2 Why Does Immigration Occur?
A Theoretical Synthesis

Douglas S. Massey

The modern history of international migration can be divided roughly into four periods. During the mercantile period, from 1500 to 1800, world immigration flows were dominated by Europe and stemmed from processes of colonization and economic growth under mercantile capitalism. Over the course of three hundred years, Europeans inhabited large portions of the Americas, Africa, Asia, and Oceania. Although the exact number of colonizing emigrants is unknown, the outflow was sufficient to establish Europe’s dominance over large parts of the world. During this period, emigrants generally fell into three classes: a relatively large number of agrarian settlers, a smaller number of administrators and artisans, and an even smaller number of entrepreneurs who founded plantations to produce raw materials for Europe’s growing mercantile economies.

Although the number of Europeans involved in plantation production was small, this sector had a profound impact on the size and composition of population in the Americas. Given preindustrial technology, plantations required large amounts of cheap labor, a demand met partially by indentured workers from East Asia; the most important source of labor, however, came from the forced migration of African slaves. Over three centuries, nearly ten million Africans were imported into the Americas, and together with European colonists, they radically transformed its social and demographic composition.

The second period of emigration, the industrial period, began early in the nineteenth century and stemmed from the economic development of Europe and the spread of industrialism to former colonies in the New World. From 1800 to 1925, more than forty-eight-million people left the industrializing countries of Europe in search of new lives in the Americas and Oceania. Of these emigrants, 85 percent went to just five destinations: Argentina, Australia, Canada, New Zealand, and the United States (the last alone receiving 60 percent). Key sending nations were Britain, Italy, Norway, Portugal, Spain, and Sweden, which all exported a large share of their potential population in the course of industrializing. Although international migrants were not exclusively European, the overwhelming majority came from that continent. Of all U.S. immigrants between 1820 and 1920, for example, 88 percent were from Europe, 3 percent were from Asia, and 8 percent came from the Americas.

The period of large-scale European emigration faltered with the outbreak of World War I, which brought European emigration to an abrupt halt and ushered in a four-decade period of limited migration. Although emigration revived somewhat during the early 1920s, by then several important receiving countries (most notably the United States) had passed restrictive immigration laws. The onset of the Great Depression stopped virtually all international movement in 1929, and except for a small amount of return migration, there was little movement during the 1930s. During the 1940s international migration was checked by the Second World War. What mobility there was consisted largely of the movements of refugees and displaced persons and was not tied strongly to the rhythms of economic growth and development; this pattern persisted well into the subsequent decade.

The period of postindustrial migration emerged during the 1960s and constituted a sharp break with the past. Rather than being dominated by outflows from Europe to a handful of former colonies, immigration became a truly global phenomenon as the number and variety of both sending and receiving countries increased and the global supply of immigrants shifted from Europe to the developing world. Whereas migration during the industrial era brought people from densely settled, rapidly industrializing areas into sparsely settled, rapidly industrializing nations, migration in the postindustrial era brought people from densely settled countries at the earliest stages of industrialization into densely settled, economically mature, postindustrial societies.
Before 1925, 85 percent of all international migrants originated in Europe, but since 1960 Europe has contributed an increasingly small fraction of emigrants to world migration flows as emigration from Africa, Asia, and Latin America has increased dramatically. The number and variety of destination countries have likewise grown. In addition to traditional immigrant-receiving nations such as Canada, the United States, Australia, New Zealand, and Argentina, countries throughout Western Europe now attract significant numbers of immigrants—notably Germany, France, Belgium, Switzerland, Sweden, and the Netherlands. During the late 1970s even longtime sending countries such as Italy, Spain, and Portugal began receiving immigrants from the Middle East and Africa, and after the rapid escalation of oil prices in 1973, several less developed but capital-rich nations in the Gulf region also began to sponsor massive labor migration. By the 1980s international migration had spread into Asia, not just to Japan but also to newly industrialized countries such as Korea, Taiwan, Hong Kong, Singapore, Malaysia, and Thailand.

By 1990, therefore, international migration had become truly a global phenomenon, and reflecting this new salience, social scientists from a variety of disciplines sought to formulate new theories of migration to supplement those developed during the earlier industrial era. Like prior efforts, these theories sought to explain why international immigration began and how it persisted across space and time; but rather than seeking to account for a massive exodus from a rapidly industrializing Europe into rapidly industrializing former colonies, they sought to explicate a more complex migration regime that involved population flows from industrializing to mature economies, a diversity of sources and destinations, vastly reduced costs of transportation, cheaper and more rapid communications, rising governmental intervention, and greater circularity of movements.

Theories of International Migration

Neoclassical Economics

The oldest and best-known theory of international migration has its roots in models developed originally to explain internal labor migration in the process of economic development (Lewis 1954; Ranis and Fei 1961). According to this theory and its extensions, international migration, like its internal counterpart, is caused by geographic differences in the supply of and demand for labor (Todaro and Maruszko 1987). A country with a large endowment of labor relative to capital will have a low equilibrium wage, while a nation with a limited endowment of labor relative to capital will be 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 or labor-surplus country to move to the high-wage or labor-scarce country. As a result of this movement, the supply of labor decreases and wages eventually rise in the capital-poor country, while the supply of labor increases and wages ultimately fall in the capital-rich country, leading, at equilibrium, to an international wage differential that reflects the costs of international movement, both 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 includes the migration of human capital, with highly skilled workers going 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 unskilled labor, therefore, must be kept conceptually distinct from the associated international flow of human capital. Even in the most aggregated macrolevel models, the heterogeneity of immigrants along skill lines must be clearly recognized, although the line between skilled and unskilled workers often is not easy to draw.

Associated with this macroeconomic theory is an accompanying microeconomic model of individual choice (Borjas 1989; Todaro 1969, 1976, 1989). 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 (Sjaastad 1962). People choose to move to where they can be most productive, given their skills, but before they can reap 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 (Todaro and Mauszko 1987).

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 (Bojjas 1989, 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 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. In theory, a potential migrant goes to wherever the expected net returns to migration are greatest.

