differentials in both earnings and employment rates, whose product determines expected earnings (the prior model, in contrast, assumed full employment).

2. Individual human capital characteristics that increase the likely rate of remuneration or the probability of employment in the destination relative to the sending country (e.g., education, experience, training, language skills) will increase the likelihood of international movement, other things being equal.

3. Individual characteristics, social conditions, or technologies that lower migration costs increase the net returns to migration and, hence, raise the probability of international movement.

4. Because of 2 and 3, individuals within the same country can display very different proclivities to migrate.

5. Aggregate migration flows between countries are simple sums of individual moves undertaken on the basis of individual cost-benefit calculations.

6. International movement does not occur in the absence of differences in earnings and/or employment rates between countries. Migration occurs until expected earnings (the product of earnings and employment rates) have been equalized internationally (net of the costs of movement), and movement does not stop until this product has been equalized.

7. The size of the differential in expected returns determines the size of the international flow of migrants between countries.

8. Migration decisions stem from disequilibria or discontinuities between labor markets; other markets do not directly influence the decision to migrate.

9. If conditions in receiving countries are psychologically attractive to prospective migrants, migration costs may be negative. In this case, a negative earnings differential may be necessary to halt migration between countries.
10 Governments control immigration primarily through policies that affect expected earnings in sending and/or receiving countries—for example, those that attempt to lower the likelihood of employment or raise the risk of underemployment in the destination area (through employer sanctions), those that seek to raise incomes at the origin (through long-term development programs), or those that aim to increase the costs (both psychological and material) of migration.

The new economics of migration

In recent years, a “new economics of migration” has arisen to challenge many of the assumptions and conclusions of neoclassical theory (Stark and Bloom, 1985). A key insight of this new approach is that migration decisions are not made by isolated individual actors, but by larger units of related people—typically families or households—in which people act collectively not only to maximize expected income, but also to minimize risks and to loosen constraints associated with a variety of market failures, apart from those in the labor market (Stark and Levhari, 1982; Stark, 1984; Katz and Stark, 1986; Lauby and Stark, 1988; Taylor, 1986; Stark, 1991).

Unlike individuals, households are in a position to control risks to their economic well-being by diversifying the allocation of household resources, such as family labor. While some family members can be assigned economic activities in the local economy, others may be sent to work in foreign labor markets where wages and employment conditions are negatively correlated or weakly correlated with those in the local area. In the event that local economic conditions deteriorate and activities there fail to bring in sufficient income, the household can rely on migrant remittances for support.

In developed countries, risks to household income are generally minimized through private insurance markets or governmental programs, but in developing countries these institutional mechanisms for managing risk are imperfect, absent, or inaccessible to poor families, giving them incentives to diversify risks through migration. In developed countries, moreover, credit markets are relatively well-developed to enable families to finance new projects, such as the adoption of new production technology. In most developing areas, in contrast, credit is usually not available or is procurable only at high cost. In the absence of accessible public or affordable private insurance and credit programs, market failures create strong pressures for international movement, as the following examples show.

**Crop insurance markets** Whenever farm households put time and money into sowing a crop, they are betting that the investment will pay off at a future date in the form of a product that can be sold for cash to purchase desired goods and services, or which can be consumed directly for subsistence. Between the
time a crop is planted and harvested, however, human or natural events may reduce or eliminate the harvest, leaving the family with insufficient income or food for subsistence. Likewise, the introduction of new agricultural technology (such as high-yielding seeds or new methods of cultivation) may alter the objective and/or subjective risks confronting farm households. Using a new seed variety may increase a farmer’s yield if the development expert is right; but if he or she is wrong, the household faces the prospect of having insufficient food or income.

In developed countries, these sorts of objective and subjective risks are managed through formal insurance arrangements, whereby agricultural producers pay a fee to a company or a government agency to insure the crop against future loss. The insuring institution assumes the risk to the future crop, and should a drought or flood destroy the harvest or a new technology backfire, it pays the producer for the insured market value of the crop, thereby guaranteeing the economic well-being of the family. If crop insurance is not available, families have an incentive to self-insure by sending one or more workers abroad to remit earnings home, thereby guaranteeing family income even if the harvest fails.

Futures markets Whenever a household sows a cash crop, it assumes that the crop, when harvested, can be sold for a price sufficient to sustain the family or improve its well-being. In making this bet, however, there is a risk that the price for the crop may drop below expected levels, leaving the family with insufficient income. In developed countries, price risk is managed through futures markets that allow farmers to sell all or part of their crop for future delivery at a guaranteed price. Investors assume the risk of loss should prices fall below the guaranteed price, and they reap the gain should prices rise above this level. Most developing countries lack futures markets, and when they exist, poor farm households generally lack access to them. Migration offers a mechanism by which farm families can self-insure against income risks arising from crop price fluctuations.

Unemployment insurance Nonfarm families, as well as many farm households, depend on wages earned by family workers. If local economic conditions deteriorate and employment levels fall, or if a family member is injured and cannot work, the household’s livelihood may be threatened by a reduction or loss of income. In wealthy countries, governments maintain insurance programs that protect workers and their families from this risk, but in poor countries such unemployment and disability programs are absent or incomplete in their coverage, again giving families incentives to self-insure by sending workers abroad.

If employment conditions in foreign and local labor markets are negatively correlated or are uncorrelated, then international migration provides a way of reducing the risk to family wages and guarantees a reliable stream of income, in the form of remittances, to support the family. Moreover, migration fulfills this
insurance function whether or not remittances are actually observed. Migrants, like formal insurance contracts, only have to pay out if losses are realized. The existence of an implicit or explicit insurance arrangement, however, can have an important effect on a household’s economic behavior, and the desire to acquire this insurance may be a primary motivation for families to participate in international migration.

Capital markets Households may desire to increase the productivity of their assets, but to do so they need to acquire capital to make additional investments. Farm families, for example, may seek to irrigate their fields, apply fertilizers, buy scientifically improved seeds, or acquire machinery, but they may lack the money to purchase these inputs. Nonfarm families may seek to invest in the education or training of household members, or to acquire capital goods that can be used to produce goods for sale on consumer markets, but again they may lack money to cover these costs. In developed countries, investments are funded either through private savings or borrowing, both of which are greatly assisted by access to a sound and efficient banking system. Borrowing can also provide protection against consumption risk if income is variable. In many developing countries, however, savings institutions are unreliable or underdeveloped, and people are reluctant to entrust their savings to them.

In poor countries the needed funds may also be difficult to borrow because the family lacks collateral to qualify for a loan, because there is a scarcity of lending capital, or because the banking system provides incomplete coverage, serving mainly the needs of the affluent. For poor families, the only real access to borrowing is often from local moneylenders who charge high interest rates, making transaction costs prohibitive. Under these circumstances, migration again becomes attractive as an alternative source of capital to finance improvements in productivity and ensure stability in consumption, and the family has a strong incentive to send one or more workers abroad to accumulate savings or to transfer capital back in the form of remittances.

A key proposition in the foregoing discussion is that income is not a homogeneous good, as assumed by neoclassical economics. The source of the income really matters, and households have significant incentives to invest scarce family resources in activities and projects that provide access to new income sources, even if these activities do not necessarily increase total income.

