1 The Theory of the Demographic Transition: Background and Logical Status

Introduction

The historical background of the theory of the demographic transition is examined in this chapter, with the purpose of putting problems currently faced by the theory in perspective. The theory originated as a means of explaining European demographic trends that did not accord with accepted Malthusian ideas, and this in itself accounts for some of its present difficulties.

The second aim of this chapter is the development of a more systematic statement of transition theory. For the moment there is no attempt to modify the theory, but only to summarize those basic arguments which can be considered a macroscopic sociodemographic theory.

The Malthusian Model

The work of Thomas Robert Malthus has continued to stimulate demographic theorizing and research for over 150 years. In a sense his contribution was a negative one, since other writers, from Marx to contemporary demographers, have mainly been interested in refuting his theory, at least in its more simplistic versions.

Malthus asserted that the size of a human population is ultimately limited by “positive” or “preventive” checks. The former include “war, disease, hunger, and whatever … contributes to shorten the natural duration of human life” (Malthus, 1803:14). The “preventive” check is abstinence from sexual relations, accomplished either by delay of marriage or continence within marriage. Malthus then made the crucial assumption that the preventive check could not operate effectively without the threat of misery. Thus Malthus concluded that good living conditions for the majority of mankind implied unchecked population growth.

Malthus estimated that a population could double every twenty-five years. Although the exact figure is not crucial, his calculation was reasonably accurate. By contrast, Malthus asserted that food supplies in the long run could increase only in an arithmetic progression. This leads directly to the “Malthusian dilemma”: with good living conditions, populations grow rapidly, but eventually the food supply is overtaxed until either the threat of misery forces use of the preventive check or else the positive checks intervene by increasing the death rate. The eventual result is always population equilibrium and human deprivation.

There are two distinct propositions upon which Malthus constructed his argument: an assumption about the determinants of natality, and an assumption about the rate at which food supplies can increase. The theory of the demographic transition deals only with the former, by suggesting a new set of conditions under which natality can be checked. Only this aspect of the Malthusian model is discussed below.

Anomalous Empirical Trends

Class and Natality

Malthus himself came to recognize the growing evidence that the birth rate could be limited by voluntary means in the absence of severe deprivation (Malthus, 1836:436 ff.). This discovery
seems to have stemmed from the observation that the upper socioeconomic classes had lower natality than others. By itself this fact probably did not suggest much reason for hope, since the upper classes were a distinct minority; however, Malthus contended that it might be possible to diffuse such admirable behavior patterns to the poor, and recommended universal education to accomplish this end (Malthus, 1836:437, 477). In recognizing the importance of class and education, Malthus was coming to appreciate the social determinants of natality. This is in contrast to his earlier formulation, which stressed biological processes and simple economic ideas. Although based on empirical observations, the inverse class-natality relationship detectable in his time was rather flimsy evidence to have so softened the gloomy pronouncements of Malthus’s early writings. Malthus’s thinking seems to have been affected as much by criticism he received, especially from Godwin (Petersen, 1971), as by his own research findings. However, as time passed, what must have appeared to him as a very minor trend grew to become a striking empirical reality.

European Natality Trends

It is ironic that Malthus presented the first edition of his Essay on Population (Malthus, 1798) at the very time when its assumptions were ceasing to hold in Europe. It is doubly ironic that the theory of the demographic transition, which arose to explain the divergence from Malthusian behavior, was not really formulated until those anomalous demographic trends were virtually complete in northwestern Europe. Theory has thus followed in the wake of empirical observations of European nations. This leads us to consider the demographic trends that came to be known as the demographic transition.

By 1800 mortality had fallen substantially in the more developed areas of Europe, generally in the northwest. This was contributing to a slow (by present standards) but persistent population growth, which presumably aggravated the concerns felt by Malthus and others. Gradually, however, scattered natality reductions became evident and by 1875 the trend was established throughout northwestern Europe. By 1900 birth rates all over Europe, with the exception of Russia and some of the eastern and southern extremities of the continent, were dropping. By the time of the Great Depression natality in western, central, and northern Europe was low enough to be at the point of long-term equilibrium with mortality, which itself was leveling off at a life expectation (at birth) of about seventy years. Although there were subsequent fluctuations, birth rates remained low in these areas. In eastern and southern Europe, including Russia, both mortality and natality decline lagged behind, so that the decline of the birth rate was not completed until fairly recent times. In a few isolated areas, particularly Albania, natality was still high, but falling, in the 1960s. Reductions in death and birth rates are among the most persistent and readily observable aspects of social change in nineteenth- and twentieth-century Europe. (See Kuczynski, 1928, 1931, and 1936 and United Nations Demographic Yearbooks, 1949-70.)