**The New Economics of Migration**

In recent years a “new economics of labor migration” has arisen to challenge many of the assumptions and conclusions of neoclassical theory (Stark and Bloom 1985). A key insight of this approach is that migration decisions are made not by isolated individual actors but within larger units of interrelated people—typically families or households but sometimes entire communities—in which people act collectively to maximize not only expected income but also minimize risks to income and maximize status within an embedded hierarchy and to overcome a variety of local market failures (Stark 1991; Taylor 1986, 1987).

In most developed countries, risks to household income are managed through institutional mechanisms. Crop insurance programs and futures markets give farmers a way of protecting themselves against natural disasters and price fluctuations, whereas unemployment insurance and government transfers protect workers against the vagaries of the business cycle and structural economic change. Private or government-sponsored retirement programs, meanwhile, protect citizens against the risk of old-age poverty through regular contributions or taxation, in essence spreading a person’s income out over time.

Households are in a better position than individuals to control risks to economic well-being by diversifying the allocation of productive resources. Although individuals may use sequential short periods of seasonal migration to diversify the allocation of their labor over time, this strategy is less feasible for international migrants, who face large sunk costs of movement. Households, however, can easily diversify their sources of income by allocating different family workers to different labor markets. Some members may work in the local economy; others may work elsewhere in the same country; and still others may work abroad. As long as economic conditions in the nonlocal labor markets are negatively or weakly associated with those in the home community, households will be able to control risks through geographic diversification. In the event that economic conditions at home deteriorate and productive activities there fail to generate sufficient income, the household can rely on migrant remittances for support.

Markets for credit and capital also tend to be complete and to function well in developed countries, giving most families a means of borrowing to smooth consumption or undertake investment. In the absence of a sound and efficient banking system, however, international migration becomes attractive as a strategy for accumulating funds that can be used in lieu of borrowing. Households simply send one or more workers abroad to accumulate savings or send them home in the form of remittances. Although most migrant savings and remittances go toward consumption, some of the funds are inevitably also channeled into productive investment.

A key proposition of the new economic model 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 increase total income. The new economics of migration also questions the assumption that income has a constant effect on utility across socioeconomic settings—that is, that a $100 real increase in income means the same thing to a person regardless of community conditions and irrespective of his or her position in the local income distribution.

Theorists of the new economics argue that households send workers abroad not only to im-
prove their incomes in absolute terms but also to increase them relative to other households, and hence to reduce their relative deprivation compared with some reference group (see Stark and Taylor 1989, 1991; Stark 1991). A household’s sense of relative deprivation depends on the incomes of which it is deprived in the reference-group income distribution. To illustrate, 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.

**Segmented Labor Market Theory**

Although neoclassical theory and the new economics of migration lead to divergent conclusions about the origins and nature of international migration, both are essentially microlevel decision models. They differ in the units assumed to make the decision (the individual or the household), the entity being maximized or minimized (income or risk), their 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 segmented labor market theory, which discounts decisions made by individuals and argues that international migration stems from the intrinsic labor demands of modern industrial societies.

Michael Piore (1979) has been the most forceful and eloquent proponent of this theoretical viewpoint, arguing that international migration is caused by a permanent labor demand 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 low-wage workers). The built-in demand for inexpensive and flexible labor stems from four fundamental characteristics of advanced industrial societies and their economies.

The first is structural inflation. Wages not only reflect conditions of supply and demand but confer status and prestige, social qualities that inhere in 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, and company job classifications) ensure 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, may demand a corresponding wage increase, thus threatening the position of cooks, who will 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.

The demand for cheap, flexible labor is also augmented by social constraints on motivation embedded within occupational hierarchies. Most
people work not only to generate income but also to accumulate social status. Acute motivational problems arise at the bottom of any job hierarchy because there is no status to be maintained and few avenues for upward mobility. The problem is inescapable and structural because the bottom cannot be eliminated from labor markets. 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 disjunction 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; 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.

The demand for immigrant labor also stems from the inherent duality of 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 because performing them well requires considerable knowledge and experience, firm-specific human capital is accumulated. 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 idling (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. The first act of secondary-sector employers during down cycles is to 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.

In their analysis of the process by which Cuban immigrants were incorporated into the United States, Alejandro Portes and Robert Bach (1985) uncovered evidence of a third employment sector that blends features of primary and secondary labor markets and also yields a structural demand for immigrant labor. Like the secondary sector, ethnic enclaves contain low-status jobs characterized by low pay, chronic instability, and unpleasant working conditions, jobs that are routinely shunned by natives. Unlike the secondary sector, however, the enclave provides immigrants with significant economic returns to education and experience, as well as the very real prospect of upward socioeconomic mobility, thus replicating features of the primary sector.

Not all immigrations produce ethnic enclaves,
however, and studies suggest they are rather hard to create (Logan, Alba, and McNulty 1994). Indeed, the ethnic enclaves identified to this point have formed under rather unusual circumstances of geographic concentration and distinctively timed, class-selective immigration (Portes and Stepick 1993; Wilson and Martin 1982). In general, an enclave economy emerges when an initial wave of elite immigrants possessing significant amounts of financial, human, social, and cultural capital concentrates disproportionately in one urban area and, after becoming established there and founding new business enterprises, employs successive waves of lower-status but aspiring immigrants from the same country.

The existence of a large, concentrated population of co-ethnics creates a demand for specialized cultural products and ethnic services that immigrant entrepreneurs are uniquely qualified to fill. In addition, their privileged access to a growing pool of low-wage immigrant labor gives them an advantage when competing with firms outside the enclave. Immigrants working in the enclave trade low wages and the acceptance of strict discipline on arrival for a greater chance of advancement and independence later on (Portes and Bach 1985). The implicit contract between employers and workers stems from a norm of ethnic solidarity (a form of cultural capital), which suffuses and supports the enclave (Portes and Manning 1986; Portes and Rumbaut 1990). At the same time, social networks and personal linkages to other entrepreneurs (a form of social capital) launch new immigrants on independent careers in small business, and once established, these new entrepreneurs are expected to help and promote other immigrants in return (cultural capital again). The prospects for enclave formation are especially enhanced if the initial immigrants are well educated and possess organizational or business skills (human capital), or if they have access to savings, credit, or government assistance (financial capital).