The new economics of migration also questions the assumption that income has a constant effect on utility for an actor across socioeconomic settings—that a $100 real increase in income means the same thing to a person regardless of local community conditions and irrespective of his or her position in the income distribution. The new economic theorists argue, in contrast, that households send workers abroad not only to improve income in absolute terms, but also to increase income relative to other households, and, hence, to reduce their relative deprivation compared with some reference group (see Stark, Taylor,
A household’s sense of relative deprivation depends on the incomes of which it is deprived in the reference-group income distribution. If $F(y)$ is the cumulative income distribution and $h[1-F(y)]$ represents the dissatisfaction felt by a household with income $y$ from not having an income that is slightly higher than $y$ (i.e., $y + \Delta$), then the relative deprivation of a household with income $y$ can be expressed conceptually as:

$$RD(y) = \int_y^{y_{\text{max}}} h[1-F(z)] \, dz$$

(2)

where $y_{\text{max}}$ is the highest income found in the community. In the simple case where $h[1-F(y)] = 1-F(y)$, this expression is equivalent to the product of two terms: the share of households with income greater than $y$, and the average difference between these higher household incomes and $y$ (Stark and Taylor, 1989).

To illustrate this concept of relative income, consider an increase in the income of affluent households. If poor households’ incomes are unchanged, then their relative deprivation increases. If household utility is negatively affected by relative deprivation, then even though a poor household’s absolute income and expected gains from migration remain unchanged, its incentive to participate in international migration increases if, by sending a family member abroad, it can hope to reap a relative income gain in the community. The likelihood of migration thus grows because of the change in other households’ incomes. Market failures that constrain local income opportunities for poor households may also increase the attractiveness of migration as an avenue for effecting gains in relative income.

The theoretical models growing out of the “new economics” of migration yield a set of propositions and hypotheses that are quite different from those emanating from neoclassical theory, and they lead to a very different set of policy prescriptions:

1 Families, households, or other culturally defined units of production and consumption are the appropriate units of analysis for migration research, not the autonomous individual.

2 A wage differential is not a necessary condition for international migration to occur; households may have strong incentives to diversify risks through transnational movement even in the absence of wage differentials.

3 International migration and local employment or local production are not mutually exclusive possibilities. Indeed, there are strong incentives for households to engage in both migration and local activities. In fact, an increase in the returns to local economic activities may heighten the attractiveness of migration as a means of overcoming capital and risk constraints on investing in those activities. Thus, economic development within sending regions need not reduce the pressures for international migration.
4 International movement does not necessarily stop when wage differentials have been eliminated across national boundaries. Incentives for migration may continue to exist if other markets within sending countries are absent, imperfect, or in disequilibria.

5 The same expected gain in income will not have the same effect on the probability of migration for households located at different points in the income distribution, or among those located in communities with different income distributions.

6 Governments can influence migration rates not only through policies that influence labor markets, but also through those that shape insurance markets, capital markets, and futures markets. Government insurance programs, particularly unemployment insurance, can significantly affect the incentives for international movement.

7 Government policies and economic changes that shape income distributions will change the relative deprivation of some households and thus alter their incentives to migrate.

8 Government policies and economic changes that affect the distribution of income will influence international migration independent of their effects on mean income. In fact, government policies that produce a higher mean income in migrant-sending areas may increase migration if relatively poor households do not share in the income gain. Conversely, policies may reduce migration if relatively rich households do not share in the income gain.

Dual labor market theory

Although neoclassical human capital theory and the new economics of migration lead to divergent conclusions about the origins and nature of international migration, both are essentially micro-level decision models. What differ are the units assumed to make the decision (the individual or the household), the entity being maximized or minimized (income or risk), assumptions about the economic context of decisionmaking (complete and well-functioning markets versus missing or imperfect markets), and the extent to which the migration decision is socially contextualized (whether income is evaluated in absolute terms or relative to some reference group). Standing distinctly apart from these models of rational choice, however, is dual labor market theory, which sets its sights away from decisions made by individuals and argues that international migration stems from the intrinsic labor demands of modern industrial societies.

Piore (1979) has been the most forceful and elegant proponent of this theoretical viewpoint, arguing that international migration is caused by a permanent demand for immigrant labor that is inherent to the economic structure of developed nations. According to Piore, immigration is not caused by push factors in sending countries (low wages or high unemployment), but by pull factors in receiving countries (a chronic and unavoidable need for foreign
workers). This built-in demand for immigrant labor stems from four fundamental characteristics of advanced industrial societies and their economies.

**Structural inflation** Wages not only reflect conditions of supply and demand; they also confer status and prestige, social qualities that inhere to the jobs to which the wages are attached. In general, people believe that wages should reflect social status, and they have rather rigid notions about the correlation between occupational status and pay. As a result, wages offered by employers are not entirely free to respond to changes in the supply of workers. A variety of informal social expectations and formal institutional mechanisms (such as union contracts, civil service rules, bureaucratic regulations, company job classifications) ensures that wages correspond to the hierarchies of prestige and status that people perceive and expect.

If employers seek to attract workers for unskilled jobs at the bottom of an occupational hierarchy, they cannot simply raise wages. Raising wages at the bottom of the hierarchy would upset socially defined relationships between status and remuneration. If wages are increased at the bottom, there will be strong pressure to raise wages by corresponding amounts at other levels of the hierarchy. If the wages of busboys are raised in response to a shortage of entry-level workers, for example, they may overlap with those of waitresses, thereby threatening their status and undermining the accepted social hierarchy. Waitresses, in turn, demand a corresponding wage increase, which threatens the position of cooks, who also pressure employers for a raise. Workers may be aided in their efforts by union representatives or contracts.

Thus the cost to employers of raising wages to attract low-level workers is typically more than the cost of these workers' wages alone; wages must be increased proportionately throughout the job hierarchy in order to keep them in line with social expectations, a problem known as structural inflation. Attracting native workers by raising entry wages during times of labor scarcity is thus expensive and disruptive, providing employers with a strong incentive to seek easier and cheaper solutions, such as the importation of migrant workers who will accept low wages.

**Motivational problems** Occupational hierarchies are also critical for the motivation of workers, since people work not only for income, but also for the accumulation and maintenance of social status. Acute motivational problems arise at the bottom of the job hierarchy because there is no status to be maintained and there are few avenues for upward mobility. The problem is inescapable and structural because the bottom cannot be eliminated from the labor market. Mechanization to eliminate the lowest and least desirable class of jobs will simply create a new bottom tier composed of jobs that used to be just above the bottom rung. Since there always has to be a bottom of any hierarchy, motivational problems are inescapable. What employers need are workers who
view bottom-level jobs simply as a means to the end of earning money, and for whom employment is reduced solely to income, with no implications for status or prestige.

For a variety of reasons, immigrants satisfy this need, at least at the beginning of their migratory careers. Most migrants begin as target earners, seeking to earn money for a specific goal that will improve their status or well-being at home—building a house, paying for school, buying land, acquiring consumer goods. Moreover, the disjuncture in living standards between developed and developing societies means that even low wages abroad appear to be generous by the standards of the home community; and even though a migrant may realize that a foreign job is of low status abroad, he does not view himself as being a part of the receiving society. Rather he sees himself as a member of his home community, within which foreign labor and hard-currency remittances carry considerable honor and prestige.

**Economic dualism** Bifurcated labor markets come to characterize advanced industrial economies because of the inherent duality between labor and capital. Capital is a fixed factor of production that can be idled by lower demand but not laid off; owners of capital must bear the costs of its unemployment. Labor is a variable factor of production that can be released when demand falls, so that workers are forced to bear the costs of their own unemployment. Whenever possible, therefore, capitalists seek out the stable, permanent portion of demand and reserve it for the employment of equipment, whereas the variable portion of demand is met by adding labor. Thus, capital-intensive methods are used to meet basic demand, and labor-intensive methods are reserved for the seasonal, fluctuating component. This dualism creates distinctions among workers, leading to a bifurcation of the labor force.