In its most elementary form, the theory of the demographic transition does not try to explain these trends, but simply summarizes them and labels the resulting empirical generalization the “demographic transition.” Even at this stage the ideas have some empirical import. They assert that a substantial mortality decline invariably precedes and is inevitably followed by a major natality decline. This might be characterized as a phase theory, because it outlines three sequential stages of demographic development: initially, high and fluctuating mortality balances high natality; next, declining mortality coupled with high natality produces rapid growth; and finally, natality drops to roughly’ balance mortality and continues to fluctuate while mortality is
relatively stable (Thompson, 1929; Coale and Hoover, 1958:12-13). However, demographers have not been content with mere description and, while theory construction has not been systematic, a theory has arisen nonetheless.

From this point one could either move directly to theory-building or else develop more detailed empirical generalizations. Demographers have done both, and while this discussion stresses the former, a few of the main empirical regularities are noted below in conjunction with theoretical development.

Common demographic patterns suggest that the values and norms governing natality are difficult to change. The tendency of mortality decline to precede natality decline is widely, but not universally, accepted (Stolnitz, 1964). Stolnitz (1964) has also indicated that future departures from high natality in non-European countries are likely to be slow in coming, and to lag well behind mortality reductions.

It has also been suggested that natality decline comes first to those parts of society most exposed to new values. This may account for early natality declines among upper classes (Cowgill, 1963). By the same token, it has been suggested that urbanites should have reduced natality sooner than rural persons, and in fact have done so. Although often considered a proven fact of the European experience, this latter assertion has recently been disputed (Carlsson, 1966).

Finally, the irreversibility of natality decline is another very important empirical finding based on the experience of European peoples (Stolnitz, 1964). There may be subsequent fluctuations of natality and at times these may be rather great in magnitude, such as the postwar natality recovery in Western nations. In the longer term, however, natality remains low, and in no case has it ever returned to pre-transition levels. The arguments used to explain the demographic transition attempt to explain this and other empirical regularities referred to above.

The Theory of the Demographic Transition

Social Structure and Demographic Behavior

It was always obvious that the demographic transition was associated with the process that is variously called “modernization,” “development,” “socioeconomic development,” “bureaucratization,” “industrialization,” “urbanization,” “progress” etc. Although early references to this association tended to be quite casual, a theoretical explanation was beginning to take shape and to become more detailed. The first statement of these ideas is usually attributed to Thompson (1929-see also Carr-Saunders, 1936).

The term “socioeconomic development” is often used quite carelessly, and left undefined. This book uses a standard, although rather broad, definition of “development” (which will also be referred to as “socioeconomic development” or “modernization”). It is assumed that the transition from a traditional agrarian to a modern urban society has several identifiable aspects upon which most social scientists would agree. Thus urbanization, rising levels of production and consumption, emergence of the money economy, advances in education, growing importance of social relationships and institutions not based on kinship, increasing secularization of life, and the shift from normative to functional integration are all aspects of development. Although each of these, and others that might be included, is a phenomenon in itself, they are interrelated (Ness, 1970), and there appears to be a common core, or a “common factor,” which makes the term “development” meaningful. One or more dimensions of this process can arise independently, but
ultimately all must occur if a society is to change, and this transformation is facilitated if all aspects are moving more or less together. These assumptions, if allowed, make it possible to speak of socioeconomic development in the broad sense. Note that, for reasons soon to become apparent, falling mortality levels have not been treated as part of the process. For the purposes of this study it is advantageous to consider mortality as a separate phenomenon, associated with development but not a part of it. However, there is no theoretical reason to consider it any more independent of development than some of the other types of social change outlined above.