Although an ethnic enclave may begin with the immigration of entrepreneurs, in order to function effectively over time it requires a steady stream of new workers willing to trade low initial wages for the possibility of later mobility, yielding an independent structural source of labor demand for immigrant workers and complementing that emanating from the secondary sector. As immigrant entrepreneurs arrive, concentrate geographically, and establish new business enterprises that rely on immigrant labor for their survival, immigration can, quite literally, generate its own demand. The problems of motivation and structural inflation inherent in 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 little chance for advancement. In the past this demand was met partially by three sets of people with social statuses and characteristics conducive to these sorts of jobs: women, teenagers, and rural-to-urban migrants.

Historically women have tended to participate in the labor force up to the time of the birth of their first child, and to a lesser extent after their children have grown. They have 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 the low wages and instability because they viewed the work as transient and the earnings as supplemental; the positions they held were not threatening to their main social status, which was 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 (by giving them clothes, cars, music, and so on); the job is just a means to an end.

Finally, rural areas of developed nations for many years provided industrial cities with a steady supply of low-wage workers. Movement from a social and economic backwater to the dynamism and excitement of the city created a sense of upward mobility and personal improvement regardless of the modesty of one’s circumstances after migrating, and even menial unskilled jobs in cities provided access to housing, food, and consumer goods that represented a real step up in the world for impoverished migrants from the countryside. As long as large reserves of rural population existed, new industrial nations could look internally to satisfy the emerging demand for unskilled, low-wage labor.
In advanced industrial societies, however, these sources of entry-level workers have shrunk over time because of four fundamental sociodemographic 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; the decline in birth rates and the extension of formal education, which have produced very small cohorts of teenagers entering the labor force; and the urbanization of society, which has eliminated farms and rural communities as potential sources for new migrants to the city. The imbalance between the structural demand for entry-level workers and the limited domestic supply of such workers has generated an underlying, long-run demand for immigrants in developed countries.

Segmented 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.

World Systems Theory

A social theory approach emerged during the 1950s in response to functionalist theories of social change and development, which held that countries develop economically by progressing through an orderly series of evolutionary stages culminating in modernization and industrialization. In contrast, historical-structural theorists argued that because political power is unequally distributed across nations, the expansion of global capitalism acts to perpetuate inequalities and reinforce a stratified economic order. Rather than experiencing an inexorable progression toward development and modernization, poor countries in reality are trapped by their disadvantaged position within an unequal geopolitical structure, which perpetuates their poverty.

Historical-structural theory reached its peak of influence during the 1960s and 1970s, and it gained particular currency among social scientists in Latin America. Theorists such as César Furtado (1965, 1970) and Fernando Cardoso and Enzo Faletto (1969, 1979) observed a deterioration in the terms of trade between wealthy capitalist countries and poor nations in the years after World War II and concluded that developing nations were being forced into dependency by structural conditions dictated to them by powerful capitalist countries. According to André Goudreault (1969) the forces of global capitalism acted to “develop underdevelopment” within the Third World. This line of historical-structural thinking became known as dependency theory and ultimately embraced a diverse group of scholars who drew inspiration from the work of Paul Baran (1957, 1973) and his conceptualization of the ideas of Marx and Lenin.

A second line of historical-structural theory emerged somewhat later and drew on the work of the dependency theorists, as well as the historiography of the French social historian Fernand Braudel (1981, 1982). Its leading exponent was Immanuel Wallerstein (1974, 1980), who undertook a comprehensive analysis of the global expansion of the capitalist system from the sixteenth century onward. Wallerstein sought to reconstruct the historical processes by which unequal political and economic structures were created and extended throughout the world, and the mechanisms by which noncapitalist and precapitalist regions were incorporated into the global market economy. He classified countries according to their degree of dependency on the dominant capitalist powers, which he termed “core” nations. Those on the “periphery” were the most dependent, whereas those on the “semiperiphery” had slightly more independence in the global marketplace. Nations in the “external arena” remained isolated and largely outside the global capitalist system. Given the scope of this work and its sweeping vision of an expanding global capitalism, this line of thought became known generally as “world systems theory” (Simmons 1989).

At first neither world systems theorists nor dependency theorists had much interest in international migration. During the 1960s and 1970s they focused instead on rural-to-urban migrants and their insertion into the informal urban economy. In contrast to economic theorists such as Michael Todaro (1969, 1976) and sociological theorists such as Everett Lee (1966), who viewed migration as a rational calculation made by individuals to se-
cure material improvement, historical-structural theorists linked migration to the macro-organization of socioeconomic relations, the geographic division of labor, and the political mechanisms of power and domination (see Singer 1971, 1975; Stern 1988).

Only after the economic recessions of the mid-1970s did observers begin to understand that international flows were not just a "temporary" aberration, and that international migration also might be linked to the structural changes that accompanied a nation's insertion into the global marketplace. As the absorption of foreigners emerged as a political issue throughout the developed world, scholars began to apply the precepts of historical-structural theory to analyze the flows of labor they were suddenly observing with new eyes. Like their intellectual forebears, theorists such as Alejandro Portes and John Walton (1981), Elizabeth Petras (1981), Saskia Sassen (1988), and Ewa Morawinska (1990) sought to explain international migration not as a product of individual or household decisions but as a structural consequence of the expansion of markets within a global political hierarchy. Although their propensions up to now have not formed a coherent theory so much as a general approach to the study of international migration, historical-structural explanations are generally synthesized under the rubric of "world systems theory."

This theory argues that the penetration of capitalist economic relations into noncapitalist or precapitalist societies creates a mobile population that is prone to migrate. Driven by a desire for higher profits and greater wealth, owners and managers of capitalist firms in core countries enter poorer nations on the periphery of the world economy in search of land, raw materials, labor, and 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.

World systems theorists posit that international migration emerges in response to the 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 proportions of the human population have been incorporated into the market economy. To explain how migration flows are generated, some of which have moved abroad, theorists have considered how the influence of global markets is exerted over land, raw materials, and labor within peripheral regions (Massey 1988).