Workers in the capital-intensive primary sector get stable, skilled jobs working with the best equipment and tools. Employers are forced to invest in these workers by providing specialized training and education. Their jobs are complicated and require considerable knowledge and experience to perform well, leading to the accumulation of firm-specific human capital. Primary-sector workers tend to be unionized or highly professionalized, with contracts that require employers to bear a substantial share of the costs of their idleness (in the form of severance pay and unemployment benefits). Because of these costs and continuing obligations, workers in the primary sector become expensive to let go; they become more like capital.

In the labor-intensive secondary sector, however, workers hold unstable, unskilled jobs; they may be laid off at any time with little or no cost to the employer. Indeed, the employer will generally lose money by retaining workers during slack periods. During down cycles the first thing secondary-sector employers do is cut their payroll. As a result, employers force workers in this sector to bear the costs of their unemployment. They remain a variable factor of production and are, hence, expendable.
Thus, the inherent dualism between labor and capital extends to the labor force in the form of a segmented labor market structure. Low wages, unstable conditions, and the lack of reasonable prospects for mobility in the secondary sector make it difficult to attract native workers, who are instead drawn into the primary, capital-intensive sector, where wages are higher, jobs are more secure, and there is a possibility of occupational improvement. To fill the shortfall in demand within the secondary sector, employers turn to immigrants.

The demography of labor supply The problems of motivation and structural inflation inherent to modern occupational hierarchies, together with the dualism intrinsic to market economies, create a permanent demand for workers who are willing to labor under unpleasant conditions, at low wages, with great instability, and facing little chance for advancement. In the past, this demand was met partially by two sets of people with social statuses and characteristics conducive to these sorts of jobs: women and teenagers.

Historically women have tended to participate in the labor force up to the time of their first birth, and to a lesser extent after children had grown. They sought to earn supplemental income for themselves or their families. They were not primary breadwinners and their principal social identity was that of a sister, wife, or mother. They were willing to put up with low wages and instability because they viewed the work as transient and the earnings as supplemental; the positions they held were unthreatening to their main social statuses, which were grounded in the family.

Likewise, teenagers historically have moved into and out of the labor force with great frequency in order to earn extra money, to gain experience, and to try out different occupational roles. They do not view dead-end jobs as problematic because they expect to get better jobs in the future, after completing school, gaining experience, or settling down. Moreover, teenagers derive their social identities from their parents and families of orientation, not their jobs. They view work instrumentally as a means of earning spending money. The money and the things that it buys enhance their status among their peers; the job is just a means to an end.

In advanced industrial societies, however, these two sources of entry-level workers have shrunk over time because of three fundamental socio-demographic trends: the rise in female labor force participation, which has transformed women’s work into a career pursued for social status as well as income; the rise in divorce rates, which has transformed women’s jobs into a source of primary income support; and the decline in birth rates and the extension of formal education, which have produced very small cohorts of teenagers entering the labor force. The imbalance between the structural demand for entry-level workers and the limited domestic supply of such workers has increased the underlying, long-run demand for immigrants.

Dual labor market theory neither posits nor denies that actors make rational, self-interested decisions, as predicted by microeconomic models. The
negative qualities that people in industrialized countries attach to low-wage jobs, for example, may open up employment opportunities to foreign workers, thereby raising their expected earnings, increasing their ability to overcome risk and credit constraints, and enabling households to achieve relative income gains by sending family members abroad. Recruitment by employers helps to overcome informational and other constraints on international movement, enhancing migration’s value as a strategy for family income generation or risk diversification.

Although not in inherent conflict with neoclassical economics, dual labor market theory does carry implications and corollaries that are quite different from those emanating from micro-level decision models:

1 International labor migration is largely demand-based and is initiated by recruitment on the part of employers in developed societies, or by governments acting on their behalf.

2 Since the demand for immigrant workers grows out of the structural needs of the economy and is expressed through recruitment practices rather than wage offers, international wage differentials are neither a necessary nor a sufficient condition for labor migration to occur. Indeed, employers have incentives to recruit workers while holding wages constant.

3 Low-level wages in immigrant-receiving societies do not rise in response to a decrease in the supply of immigrant workers; they are held down by social and institutional mechanisms and are not free to respond to shifts in supply and demand.

4 Low-level wages may fall, however, as a result of an increase in the supply of immigrant workers, since the social and institutional checks that keep low-level wages from rising do not prevent them from falling.

5 Governments are unlikely to influence international migration through policies that produce small changes in wages or employment rates; immigrants fill a demand for labor that is structurally built into modern, post-industrial economies, and influencing this demand requires major changes in economic organization.

World systems theory

Building on the work of Wallerstein (1974), a variety of sociological theorists has linked the origins of international migration not to the bifurcation of the labor market within particular national economies, but to the structure of the world market that has developed and expanded since the sixteenth century (Portes and Walton, 1981; Petras, 1981; Castells, 1989; Sassen, 1988, 1991; Morawska, 1990). In this scheme, the penetration of capitalist economic relations into peripheral, noncapitalist societies creates a mobile population that is prone to migrate abroad.

Driven by a desire for higher profits and greater wealth, owners and managers of capitalist firms enter poor countries on the periphery of the world
economy in search of land, raw materials, labor, and new consumer markets. In the past, this market penetration was assisted by colonial regimes that administered poor regions for the benefit of economic interests in colonizing societies. Today it is made possible by neocolonial governments and multinational firms that perpetuate the power of national elites who either participate in the world economy as capitalists themselves, or offer their nation’s resources to global firms on acceptable terms.

According to world systems theory, migration is a natural outgrowth of disruptions and dislocations that inevitably occur in the process of capitalist development. As capitalism has expanded outward from its core in Western Europe, North America, Oceania, and Japan, ever-larger portions of the globe and growing shares of the human population have been incorporated into the world market economy. As land, raw materials, and labor within peripheral regions come under the influence and control of markets, migration flows are inevitably generated, some of which have always moved abroad (Massey, 1989).

**Land** In order to achieve the greatest profit from existing agrarian resources and to compete within global commodity markets, capitalist farmers in peripheral areas seek to consolidate landholding, mechanize production, introduce cash crops, and apply industrially produced inputs such as fertilizer, insecticides, and high-yield seeds. Land consolidation destroys traditional systems of land tenure based on inheritance and common rights of usufruct. Mechanization decreases the need for manual labor and makes many agrarian workers redundant to production. The substitution of cash crops for staples undermines traditional social and economic relations based on subsistence (Chayanov, 1966); and the use of modern inputs produces high crop yields at low unit prices, which drives small, noncapitalist farmers out of local markets. All of these forces contribute to the creation of a mobile labor force displaced from the land with a weakened attachment to local agrarian communities.

**Raw materials** The extraction of raw materials for sale on global markets requires industrial methods that rely on paid labor. The offer of wages to former peasants undermines traditional forms of social and economic organization based on systems of reciprocity and fixed role relations and creates incipient labor markets based on new conceptions of individualism, private gain, and social change. These trends likewise promote the geographic mobility of labor in developing regions, often with international spillovers.

**Labor** Firms from core capitalist countries enter developing countries to establish assembly plants that take advantage of low wage rates, often within special export-processing zones created by sympathetic governments. The demand for factory workers strengthens local labor markets and weakens traditional productive relations. Much of the labor demanded is female, however, and the resulting feminization of the workforce limits opportunities for men; but since
the new factory work is demanding and poorly paid, women tend only to work a few years, after which time they leave to look for new opportunities. The insertion of foreign-owned factories into peripheral regions thus undermines the peasant economy by producing goods that compete with those made locally; by feminizing the workforce without providing factory-based employment opportunities for men; and by socializing women for industrial work and modern consumption, albeit without providing a lifetime income capable of meeting these needs. The result is the creation of a population that is socially and economically uprooted and prone to migration.