Explaining Mortality Decline

The present study is concerned with natality, but a few comments about the transition-theory explanation of mortality decline are in order. The factors accounting for the decrease in the death rate are often considered nonproblematic. Furthermore, there is no real contradiction here between transition theory and the Malthusian model. It is usually assumed that individuals and cultures value life under most circumstances, and will seek to prolong it. They are therefore not hesitant to apply the technology of death control as it improves during development. Similarly, they are not reluctant to avail themselves of the increasing food supplies that are generated by the modernization of the agricultural sector. This improves nutrition and health and thus reduces mortality. There has been considerable debate among demographers over which factors (e.g., personal and environmental hygiene, public health measures, vaccination, food supply) were most important in lowering mortality at various periods (Helleiner, 1965). Some writers have even suggested that mortality could not have declined as early as was previously believed, and thus could not have accounted for the increased rate of European population growth observed in the early nineteenth century (Habakkuk, 1958; Petersen, 1960). In what follows, these problems will only be considered to the extent that they impinge directly on natality trends.

Explaining Natality Decline

Natality and Social Change. This book is concerned with natality, and in particular with those factors that are associated with natality decline during socioeconomic development. It is in this area that transition theory has made its greatest and most original contribution while stimulating research and theoretical discussion.

Malthusian thinking inextricably linked “good times” with high birth rates. On the other hand, the experience of the demographic transition in Europe suggested the opposite—in terms of wealth, education, social reform, health, and in almost every conceivable way life was generally becoming more pleasant and comfortable all the while the birth rate was falling to levels never before recorded except in periods of dire distress. It has been noted that Malthus became aware of the possibility of such behavior after observing class differentials. As early as 1864, Mill (1864:205-12) recognized that whole societies were beginning to control natality consciously, apparently to protect their rising standard of living—at least this was his interpretation. Mill not only magnified the cautious optimism of the later Malthus, but went on to anticipate the relationship between socioeconomic development and natality control, while suggesting an explanation for it. Later, when the European natality decline was largely past history, many other explanations were proposed, and despite their variety and lack of coherence, all of them are generally considered to be part of the theory of the demographic transition.

The Role of Contraceptive Technology. Although it is widely agreed that socioeconomic development is somehow responsible for natality decline, there has been some disagreement as to
how and why this comes about (Robinson, 1964). Early explanations tended to stress the increasing availability of modern contraceptive devices that accompanies development (Carr-Saunders, 1936:105 ff.; Fairchild, 1939:127-51; Bogue, 1967). Critics of this approach pointed out that traditional means of birth limitation, such as coitus interruptus, are often available long before socioeconomic development, and that these methods, while unreliable at the individual level, can reduce the aggregate birth rate considerably. They also noted that the major portion of the European natality decline antedated the widespread use of modern contraceptives (Notestein and Stix, 1940:148 ff.). Thus, one would be hard put to explain why the French birth rate dropped almost continuously from the Napoleonic period onwards to the 1930s on the basis of contraceptive technology. Most demographers have taken the more tenable view that contraceptive technology is a facilitator rather than the basic cause of the natality decline associated with development (Robinson, 1964:385).

**The Role of Values, Norms, and Mortality.** Other explanations proposed that the cause of the modern decline of natality was to be found in “new patterns of living and new values” (Notestein and Stix, 1940:149-50). Changing values, occupational and residential composition, and increasing education were specifically mentioned by these authors. Under the heading of “values” Notestein and Stix pointed out that development tends to weaken the traditional fatalist orientation, and thus leads people to feel that many aspects of their lives, including natality, can be regulated. They suggested, as Malthus had, that education helps to spread the new attitude throughout society. They also regarded urbanization as a multifaceted cause of reduced birth rates; urbanism tends to foster a more highly skilled and better educated work force, which tends to make the laboring class more similar to the upper classes. It was also suggested that urban women are less restricted by traditional sex roles and are more often employed outside the home (Notestein and Stix, 1940:148 ff.). Stolnitz (1964:33-34) pointed out that the employment of women means that their lost wages and/or the price of child care must be added to the economic cost of children. What is outlined above, then, is a sociodemographic theory of sorts. Changes in values and social structure are linked, more or less clearly, to falling natality. This orientation takes the early class differentials in natality very seriously, and tends to assume that the motivation and ability to control births diffuses to the lower classes as they acquire some of the benefits, social attributes, and values once confined to the upper classes.