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 because the use of modern inputs produces high crop yields at low unit prices, small, noncapitalist farmers are driven 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.

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 norms 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.

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; since the new factory work is demanding and poorly paid, however, women tend to work only 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 local 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 into 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, military, 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.

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 but promote the movement of people by reducing the costs of movement along certain international pathways. Because investment and globalization are inevitably accompanied by the buildup of a transportation and communication infrastructure, the international movement of labor generally follows the international movement of goods and capital in the opposite direction.

The creation and perpetuation of a global trading regime requires an underlying system of international security. Core capitalist nations have both an economic interest in geopolitical order and the military means of preserving it; most of the leading powers maintain relatively large armed forces to deploy as needed to preserve the integrity of the global capitalist system. Threats to the stability of that system are often met by military force projected from one or more of the core nations. After 1945, for example, the threat of Communist expansion in Europe presented such a threat to a weakened capitalism that Britain, France, and the United States stationed large numbers of troops permanently on bases throughout the continent. They also periodically dispatched military forces to various hot spots that cropped up in Africa, the Middle East, Asia, and Latin America over the course of the forty-year cold war with the Soviet Union.

As the leading economic and political power in the postwar capitalist world, the United States has played the most important role in preserving international peace and security in the postwar world, maintaining a uniquely large military establishment and frequently dispatching its armed forces to counter leftist insurgencies, turn back Communist invasions, or quell outbreaks of violence that threaten the capitalist order. In the years since 1945 the United States has intervened covertly or overtly in at least a dozen countries: Iran, Guatemala, Nicaragua, Cuba, the Dominican Republic, Haiti, Grenada, Chile, Somalia, and, of course, Vietnam, Cambodia, and Korea. Most recently it led a multinational force to restore order (and the flow of oil) in Kuwait. In order to retain its capacity to project military power whenever and wherever it is needed, the United States maintains many large military bases in far-flung areas throughout the world.

Each military base and armed intervention, however, creates a range of social and political connections that promote the subsequent movement of immigrants. Soldiers often acquire local spouses who seek to accompany them home when their tour of duty ends, and in the United States (and many other countries) spouses have a privileged claim on entry by virtue of their marriage to a citizen. Spouses, in turn, may seek to sponsor the immigration of their brothers, sisters, mothers, fathers, and minor children. These people have their own claims on entry by virtue of their kinship with a legal resident and, if the spouse ultimately naturalizes, by virtue of their kinship with a citizen.

Large-scale military operations also involve the hiring of numerous support personnel from the local civilian population, creating personal relationships, political debts, and moral obligations that may be invoked to gain access to immigrant visas when the military departs or to obtain refugee status in the event that the client state collapses. If the military presence is long-term, moreover, a host of commercial and service establishments inevitably grow up around the base, further expanding the range of personal interactions and social debts and transmitting new linguistic codes and cultural conventions into the local population. Intensive contact with U.S. troops not only increases the odds of marriage but also inculcates a knowledge of English and U.S. culture that raises the
potential rewards of working in the United States and increases the motivation to do so. For these reasons, therefore, significant military deployments are typically accompanied by sizable return flows of immigrants (Donato 1991; Jasso and Rosenzweig 1990; Schmidl 1997).

Finally, economic globalization creates ideological and cultural links between core capitalist countries and their peripheries through other means besides military intervention. In many cases, these cultural links are long-standing, reflecting a colonial past when 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 U.S. 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 consumer 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 and communication infrastructure to channel international migration to particular core countries.

The world economy is managed from a relatively small number of urban centers where 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 the services of unskilled workers (busboys, gardeners, 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 provide. 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.

Social Capital Theory

The economist Glenn Loury (1977) introduced the concept of social capital to designate a set of intangible resources in families and communities that help to promote social development among young people, but Pierre Bourdieu (1986) pointed out its broader relevance to human society. According to Bourdieu and Loïc Wacquant (1992, 119), "social capital is the sum of the resources, actual or virtual, that accrue to an individual or a group by virtue of possessing a durable network of more or less institutionalized relationships of mutual acquaintance and recognition."

The key characteristic of social capital is its convertibility: it may be translated into other forms of capital, notably financial capital—in this case, foreign wages and the remittances they permit (Harker, Mahar, and Wilkes 1990). People gain access to social capital through membership in networks and social institutions and then convert it into other forms of capital to improve or maintain their position in society (Bourdieu 1986; Coleman 1990). Although Alejandro Portes and Julia Sassenbrenner (1993) point out that social capital may have negative as well as positive consequences for the individual, theorists generally have emphasized the positive role it plays in the acquisition and accumulation of other forms of capital (see Coleman 1988, 1990), an emphasis that has been particularly strong in work on migrant networks.

Migrant networks are sets of interpersonal ties that connect migrants, former migrants, and non-
migrants 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 on to gain access to various kinds of financial capital: foreign employment, high wages, and the possibility of accumulating savings and sending remittances.

Beginning in the 1920s, sociologists recognized the importance of networks in promoting international movement (see Gamio 1930; Thomas and Znaniecki 1918–20). Drawing on social ties to relatives and friends who have migrated before, nonmigrants gained access to knowledge, assistance, and other resources that facilitated movement (Choldin 1973). Charles Tilly and Charles H. Brown (1967) referred to these ties as the “auspices” of migration; others have labeled them “migration chains” (MacDonald and MacDonald 1974); and Mildred Lev and Walter Wadycki (1973) have called them a “family and friends effect.” Edward Taylor (1986, 1987) characterizes them as a form of economic “migration capital.” My associates and I (1987, 170) appear to have been the first to identify migrant networks specifically as a form of social capital.

Following Coleman’s (1990, 304) dictum that “social capital . . . is created when the relations among persons change in ways that facilitate action,” My associates and I (1987) identified migration itself as the catalyst for this change in the nature of social relations. Everyday ties of friendship and kinship provide few advantages, in and of themselves, to people seeking to migrate abroad. Once someone in a personal network has migrated, however, the ties are transformed into a resource that can be used to gain access to foreign employment and all that it brings. Each act of migration creates social capital among people to whom the new migrant is related, thereby raising the odds of their migration (Massey et al. 1987; Massey, Goldring, and Durand 1994).