The same capitalist economic processes that create migrants in peripheral regions simultaneously attract them to developed countries. Although some people displaced by the process of market penetration move to cities, leading to the urbanization of developing societies, inevitably many are drawn abroad because globalization creates material and ideological links to the places where capital originates. The foreign investment that drives economic globalization is managed from a small number of global cities, whose structural characteristics create a strong demand for immigrant labor.

Material links In order to ship goods, deliver machinery, extract and export raw materials, coordinate business operations, and manage expatriate assembly plants, capitalists in core nations build and expand transportation and communication links to the peripheral countries where they have invested. These links not only facilitate the movement of goods, products, information, and capital, they also promote the movement of people by reducing the costs of movement along certain international pathways. Because investment and globalization are inevitably accompanied by the build-up of a transportation and communication infrastructure, the international movement of labor generally follows the international movement of goods and capital in the opposite direction.

Ideological links The process of economic globalization creates cultural links between core capitalist countries and their hinterlands within the developing world. In many cases, these cultural links are longstanding, reflecting a colonial past in which core countries established administrative and educational systems that mirrored their own in order to govern and exploit a peripheral region. Citizens of Senegal, for example, learn French, study at lycées, and use a currency directly tied to the French franc in economic transactions. Likewise, Indians and Pakistanis learn English, take British-style degrees, and join with others in a transnational union known as the British Commonwealth. Even in the absence of a colonial past, the influence of economic penetration can be profound: Mexicans increasingly study at US universities, speak English, and follow American consumer styles closely.

These ideological and cultural connections are reinforced by mass communications and advertising campaigns directed from the core countries. Television
programming from the United States, France, Britain, and Germany transmits
information about lifestyles and living standards in the developed world, and
commercials prepared by foreign advertising agencies inculcate modern con-
sumer tastes within peripheral peoples. The diffusion of core country languages
and cultural patterns and the spread of modern consumption patterns interact
with the emergence of a transportation/communication infrastructure to channel
international migration to particular core countries.

Global cities The world economy is managed from a relatively small
number of urban centers in which banking, finance, administration, professional
services, and high-tech production tend to be concentrated (Castells, 1989;
Sassen, 1991). In the United States, global cities include New York, Chicago, Los
Angeles, and Miami; in Europe, they include London, Paris, Frankfurt, and
Milan; and in the Pacific, Tokyo, Osaka, and Sydney qualify. Within these global
cities, a great deal of wealth and a highly educated workforce are concentrated,
creating a strong demand for services from unskilled workers (busboys, gar-
deners, waiters, hotel workers, domestic servants). At the same time, the shifting
of heavy industrial production overseas; the growth of high-tech manufacturing
in electronics, computers, and telecommunications; and the expansion of service
sectors such as health and education create a bifurcated labor market structure
with strong demand for workers at both the upper and lower ends, but with
relatively weak demand in the middle.

Poorly educated natives resist taking low-paying jobs at the bottom of the
occupational hierarchy, creating a strong demand for immigrants. Meanwhile,
well-educated natives and skilled foreigners dominate the lucrative jobs at the
upper tier of the occupational distribution, and the concentration of wealth
among them helps to fuel the demand for the type of services immigrants are
most willing to meet. Native workers with modest educations cling to jobs in the
declining middle, migrate out of global cities, or rely on social insurance
programs for support.

World systems theory thus argues that international migration follows the
political and economic organization of an expanding global market, a view that
yields six distinct hypotheses:

1 International migration is a natural consequence of capitalist market
formation in the developing world; the penetration of the global economy into
peripheral regions is the catalyst for international movement.

2 The international flow of labor follows the international flow of goods
and capital, but in the opposite direction. Capitalist investment foments changes
that create an uprooted, mobile population in peripheral countries while simulta-
neously forging strong material and cultural links with core countries, leading
to transnational movement.

3 International migration is especially likely between past colonial powers
and their former colonies, because cultural, linguistic, administrative, invest-
ment, transportation, and communication links were established early and were
allowed to develop free from outside competition during the colonial era, leading to the formation of specific transnational markets and cultural systems.

4 Since international migration stems from the globalization of the market economy, the way for governments to influence immigration rates is by regulating the overseas investment activities of corporations and controlling international flows of capital and goods. Such policies, however, are unlikely to be implemented because they are difficult to enforce, tend to incite international trade disputes, risk world economic recession, and antagonize multinational firms with substantial political resources that can be mobilized to block them.

5 Political and military interventions by governments of capitalist countries to protect investments abroad and to support foreign governments sympathetic to the expansion of the global market, when they fail, produce refugee movements directed to particular core countries, constituting another form of international migration.

6 International migration ultimately has little to do with wage rates or employment differentials between countries; it follows from the dynamics of market creation and the structure of the global economy.

The perpetuation of international movement

Immigration may begin for a variety of reasons—a desire for individual income gain, an attempt to diversify risks to household income, a program of recruitment to satisfy employer demands for low-wage workers, an international displacement of peasants by market penetration within peripheral regions, or some combination thereof. But the conditions that initiate international movement may be quite different from those that perpetuate it across time and space. Although wage differentials, relative risks, recruitment efforts, and market penetration may continue to cause people to move, new conditions that arise in the course of migration come to function as independent causes themselves: migrant networks spread, institutions supporting transnational movement develop, and the social meaning of work changes in receiving societies. The general thrust of these transformations is to make additional movement more likely, a process known as cumulative causation.

Network theory

Migrant networks are sets of interpersonal ties that connect migrants, former migrants, and nonmigrants in origin and destination areas through ties of kinship, friendship, and shared community origin. They increase the likelihood of international movement because they lower the costs and risks of movement and increase the expected net returns to migration. Network connections constitute a form of social capital that people can draw upon to gain access to foreign employment. Once the number of migrants reaches a critical threshold,
the expansion of networks reduces the costs and risks of movement, which causes the probability of migration to rise, which causes additional movement, which further expands the networks, and so on. Over time migratory behavior spreads outward to encompass broader segments of the sending society (Hugo, 1981; Taylor, 1986; Massey and García España, 1987; Massey, 1990a, 1990b; Gurak and Caces, 1992).

Declining costs The first migrants who leave for a new destination have no social ties to draw upon, and for them migration is costly, particularly if it involves entering another country without documents. After the first migrants have left, however, the potential costs of migration are substantially lowered for friends and relatives left behind. Because of the nature of kinship and friendship structures, each new migrant creates a set of people with social ties to the destination area. Migrants are inevitably linked to nonmigrants, and the latter draw upon obligations implicit in relationships such as kinship and friendship to gain access to employment and assistance at the point of destination.

Once the number of network connections in an origin area reaches a critical threshold, migration becomes self-perpetuating because each act of migration itself creates the social structure needed to sustain it. Every new migrant reduces the costs of subsequent migration for a set of friends and relatives, and some of these people are thereby induced to migrate, which further expands the set of people with ties abroad, which, in turn, reduces costs for a new set of people, causing some of them to migrate, and so on.

Declining risks Networks also make international migration extremely attractive as a strategy for risk diversification. When migrant networks are well-developed, they put a destination job within easy reach of most community members and make emigration a reliable and secure source of income. Thus, the self-sustaining growth of networks that occurs through the progressive reduction of costs may also be explained theoretically by the progressive reduction of risks. Every new migrant expands the network and reduces the risks of movement for all those to whom he or she is related, eventually making it virtually risk-free and costless to diversify household labor allocations through emigration.