One of the persistent problems of transition theory is that various authors rake essentially similar arguments but stress different points or use different terminology. For instance, economic terms can often be substituted for sociological concepts or used in conjunction with them. Thus development may cause natality decline partly by raising the cost and lowering the economic value (of children (Coale, 1969). A sociologist could also point out, however, that the adoption of prohibitions against child labor, a normative change, is a part of this process, or it could be suggested that it is a decline in the status value of large families that is crucial (Davis, 1963). Similarly, one could argue that rising status of women results in a greater opportunity to choose some role other than wife-mother for a greater portion of the life cycle. When value changes are involved, the confusion extends even further. Various authors have stated that natality falls partly because of the emergence of “nonfatalist” attitudes (Cowgill, 1963), “nontraditional” values (Stolnitz, 1964), “secular-rational” attitudes (Coale, 1969), “work orientation,” and “individualism” (Clifford, 1971). Although the many shades of meaning and terminologies can confuse presentations of transition theory, the fact that so many different approaches lead to similar formulations is encouraging. An interrelated nexus of social, economic, and
psychological changes that accompany socioeconomic development makes large families far less desirable for parents.

The declining importance of kinship structures that accompanies development has often been accorded a special theoretical position in the explanation of natality decline. In traditional societies, extended kinship systems help to place a high social and economic premium on the bearing and raising of large families (Lorimer, 1954:58-90, 151-203). As development occurs, the importance of extended kinship systems is reduced by the growth of other institutions, and the nuclear family becomes the basic family unit. Freedman (1963) has developed this general argument further by noting that the process involves participation in larger, more specialized institutions; in fact he defines “development” as “the shift from major dependence on relatively self-contained local institutions, to dependence upon larger social, economic, and political units.” Freedman (1963) suggested that this places the social and economic costs of children more directly on the parents, and at the same time makes more than a small number of children irrelevant to the “goals men seek in the more developed society.”

Some, but not all, authors have postulated that mortality decline itself causes natality reduction (Freedman, 1963; Coale, 1969). This mechanism has usually been viewed as somewhat distinct from the impact of socioeconomic development, although mortality is associated with development. It is usually argued that with lower infant and child mortality, fewer births are necessary to assure a given ultimate family size. Thus, with high mortality many births might be required to guarantee with reasonable probability that one son will survive to support his parents in their old age. Unfortunately, this idea, which is sometimes called the concept of “insurance births,” makes an implicit assumption about rationality. It seems to imply that persons, even in traditional societies, have some fairly specific ideas about the ideal or desirable ultimate family size, and that they take mortality into account when deciding how many children to have.

There is another way of explaining the causal link between mortality and natality, however, which does not put so much of a burden on the individual. The lowering of infant and child mortality without offsetting birth control can greatly increase household size, and play havoc with social structures designed to accommodate lower survivorship (Freedman, 1963:226-27). One example is ‘inheritance systems-arrangements that may have been serviceable in the past may become hopelessly overloaded when too many children survive to adulthood. This can be especially crucial when it is land that is involved. Thus, if individuals control their natality in response to lower mortality, it may be the result not of an understanding of demographic trends and their implications, but of a simple defensive reaction to the burdens of families larger than existing social and economic structures can readily handle.

Organizing the Theory as It Pertains to Natality

The above factors can be grouped on the basis of causal priority and the type of motivation through which they are assumed to operate. First it is generally agreed that socioeconomic development is the basic cause of natality decline. This covers at least four interrelated phenomena: urbanization, education, nonkinship institutions, and consumption levels (standard of living). These are said to facilitate intervening modifications of individual behavior via social, economic, and/or psychological mechanisms. The social cluster includes the relaxation of sex-role restrictions on women, the decreased predominance of extended kinship systems, and the reduced status value of children, especially male heirs. The central factor certainly seems to be kinship-as development proceeds, formal organizations based on secondary association take
over many of the economic, educational, and political functions previously served by kin groups. Economic variables of importance include the reduced labor value of children, the increased cost of raising children, the emergence of competing consumer goods, and the money value of the wife’s labor outside the home. Psychological factors are more difficult to enumerate, since attitudinal changes can be expressed in many ways, but the following orientations seem to facilitate natality decline -- nonfatalism, materialism, rationality, and secularism.