The first migrants who leave for a new destination have no social ties to draw on, 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 on the obligations implicit in relationships such as kinship and friendship to gain access to employment and assistance at the point of destination.

Networks make international migration extremely attractive as a strategy for risk diversification or utility maximization. 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 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.

Once international migration has begun, private institutions and voluntary organizations also tend to arise to satisfy the demand created by a growing 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 (Hagan and Gonzalez-Baker 1993).

Jon Goss and Bruce Lindquist (1995) point to migrant institutions as a structural complement to migrant networks, arguing that interpersonal ties are not the only means by which international movement is perpetuated. Building on ideas put forth by Anthony Giddens (1990), they argue (1995, 345) that international migration is best examined not as a result of individual motivations and structural determinations, although these must play a part in any explanation, but as the articulation of agents with particular interests and playing specific roles within an institutional environment, drawing knowledgeably upon sets of rules in order to increase access to resources.
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 (Prothero 1990). 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 (Christiansen 1996). Over time individuals, firms, and organizations become well known to immigrants and institutionally stable, constituting another form of social capital that migrants can draw on to gain access to foreign labor markets. Recruiting agents can at times be active in creating new flows of migration from areas of labor surplus to areas of labor scarcity.

The recognition of a gradual buildup of institutions, organizations, and entrepreneurs dedicated to arranging immigrant entry, legal or illegal, yields hypotheses that are once again quite distinct from those emanating from macrolevel decision models. The theory of social capital accepts the view of international migration as an individual or household decision 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.

**Cumulative Causation**

The theory of cumulative causation argues that over time international migration tends to sustain itself in ways that make additional movement progressively more likely, a process that was first identified by Gunnar Myrdal (1957) and that I later reintroduced to the field (Massey 1990). Causation is cumulative in the sense 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 eight ways that migration is affected in this cumulative fashion: the expansion of networks, the distribution of income, the distribution of land, the organization of agriculture, culture, the regional distribution of human capital, the social meaning of work, and the structure of production. Feedbacks through other variables are also possible but have not been systematically treated.

As just discussed, once the number of network connections in an origin area reaches a critical threshold, migration tends to become self-perpetuating because each act of migration creates the social structure needed to sustain it. Every new migrant reduces the costs and risks of subsequent migration for a set of friends and relatives, and some of these people are thereby induced to migrate, thus further expanding the set of people with ties abroad and, in turn, reducing costs for a new set of people, some of whom are now more likely to decide to migrate, and so on. Over time migratory behavior spreads outward to encompass broader segments of the sending society (Hugo 1981; Massey 1990; Massey, Goldring, and Durand 1994; Taylor 1986).

The new economics of migration argues that as a household's sense of relative deprivation increases, so does the motivation to migrate. After one or two households have begun participating in foreign wage labor, the resulting remittances increase their incomes greatly. Given the costs and risks associated with international movement, moreover, the first households to migrate tend to be located in the middle or upper ranges of the local income hierarchy (Massey, Goldring, and Durand 1994). 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; income inequality is thus further exacerbated and the sense of relative deprivation among nonmigrants increases, inducing still more families to migrate, and so on (Stark 1991; Stark and Taylor 1989; Taylor 1992).

An important spending target for many migrants, especially those from rural communities, is the purchase of land. But land is often 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 out-migration. The more out-migration there is, the more people there are with access to the funds necessary to buy land, leading to additional purchases by migrants and more land withdrawn from production, creating land shortages and price infla-
tion and still more pressure for out-migration (Mines 1984; Reichert 1981; Rhoades 1978; Wiest 1984).

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 out-movement (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.

As migration grows in prevalence within a community or nation, 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 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 (Alarcón 1992; Brettell 1979; Goldring 1996a; Massey et al. 1987; Rouse 1989, 1991).

Migration is a selective process that often tends, initially at least, to draw relatively well-educated, skilled, productive, and highly motivated people away from sending communities, although, as pointed out earlier, migration tends to become less selective over time as the costs and risks fall because of network formation; moreover, this initial selectivity depends critically on the characteristics of migrant labor markets (Taylor 1987). Sustained out-migration thus may lead 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 (Greenwood 1981, 1985; Greenwood, Hunt, and McDowell 1987; Myrdal 1957). 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.

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 stigmatized and viewed as culturally inappropriate for native workers (Böhringer 1972, 1984; 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.

In any bounded population, of course, processes of cumulative causation cannot continue ad infinitum. If migration continues long enough, networks eventually reach a point of numerical saturation within any particular community. More and more community members reside in branch settlements overseas, and virtually all of those at home are connected to someone living abroad or having substantial foreign experience. When networks reach such a high level of elaboration, the costs of migration do not fall as sharply with each new migrant, and migration loses its dynamic momentum for growth. The prevalence of migration in the community approaches an upper limit, and migratory experience becomes so diffused that the stock of potential new migrants becomes very small and is increasingly composed of women, children, and the elderly.

If migration continues long enough, local labor shortages and rising wages in the home commu-
nity may further dampen the pressures for emigration (Gregory 1986), causing the rate of entry into the international migrant workforce to trail off (Hatton and Williamson 1994a). Observed at the national level, this trend may be difficult to detect because new communities are continuously incorporated into the migratory stream. As the rate of out-migration decelerates in places with longer histories of migration, new areas are drawn into transnational circuits and their rates of migration begin to accelerate. As a result, the total outflow from the nation as a whole may continue to grow as migration spreads from place to place.

Nevertheless, migratory experience eventually becomes widely diffused even across communities, and observers have identified the historical emergence of a characteristic “migration curve” in national populations that have made the transition from emigration to immigration. According to Sune Ackerman (1976), this curve starts at low levels and rises to a peak before declining, yielding what Philip Martin and Edward Taylor (1996) have called a “migration hump” that countries experience in the course of economic development. Timothy Hatton and Jeffrey Williamson (1994b, 9–10) note that “the upswing of the emigration cycle usually coincide[s] with industrialization and rising real wages at home [as] demographic forces, industrialization, and the mounting stock of previous emigrants abroad all serve to drive up the emigration rate. As these forces weakened [historically], the narrowing gap between home and foreign wages began to dominate and emigration receded.”