This dynamic theory accepts the view of international migration as an individual or household decision process, but argues that acts of migration at one point in time systematically alter the context within which future migration decisions are made, greatly increasing the likelihood that later decisionmakers will choose to migrate. The conceptualization of migration as a self-sustaining diffusion process has implications and corollaries that are quite different from those derived from the general equilibrium analyses typically employed to study migration:

1 Once begun, international migration tends to expand over time until network connections have diffused so widely in a sending region that all people
who wish to migrate can do so without difficulty; then migration begins to decelerate.

2 The size of the migratory flow between two countries is not strongly correlated to wage differentials or employment rates, because whatever effects these variables have in promoting or inhibiting migration are progressively overshadowed by the falling costs and risks of movement stemming from the growth of migrant networks over time.

3 As international migration becomes institutionalized through the formation and elaboration of networks, it becomes progressively independent of the factors that originally caused it, be they structural or individual.

4 As networks expand and the costs and risks of migration fall, the flow becomes less selective in socioeconomic terms and more representative of the sending community or society.

5 Governments can expect to have great difficulty controlling flows once they have begun, because the process of network formation lies largely outside their control and occurs no matter what policy regime is pursued.

6 Certain immigration policies, however, such as those intended to promote reunification between immigrants and their families abroad, work at cross-purposes with the control of immigration flows, since they reinforce migrant networks by giving members of kin networks special rights of entry.

Institutional theory

Once international migration has begun, private institutions and voluntary organizations arise to satisfy the demand created by an imbalance between the large number of people who seek entry into capital-rich countries and the limited number of immigrant visas these countries typically offer. This imbalance, and the barriers that core countries erect to keep people out, create a lucrative economic niche for entrepreneurs and institutions dedicated to promoting international movement for profit, yielding a black market in migration. As this underground market creates conditions conducive to exploitation and victimization, voluntary humanitarian organizations also arise in developed countries to enforce the rights and improve the treatment of legal and undocumented migrants.

For-profit organizations and private entrepreneurs provide a range of services to migrants in exchange for fees set on the underground market: surreptitious smuggling across borders; clandestine transport to internal destinations; labor contracting between employers and migrants; counterfeit documents and visas; arranged marriages between migrants and legal residents or citizens of the destination country; and lodging, credit, and other assistance in countries of destination. Humanitarian groups help migrants by providing counseling, social services, shelter, legal advice about how to obtain legitimate papers, and even insulation from immigration law enforcement authorities. Over time, individuals, firms, and organizations become well-known to immigrants
and institutionally stable, constituting another form of social capital that migrants can draw upon to gain access to foreign labor markets.

The recognition of a gradual build-up of institutions, organizations, and entrepreneurs dedicated to arranging immigrant entry, legal or illegal, again yields hypotheses that are also quite distinct from those emanating from micro-level decision models:

1. As organizations develop to support, sustain, and promote international movement, the international flow of migrants becomes more and more institutionalized and independent of the factors that originally caused it.

2. Governments have difficulty controlling migration flows once they have begun because the process of institutionalization is difficult to regulate. Given the profits to be made by meeting the demand for immigrant entry, police efforts only serve to create a black market in international movement, and stricter immigration policies are met with resistance from humanitarian groups.

Cumulative causation

In addition to the growth of networks and the development of migrant-supporting institutions, international migration sustains itself in other ways that make additional movement progressively more likely over time, a process Myrdal (1957) called cumulative causation (Massey, 1990b). Causation is cumulative in that each act of migration alters the social context within which subsequent migration decisions are made, typically in ways that make additional movement more likely. So far, social scientists have discussed six socioeconomic factors that are potentially affected by migration in this cumulative fashion: the distribution of income, the distribution of land, the organization of agriculture, culture, the regional distribution of human capital, and the social meaning of work. Feedbacks through other variables are also possible, but have not been systematically treated (Stark, Taylor, and Yitzhaki, 1986; Taylor, 1992).

The distribution of income As we have already noted, people may be motivated to migrate not only to increase their absolute income or to diversify their risks, but also to improve their income relative to other households in their reference group. As a household’s sense of relative deprivation increases, so does the motivation to migrate. Before anyone has migrated from a community, income inequality within most poor, rural settings is not great because nearly all families live close to the subsistence level with minimal outside incomes. After one or two households have begun participating in foreign wage labor, however, remittances increase their incomes greatly. Given the costs and risks associated with international movement, moreover, the first households to migrate are usually located in the middle or upper ranges of the local income hierarchy.

Seeing some families vastly improve their income through migration makes families lower in the income distribution feel relatively deprived, inducing some of them to migrate, which further exacerbates income inequality and
increases the sense of relative deprivation among nonmigrants, inducing still more families to migrate, and so on. Income inequality and relative deprivation go through a series of phases, being low at first, then high as the rate of outmigration accelerates, then low again as a majority of households participate in the migrant workforce, reaching a minimum when practically all families are involved in foreign wage labor (Stark, Taylor, and Yitzhaki, 1986; Stark and Taylor, 1989; Stark, 1991; Taylor, 1992).

The distribution of land An important spending target for migrants from rural communities is the purchase of land. But land is purchased by migrants abroad typically for its prestige value or as a source of retirement income rather than as a productive investment. International migrants are likely to use their higher earnings to purchase farmland, but they are more likely than nonmigrants to let the land lie fallow since foreign wage labor is more lucrative than local agrarian production. This pattern of land use lowers the demand for local farm labor, thereby increasing the pressures for outmigration. The more outmigration, the more people have access to the funds necessary to buy land, leading to additional purchases by migrants and more land withdrawn from production, creating still more pressure for outmigration (Rhoades, 1978; Reichert, 1981; Mines, 1984; Wiest, 1984).

The organization of agrarian production When migrant households do farm the land they own, moreover, they are more likely than nonmigrant families to use capital-intensive methods (machinery, herbicides, irrigation, fertilizers, and improved seeds) since they have access to capital to finance these inputs. Thus migrant households need less labor per unit of output than nonmigrant households, thereby displacing local workers from traditional tasks and again increasing the pressures for outmovement (Massey et al., 1987). The more migration, the greater the capitalization of agriculture and the greater the displacement of agrarian labor, leading to still greater migration.

The culture of migration As migration grows in prevalence within a community, it changes values and cultural perceptions in ways that increase the probability of future migration. Among the migrants themselves, experience in an advanced industrial economy changes tastes and motivations (Piore, 1979). Although migrants may begin as target earners seeking to make one trip and earn money for a narrow purpose, after migrating they acquire a stronger concept of social mobility and a taste for consumer goods and styles of life that are difficult to attain through local labor. Once someone has migrated, therefore, he or she is very likely to migrate again, and the odds of taking an additional trip rise with the number of trips already taken (Massey, 1986).

At the community level, migration becomes deeply ingrained into the repertoire of people’s behaviors, and values associated with migration become
part of the community’s values. For young men, and in many settings young women as well, migration becomes a rite of passage, and those who do not attempt to elevate their status through international movement are considered lazy, unenterprising, and undesirable (Reichert, 1982). Eventually, knowledge about foreign locations and jobs becomes widely diffused, and values, sentiments, and behaviors characteristic of the core society spread widely within the sending region (Massey et al., 1987; Alarcón, 1992).