Usually these variables are thought to mediate between socioeconomic development and “desired or ideal family size.” The latter are hypothetical constructs which appear, though often implicitly, in most explanations. This is not a necessary assumption; even in developed societies family size is not always the result of planned, conscious decision. Of course this problem can be avoided by making “desired or ideal family size” a norm, and thus avoiding the assumption of individual decision.

Mortality is also assumed to affect natality, but without any apparent impact, on the ultimate family size desired or thought to be ideal. Rather mortality, especially that of infants and children, is thought to mediate between ideal-desired family size and the practice of birth control. (This assumption will be modified in a subsequent chapter.)

Finally, natality decline is caused by the implementation of some type of birth control. This in turn is affected by desired-ideal family size and mortality ‘levels, although, following arguments made above, it is typically assumed that birth-control technology has an independent, but not overriding effect. Delay of marriage or abstinence from sex might also be viewed as means of birth limitation.

Trying to illustrate such a system with even a primitive causal diagram is very difficult, since almost every set of variables might potentially affect every other. Nevertheless a crude attempt is made below (Figure 1-1). This model does not pretend to be complete, but only a summary of some of the major relationships suggested by transition theory.

Transition theorists generally portray causation as moving from the upper left to the lower right. We cannot, however, rule out feedback. Some of these possible “reverse” relationships are indicated by the double-headed arrows, but even these do not exhaust the conceivable complications. For instance, it is very probable that natality itself affects the development variables, that desired size affects birth-control technology, etc.. These relationships are excluded from the diagram for the sake of legibility.

In general, the theory of the demographic transition is not well explicated, but is weighted down with numerous supplementary arguments drawn from different disciplines. The only clear implication of the theory is the prediction that socioeconomic development will lead to natality decline, and that this will tend to occur sometime after a major decline in mortality.

*The Demographic Transition as a Sociodemographic Theory*

The theory of demographic transition is not a set of logically interrelated propositions from which testable empirical statements can be formally derived. Nevertheless, it does make empirical assertions and propose causal arguments.
Some versions of the “theory” are in fact empirical generalizations. However, if one thinks of these generalizations as predictions of what “should” happen in societies that have not yet experienced the demographic transition, then one begins to have a theory of phases of demographic movements. The present transition theory also consists of explanations of the above empirical regularities. At the first level this simply involves the notion that socioeconomic development reduces the birth rate, but more detailed arguments are incorporated to explicate the relationship between socioeconomic structure and natality. Unfortunately these additional arguments are not always complementary or cumulative. Rather, each more or less independently attempts to make the development-natality-decline association plausible. Since the theory arose specifically to explain the previously anomalous natality declines, this tendency is understandable, but little seems to have been added to precision or the predictive capacity in the process.

The Theory of the Demographic Transition as a Paradigm

Disappointment with the theory should not obscure its great contribution. Sometime between the first and most simplistic version of Malthus and the statement of the theory of the demographic transition, a great discovery was made: Natality is a function of economic, social, psychological, and cultural factors. Despite its shortcomings, the theory has provided a framework for the scientific study of human natality, and we might consider whether the “theory” is in fact a paradigm.

Thomas Kuhn (1962) has discussed the nature and import of scientific paradigms at some length. He states that a paradigm is attained after some striking breakthrough that is sufficiently
unprecedented to attract an enduring group of adherents, solve some major problems, and yet be “sufficiently open-ended to leave all sorts of problems for the … practitioner to resolve” (Kuhn, 1962:10). Often such a paradigm arises in reaction to a set of data, anomalous relative to previously existing theory.