**A SYNTHETIC THEORETICAL ACCOUNT**

Because the above theories posit causal mechanisms operating at multiple levels of aggregation, the various explanations are not necessarily contradictory. It is entirely possible for individuals to engage in cost-benefit calculations; for households to minimize risk and overcome barriers to capital and credit; for both individuals and households to draw on social capital to facilitate international movement; and for the socioeconomic context within which migration decisions are made to be determined by structural forces operating at the national and international levels, often influenced by migration itself. Thus, a synthetic approach to theory construction would seem to be in order.

I recently participated in two major intellectual projects that sought to achieve precisely this kind of synthesis. The first was the Committee on South-North Migration of the International Union for the Scientific Study of Population (IUSSP), which undertook a review of empirical research on population movements into the world’s five principal immigrant-receiving areas: North America, Western Europe, the Persian Gulf, Asia and the Pacific, and the Southern Cone of South America. In the course of an exhaustive, five-year, system-by-system analysis, the committee reviewed every available piece of empirical information published since around 1960 to evaluate the efficacy of propositions associated with the various theoretical models, yielding a bibliography that ultimately contained more than nine hundred entries (For the results of this review, see Massey et al. [1998]; interim findings were published in Massey et al. [1993, 1994] and in Taylor et al. [1996a, 1996b]).

A second project sought to achieve a synthetic understanding of the forces producing and sustaining international migration by analyzing the specific case of Mexico-U.S. migration, the largest sustained flow of immigrants anywhere in the world. This work grows out of an ongoing study that has surveyed documented and undocumented Mexican migrants for more than a decade (see Durand and Massey 1992; Massey et al. 1987; Massey et al. 1994; Massey 1998). With Kristin Espinosa I used these data to test competing theoretical explanations directly against one another within a common statistical framework. We drew on life histories compiled for some thirty-seven hundred household heads originating in twenty-five separate communities to estimate longitudinal models predicting the likelihood of initial, repeat, and return migration between Mexico and the United States (Massey and Espinosa 1997). So as to broaden the study’s inferential base, these communities ranged from rural hamlets to large urban areas and included a variety of economic bases: agrarian villages, commercial agricultural towns, industrial cities, mining towns, fishing villages, tourist cities, as well as a number of diverse metropolitan economies. The resulting data have been shown to be remarkably representative of the entire population of Mexican migrants to the United States (Massey and Zenteno forthcoming).

Together these two efforts provide an unusually firm basis for constructing an empirically grounded, synthetic theory of international migration. Unlike earlier efforts, the IUSSP evaluation is comparative, considering all disciplines, all contemporary migration systems, and the entire world literature on immigration. Likewise, although the Massey
and Espinosa (1997) evaluation focuses on just one case, it is comprehensive in the sense that it examines variables operating at multiple levels simultaneously and employs a fully dynamic longitudinal model for statistical analysis. Although, owing to limitations of space, I cannot cite specific studies and statistical estimates to buttress each point, the ensuing synthesis is nonetheless supported by the accumulated record of empirical research and consistent with the best empirical work done to date.

International migration originates in the social, economic, cultural, and political transformations that accompany the penetration of capitalist markets into nonmarket and premarket societies (as hypothesized under world systems theory). In the context of a globalizing economy, the entry of markets and capital-intensive production technologies into peripheral regions disrupts existing social and economic arrangements and brings about a displacement of people from customary livelihoods, creating a mobile population of workers who actively search for new ways of achieving economic sustenance. International migrants tend not to come from poor, isolated places that are disconnected from world markets, but from regions and nations that are undergoing rapid change and development as a result of their incorporation into global trade, information, and production networks. In the short run, therefore, international migration does not stem from a lack of economic development, but from development itself.

One means by which people displaced from traditional livelihoods seek to ensure their economic well-being is by selling their services on emerging national and international labor markets (neoclassical economics). Because wages are generally higher in urban than in rural areas, much of this process of labor commodification is expressed in the form of rural-to-urban migration. This movement occurs even when the probability of obtaining an urban job is low, because when multiplied by high urban wages, even low employment probabilities yield expected incomes above those prevailing in rural areas, where wages and employment are both low. According to the neoclassical model, if the difference between incomes expected in urban and rural sectors exceeds the costs of movement between them, as is typical, people will migrate to cities to reap higher lifetime earnings.

Wages are even higher, of course, in developed countries overseas, and the larger size of these international wage differentials prompts some people displaced in the course of economic development to offer their services on international labor markets by moving abroad for work. International wage differentials are not the only factor motivating people to migrate, however, or even the most important. People displaced in the course of economic growth do not move simply to reap higher lifetime earnings by relocating permanently to a foreign setting (although some do). Rather, households struggling to cope with the jarring transformations of economic development use international migration as a means of overcoming market failures that threaten their material well-being (the new economics of labor migration).

In most developing countries, labor markets—both rural and urban—are volatile and characterized by wide oscillations and structural limitations that render them unable to absorb streams of workers being displaced from precapitalist or noncapitalist sectors. Since national insurance markets are rudimentary and government unemployment insurance programs are limited or nonexistent, households cannot adequately protect themselves from risks to their well-being stemming from under- or unemployment. Thus, the lack of access to unemployment insurance creates an incentive for families to self-insure by sending one or more members overseas for work. By allocating members to different labor markets in multiple geographic regions—rural, urban, and foreign—a household can diversify its labor portfolio and reduce risks to income, as long as conditions in the various labor markets are weakly or negatively correlated (new economics of migration).

Household members who remain behind to participate in the ongoing structural transformation of agriculture, meanwhile, generally lack access to insurance markets for crops and futures. As households shift from subsistence to commercial farming, they are forced to adopt new production methods that make use of untested technologies, unfamiliar crops, and untied inputs. As they plunge into the unknown world of capitalist production, the lack of insurance or futures markets leaves them vulnerable to economic disaster should these new methods fail, providing yet another incentive for families to self-insure against risk through international migration. Should crops fail or commodity prices fall precipitously, households with at least one worker employed overseas will not be left without a means of subsistence (new economics of migration).