The regional distribution of human capital Migration is a selective process that tends, initially at least, to draw relatively well-educated, skilled, productive, and highly motivated people away from sending communities (as pointed out earlier, however, migration tends to become less selective over time as the costs and risks fall because of network formation). Sustained outmigration thus leads to the depletion of human capital in sending regions and its accumulation in receiving areas, enhancing the productivity of the latter while lowering that of the former. Over time, therefore, the accumulation of human capital reinforces economic growth in receiving areas while its simultaneous depletion in sending areas exacerbates their stagnation, thereby further enhancing the conditions for migration (Myrdal, 1957; Greenwood, 1981, 1985; Greenwood, Hunt, and McDowell, 1987). Programs of school construction and educational expansion in sending areas reinforce this cumulative migration process because raising educational levels in peripheral rural areas increases the potential returns to migration and gives people a greater incentive to leave for urban destinations at home or abroad.

Social labeling Within receiving societies, once immigrants have been recruited into particular occupations in significant numbers, those jobs become culturally labeled as “immigrant jobs” and native workers are reluctant to fill them, reinforcing the structural demand for immigrants. Immigration changes the social definition of work, causing a certain class of jobs to be defined as stigmatizing and viewed as culturally inappropriate for native workers (Böhning, 1972; Piore, 1979). The stigma comes from the presence of immigrants, not from the characteristics of the job. In most European countries, for example, jobs in automobile manufacturing came to be considered “immigrant jobs,” whereas in the United States they are considered “native jobs.”

Viewing international migration in dynamic terms as a cumulative social process yields a set of propositions broadly consistent with those derived from network theory:

1 Social, economic, and cultural changes brought about in sending and receiving countries by international migration give the movement of people a powerful internal momentum resistant to easy control or regulation, since the feedback mechanisms of cumulative causation largely lie outside the reach of government.
2 During times of domestic unemployment and joblessness, governments find it difficult to curtail labor migration and to recruit natives back into jobs formerly held by immigrants. A value shift has occurred among native workers, who refuse the “immigrant” jobs, making it necessary to retain or recruit more immigrants.

3 The social labeling of a job as “immigrant” follows from the concentration of immigrants within it; once immigrants have entered a job in significant numbers, whatever its characteristics, it will be difficult to recruit native workers back into that occupational category.

Migration systems theory

The various propositions of world systems theory, network theory, institutional theory, and the theory of cumulative causation all suggest that migration flows acquire a measure of stability and structure over space and time, allowing for the identification of stable international migration systems. These systems are characterized by relatively intense exchanges of goods, capital, and people between certain countries and less intense exchanges between others. An international migration system generally includes a core receiving region, which may be a country or group of countries, and a set of specific sending countries linked to it by unusually large flows of immigrants (Fawcett, 1989; Zlotnik, 1992).

Although not a separate theory so much as a generalization following from the foregoing theories, a migration systems perspective yields several interesting hypotheses and propositions:

1 Countries within a system need not be geographically close since flows reflect political and economic relationships rather than physical ones. Although proximity obviously facilitates the formation of exchange relationships, it does not guarantee them nor does distance preclude them.

2 Multipolar systems are possible, whereby a set of dispersed core countries receive immigrants from a set of overlapping sending nations.

3 Nations may belong to more than one migration system, but multiple membership is more common among sending than receiving nations.

4 As political and economic conditions change, systems evolve, so that stability does not imply a fixed structure. Countries may join or drop out of a system in response to social change, economic fluctuations, or political upheaval.

Evaluation of theories

Because theories proposed to explain the origins and persistence of international migration posit causal mechanisms at many levels of aggregation, the various explanations are not necessarily contradictory unless one adopts the rigid position that causes must operate at one level and one level only. We find no a priori grounds for such an assertion. As stated earlier, it is entirely possible that
individuals engage in cost-benefit calculations; that households act to diversify labor allocations; and that the socioeconomic context within which these decisions are made is determined by structural forces operating at the national and international levels (Papademetriou and Martin, 1991). Thus, we are skeptical both of atomistic theories that deny the importance of structural constraints on individual decisions, and of structural theories that deny agency to individuals and families.

Rather than adopting the narrow argument of theoretical exclusivity, we adopt the broader position that causal processes relevant to international migration might operate on multiple levels simultaneously, and that sorting out which of the explanations are useful is an empirical and not only a logical task. Each model must be considered on its own terms and its leading tenets examined carefully to derive testable propositions. Only then can we clearly specify the data and methods required to evaluate them empirically.

The neoclassical economic model yields a clear empirical prediction that, in principle, should be readily verifiable: that the volume of international migration is directly and significantly related, over time and across countries, to the size of the international gap in wage rates. Regression analyses testing the theories of Lewis (1954) and Ranis and Fei (1961) should therefore contain transnational wage differentials as the leading predictor, with geographic distance between countries perhaps entered as a proxy for the costs of movement.

Later refinements of the neoclassical model, however, suggest that the pertinent factor in migration decisionmaking is the expected earnings gap, not the absolute real-wage differential (Todaro, 1969, 1976; Todaro and Maruszko, 1987). At any point in time, expected earnings are defined as real earnings in the country under consideration multiplied by the probability of employment there. Although typically estimated as one minus the unemployment rate, the likelihood of employment is probably more appropriately measured as one minus the underemployment rate, given the pervasiveness of sporadic, part-time employment in low-skill jobs within developing regions. The key predictor of international migratory flows is thus an interaction term that cross-multiplies wages and employment probabilities. A statistical test for the significance of this interaction term, compared to a regression model where real wages alone appear, constitutes a critical test comparison between the Ranis–Fei and the Todaro versions of neoclassical theory. (See Todaro, 1980, and Greenwood, 1985, for reviews of the substantial empirical research literature testing the Todaro model.)

A logical corollary of both models, however, is that international movement should not occur in the absence of an international gap in either observed or expected wages, and that movement between countries should cease when wage differentials have been erased (net of the costs of movement, monetary and psychological). International flows that occur in the absence of a wage gap, or that end before a gap has been eliminated, represent anomalous conditions that constitute prima facie evidence challenging the assumptions of neoclassical economic theory.
At the individual level, the Todaro model and its successors predict that individual and household characteristics that are positively related to the rate of remuneration or the probability of employment in destination areas will increase the probability of migration by raising the expected returns to international movement. Hence, the likelihood of emigration is predicted to be reliably related to such standard human capital variables as age, experience, schooling, marital status, and skill. The propensity for international migration is also expected to vary with a household’s access to income-generating resources at home (such as owning land or supporting a business enterprise), since these will affect the net return to movement.

Since human capital variables that affect rates of employment and remuneration in destination areas also tend to affect wage and employment rates in places of origin, a key empirical issue is where the effect of human capital is greater, at home or abroad. Given the fact that international migration involves a change of language, culture, and economic system, human capital acquired at home generally transfers abroad imperfectly (see Chiswick, 1979). In this case, international migrants may be negatively selected with respect to variables such as education and job experience.

Among rural Mexicans, for example, the economic returns to schooling have historically been greater in urban areas of Mexico than in the United States. Whereas an undocumented migrant with a secondary education gets the same minimum-wage job in Los Angeles as one with no schooling at all, that education would qualify the same person for a clerical or white collar job in Mexico City, thereby raising the likelihood of rural-urban migration and lowering the probability of international movement (Taylor, 1987).

This pattern of negative selectivity cannot be hypothesized universally, however, since selection on human capital variables depends on the transferability of the skill or ability under consideration, which itself is determined by social, economic, and historical conditions specific to the countries involved. In general, any social change that affects the market value of human capital in either society has the potential of shifting the size and direction of the relationship between specific predictor variables and the likelihood of international movement.