The theory of the demographic transition fits Kuhn’s criteria fairly well. The theory did arise after an anomalous series of events—the European natality decline. The Malthusian model could not explain what was happening, so the way was cleared for new ideas. The real “breakthrough” consisted of linking social change with demographic change and proposing new sociological, cultural, psychological, and refined economic arguments to further explain the connection. Previously these factors were not clearly recognized as causes of natality. The theory of the demographic transition not only suggested that they were, but also indicated how and in what directions they ought to operate. It can be argued, of course, that the so-called “explanations” were not logically tight, and certainly that the theory was and is not as determinate as theories typically found in the natural sciences. Kuhn (1962:27-34) points out, however, that even Newton’s classic theoretical work Principia, and many other efforts of similar stature, were rough and cumbersome in their initial statement. It required a great deal of reformulation and revision to bring these theories to the elegant and articulate level at which we now encounter them. If the explanations of the, various statements of the theory of the demographic transition are not rigorous, they are at least plausible in the light of modern social science. In any case, theme most important function of the theory has been making further demographic research possible by telling investigators what variables to consider and what relationships are reasonable. According to Kuhn, this is the primary function of a paradigm, for it makes what he calls “normal science” a viable enterprise (Kuhn, 1962:3542).

The theory’s weakness, both as a theory and as a paradigm, is that it leaves too much unstated and too many ideas unintegrated. It still requires a great deal of refinement. Unfortunately, demographers have often been more interested in finding an exception to some of the empirical generalizations more or less implied by transition theory than in improving the logical structure. For this reason and others, the paradigm suggested by transition theory has remained rather crude.

**Specific Problems Currently Faced by the Theory**

Recently there has been a resurgence of interest in transition theory, for at least three reasons. First, there has been an unprecedented increase in world population, due largely to declining mortality in the developing areas. Second, conditions in these nations now seem to resemble in certain respects those which preceded the massive natality decline in Europe. Third, quantitative data are increasingly available to measure not only demographic variables but also many aspects of socioeconomic structure. Circumstances thus seem ripe for a refinement and extension of the theory, once we have summarized its current problems.

Most difficulties arise from the disagreement among investigators as to the content of the theory. This is understandable, since the theory has certain aspects of an empirical generalization, a true sociodemographic theory, and a paradigm. A given datum may support, or be irrelevant to or inconsistent with the theory, depending upon how it is interpreted. For the present we will use the model presented in Figure 1-1 and the ideas underlying it as a guide in determining the predictive capacity of transition theory. (This model still needs to be made more explicit and
Sometimes the theory is faulted incorrectly because it is interpreted as the most simplistic empirical generalization. The mere fact that the particular demographic trends previously observed in Europe are not found elsewhere does not disprove the theory, but this is often taken to be so. The theory itself should not, as a theory, predict that the same trends of the dependent variables will occur everywhere, but only that the predicted relationships with the independent variables will hold true.

On the other hand, the fact that the theory has certain characteristics of a paradigm sometimes leads researchers to expect too much of it. No social theory will explain all salient features of a dependent variable. The theory of the demographic transition does not claim to exhaust all the possible determinants of natality, but only to isolate some strong relationships that were responsible for major trends in the past. Since the theory was formulated with the object of explaining the demographic trends in nineteenth- and early twentieth-century Europe, probable determinants of natality that did not help explain these particular trends did not receive much attention. This does not imply that other causal factors are negligible or that transition theory is wrong in leaving them out. It does warn the researcher that the independent variables embedded in transition theory may not completely determine natality.

The increasing popularity of statistical analyses has also resulted in some difficulties for transition theory. A regression or any other mathematical model represents some kind of hypothesized system. Given that many researchers do not know what the theory of the demographic transition states, they run a serious risk in using a mathematical model to test it. Standard mathematical procedures will give sensible answers only to questions that are appropriately posed. Unfortunately, some investigators have forsaken theory completely in favor of a blind regression of natality on a variety of measures, some socioeconomic and some demographic, without any consideration of the assumed causal system.

It is hoped that in what follows, some of these problems can be alleviated or avoided while the theory is revised and refined. This new model will then be tested with data from the Latin American region. However, for the moment the theory will be left as it stands, a partially explicated theory linking socioeconomic structure, mortality, and natality. In the next two chapters empirical experiences of both Europe and the more recently developing areas will be examined. It is to be hoped that this will help to indicate the ways in which the theory must be altered.