Another failure common to developing countries occurs in capital and consumer credit markets. Families seeking to engage in new forms of
agriculture or looking to establish new business enterprises need capital to purchase inputs and begin production. The shift to a market economy also creates new consumer demands for expensive items such as housing, automobiles, and appliances. The financing of both production and consumption requires cash, but the weak and poorly developed banking industries characteristic of most developing nations cannot meet the demands for loans and credit, giving households one final motivation for international labor migration. By sending a family member temporarily abroad for work, a household can accumulate savings and overcome failures in capital and consumer credit markets by self-financing production or consumption (new economics of migration).

Whereas the rational actor posited by neoclassical economics takes advantage of geographic disequilibria in labor markets to move abroad permanently to achieve higher lifetime earnings, the rational actor assumed by the new economics of labor migration seeks to cope with failures in insurance, futures, capital, and credit markets by moving overseas temporarily to repatriate earnings in the form of regular remittances or lump-sum transfers. In this way the household of this rational actor controls risk by diversifying sources of income and self-finances production or consumption by acquiring alternate sources of capital.

Thus, although wage differentials, the favored explanatory factor of neoclassical economics, may account for some of the historical and temporal variation in international migration, failures in capital, credit, futures, and insurance markets, key factors hypothesized by the new economics of labor migration, are also powerful causes. In purely theoretical terms, wage differentials are neither necessary nor sufficient for international migration to occur. Even with equal wages across labor markets, people may have an incentive to migrate if other markets are inefficient or poorly developed.

Although the early phases of economic development in poor nations may create a mobile population seeking to earn more money, self-insure against risk, or self-finance production or consumption, postindustrial patterns of economic growth in wealthy nations yield a bifurcation of labor markets. Jobs in the primary sector provide steady work and high pay for native workers, while those in the secondary sector offer low pay, little stability, and few opportunities for advancement, repelling natives and generating a structural demand for immigrant workers (segmented labor market theory). The process of labor market bifurcation is most acute in global cities, where a concentration of managerial, administrative, and technical expertise leads to a concentration of wealth and a strong ancillary demand for low-wage services (world systems theory). Unable to attract native workers, employers turn to immigrants and often initiate immigrant flows directly through formal recruitment (segmented labor market theory).

Although instrumental in initiating immigration, recruitment becomes less important over time because the same processes of economic globalization that create mobile populations in developing regions and generate a demand for their services in global cities also create links of transportation and communication, as well as politics and culture, to make the international movement of people possible, even likely (world systems theory). Immigration is also promoted by foreign policies and military actions that core capitalist nations undertake to maintain international security, protect foreign investments, and guarantee access to raw materials; these foreign entanglements create links and obligations that generate ancillary flows of refugees, asylees, and military dependents (world systems theory).

Eventually labor recruitment becomes superfluous, for once begun, immigration displays a strong tendency to continue through the growth and elaboration of migrant networks (social capital theory). The concentration of immigrants in certain destination areas creates a "family and friends effect" that channels immigrants to the same places and facilitates their arrival and incorporation. An enclave economy may form if enough migrants arrive under the right conditions, further augmenting the demand for immigrant workers (segmented labor market theory).

The spread of migratory behavior within sending communities sets off ancillary structural changes, shifting distributions of income and land and modifying local cultures in ways that promote additional international movement. Over time the process of network expansion becomes self-perpetuating because each act of migration creates social infrastructure capable of promoting additional movement (the theory of cumulative causation). As receiving countries implement restrictive policies to counter rising tides of immigrants, they create a lucrative niche into which enterprising agents, contractors, and other middlemen move to create migration-supporting institutions that also serve to connect areas of labor supply and demand for purposes of profit, providing migrants with an-
other resource capable of supporting and sustain-
ning international movement (social capital theory).

During the initial phases of emigration from any
sending country, the effects of capitalist penetra-
tion, market failure, social networks, and cumula-
tive causation dominate in explaining the flows,
but as the level of out-migration reaches high
levels and the costs and risks of international
movement drop, movement is increasingly deter-
mined by international wage differentials (neo-
classical economics) and labor demand (seg-
mented labor market theory). As economic growth
in sending regions occurs, international wage gaps
gradually diminish, and well-functioning markets
for capital, credit, insurance, and futures arise,
progressively lowering the incentives for emigra-
tion. If these trends continue, the country ulti-
mately becomes integrated into the international
economy as a developed, capitalist nation, where-
on it undergoes a migration transition: net out-
migration progressively winds down, and the for-
mer sending nation itself becomes an importer of
labor.

MIGRATION THEORY RECONSIDERED

All theories thus play some role in accounting for
contemporary patterns and processes of interna-
tional migration, even though different models
may be more relevant to explaining particular mi-
gration flows, and different explanations carry dif-
ferent weights in different world regions depend-
ging on the local circumstances of history, politics,
and geography. Generalizing across all theories, I
conclude that a satisfactory theoretical account of
international migration must contain at least four
elements: a treatment of the structural forces that
promote emigration from developing countries; a
characterization of the structural forces that attract
immigrants into developed nations; a considera-
tion of the motivations, goals, and aspirations of
the people who respond to these structural forces
by becoming international migrants; and a treat-
ment of the social and economic structures that
arise to connect areas of out- and in-migration.

Any theoretical explanation that embraces just one
of these elements will necessarily be incomplete
and misleading and will provide a faulty basis for
understanding international migration and devel-
oping policies to accommodate it.

Each theory specified to date focuses on just
one or two of these four elements, so all are neces-
sary to build a comprehensive, integrated under-
standing of international migration in the late
nineteenth century. The leading theoretical treat-
ment of the forces that promote emigration from
developing countries is world systems theory. To-
gether, world systems theory, segmented labor
market theory, and neoclassical macroeconomics
offer explanations for why developed countries at-
tract immigrants. Social capital theory and world
systems theory explain how structural links emerge
to connect areas of origin and destination. Neo-
classical economics and the new economics of la-
bor migration deal with the motivations of the
people who become international migrants, and
the theory of cumulative causation describes how
international migration promotes changes in per-
sonal motivations and origin, destination, and in-
tervening structures to give immigration a self-per-
petuating, dynamic character.