Thus it is nearly impossible, a priori, to predict the direction of the relationship between an individual background variable and the probability of migration, and it is consequently difficult to derive a convincing test of neoclassical economic theory at the micro level in a reduced-form regression— that is, one in which the probability of migration is modeled directly as a function of individual and household variables. In general, the only universal prediction that can be offered is that human capital should somehow be reliably related to the likelihood of international movement, but the strength and direction of the relationship is impossible to know in the absence of historical information about the countries involved. Only after the historical circumstances have been clearly...
specified and their influence on the returns to specific forms of human capital clarified, can a critical test of the neoclassical microeconomic model be formulated.

A more formal alternative is to model the probability of migration structurally as a function of the expected income differential, and simultaneously model the expected-income differential as a function of individual and household variables. In this way, the effects of individual background variables on migration through their influence on the expected-earnings differential can be tested explicitly. In addition, the possible effects of these variables on migration independent of their influence on expected earnings can be explored (Taylor, 1986). In the absence of structural tests, it is difficult to falsify microeconomic theory by examining individual regressions. The only evidence that could conceivably cast serious doubt on the validity of the human capital theory of migration would be the complete absence of a relationship between human capital and migration.

In contrast to neoclassical economic theory, the new economics of migration focuses on the household or family, rather than the individual, as the relevant decisionmaking unit; and it posits that migration is a response to income risk and to failures in a variety of markets (insurance, credit, labor), which together constrain local income opportunities and inhibit risk-spreading. The most direct test of this theory would be to relate the presence or absence of such market imperfections to households’ propensities to participate in international migration. If the new economics of migration is correct, households confronted by the greatest local market imperfections should be most likely to adopt an international migration strategy, other things being equal.

Unfortunately, other things generally are not equal. Typically there is a high correlation between market imperfections and other variables (namely low wages and incomes) that are the focus of the neoclassical (human capital) migration model. The greatest challenge of this direct test, then, is to isolate the influence of market imperfections and risk on international migration from the role of other income and employment variables.

One of the most distinguishing contributions of the new economics of migration is its integration of migration decisionmaking with migrants’ remittance behavior and households’ remittance use—aspects of migration that hitherto have been treated separately in the literature. If risks to income and a desire to overcome local constraints on production are the driving forces behind migration, then the outcomes of migration (e.g., the patterns and uses of remittances) should reflect this fact. A number of indirect tests of the new economics model are available.

If risk diversification is the underlying motivation, then migrant remittances should be greatest in households most exposed to local income risks and in periods when this risk is most acute (e.g., during a severe drought, as demonstrated by Lucas and Stark, 1985). If a primary motivation of migration is
to overcome risk and credit constraints on local production stemming from market failures, then migration and remittances should positively influence local income-generating activities (Lucas, 1987; Taylor, 1992). Such findings would provide evidence in favor of the new economics of migration, because positive effects of migration on local production activities are ruled out by neoclassical economic theory, as are risk effects. Neoclassical theory focuses on an individual’s maximization of expected income and assumes that markets are complete and well-functioning.

The new economics of migration also places migration within a broader community context, specifically linking a household’s migration decision to its position in the local income distribution. The theory of relative deprivation predicts that a household’s odds of sending migrants abroad are greater the larger the amount of income earned by households above it in the reference income distribution, and more generally, the greater the income inequality in the reference community. A systematic test of this proposition requires a multi-level statistical model that not only contains the usual individual and household-level predictor variables, but also incorporates the community characteristic of income inequality, or an operational measure of relative income. Stark and Taylor (1989) found that relative income was more significant than absolute income in explaining international labor migration within a sample of rural Mexican households, except at the two extremes of the income distribution.

The new economic model can also be tested at the aggregate level. Unlike the neoclassical model, risk diversification allows for movement in the absence of international differences in wages or employment rates, because it links migration not just to conditions in the labor market but to failures in the capital and insurance markets as well. In order to test this conceptualization, regressions predicting international population movements should contain, as independent variables, indicators of the presence or absence of insurance programs (e.g., crop insurance and unemployment insurance), the presence or absence of key markets (e.g., futures and capital markets), levels of market coverage (per capita measures of market participation), and transaction costs (e.g., insurance and interest rates). In general, deficiencies in these ancillary markets are predicted to increase the size of international flows and to raise the likelihood that particular households send migrants abroad, holding constant conditions in the labor market.

Although dual labor market theory posits a bifurcated occupational structure and a dual pattern of economic organization for advanced industrial societies, in practice it has proved difficult to verify this segmented market structure empirically (Cain, 1976; Hodson and Kaufman, 1982). Usually the distinction between “primary” and “secondary” sectors is arbitrary, leading to great instability in empirical estimates and a high degree of dependency of results on the decision rule chosen to allocate jobs to sectors (Tolbert, Horan, and Beck, 1980; Hodson and Kaufman, 1981; Horan, Tolbert, and Beck, 1981; but see Dickens and Lang, 1985, for an exception to this criticism).
Rather than attempting to verify the empirical structure of the labor market, therefore, a more efficacious strategy might be to focus on the theory’s predictions regarding patterns of international movement, which are quite specific and objectively testable. Piore and others argue that immigration is driven by conditions of labor demand rather than supply. In statistical models that regress secular trends in international migration on changing market conditions in sending and receiving countries, one should therefore observe a higher degree of explanatory power among receiving-country indicators compared with those for sending countries. If real wages and employment conditions are entered into an equation predicting movement between Turkey and Germany, for example, German indicators should dominate in terms of predictive power.

Being demand-based, the dual labor market approach also predicts that international flows of labor begin through formal recruitment mechanisms rather than individual efforts. In principle, it should be easy to verify this proposition simply by listing the major international migration flows that have emerged since 1950 and documenting which ones were initiated by formal recruitment procedures, either public or private. If most or all of the flows are traceable to some sort of recruitment program, then a key prediction of dual labor market theory will have been sustained. In his book, Piore does not undertake this exercise; he refers only to several cases that happen to be consistent with his theory (for an example of such an exercise, however, see Massey and Liang, 1989).

One last prediction of dual labor market theory is that secondary-sector wages are flexible downward, but not upward. Over time, therefore, fluctuations in wage rates in jobs filled by immigrants should not be strongly related to fluctuations in labor supply and demand. During periods of low labor immigration and high labor demand, wages in receiving countries should not rise to attract native workers because of institutional rigidities, but during periods of high immigration and low demand there is nothing to prevent wages from falling in response to competitive pressure. We thus expect an interaction between changes in wage rates and whether or not immigration was contracting or expanding during the period: the effect is expected to be zero in the former case and negative in the latter. We also expect a widening wage gap between these jobs and those held by native workers over time.

Although world systems theory constitutes a complex and at times diffuse conceptual structure, it yields several relatively straightforward and testable propositions, the first of which is that international flows of labor follow international flows of capital, only in the opposite direction. According to Sassen and others, emigrants are created by direct foreign investment in developing countries and the disruptions that such investment brings. Thus, we should observe that streams of foreign capital going into peripheral regions are accompanied by corresponding outflows of emigrants.

This basic migratory process should be augmented by the existence of ideological and material ties created by prior colonization as well as ongoing
processes of market penetration. If one were to specify a model of international migration flows to test world systems theory, therefore, one would want to include indicators of prior colonial relationships, the prevalence of common languages, the intensity of trade relations, the existence of transportation and communication links, and the relative frequency of communications and travel between the countries.

Finally, world systems theory specifies not only that international migration should flow from periphery to core along paths of capital investment, but also that it is directed to certain “global cities” that channel and control foreign investment. Although the theory does not provide specific criteria for defining a “global city,” a set of operational criteria might be developed from information about capital assets and corporate headquarters. One could then examine the relative frequency of movement to global cities, as opposed to other places within the developed or developing world.

Network theory leads to a series of eminently testable propositions. According to Piore, Massey, and others, once someone has migrated internationally, he or she is very likely to do so again, leading to repeated movements over time. Thus the likelihood of an additional trip should increase with each trip taken; the probability of transnational migration should be greater among those with prior international experience than among those without it; and the likelihood of additional migration should increase as the amount of foreign experience rises.

A second proposition is that controlling for a person’s individual migrant experience, the probability of international migration should be greater for individuals who are related to someone who has prior international experience, or for individuals connected to someone who is actually living abroad. Moreover, the likelihood of movement should increase with the closeness of the relationship (i.e., having a brother in Germany is more likely to induce a Turk to migrate there than having a cousin, a neighbor, or a friend); and it should also rise with the quality of the social capital embodied in the relationship (having a brother who has lived in Germany for ten years is more valuable to a potential emigrant than having one who has just arrived, and having one who is a legal resident is better than having one who lacks residence documents).

Another hypothesis stems from the recognition that international movement requires migrants to overcome more barriers than does internal movement. In addition to the normal costs of travel and searching for work are the costs of learning and adapting to a new culture, the costs of acquiring appropriate documentation, and, if acquiring legal papers is impossible, of evading arrest and deportation. In general, the greater the barriers to movement, the more important should network ties become in promoting migration, since they reduce the costs and risks of movement. We should thus observe that network connections are systematically more powerful in predicting international migration than internal migration. Taylor (1986) finds this differentiated effect of migration networks for a sample of Mexican households.
Within households, we should also be able to detect the effect of social capital on individual migration behavior. In general, members of households in which someone has already migrated abroad should display higher probabilities of movement than those from households that lack migratory experience. If network theory is correct, for example, a common vector by which migratory behavior is transmitted is from fathers to sons (Massey et al., 1987). Dependent sons whose fathers are active or former international migrants should be more likely to emigrate than those whose fathers lack foreign experience.

Finally, at the community level, one should be able to observe the effect of the prevalence of network ties. People should be more likely to migrate abroad if they come from a community where many people have migrated and where a large stock of foreign experience has accumulated than if they come from a place where international migration is relatively uncommon (Massey and García España, 1987). Moreover, as the stock of social ties and international migrant experience grows over time, migration should become progressively less selective and spread from the middle to the lower segments of the socioeconomic hierarchy. In general, then, individual or household migration decisions need to be placed within a local setting, suggesting the need for multi-level analytic models incorporating indexes of network connections within the community.

Institutional theory argues that disparities between the supply of and demand for entry visas into core receiving societies create a lucrative niche for entrepreneurs to provide licit and illicit entry services, and that the exploitation that results from this disparity will also prompt humanitarian organizations to intervene on immigrants’ behalf. The establishment and growth of institutions dedicated to facilitating immigration constitutes another form of social infrastructure that persists over time and increases the volume of international population movements.

Although it may be feasible through case studies to document such institutional development and its effect on immigration, it is more difficult to link institutions to aggregate population flows or micro-level migration decisions in an analytically rigorous fashion. On special surveys, migrants and nonmigrants might be asked whether they are aware of institutions providing support to immigrants, and responses to this question may be used to predict the likelihood of movement. Or the presence of such organizations might be documented across communities and used to predict the rate of outmigration at the community level, or, in a multi-level model, the probability of emigration at the individual or household level.

Lastly, the theory of cumulative causation states the general hypothesis that migration sustains itself in such a way that migration tends to create more migration. This hypothesis follows from the proposition that individual or household decisions are affected by the socioeconomic context within which they are made, and that acts of migration at one point in time affect the context within which subsequent decisions are made. Migration decisions made by families and individuals influence social and economic structures within the
community, which influence later decisions by other individuals and households. On balance, the changes at the community level increase the odds of subsequent movement, leading to migration’s cumulative causation over time (Massey et al., 1987; Massey, 1990b).

The systematic testing of this theory poses substantial data demands. In order to test for cumulative causation at the aggregate level using cross-sectional data, complicated recursive systems of structural equations must be specified, and these typically require instrumental variables that are difficult to define and identify, especially in international data sets. Ideally the theory should be tested using multi-level longitudinal data, which contain variables defined at the individual, household, community, and perhaps even national levels, all measured at different points in time. Only with such a data set can the reciprocal feedback effects of individual or household decisions on social structure be discerned and measured.

The theory of cumulative causation, while in many ways still rudimentary in its development, does point to several factors as particularly important in channeling the feedback between individual behavior and community structure. The first factor is migrant networks, suggesting the need to gather detailed information about kin and friendship ties between migrants and nonmigrants. A second factor is income equality, which requires the accurate measurement of household income. A third is land distribution, which requires detailed data on land tenure and ownership. A fourth, pertaining only to rural areas, is the nature of agrarian production, which requires information on the use of irrigation, machinery, hired labor, herbicides, pesticides, and improved seeds by both migrant and nonmigrant families. The last and perhaps most difficult factor to measure in testing for cumulative causation is culture, which requires information about beliefs, values, and normative practices.

Ideally all of these factors should be measured longitudinally, although in some cases—culture, for example—this would be next to impossible. Given the difficulty of securing longitudinal information on changes in the prevalence of migrant networks, the degree of income inequality, the skewness of land distribution, and the capital intensiveness of agricultural production, an alternative strategy might be to rely on geographic diversity in these factors across communities, specifying recursive structural equation systems to model the feedbacks, but this approach raises serious technical issues with respect to identification and instrumentation.

The final conceptual scheme we discussed was the systems perspective, which argues that causal forces operating at a variety of levels lend a degree of permanence to international flows and over time lead to the emergence of stable migration systems. These systems are characterized by relatively large flows of migrants between member countries compared to flows from outside the system. Verifying the existence of such systems is a straightforward empirical matter of establishing some threshold of intensity for inclusion of a flow within a systemic
structure, and then applying it to identify those prevailing in the world today. Some efforts along these lines have already been attempted (Zlotnik, 1992).

Conclusion

Theories developed to understand contemporary processes of international migration posit causal mechanisms that operate at widely divergent levels of analysis. Although the propositions, assumptions, and hypotheses derived from each perspective are not inherently contradictory, they nonetheless carry very different implications for policy formulation. Depending on which model is supported and under what circumstances, a social scientist might recommend that policymakers attempt to regulate international migration by changing wages and employment conditions in destination countries; by promoting economic development in origin countries; by establishing programs of social insurance in sending societies; by reducing income inequality in places of origin; by improving futures or capital markets in developing regions; or by some combination of these actions. Or one might advise that all of these programs are fruitless given the structural imperatives for international movement growing out of market economic relations.

Whatever the case, given the size and scale of contemporary migration flows, and given the potential for misunderstanding and conflict inherent in the emergence of diverse, multi-ethnic societies around the world, political decisions about international migration will be among the most important made over the next two decades. Likewise, sorting out the relative empirical support for each of the theoretical schemes and integrating them in light of that evaluation will be among the most important tasks carried out by social scientists in ensuing years. We hope that by explicating the leading theories of international migration and by clarifying their underlying assumptions and key propositions, we have laid the groundwork for that necessary empirical work.

Note

The authors are members of the IUSSP Committee on South-North Migration, which is currently undertaking a systematic examination of theories of international migration and the evidence supporting them. The Committee is chaired by Douglas S. Massey, who took primary responsibility for writing the text of this presentation, but the ideas, concepts, and conclusions expressed in the article are the collective work of all committee members. The committee welcomes comments and criticisms from interested readers.

References


DOUGLAS S. MASSEY ET AL.