In this seemingly complete theoretical account
of the forces that produce and shape international
migration, however, one major factor has been
sorely neglected: the state. Even though govern-
ments may not be able to control fully the power-
ful forces promoting and sustaining international
migration, state policies clearly have an influence
in determining the size and composition of the
flows. Yet among the theories reviewed here, sur-
prisingly little attention has been devoted to na-
tion-states or their governments as active agents
whose behavior shapes, if not controls, interna-
tional population movements. segmented labor
market theory considers the state relevant only in-
sofar as it acts on behalf of employers to establish
labor recruitment programs. world systems theory
treats the state primarily as a handmaiden to cap-
talist interests, projecting military and political
power to expand markets, acquire raw materials,
and guarantee free trade. Social capital theory
mentions the state only insofar as its use of family
reunification criteria in immigrant admissions rein-
forses the operation of migrant networks. The re-
main ing theoretical paradigms—neoclassical eco-
nomics, the new economics of labor migration,
and the theory of cumulative causation—do not
deal directly with the state at all.

In general, therefore, contemporary theories of
international migration do not consider the state
to be a significant independent actor capable of
shaping international migration for its own pur-
poses, or for the purposes of the politicians and
bureaucrats who administer it. To the extent that
the state is discussed at all, moreover, attention
has focused primarily on immigrant-receiving na-
tions; little has been said about the interests and
behavior of governments or politicians in regions of origin. Thus, the role of the state in initiating and promoting (or stopping and preventing) international migration is remarkably undertheorized and little studied.

Although scholars have surveyed national immigration policies (see Cornelius, Martin, and Hollifield 1994a; Dib 1988; Kubat 1979b; Papademetriou 1996), conducted case studies of specific state agencies (Calavita 1992), and compiled legislative histories of immigration law in specific countries (Hutchinson 1981), with the recent exception of the as yet unpublished work of Eytan Meyers (1995), they have not attempted to theorize the behavior of governments or the actions of politicians themselves. Hypotheses concerning the interests, role, and behavior of the state constitute a missing link in theories of international migration.

The resurgence of massive global migration in the late twentieth century thus offers many formidable challenges to social scientists, policymakers, and the public. For theoreticians, a principal challenge is to model the behavior of nation-states and political actors, filling a void in the general theory of international migration constructed here. Although the field seems to be on the verge of developing a good theoretical model of the structural forces that promote migration from developing to developed nations, as well as a cogent conceptual understanding of the transnational structures that support this movement and of the personal motivations of the migrants who respond to these structural dynamics, what theorists lack at this point is an adequate account of the motivations, interests, and behavior of the political actors who employ state power to influence these processes, and how their interventions affect outcomes at the individual and aggregate levels.

For empirical researchers perhaps the greatest challenge is to design studies that are more closely connected to theory. In considering empirical research from around the world, its most striking feature is the degree to which it is unconnected not just to a particular theory but to any theory at all. If our knowledge of international migration is to advance, researchers must make greater efforts to familiarize themselves with the principal contemporary theories of international migration and to formulate research designs capable of testing their leading propositions. What is especially needed at this juncture are studies that simultaneously test the propositions of several theories, so that the relative efficacy of different explanations can be directly compared and contrasted.

Given the current state of knowledge, the "truth" of a theory is less important than how well it compares with others in explaining international migration. Only a handful of studies anywhere test more than two theories at once, let alone examine all the theoretical paradigms reviewed here; at this point in time the research literature is far too restricted geographically to serve as a satisfying basis for generalization. Too many of the rigorous theoretical tests conducted so far have been based on samples of Mexican rural areas, and rural- and urban-origin emigrants may differ in important ways that are not yet well understood, whereas the case of Mexico-U.S. migration is unique by any criterion.

Perhaps the most profound challenge of all will be faced by citizens and policymakers in migrant-sending and -receiving countries. Inhabitants of the latter will have to move beyond the psychological denial that characterizes their approach to immigration policy. They must develop policies that recognize the inevitability of labor flows within a globalized economy, supported by well-established regional networks of trade, production, investment, and communications. Attempts to suppress population flows that are a natural consequence of a nation's participation in these economic networks will not be successful, but they will present grave threats to individual rights, civil liberties, and human dignity.

In sending societies, meanwhile, citizens and policymakers face different but equally vexing issues. Rather than passively acquiescing to emigration and simply waiting for remittances to materialize, developing countries must actively plan to derive benefits from what is potentially an important engine of economic growth. Unless concrete steps are taken to attract migradollars (migrant remittances and savings) and channel them into productive ends, to capture remittances for purposes of foreign exchange, and to forestall the loss of human capital through emigration, development outcomes are likely to be disappointing.

Finally, the globalization of capital and labor markets and the internationalization of production pose strong challenges to the very concept of the nation-state and the idea of national sovereignty, requiring political leaders and citizens in both sending and receiving nations to move beyond nineteenth-century conceptions of territory and citizenship to more expansive notions that embrace the transnational spaces that are currently being formed throughout the world as a result of massive circular migration. These changes are es-
especially daunting because they occur at a time when the forces of globalization are also producing downward pressure on wages and incomes and rising inequality throughout the world.

These are formidable challenges indeed, but they will have to be met, for international migration will surely continue. Barring an international catastrophe of unprecedented proportions, immigration will most likely expand and grow, for none of the causal forces responsible for immigration show any sign of moderating. The market economy is expanding to ever farther reaches of the globe, labor markets in developed countries are growing more rather than less segmented, transnational migration and trade networks are expanding, large stocks of migration-related human and social capital are forming in sending countries everywhere, and the power of the nation-state is faltering in the face of this transnational onslaught. The twenty-first century will be one of globalism, and international migration will no doubt figure prominently within it.

This chapter draws extensively on collaborative research undertaken with Joaquin Arango, Graeme Hugo, Ali Kourouchi, Adela Pellegrino, and J. Edward Taylor (see Massey et al. 1998)